

# 視覚パターン認識とネオコグニトロン

福島 邦彦

ファジィシステム研究所

fukushima@m.ieice.org

<http://personalpage.flsi.or.jp/fukushima/>

## 1. はじめに

ネオコグニトロンは、変形に強い視覚パターン認識能力を持つ人工神経回路である<sup>1,2)</sup>。最近、深層学習やdeep CNN (convolutional neural network) が注目を集めているが、ネオコグニトロンもそのような階層型多層神経回路の一種で、パターン認識能力を学習によって獲得していく。ネオコグニトロンの歴史は古いが、現在に至るまで種々の改良が加えられ発展を続けている。現在広く用いられているdeep CNNとの相違点に注目して、最近のネオコグニトロンを紹介する。

## 2. ネオコグニトロンの概要

### 2.1 ネオコグニトロンの回路構造

ネオコグニトロンは図1に示すように、入力層 $U_0$ の後ろに、特徴抽出を行うS細胞の層 $U_s$ と、位置ずれを許容する働きを持つC細胞の層 $U_c$  (pooling層)とを交互に階層的に配置した多層神経回路である。各層は、複数の細胞面(sub-layer)から構成されている。細胞面内の細胞は、同一の入力結合を共有している。ちなみにS細胞とC細胞は、大脳視覚野の単純型細胞 (simple cell) と複雑型細胞 (complex cell) に似ているので、その頭文字をとって名づけられた。

S細胞は学習によって変化する可変入力結合を持つ。学習を終わると、S細胞は特徴抽出細胞として働くようになり、入力層に与えられる

刺激パターンの局所的な特徴を抽出する。細胞面内のすべてのS細胞は結合を共有しているため、各細胞面はconvolutionの演算を行っていることになる。あるいは、前層の出力に空間フィルタを掛けていると解釈することもできる。

### 2.2 特徴の抽出と統合

ネオコグニトロンでは、個々の細胞は一つ前の層の比較的狭い範囲にある細胞だけから結合を受けている。その結果、下位の層のS細胞は小さい受容野を持ち、入力パターンの局所的な特徴を抽出する。例えば文字 'A' を学習したネオコグニトロンの下位の層 $U_{s2}$ には、図2に示すように、文字 'A' の種々の部分パターン、すなわち局所の特徴に反応するS細胞が並んでいる。

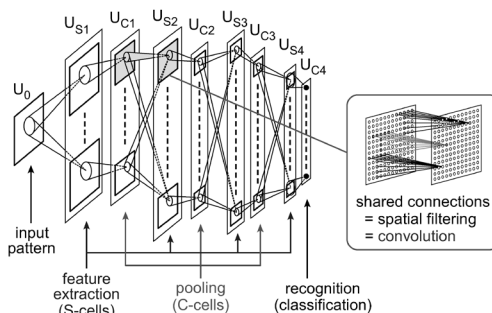


図1 ネオコグニトロンの回路構造。

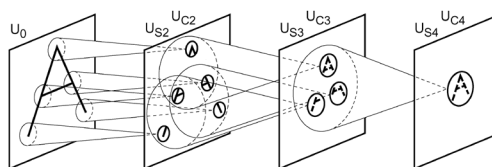


図2 ネオコグニトロンにおけるパターン認識の原理<sup>1)</sup>。

このような局所的特徴を抽出するS細胞が、どのような組合せで反応しているかを見ているのが、その次の層 $U_{S3}$ のS細胞である。この細胞は、前層のS細胞に比して、もう少し大きな受容野を持っていて、大局的な特徴を抽出している。

そのもう一段上には、この大局的な特徴を抽出する細胞の出力を集めて、もっと大きい特徴をみている細胞がある。つまり、この細胞は、大局的な特徴をみる前層の細胞の反応を介して、間接的に入力層全体の情報を捕らえていることになる。

