

## 階層化された素子群の自己組織化

別府智彦

佐瀬幹哉

小杉幸夫

東京工業大学総合理工学研究科

あらまし 近年盛んに研究の行われているニューラルネットの中でも自己組織特徴マップは入力情報のトポロジカルな関係を保存した写像を教師信号等を必要とせず自己組織的に獲得するという点で魅力的なモデルである。本研究では特徴マップの学習サイクルを高速に行うことを目的とし、生体内の神経系や社会組織との類推から、特徴マップ上のニューラルユニットの機能を階層化し、小数の上位ユニットによって残りの下位ユニットを制御する方法が学習サイクルの高速化に有効であることを示す。同時に獲得された写像関係の良否を評価するため写像の歪みに注目した評価基準を提案し、それについても論ずる。

### Self-Organizing Feature Map Using Classified Neural Units

Tomohiko BEPPU, Mikiya SASE and Yukio KOSUGI

The Graduate School at Nagatsuta, Tokyo Institute of Technology

4259 Nagatsuta, Midori-ku, Yokohama 227, Japan

Abstract We propose a new architecture for Kohonen type Self-Organizing Feature Map. The network consists of a few "Upper Class Units" and the rest, called "Lower Class Units". The Upper Class Units act as controller of the others (like opinion leaders in society). With this architecture, we can get a topographic preserving maps as well as conventional one and reduce the load in learning process, resulting faster convergence. Quality measures for evaluating the topographic preserving maps are also proposed.

## 1. はじめに

Kohonenらの提案する自己組織特徴マップ[1]は、大脳に於ける神経活動の生理学的な観察に基づいた単純な写像学習のモデルとして興味深いものである。その学習は教師信号等を必要とせずに行われ、多次元の入力情報の位相関係を保存するように二次元のニューラルネット上で自己組織的に行われると言う点でたいへん魅力的である。このような特徴を活かしロボット・マニピュレータの視覚-運動制御[2]、画像信号の圧縮[3]、不規則信号の処理[4]等の応用が考えられている。しかし、ここでも他のニューラルネット同様、実用化に際しては多数のユニットを並列的に駆動しなければならないため高速に動作することが求められる。

一方、多数の動作単位を並列的に駆動して問題を解決するという方法は社会活動や生物の神経系の活動に於いてもしばしば見られる現象でもある。このように多数の構成要素をうまく並列的に作動させ、目的にかなった動作をさせようとするとき、それら全ての影響力を等しくあつかうよりも、それらの中の一部（オピニオンリーダー）によって残りの多数を指導するようにした方が効率が良いであろうことは容易に想像できる。

たとえばC. W. ミルズは彼の著書“パワーエリート”において、アメリカの重要な政策決定が比較的小数のパワーエリートに支配されると主張し、また日本に於いてもある一定数のリーダー、エリートによって社会が運営されていることは様々に指摘されている。[5]

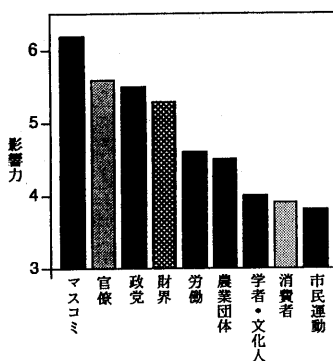


Fig.1 日本における影響力の階層[5]

生理学的にみても、たとえば小脳皮質に於いて、バスケット細胞に集まる神経の数に比べプルキンエ細胞に集まる神経はおおよそ 1/100 であり、逆にプルキンエ細胞から伸びる神経の数はバスケット細胞か

らのそれに比べおおよそ 1.6 倍である。[6]

本研究ではこうした考えを多数の素子の集合である自己組織特徴マップにあてはめることで、高速な学習が可能となることを示す。

## 2. 特徴マップの構成と動作

特徴マップはN次元のベクトルとして与えられる入力データXに対しN個のユニットからなる入力層と2次元的に配置された十分に大きいマップ層から構成される。マップ層上のユニットはそれぞれがN次元の荷重ベクトルmを持ち、それによって入力層の全てのユニットに接続されている。

