

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks



伊達 章 (Akira DATE Ph. D.)

宮崎大学工学部 准教授

(Associate Professor, Faculty of Engineering,
University of Miyazaki)

電子情報通信学会 日本神経回路学会 日本神経科学学会
研究専門分野: 計算論的神経科学 パターン情報処理

あらまし 画像や音声など高次元データ信号のもっている構造を抽出する手法、そのパターン認識との関係について議論する。パターン認識は、人間には比較的容易にできるのに、コンピュータに行わせようとすると難しい。最近、産業界の利用するパターン認識の技術が、従来手法から Deep Learning を用いた手法に置き換わりつつある。手法は、25 年前に流行したニューラルネットワークを用いたバックプロパゲーションというから驚きである。昔との違いは、まずは学習ということを忘れ、一層ずつ順に、自己組織のモデルを使って回路のパラメータを教師なし学習させればよいという。教師付き学習の前に教師なし学習をするという準備が、どれほどいいのか、これは簡単には答えられないが、自己組織のモデルが重要な役割を果たしているのは間違いない。本稿では、筆者らが研究してきた自己組織のモデルとその応用例を紹介する。特に、Hinton らの制約付きボルツマンマシンの研究に触発され実施した、トポグラフィックマッピング形成についての研究を紹介する。

1. はじめに

画像認識、音声認識など、パターン認識は、人間には比較的容易にできるのに、コンピュータに行わせようとすると難しい。最近、Microsoft や Google をはじめとする産業界の利用するパターン認識の技術が、従来手法から Deep Learning を用いた手法に置き換わりつつある。この手法は、扱うデータ数やモデルの大きさが桁違いに大きくなっているなどの違いはあるが、基本的には約 25 年前に活発に研究された多層の神経回路モデルを使った情報処理である。この学習方法は、「例題からの学習」と呼ばれ、入力データとそれに対応する望ましい答が例題としてあらかじめ与えられている。例題を学習し、出会ったことのないに対しても正しい答えを出力しようというのが問題である。例題は豊富にあるものの、どういう仕組みで識別のアルゴリズムを構築したらよいかわからない問題に対して有効な手法である。しかし、神経回路モデルをはじめ非線形を用いる手法では、局所解が多数あってそこに落ち込むという問題点がある。よほど工夫しないと、大規模な問題に対してはうまく動かない。そこに巧妙な仕掛けをもつサポートベクトル機械 (SVM) が出現したこともあり、神経回路モデルは産業界では使われなくなった。

層状の神経回路モデルが再び注目されているには理由がある。さまざまなパターン認識・識別のコンテストで、Deep Learning を用いた手法が優勝するようになったのである。では、25 年前とはなにが違うのか。10 層の回路を使う場合を考えよう。この 10 層の回路の入力層と出力層に、例題と答を与えて一挙に学習させてもうまく動かない。この原因が、調節可能な回路のパラメータ (素子間の結合係数) の数が膨大であり、多数ある局所解の一つに収束してしまうことであることはまず間違いない。パラメータの初期値が最適解の比較的近くなければ、うまくいかない。ところが、2006 年頃にブレークスルーが起きた [1]。大きな回路を使い一挙に学習するのではなく、とりあえずパターン識別の問題は後回しにして、1 段ずつ順に、教師なし学習でパラメータを調整していくのである (これは

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

pre-training と呼ばれている)。こうして回路を大雑把に作り、そのあと最後に誤差逆伝搬法で学習してみるのである。こうすると、高性能のパターン認識マシンが完成する。どうしてこれで高性能マシンが作成できるのかよくわからないが、事実として、さまざまなベンチマークタスクで従来手法を上回る性能を示している。

本稿では、これまで筆者らが行なってきた神経回路の自己組織化モデル（教師なし学習）の研究を紹介する。特に、Deep Learning の発端となった Hinton らの研究に触発され、実施した研究を紹介する。これは、倉田耕治博士が開発したボルツマンマシンを用いたトポグラフィックマッピング形成のモデルについての研究であり、やはり、約 25 年前に提案されたものである[2]。当時は、学習のダイナミクスを追いかけるコンピュータシミュレーションができなかったが、最近、学習のダイナミクスを追いかけるができるようになった。以下では、まずはじめに、自己組織神経回路モデルを使う目的、ボルツマンマシンとは何か、どういう発展があったか、このあたりについて、できるだけ数式を使わない説明を試みる。