### 2.3 変形に強いパターン認識

C細胞は、いくつかのS細胞から固定結合を受け取っていて、そのうちのどれか一つのS細胞が出力を出せば、C細胞も出力を出す。これらのS細胞はいずれも同じ形の特徴を抽出するが、特徴を抽出してくる場所はS細胞ごとに少しずつ異なる。入力パターンの位置ずれや変形のために、ある特徴の位置が入力層上で移動すると、最初とは異なる別のS細胞がその特徴を捕らえて反応するかもしれないが、これらのS細胞の出力を全部まとめて受け取っているC細胞は、出力を出し続ける。このようにしてC細胞は、S細胞が抽出した特徴の、位置ずれの影響を吸収するのである。

多層回路内にはS細胞の層とC細胞の層とが交互に繰り返して階層的に結合されているので、入力パターンの情報は、多層回路内の各段でS細胞による特徴の抽出とC細胞による特徴の位置ずれの許容化を繰り返しながら上位段に送られていく。この過程で、下位段で抽出された局所的な特徴は、次第に大局的な特徴に組み上げられていく。このとき、入力パターンの拡大や縮小をはじめとする種々の変形にともなう局所的特徴の相対的な位置ずれの影響は、C細胞の働きによって少しずつ吸収されるので、最終的には入力パターンのかなりの変形に対しても影響されない出力を得ることができるのである。

ちなみに、C細胞によるこのような操作は、

CNNでpoolingと呼んでいる操作とほぼ同じである。しかしネオコグニトロンでは、max poolingではなく、平均化操作によってpoolingを行っている。C細胞によるこの操作は、S細胞層の反応を、それに続くC細胞層で空間的にぼかしていると解釈することもできる。ぼかすことによって、変形や位置ずれの影響を受けにくくしているのである。またネオコグニトロンの各段では、演算コストを減らすために細胞の間引き(down sampling)を行っているが、ぼかし操作は空間周波数の高周波成分を除去するlow-pass filterの役割も果たし、粗いdown samplingによって発生するaliasing noiseの抑制にも貢献している。更にまた、平均化操作によって加算的なランダムノイズを軽減する働きも生じる。

### 2.4 S細胞による特徴抽出

S細胞は、複数個のC細胞から興奮性結合を受け取っている(図3)。ここでベクトル表示を用いて、C細胞の出力を $x$ 、興奮性結合の強度を $a$ と表す。興奮性結合 $a$ は、 $a = X/\|X\|$ で与えられる。 $X$ は、S細胞が学習した学習ベクトルである。このS細胞はまた、V細胞と名づけられた細胞を介して強度 $-\theta$  ( $0 \leq \theta < 1$ )の抑制性結合を受け取っている。V細胞は、C細胞の出力の二乗平均値(ノルム) $\|x\|$ を計算している。

これらの入力を受けてS細胞の出力は、

$$u = \|x\| \cdot \frac{\varphi[s - \theta]}{1 - \theta}, \quad s = \frac{(X, x)}{\|X\| \cdot \|x\|} \quad (1)$$

となる。 $\varphi[\ ]$ は、 $\varphi[x] = \max(x, 0)$ で定義されるrectified linear(半波整流型)関数である。多次元特徴空間内で $s$ は、 $X$ と $x$ との内積(ノ

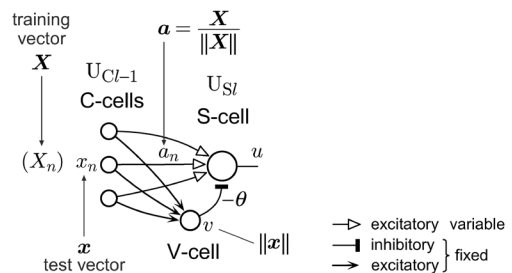


図3 S細胞に収束する結合<sup>3)</sup>。

ルムで正規化された内積)で定義される類似度を表している。類似度 $s$ が $\theta$ よりも大きければ、S細胞は0でない出力を出す<sup>2)</sup>。このように、 $\theta$ はS細胞の閾値を決定する。多次元空間内で $s > \theta$ を満足する領域を、このS細胞の反応の許容領域と呼ぶ。また $X$ を、このS細胞の参照ベクトルと呼ぶ。これは、このS細胞に対する最適特徴(最も強く反応する特徴)を表している。