その動作は以下のとおりである。

### 【初期値設定】

各々のユニットが持つ荷重ベクトルmの初期値を設定する。この設定には乱数を用いその値が適当に分散するようにする。

### 【入力パターン提示】

入力ユニットに何らかの入力データXを与える。

### 【発火ユニットの探索】

発火するユニットを“全てのユニット”の中から求める。

入力データXとユニットの荷重ベクトルmとのユークリッド距離を式(1)によって求め、この式によって求めた距離Dが最も小さい値をもつユニットCを発火するユニットとする。

$$D = \|\bar{X}(t) - \bar{m}_{i,j}(t)\| \quad (1)$$

$$\|\bar{X}(t) - \bar{m}_c(t)\| = \min_{i,j} \{\|\bar{X}(t) - \bar{m}_{i,j}(t)\|\} \quad (2)$$

### 【荷重ベクトルmの更新】

ここでは発火したユニットCの近傍 $N_c$ にある全てのユニットが持つ荷重ベクトルを同時に更新する。更新は次の式に従い、それぞれのユニットの持つ荷重ベクトルが入力ベクトルに近づくように行われる。

$$\left\{ \begin{array}{l} \text{for } i \in N_c \\ \bar{m}_{i,j}(t+1) = \bar{m}_{i,j}(t) + \alpha(t) [\bar{X}(t) - \bar{m}_{i,j}(t)] \\ \text{otherwise} \\ \bar{m}_{i,j}(t+1) = \bar{m}_{i,j}(t) \end{array} \right\} \quad (3)$$

**【ユニットの近傍の修正】**

ユニットの近傍  $N_c$  の範囲を狭める。

**【入力パターンの提示】**

さらに新しいデータを与え学習を進める。

こうした学習サイクルは特別な教師信号を必要とせずに行われ、その結果二次元ニューラルネット上の各々のユニットが持つ荷重ベクトルは互いに入力信号のトポジカルな特徴を保存した写像関係を形成するように変更される。

**3. 機能階層化特徴マップの構成と動作**

新しく提案する機能階層化特徴マップの構成と動作を図(2),(3)に示す。

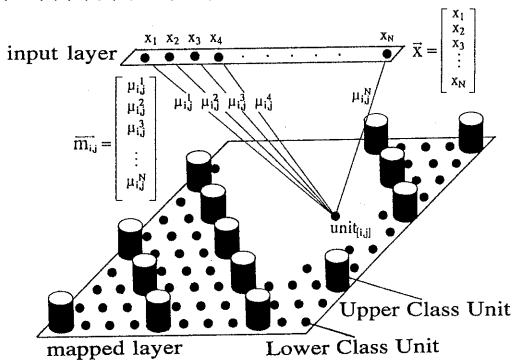


Fig.2 Construction of the Classified Feature Map

機能階層化特徴マップも特徴マップと同様に入力層とマップ層によって構成されるが、マップ層は”上位ユニット”と”下位ユニット”の二種類に機能の階層化されたユニットによって構成される。このとき二種類のユニットはマップ層上に一様に配置される。

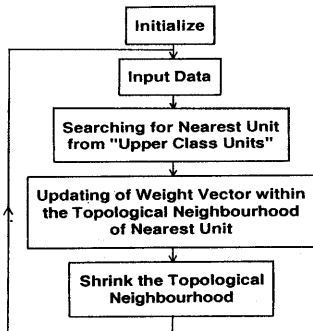


Fig.3 Operation of the Classified Feature Map

ユニットの機能の階層化は前述の特徴マップの学習動作に対し、発火するユニットの探索動作を以下のように変更する事によって行う。

式(1)を用い、入力データ  $X$  と上位ユニットの持つ荷重ベクトル  $m$  とのユークリッド距離を求め、その距離  $D$  がもっとも小さい値を持つ上位ユニットを求め、それを発火するユニットとする。ここで行われる”発火するユニットの探索”では下位ユニットの荷重ベクトルは考慮されない。

この変更により、発火するユニットの探索は上位ユニットの中からのみ行えば良くなるため、従来のように全てのユニットの中から探索する場合に比べ探索すべきユニットの数を減らすことができ、探索に必要な時間を非常に短くすることができる。また荷重ベクトルの変更、つまり学習は従来の特徴マップと同様に発火したユニットの近傍にある全てのユニットに対して、その機能の階層に関わらず行われるため、特徴マップ上に獲得される写像は従来からの方法と同等なものを得ることができる。

この変更によって得られる学習の高速性は、繰り返しによって累積的に働くため、さまざまな環境に対して多数回の学習によって適応することを求められる場合にとくに有利になると思われる。

**4. シミュレーション結果**

入力層に3ユニット、マップ層に  $41 \times 41$  ユニットの持つニューラルネットを用い実験を行った。マップ層上の上位ユニットの構成比は表(1)に示すように 8.6% から 100% までの4種類のものに対し実験を行った。