2. 高次元空間に埋め込まれた低次元構造の抽出

パターン認識は、人にとっては無意識のうちに行ってしまうため、問題の難しさが一般人には理解しにくい。なぜそこに書いてある文字が読めるのか、と問われても困惑するだけかもしれない。たとえば、手書きで「2」と書いてみよう。これを白黒の 100×100 ピクセルの画像として読み込んだとする。これは数字が 1 万個ならんだ 1 万次元ベクトルである。出現する可能性のある、ありとあらゆる画像の種類は、 $2^{10000} = 2^{10^{1000}} \approx 10^{3^{1000}} = 10^{3000}$ 通りある。何度も手書きで「2」を書いてみよう（大きさは正規化して 100×100 にしておく）。1000 枚くらい作ったとして、1 万次元の空間に、これらの 1000 枚がどのような位置関係に分布しているか想像してみよう。数字「4」の分布との関係はどうか。文字に限らず、100×100 ピクセルの風景の画像パッチを大量に集め、その分布を調べてみよう。

白黒白黒... という画像などは、自然界には存在しないため、まったく出現しない。そんな画像は多い。つまり、画像や音声データは高次元の空間に分布しているが、分布には偏りがある。高次元の中に埋め込まれた低次元の部分空間と言っていいだろう。データの分布に沿った良い座標系を取ることで、情報をあまり失わずに観測データを低次元の空間に表現することができる。この高次元に埋め込まれている構造を、人間は意識せずに取り出せるようになってきている。この入力信号のもっている構造を取り出す方法が、1960 年代から活発に研究されてきた[3][4][5]。高次元空間の超平面上、またはそれに近い位置にデータが（まっすぐに）分布している場合は、主成分分析（PCA; Principal Component Analysis）が有力な方法である。ただし、まっすぐか曲がっているか、通常は前もって知ることができないのが問題である。PCA は線形の方法としての限界があり、そこに神経回路モデルを用いる非線形の方法の存在価値がある。

高次元観測データの例を 2 つ紹介しよう。部屋の中をロボットが動いている。ロボットの前面（顔）にはカメラをつけておく。大量の観測データが得るのは簡単である（図 1）。観測される画像には、どんな規則性があるだろうか。5 枚の画像（図 1）は観測画像の例である。ロボットが平面上を動いているとして、その位置を (u, v) とする。また、ロボットの顔が向いている方向を θ としよう。そうすると、観測画像 I を $I = I(u, v, \theta)$ と見ることはできないだろうか。ロボットが前進すると観測データも変化する。つまり、高次元空間の中に (u, v, θ) を座標系とする曲がった座標があると見るのである。

この問題に対し、筆者らは、単純な自己組織化モデルを組み合わせ、情報を分離抽出する回路を開発した[6]。図 2 に示しているのは、ロボットが見ている画像から、ロボットの位置や向いている方向を抽出する神経回路モデルである。これは、Kohonen の自己組織化マップ (SOM) とニューラルガスとよばれるモデルを組み合わせで実現した。具体的には、2 次元の位置情報を表現する部分には 2 次元の SOM（図 2 では r_1, r_2

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

軸)、それと直交する方向に 1 次元の向きを表現するための素子コラム (r_3) を配置した。SOM とニューラルガスの直積モデルでは素子面内では順序学習、素子コラム内では近傍学習とよばれる学習を適用することで、分離抽出を実現した。

顔画像に含まれる構造を抽出する例についても紹介

しておこう。人が他人の顔を認識するとき、主として二つの情報が重要である。一つはいうまでもなく、その顔の主は誰かということであり、もう一つは、その顔の伝える感情、つまり表情である。顔画像自体は高次元の情報であるが、これを人物と表情の組合せと考えるならば、2次元の構造が隠れていることになる。

