

## 機能マップの自己組織化への統計力学的アプローチ

黒岩 丈介† 三宅 章吾‡ 猪苗代 盛§ 阿曾 弘具¶

† 広島大学総合科学部数理情報科学コース, E-mail: jou@mis.hiroshima-u.ac.jp .

‡ 東北大学大学院工学研究科応用物理専攻, E-mail: miyake@camp.apph.tohoku.ac.jp .

§ 日本大学工学部情報工学科, E-mail: inawa@inawa.ce.nihon-u.ac.jp .

¶ 東北大学大学院工学研究科電子・通信工学専攻, E-mail: aso@aso.ecei.tohoku.ac.jp .

## 1 はじめに

大脳皮質の一次視覚野には、特定の方位の線方に対して選択的に反応する、方位選択性ニューロンが存在する [1]. このようなニューロンは視覚入力 of 最も単純な特徴抽出を行なっているものと考えられる. 近年、光計測技術の進歩により、広範囲のニューロンの方位刺激に対する反応のようす (orientation preference map) が生理学的実験により明らかとなってきた [2, 3]. 理論的には、どのようなメカニズムでこのような orientation preference map が形成されるのかということが問題となる.

これまでに我々は、(i) 非線形出力特性を有するニューロンと、(ii) 外部刺激を入力刺激として用いる、という 2 点をモデルの構成の基礎として、orientation preference map の自己組織化の問題への統計力学的な定式化を与えた [4, 5]. 更に、生理学的実験結果と類似の orientation preference map を自己組織化する際に有効な学習法とし、クラスター学習則を提案した [6] (他の研究は [7] 参照).

本報告の目的は、orientation preference map の自己組織化の問題を統計力学的な視点から捉え、クラスター学習則や抑制性ニューロンプールの役割りを明らかにすることである.

## 2 モデルの概要

抑制性ニューロンプール [8] を有する、2 層神経回路網モデルを用いる. 入力層及び出力層は、 $L$  及び  $N$  個の McCulloch and Pitts タイプの形式ニューロンから成る. 出力層のニューロン  $i$  は、次の 3 種類の入力を受け取る. (i) 興奮性のシナプス結合  $s_{ik}$  を通じて、入力層のニューロン  $k$  からの外部入力  $X_k$ . (ii) 抑制性のシナプス結合  $s_i$  を通じて、抑制性ニューロンプールからの抑制性の一定入力  $X_0 (= 1)$ . (iii) 側方結合  $w_{ij}$  を通じて、近隣のニューロン  $j$  の出力  $z_j$  のフィードバックの寄与.

自己組織化は、シナプス結合  $s_{ik}$  と  $s_i$  をクラスター学習則 [6] によって変化させることで達成される. 従来の Hebb タイプの学習では入力層と出力層のニューロンの同時発火に基づいてシナプス結合を調整していたのに対し、クラスター学習則では出力層の近隣のニューロンの発火の影響も考慮する. つまり、クラスター学習は以下のように記述される.

$$\tau \frac{ds_{ik}}{dt} = -s_{ik} + c_1 z_i X_k + c'_1 \sum_j e_{ij} z_j X_k \quad (1a)$$

$$\tau \frac{ds_i}{dt} = -s_i + c_2 z_i X_0 + c'_2 \sum_j e_{ij} z_j X_0 \quad (1b)$$

ここで、 $e_{ij} (\geq 0)$  は出力層のニューロン  $j$  から  $i$  への寄与率であり、 $\tau$  は学習の時定数である. また、 $c_1$ ,  $c_2$ ,  $c'_1$ ,  $c'_2$  は学習効率を表わす.

側方結合  $w_{ij}$  は学習の間変化しないものとする. ここでは、以下に与えるようなメキシカンハットタイプのものを用いる.

$$w_{ij} = (E + I) \exp\left(-\frac{|i-j|^2}{2r_E^2}\right) - I \exp\left(-\frac{|i-j|^2}{2r_I^2}\right) \quad (2)$$

ここで、 $E$  は興奮性の強度、 $I$  は抑制性の強度、 $r_E$  は興奮性結合のレンジ、 $r_I$  は抑制性結合のレンジである.