	Population of Upper Class Unit	Relative Population of Upper Class Unit (%)	Population of Lower Class Unit	Relative Population of Lower Class Unit (%)
Conventional Method	1681	100.0	0	0.0
Example 1	841	50.0	840	50.0
Example 2	394	23.4	1287	76.6
Example 3	145	8.6	1536	81.4

Table 1 Population of  $41 \times 41$  Neural Units

入力データとして式(4)のような3次元の空間座標を与えた。これらは3次元空間内の2次元平面を現しており、3次元ベクトルではあるが2次元としての性質を持つため、2次元の特徴マップニューラルネット上にも十分に写影されると思われる。

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \text{random value (0~1)} \\ \text{random value (0~1)} \\ X \end{bmatrix} \quad (4)$$

実験によって得られた処理速度向上の結果を図(4)に示す。先に示したように一回の学習サイクルは発火するユニットの探索と、その近傍のユニットが持つ荷重ベクトルの更新の二つに大きく分けられる。学習の初期にはユニットの近傍の範囲が大きいいため荷重ベクトルの更新を要するユニット(下位のユニットを含む)の数が多し。そのため、学習に要する時間全体の中で、更新に要する時間の割合が大きく、探索すべきユニットが減少したことによる学習速度の向上は余りはっきりと現れない。学習が進むにつれユニットの近傍範囲が狭まり、一回の学習サイクルに於ける探索時間の比率が増すため、機能の階層化による探索ユニット数の減少の効果が顕著に現れ、学習が高速に行われるようになる。

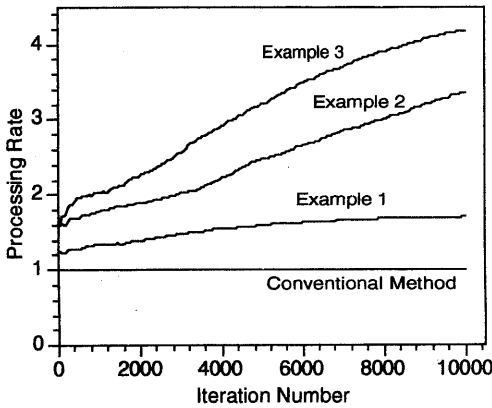


Fig.4 Processing Rate v.s. Iteration Number

またこの学習によって得られる写像と従来のマップによって得られる写像を図(5),(6)に示すが、機能階層化特徴マップによって得られる写像の品質も上位ユニットの割合が適当であれば従来のものと同等であることがわかる。上位ユニットの割合が極端に少ない場合、その学習はより高速に行われるが、図(7)(上位ユニットの割合が0.77%のもの)のように適当な写像関係を獲得できなくなる。

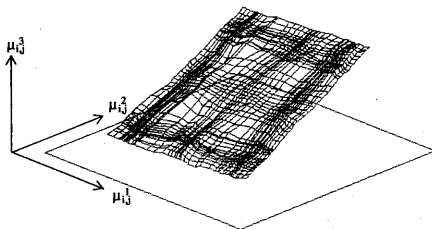


Fig.5 Distribution of Weight Vector (Conventional type)

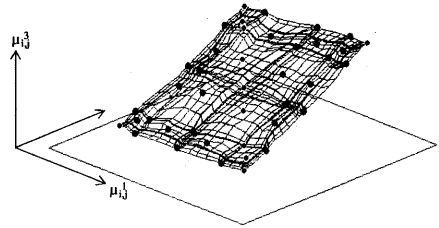


Fig.6 Distribution of Weight Vector (Example 3)

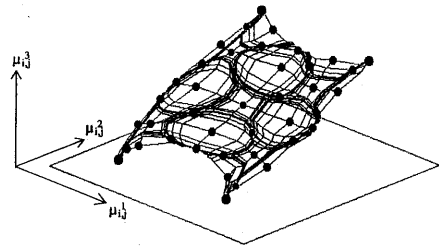


Fig.7 Distribution of Weight Vector (with less Upper Class Units)

## 5. 写像関係の評価基準

先に述べたように機能階層化特徴マップは従来の特徴マップに比べ非常に高速に動作させることができる。ここでは機能階層化特徴マップが高速な動作だけでなくそのマップ上につくる写像関係の品質の上でも従来の特徴マップと同等である事を示す。