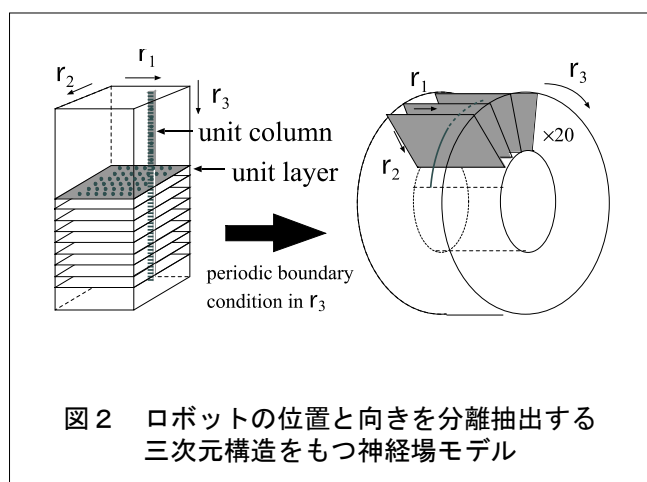
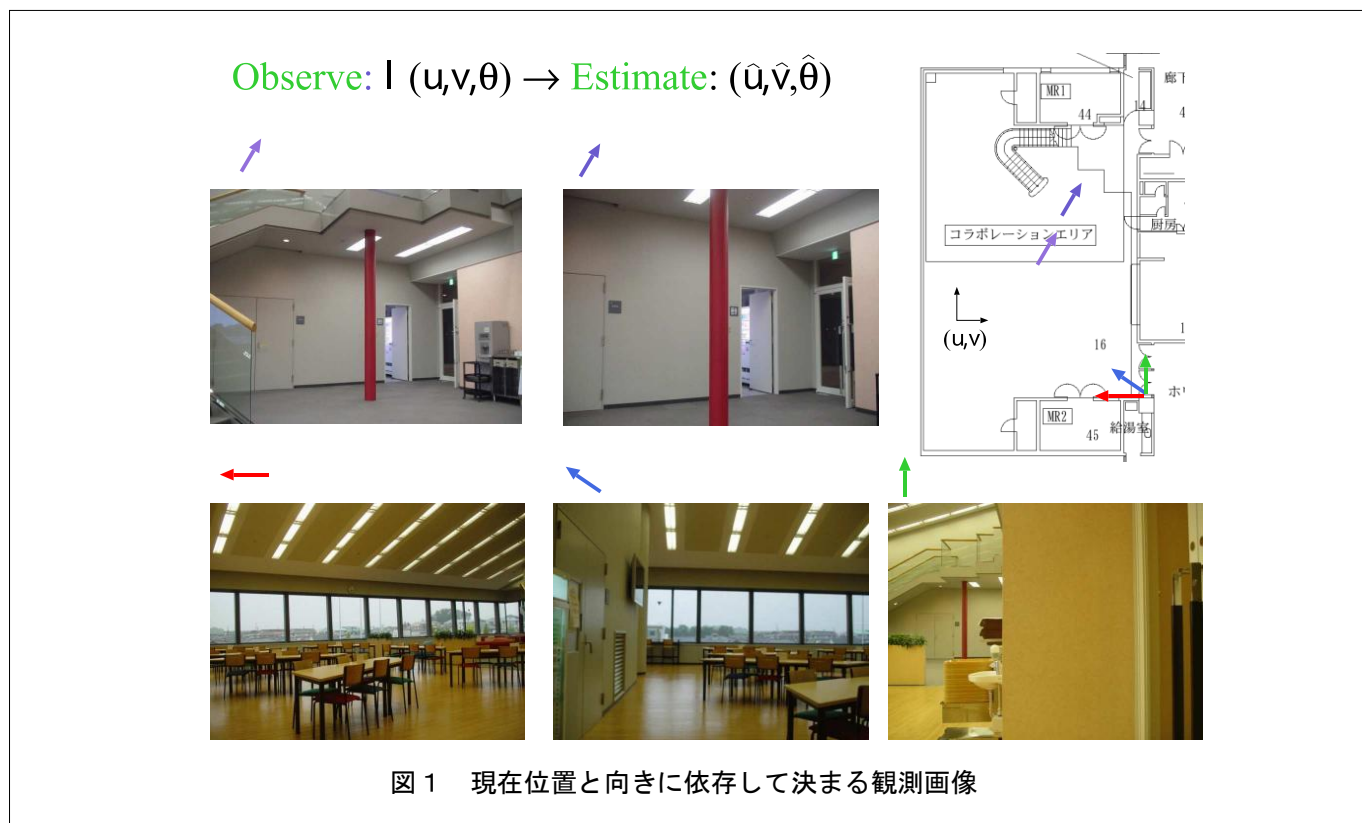


図 3 は顔表情画像データベース (JAFPE) [7]に含まれている日本人女性の顔画像の一部である。このデータベースには、10人の女性が7種類の表情をしたときの顔画像が含まれている。このうち、4人の女性の4表情、16枚を例示した。それぞれの顔画像は、 $64 \times 64 = 4096$ 次元のベクトルである。この場合も、4096次元の空間に各画像が分布している。Aさんの顔画像とBさんの顔画像が大量に得られたとして、Aさんの画像の分布とBさんの画像の分布とは、どのような関係にあるだろうか、「人物」と「表情の種類」を決めれば画像が決まると思えば、高次元空間に「人物」と「表情」という曲がった「軸」を想像することは難しくな

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

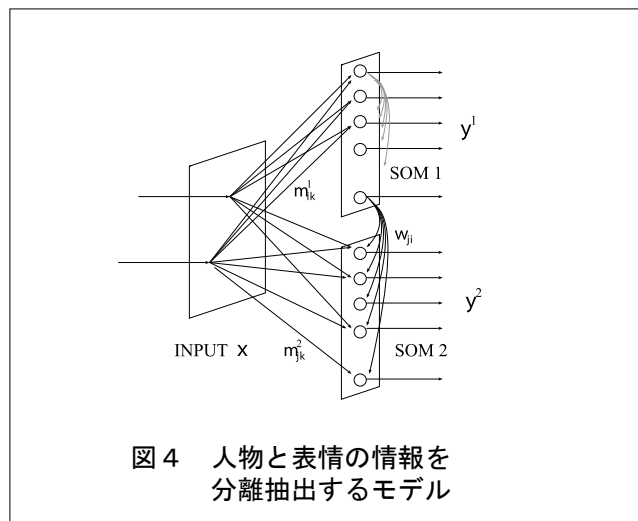
い。われわれは、笑っていても怒っていてもそれが誰の顔か分かるし、まったく別の人物、例えば、はじめて見る人間でも、笑った顔は同じ感情を表現していると理解できるのであるから、脳内では人物と表情の二つの情報は分離して抽出されていると考えられる。