### 3. Add-if-Silent 則による中間層の学習

学習は、下位の層から順次進めていく。この中間層のS細胞の学習に、AiS則(Add-if-Silent 則)を用いるのである<sup>2,3)</sup>。

もし、前シナプス側のC細胞が出力を出しているにもかかわらず、後シナプス側のすべてのS細胞が無反応であれば、新しいS細胞を作成して回路に追加する。新しく追加したS細胞の入力結合は、前シナプス側の細胞の出力に比例した強度とする。したがって、その時点の前シナプス側細胞の反応に対して選択的に反応するS細胞が作られる。

S細胞がいったん形成されて細胞層に追加された後は、その後どのような学習パターンが与えられても、その入力結合は変化させない。すなわち、AiS則による学習は、多数の学習ベクトルの集合の中から、参照ベクトルを選び出す操作である。学習パターンの集合に含まれる各パターンを1回ずつ提示するだけで学習は完了する。このようにAiS則は、多くの深層学習で用いられている手法とは異なり、繰り返し演算で結合強度を補正する必要がなく、高速に学習を進めることができる。

AiS則のもとでは、すでに存在するS細胞の許容領域内に新しい細胞が形成されることはない。したがって学習時に十分多数の学習ベクトルが提示されたならば、S細胞は、その参照ベクトルがベクトル空間内に一様に分布するように作られる。

このように簡単なAiS則で、なぜ高い認識率が得られるのであろうか。入力パターンを最終的に識別するのは、多層回路の中間層ではなく

最上位層である。中間層は、入力層に提示されたパターンを、単一細胞の反応によってではなく、多数の細胞の反応の集合によって、正確に表現していれば十分である。したがって個々の学習パターンに応じたtuningを必ずしも必要でないのである。

AiSでは、学習時に、一つの特徴に対して1個(あるいはごく少数)の細胞だけしか反応しない状態を作ることによって冗長な細胞の発生を防いでいる。しかし学習終了後の認識時にもこのような状態が保たれていると、回路は変形パターンに対して汎化能力を発揮できない。そこで認識時には、閾値を学習時よりも低い値に設定して、population codingのような状態を作る。入力パターンの状態を、多数の細胞の集合で表現することになるので、入力パターンが多少変形しても、細胞の集合としての反応状態はあり変動せず、変形パターンに対しても頑強な認識が可能になる。

## 4. 最上位層

### 4.1 内挿ベクトル法によるパターン識別

最上位層(すなわち最深層)では、中間層で抽出された特徴をもとに、入力パターンを認識(識別)する。最上位層のS細胞には、その細胞が学習した学習パターンのクラス名を示すラベルが付けられている(学習法は4.2節で後述)。

初期のネオコグニトロンでは、テストベクトル(認識すべきベクトル)に最も近い参照ベクトル(つまり最大出力S細胞)のラベルを認識結果としていた。しかし最近のネオコグニトロンでは、内挿ベクトル法(interpolating-vector)と名づけた手法を用いている<sup>2,4,5)</sup>。内挿ベクトル法では、複数個の細胞の反応を組み合わせることによって認識結果を決定する。図4のように、同一のラベルを持つ参照ベクトルの対を結ぶすべての直線の中で、テストベクトルとの距離が最も近いもの(すなわち類似度が最も大きいもの)を選び、その直線のラベルを認識結果とするのである。その結果、小さな規模の回

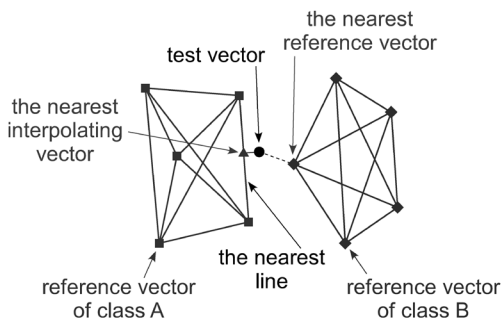


図4 内挿ベクトル法の原理<sup>2)</sup>。

路で、従来よりも高い認識率を得ることができるようになった。

内挿ベクトル法では、ベクトルの対を結ぶ直線の代わりに、三つ組みで形成される平面を用いることもできる<sup>5)</sup>。両者を区別するために、前者をInt-2、後者をInt-3と記す。Int-3のほうがInt-2に比して計算量は増加するが認識率が向上するので、通常Int-3用いている。ちなみに、Int-2でさえも、WTAやSVM (support vector machine)よりも、はるかに高い認識率を達成できる。