このような入力情報のトポグラフィックな特徴を保存するニューラルネットの学習を評価する基準としてこれまでに提案されたものとしては、Kohonenらの示した"Index of Disorder"[7]、Tilleryらの示した"Relative Rotation of Vector"[8]等がある。特に"Relative Rotation of Vector"は入力情報のアプリオリな知識に依存しないという点で魅力的であるが、後述のように実際にここで評価基準として使用するには有効なものではない。

新しく提案する評価基準としては本研究でのこの評価基準の使用目的が一様に分布した入力空間を学習したニューラルネット上に獲得される写像関係を評価する事である事を考え、学習によって得られる写像の歪みに注目したものである。

### 距離の歪み (Distance Distortion)

マップ層上の任意の隣あった一組のユニットが持つ各々の荷重ベクトルの間のユークリッド距離を式(5)によって求め、これをユニット間の距離とする。

このユニット間の距離は一樣に分布した入力を十分に学習した後は隣接領域については等しくなる事が期待される。そのためこの変化分を式(6)、(7)に従ってすべてのユニットに対し加算、平均したものを距離の歪みとする。

$$\begin{cases} d_i = |\bar{m}_{i,j} - \bar{m}_{i+1,j}| \\ d_j = |\bar{m}_{i,j} - \bar{m}_{i,j+1}| \end{cases} \quad (5)$$

$$\begin{cases} \xi_{d,i} = \sum_j \sum_l (d_i - d_{i+1}) \\ \xi_{d,j} = \sum_l \sum_j (d_j - d_{j+1}) \end{cases} \quad (6)$$

$$\xi_d = \sqrt{\xi_{d,i}^2 + \xi_{d,j}^2} \quad (7)$$

### 角度の歪み (Angular Distortion)

マップ層上の任意のユニットに対しそれぞれ行方向、列方向に隣あった二つのユニットの三つのユニットが持つ荷重ベクトルがなす角を式(8)によって求め、これをユニットの角度とする。

このユニットの角度もさきに述べたユニットの距離と同様に一樣に分布した入力を学習した後は等しくなる事が期待される。そこでこの変化分を式(9)、(10)に従って加算、平均したものを距離の歪みとする。

$$\theta_{i,j} = \text{Cos}^{-1} \frac{|\bar{m}_{i+1,j} - \bar{m}_{i,j}| |\bar{m}_{i,j+1} - \bar{m}_{i,j}|}{\langle \langle \bar{m}_{i+1,j} - \bar{m}_{i,j}, \bar{m}_{i,j+1} - \bar{m}_{i,j} \rangle \rangle} \quad (8)$$

$$\begin{cases} \xi_{a,i} = \sum_j \sum_l (\theta_{i,j} - \theta_{i+1,j}) \\ \xi_{a,j} = \sum_l \sum_j (\theta_{i,j} - \theta_{i,j+1}) \end{cases} \quad (9)$$

$$\xi_a = \sqrt{\xi_{a,i}^2 + \xi_{a,j}^2} \quad (10)$$

従来から提案されている評価基準と、本研究で用いた二つの評価基準を当てはめた例を図(8)に、また図(9)、(10)にはそれぞれ12秒後と40秒後に得られた写像を示す。

図(9)、(10)に見られるように写像は学習するに従い入力の特徴((4)で与えられる平面上での一樣な分布)を明らかによく学習しているが、従来からの基

準、特に"Relative Rotation of Vector"では学習によって写像関係が改善されている事がわからない。それに対し、今回提案した二つの基準は写像が学習し分散していく過程をよく表している事がわかる。

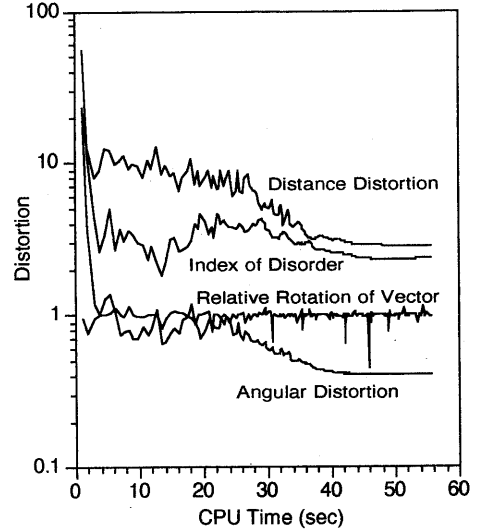


Fig.8 Map Distortion v.s.CPU Time Example 3

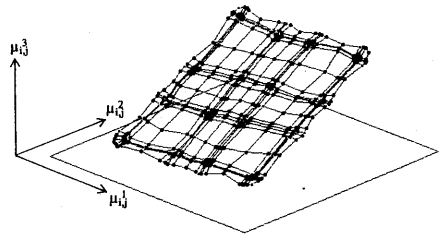


Fig.9 Distribution of Weight Vector (after 12 sec)