この問題に対して、筆者らは自己組織化マップ (SOM) を組合せて使う神経回路モデルを開発し、画像に含まれる信号の構造を利用して、個人と表情という二つの情報を分離して抽出することを試みた[8]。具体的には、1次元の素子の配列からなる層を2枚用意したモデルを用いた (図4)。入力信号と各層の素子とは、Hebb型とよばれる通常の自己組織の学習をおこなうが、各層の素子間には反 Hebb型 (互いに抑制) の学習を導入し、第1層と第2層の情報表現が相関をもたないようにすることで実現した。

3. 神経回路の自己組織化モデル

教師信号なしに入力信号に含まれている構造を取り出すモデルは、自己組織化モデルとよばれている。前節で紹介したように、神経回路モデルはこの問題と相性が良い。ここでは、次節の準備のため、古典的なト



ポグラフィックマッピングの形成という問題を簡単に説明しておこう[2][5][9][10][11]。トポグラフィックマッピングの自己組織モデルは、網膜と大脳皮質領の間にみられる連続的な対応関係や、一次視覚野の細胞の方向選択性の連続的な変化など、脳の中で近くの細胞同士が同じような機能をもつように配列している現象を、自己組織による結合形成により説明しようとするものである。モデルにはいくつかの種類があるが、次のような共通した特徴がある。

- 1) 出力層では近くの細胞同士は互いに興奮性の結合で結ばれ、少し離れた細胞同士は逆に抑制性の結合で結ばれている。このため、一つの細胞が発火すると、その周囲のある範囲の細胞がまとまって発火する。
- 2) 入力層と出力層の間の結合は、Hebb則 (同時に興奮した細胞間の結合効率がより大きくなる) により形成される。

Hintonらによって提案されたボルツマンマシンも、自己組織のモデルとみることができるので説明しておこう[12]。ボルツマンマシンは、神経興奮のダイナミクスと学習のダイナミクスの関わり合いが、単純な方程式で表される理論的に非常に見通しがよい神経回路モデルである。各素子を自由に動作させたとき入力層に現れるパターンの確率分布を、外部から与えられる入力信号の確率分布に近づけていく学習則が知られている。倉田は、このボルツマンマシンを使ったトポグ

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

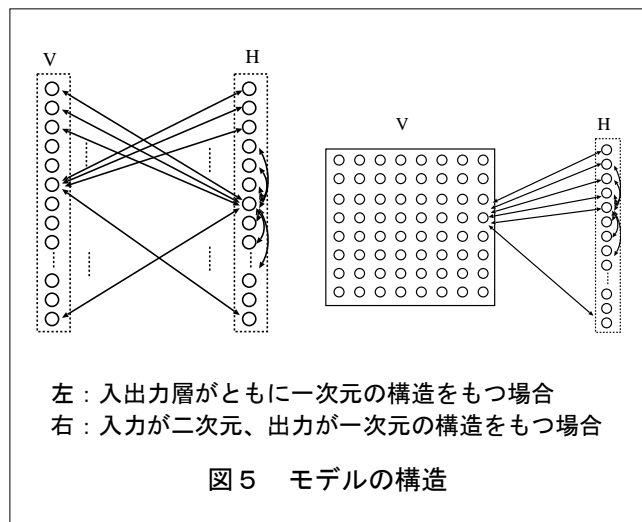
ラフィックマッピングのモデルを約 25 年前に提案した[2]。個々の細胞の興奮のダイナミクスと学習（結合係数）のダイナミクスの関わり合いの解析が、簡潔な形で表されるのがボルツマンマシンの特徴であるが、学習法をそのまま計算機で実行しようとする、膨大な時間がかかる（現実的には動かない）という欠点がある[13]。