#### 4.2 最上位層の学習

参照ベクトルは教師あり学習によって作成するのであるが、内挿ベクトル法での認識に要する計算量は、参照ベクトル数が増えると急激に増加する。従って、できるだけ少ない数の参照ベクトルで高い認識率を得られるような学習法が求められる。ここではmarginized Winner-Take-All (mWTA)と名づけた学習法を紹介しよう<sup>6)</sup>。

中間層の学習と同様に、S細胞は学習時に発生させる。学習ベクトルに対して最大出力を出したS細胞が、学習ベクトルと異なるラベルを持っていれば、識別は間違いであると判定し、新しいS細胞を発生させる。ただし最大出力細胞（すなわち学習ベクトルに最も近い参照ベクトル）の探索において、学習ベクトルとは異なるラベルをもつ参照ベクトル（S細胞）にハンディキャップを持たせて、実際の距離よりも一定割合だけ近くにあると仮定して探索を行うのである。このようなハンディキャップのもとで

も、最も近い参照ベクトルが学習ベクトルと同じラベルを持つ場合にのみ、mWTAによる識別が正しいと定義する。

学習時には、学習ベクトル $x$ が提示されるごとにmWTAで識別してみて、もし識別結果に誤りがあれば、新しいS細胞（参照ベクトル）を発生させ、 $x$ を新しいS細胞の参照ベクトルとする。

中間層とは異なり最上位層では、参照ベクトルのtuningが有効である。tuningでは、mWTAによる識別結果が正しかった場合に、内挿ベクトル法Int-3で選ばれた平面（最も距離の近い平面）を形成する3個の参照ベクトルを、いずれも学習ベクトルに近づける方向に移動させる<sup>5,6)</sup>。

しかし学習を開始したばかりで、まだ十分な数の参照ベクトルが発生していない状態では、たとえある学習ベクトルが正しく認識されたとしても、そのときに使用された参照ベクトルと学習ベクトルの類似度が高いとは、必ずしも言えない。この段階でtuningを行おうとして参照ベクトルに学習ベクトルを加算してしまうと、類似度の低いベクトルが加算され、あまり役に立たない参照ベクトルが作られる恐れがある。そこで、学習過程を2段階に分けて、第1段階ではtuningは行わず、mWTAによってS細胞（参照ベクトル）の発生だけを行ない、学習がある程度進んだ第2段階目になって始めてtuningを開始する。

#### 5. むすび

ネオコグニトロンによるパターン認識の基本的な考え方と、最近の研究状況を紹介したが、ネオコグニトロンの考え方を更に発展させた種々のシステムも開発されている。それらの概要は、例えば文献2などを参照していただきたい。

#### 文 献

- 1) K. Fukushima: Neocognitron: A self-organizing neural network model for a mechanism of

- pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36**, 193–202, 1980.
- 2) K. Fukushima: Artificial vision by multi-layered neural networks: Neocognitron and its advances. *Neural Networks*, **37**, 103–119, 2013.
  - 3) K. Fukushima: Training multi-layered neural network neocognitron. *Neural Networks*, **40**, 18–31, 2013.
  - 4) K. Fukushima: Interpolating vectors for robust pattern recognition. *Neural Networks*, **20**, 904–916, 2007.
  - 5) K. Fukushima and H. Shouno: Deep convolutional network neocognitron: Improved interpolating-vector. *IJCNN 2015*, Killarney, Ireland, 1603–1610, 2015.
  - 6) K. Fukushima: Margined winner-take-all: new learning rule for pattern recognition. *IJCNN 2016*, Vancouver, Canada, 977–984 (N-16157), 2016.