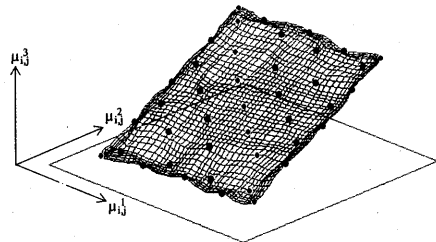


Fig.10 Distribution of Weight Vector (after 40sec)

### 6. まとめ

自己組織特徴マップに於いて特徴マップ層上の各素子の機能を階層化することによって特徴マップの

学習を高速に行う事ができることを示した。ここで提案した機能階層化特徴マップでは学習が進行しユニットの近傍範囲が狭まった後にも個々のユニットの機能は階層化されたままであるが、ユニットの近傍範囲が狭まり個々のユニットの荷重ベクトル更新時に於ける独立性が高まった後には、従来の特徴マップのように発火ユニットの探索を全てのユニットの中から行う方がより良い写像を得られると思われる。さきに述べた例で言えば、社会全体が混乱している間は少数によって全体を指導させ、全体が整ってきたところでそれぞれが自由に動けるようにすることに対応する。こうした考えを図(7)に示した例に適用し、上位ユニットが極端に少ない場合にも学習が効果的に行われるようになった例を図(11)に示す。

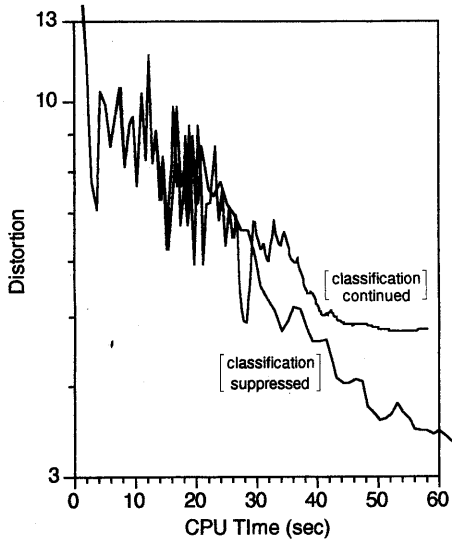


Fig.11 Distance Distortion v.s. CPU Time

今回は機能の階層化を特徴マップに対して当てはめたわけだが、他の大規模なニューラルネットに応用する事も考えられる。ニューラルネットの特徴である高い可塑性や自己適応機能、曖昧な判断等を実現するためには、多数の素子を並列的に駆動する必要性は避けて通れないものと思われるが、そうしたとき本研究で示した素子の機能の階層化が有効に活用できるのではないだろうか。

#### 参考文献

- [1] T.Kohonen;"Self-Organization and Associative Memory(second edition)",pp119-157, Springer-Verlag,1988  
 [2] T.M.Martinets,H.J.Ritter,K.J.Schulten;

"Three-Dimensional Neural Net for Learning Visuomotor Coordination of a Robot Arm", IEEE Trans.Neural Networks,Vol.1, pp131-136,1990

[3] N.M.Nasrabadi,Y.Feng,"Vector Quantization of Images Based Upon the Kohonen Self-Organizing Feature Maps"Proc.IEEE 2nd International Conference of Neural Networks,Vol.1, pp101-108,1988

[4] 本間亨,小杉幸夫;"ニューラルネットを用いた誘発電位波形処理",電子情報通信学会論文誌 D-II Vol.J73-D-II,pp1322-1327,1990

[5] 三宅一郎,綿貫讓治,嶋澄,蒲島郁夫;"平等をめぐるエリートと対抗エリート",創文社,1985

[6] たとえば K.N.Leibovic;"Information Processing in the Nervous System", Springer-Verlag,1969

[7] Ref.[1],pp143-155

[8] S.I.Tillery,N.Combs;"The Problem of Internal Representations in Neural Nets:Concepts, Implications,and a Proposed Metric",Proc.IEEE First International Conference on Neural Networks,Vol.II,pp585-590,1987