4. ボルツマンマシン神経場

ここでは、2層構造をもつボルツマンマシン、特に、結合が層間を結ぶ素子間にしか存在しない場合を考える。回路の状態を $\mathbf{x} = (x_0, \mathbf{x}_v, \mathbf{x}_h)$ と書く。このモデルは、制約付きボルツマンマシン（Restricted Boltzmann Machine, RBM）と呼ばれている。ある入力 x_v が与えられた場合の H 層の状態の確率分布 $p(\mathbf{x}_h | \mathbf{x}_v)$ を考えてみよう。H 層内において素子間の結合がない場合は、入力 x_v が与えられた条件のもとでは、H 層の各素子の発火確率は独立であるので、 $p(\mathbf{x}_h | \mathbf{x}_v) = \prod_{n_1+1}^{n_1+n_2} p(x_i | \mathbf{x}_v)$ と書ける (n_1, n_2 はそれぞれ V 層、H 層の素子の数)。従って、単に H 層の各素子の状態を更新するだけで、確率分布 $p(\mathbf{x}_h | \mathbf{x}_v)$ にしたがう状態を一つ確率的に選ぶことが容易にできる。多層神経回路を用いた教師あり学習をおこなう際、結合係数の初期値をあらかじめこの方法で学習して求めた値を使うことで、高い性能をもつ回路を実現できることが知られている[1]。

筆者らが研究しているボルツマンマシン神経場を紹介しよう[2][13]。このモデルは、やはり V 層と H 層の 2 層からなる (図 5)。V 層内の素子どうしに結合がないことは、制約付きボルツマンマシンと同じであるが、H 層内の素子間には、隣り合う連続する数個 (k_2 個とする) の素子のみが興奮するような側抑制型の内部結合がある。また、簡単のため、V 層、H 層ともに円環の構造をもたせている。入力信号としては、隣り合った k_1 個の素子が発火するパターンが与えられる。このようなパターンは n_1 通りあるが、そのうちのどれかが等確率で選ばれる。学習は層間の素子のみでおこり、H 層内の側抑制型の結合は変化しない。このモデルが

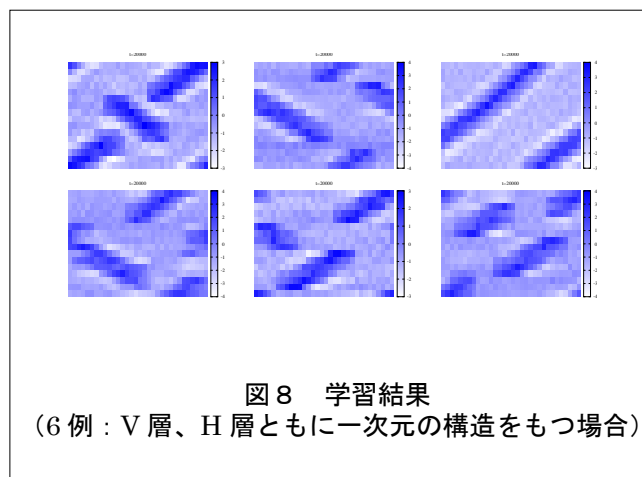
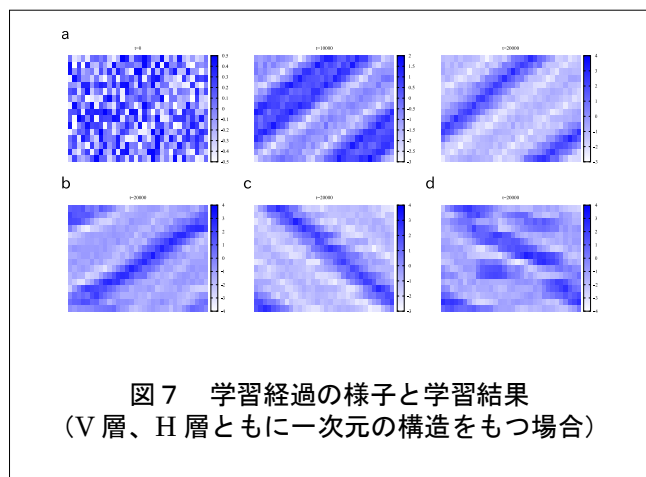
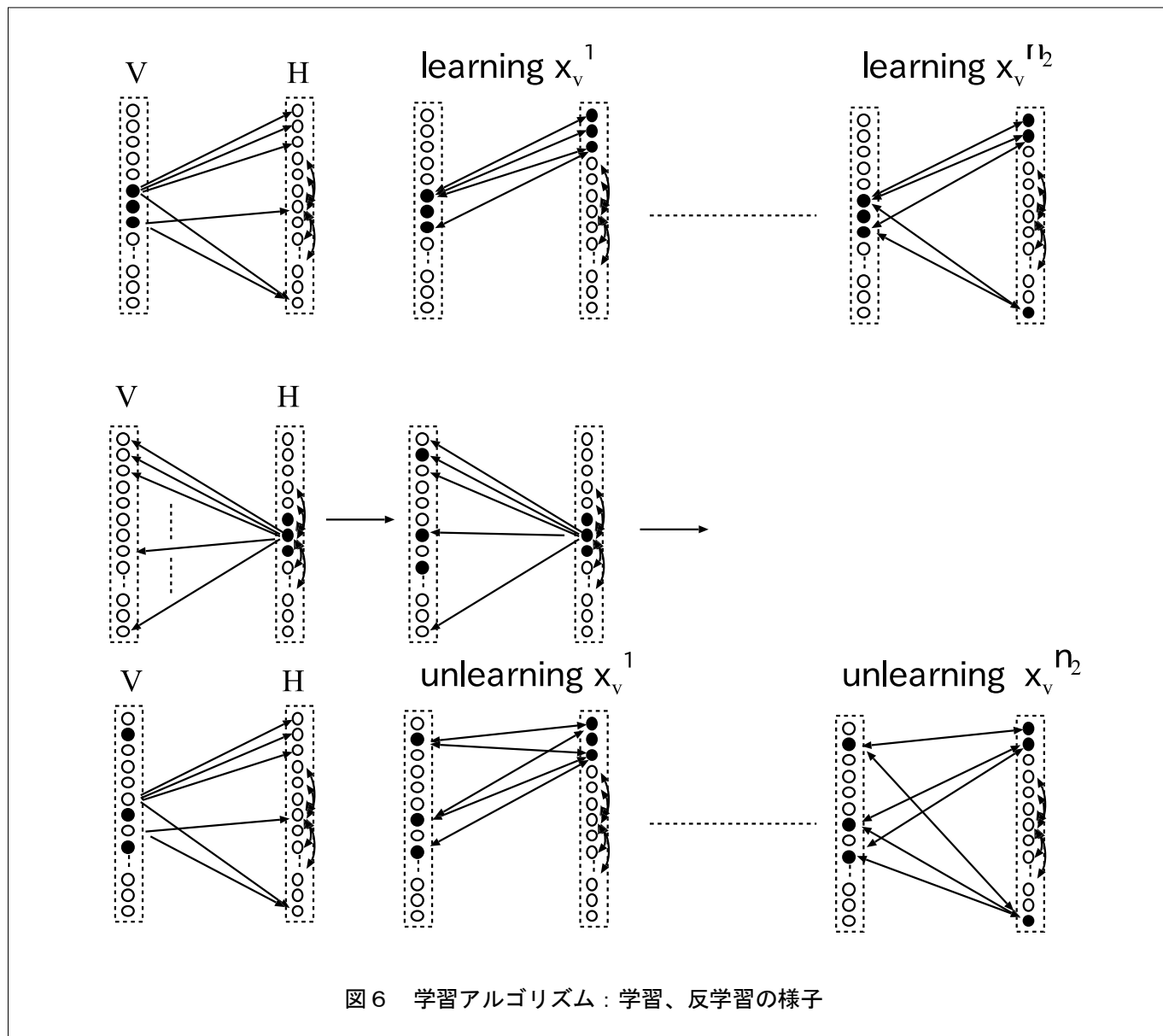
これまでの自己組織のモデルと大きく違う点は、反学習を行う点と二つの層の間で結合が双方向的である点である (学習アルゴリズムの様子は図 6、詳細は文献[13])。



V 層、H 層ともに 1 次元の構造をもつ場合について、コンピュータシミュレーションを行った結果を紹介しよう ($n_1 = 32, n_2 = 16$, 図 7, 8)。入力信号には、隣接する 6 個の素子が同時に発火するパターンを与え ($k_1 = 6$)、H 層には連続する 7 個 ($k_2 = 7, t = 10000$ 以降は 3 個) の素子が同時に興奮するような構造をもたせた。層間の結合係数 w_{ij} の初期値には、一様分布 $U[-0.5, 0.5]$ に従う乱数を与えた。図 7 は、入出力間に連続写像が形成された学習過程の例である。横軸、縦軸は、それぞれ V 層、H 層の素子番号、結合係数 w_{ij} の値が濃淡値で示されている (w_{ij} の値が大きいほど色が濃い)。学習前の状態 ($t = 0$, 図 7 左上) と学習経過の様子 ($t = 10000, 20000$, 図 7 上段中および右) を示す。同条件で異なる乱数列を用いて実験をおこなった結果 (3 例, $t = 20000$, 図 7 下段)、同じ条件でも使用する乱数列が異なると、連続写像が形成されない場合がある。この場合でも、局所的には連続写像が形成されていることがわかる。H 層で発火する素子の数を少なくすると ($k_2 = 4, t = 10000$ 以降は 3 個)、大域的な連続写像が形成されない場合がおこる。図 8 には $t = 20000$ 時点での学習結果 6 例を示している。

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

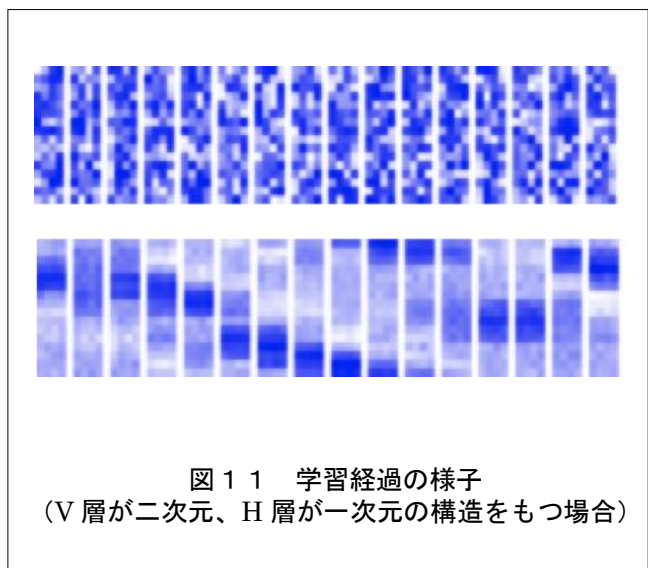
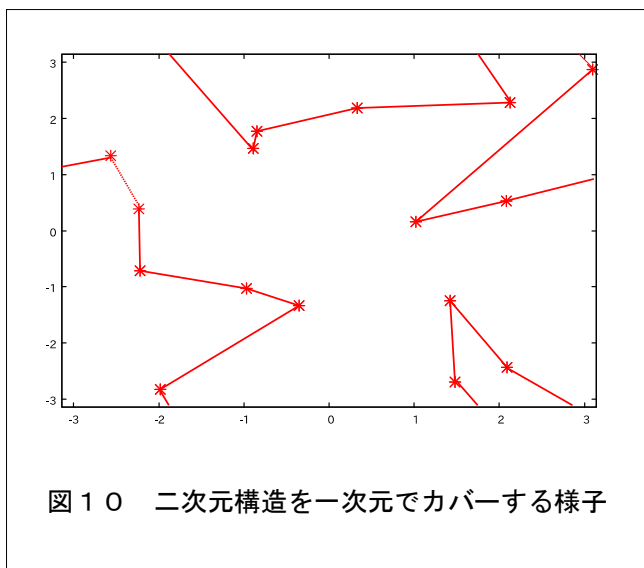
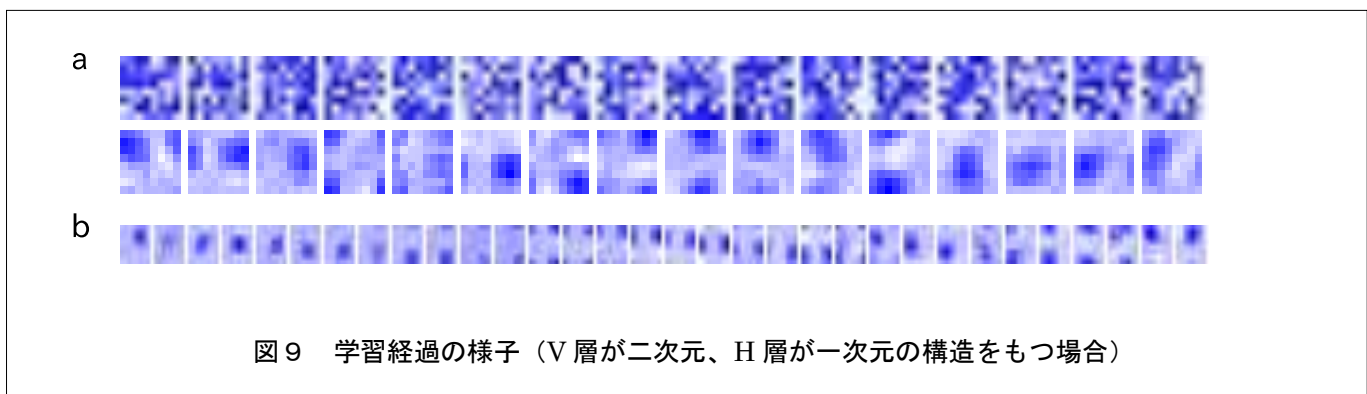


自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

V層が8×8の2次元、H層が1次元の構造をもつ場合 ($n_1=64, n_2=16$) について、コンピュータシミュレーションをおこなった結果を紹介しよう (図9)。出力層の回路の構造、学習アルゴリズムは、入力が1次元の場合と同じである。入力として、V層内で隣り合った3×3個の素子が発火するパターンを等しい確率で与えた。また、H層では連続する5個 ($k_2=5, t=10000$ 以降は2個) の素子が同時に興奮するような構造をもたせた。入出力間に連続写像が形成された例を図9に示す。H層の各素子がV層の各素子に対してもつ強度を濃淡値で表現している。上から順に学習前の状態 ($t=0$)、学習後 ($t=20000$) を示す。同条件でH層の素子数を2倍にした場合の学習結果 ($n_2=32, t=20000$) を図9最下段に示した。

また、得られた各素子をもつ結合荷重の重心を計算し、入力信号空間中に表示した結果を図10に示す。16個のH層の各素子が入力信号空間にもつ重心をプロットし、H層の隣接関係に応じて線で結んでおり、1番目の素子と16番目の素子も点線で結んでいる。2次元の構造を1次元でカバーしている様子がわかる。最後に、V層の構造を正方形ではなく細長い長方形 (4×16 、素子数 $n_1=16$ は先の図9の実験と同じ点に注意) に変えて実験をおこなった例を図11に示す。図は上から順に学習前の状態 ($t=0$)、学習後 ($t=20000$) の状態を示す。この場合も、入出力間に連続写像が形成されていることがわかる。



自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

5. おわりに

神経回路の自己組織化モデルが、非線形の主成分分析器として使え、高次元空間に埋め込まれた低次元構造の抽出という問題に応用できる事を紹介した。自己組織化モデルは、甘利のモデル、Kohonen の SOM, Hinton のボルツマンマシンなど、いくつかの古典的なモデルが存在する。それぞれのモデルは単純であるが、その相違についてよく分かっていないことが多い。今後この点について、より詳しく研究を進め、共通点と違いを整理したい。

機械によるパターン認識は、その性能が向上したとはいえ、まだまだである。たとえば、Google Map のストリートビューでは、人物の顔やナンバープレートにはプライバシー保護のためのぼかしがかけられている。時には、車のタイヤとホイール部分が顔と誤認識され、ぼかしがかけられている場合が少なくない。車の「タイヤ」を「人の顔」と誤認識する人間はまずいない。小さな子どもでも間違えないことを間違えてしまう。与えられた画像を様々な大きさの画像で切り取り、その局所的な部分の「顔らしさ」を計算、その値があらかじめ設定したしきい値を超えると、顔だと判定しているのだろう。タイヤを顔と誤認識させないためには、このしきい値を大きくする必要があるのだが、そうすると、今度は本当の顔が顔と認識されなくなる。悲しい話ではあるが、これは機械と人間の ROC gap と呼ばれており[14]、人間のパターン認識能力は、よくわからないけどすばらしいと言える。これは、車本体とタイヤの関係、顔は胴体の上にあるなど、物事の関係が表現できていないことに問題がある。Deep Learning で学習させると、どの辺まで自動で抽出できるのだろうか。人のパターン認識は仕組みが複雑で、その解明は一筋縄でゆくものではないが、この辺が解決の鍵だと考えている。

参考文献

- [1] G.E. Hinton and R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, pp.504-507, 2006.
- [2] 倉田耕治, "ボルツマン・マシンを応用したトポグラフィック・マッピングの形成モデル," 電子通信学会技術研究報告, MBE87-154, 1988.
- [3] J.W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, vol.C-18, pp.401-409, 1969.
- [4] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, Inc., New York, 1973.
- [5] T. Kohonen, "Self-organized formation of topologically correct feature map," *Biological Cybernetics*, vol.43, pp.59-69, 1982.
- [6] 伊達章, 倉田耕治, "SOM とニューラルガスの直積モデルによるロボットの位置と向き情報の分離," 電子情報通信学会論文誌, pp.1529-1538, 2004.
- [7] M.J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," *Proc. of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp.200-205, 1998.
- [8] 伊達章, 倉田耕治, "連続自己組織化マップによる表情と個人情報の分離, 画像の認識・理解シンポジウム (MIRU2004) 論文集 I, pp.613-617, 2004.
- [9] C. von der Malsburg, "Self-organization of orientation sensitive cells in the striate cortex," *Kybernetik*, vol.14, pp.85-100, 1973.
- [10] D.J. Willshaw and C. von der Malsburg, "How patterned neural connections can be set up by self-organization," *Proceedings of the Royal Society of London Series - B: Biological Sciences*, vol.194, pp.431-445, 1976.
- [11] 甘利俊一, 神経回路網の数理, 産業図書, 東京, 1978.

自己組織神経回路モデルによる情報表現の獲得

Extraction of Information Representation by Self-organizing Neural Networks

- [12] D.H. Ackley, G.E. Hinton, and T.J. Sejnowski, "A learning algorithm for {Boltzmann} machines," *Cognitive Science*, vol.9, pp.147-169, 1985.
- [13] 伊達章, 倉田耕治, "ボルツマンマシンを応用したトポグラフィックマッピングの形成モデルについて," *電子通信学会技術研究報告*, NC2012-169, 2013.
- [14] Y. Jin and S. Geman, "Context and hierarchy in a probabilistic image model," *Proceeding of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.2145-2152, 2006.

この研究は、平成20年度SCAT研究助成の対象として採用され、平成21～23年度に実施されたものです。