入力パターンとして、簡単化のため 15 方位の棒状の入力を用いる (図 1). 各入力パターン  $X^\mu$  ( $\mu = 1, 2, \dots, 15$ ) は等確率で入力層に提示されるものとする. この際、各々の入力パターンが提示される時間より学習の時定数  $\tau$  が十分大きいならば、学習方程式 (1) はアンサンブル近似で近似できる [4, 6]. 更に、十分長い間各々の入力を提示し続けるならば、自己組織化は達成され、シナプス結合の定常値が与えられる. この定常値を用いたニューロンの定常出力は、

$$z_i^\mu = f\left(\sum_j w_{ij} z_j^\mu + \frac{1}{P} \sum_\nu (c_1 v_{\mu\nu} - c_2 X_0 X_0) z_i^\nu + \frac{1}{P} \sum_\nu \sum_j (c'_1 v_{\mu\nu} - c'_2 X_0 X_0) e_{ij} z_j^\nu - u_{th}\right) \quad (3)$$

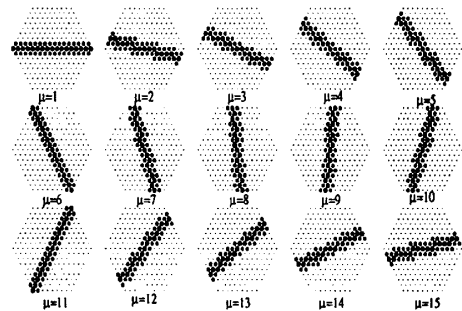


図 1: 15 種類の棒状の入力パターン. (●) の大きさが、入力の大さ 1.0 及び 0.5 に対応している.

表 1: 各パラメータ値.

|                                                            |
|------------------------------------------------------------|
| $L_x = 17, L_y = 17, N_x = 42, N_y = 42,$                  |
| $T = 7.0, u_{th} = 2.0, X_0 = 1.0, c_1 = 1.5$              |
| $(c_2/c_1) = 12.0, c'_1 = 1.5, (c'_2/c'_1) = 12.0,$        |
| $E = 0.96, I = 2.04, \sigma_E^2 = 1.73, \sigma_I^2 = 2.5,$ |

のように与えられる [6]. ここで,  $u_{th}$  はニューロンの活動の閾値であり, 出力関数  $f(x)$  は,

$$f(x) = \frac{1}{1 + \exp(-2\beta x)} \quad (4)$$

なるシグモイドタイプのものを用いる. ここで,  $\beta/2$  ( $\beta = 1/T$ ) は,  $x = 0$  での傾きを与える. また,  $v_{\mu\nu}$  は異なる入力  $X^\mu$  と  $X^\nu$  の間の空間相関であり,

$$v_{\mu\nu} = \sum_k X_k^\mu X_k^\nu \quad (5)$$

のように与えられる. ここで, 自己相関  $v_{\mu\mu}$  は異なるパターン間の相関  $v_{\mu\nu}$  ( $\mu \neq \nu$ ) に比べて非常に大きな値となることが自己組織化に於て重要となる.

### 3 Orientation preference map

$\mu$  番目の入力に対するニューロンの出力を, 直接方程式 (3) から求める代りに, 自己無撞着モンテカルロ法によって計算した [4, 5, 6]. シミュレーションの際に用いた各パラメータの値は表 1 に与える. 但し,  $e_{ij}$  は側方結合  $w_{ij}$  の正の部分と同じ値を用いた.

各入力に対する各ニューロンの出力から得られた orientation preference map を図 2 に与える. 図では, 直線の長さで出力の大きさ, 直線の方角で反応する入力の方角を表わしている. 得られた結果には, (i)  $\pm 1/2$  pinwheel singularity, (ii) linear zone, (iii) fracture, (iv) saddle point が存在する. また, 各 pinwheel singularity 間では反応する方位が滑らかに変化し, 方位変化が全体として滑らかである. このように, クラスター学習によって得られた結果は, 生理実験結果と非常に類似したものである. 同じ計算を,  $e_{ij} = 0$  とした通常の Hebb 学習則によって行くと, (i) pinwheel singularity が形成されにくくなる, (ii) fracture が非常に多く存在し, 滑らかな方位変化が見られなくなる, という結果となる. このことは, クラスター学習が orientation preference map の自己組織化に於て重要な役割を担っていることを示している. 次に, このクラスター学習則がどのような効果を持っているかについて, 統計力学的観点から考察する.

### 4 統計力学的観点からの考察

我々はこれまでに, 定常出力を記述する方程式 (3) が 3 次元 Ising スピン空間における平均場方程式に対応していることを示し, 平均場近似の範囲内

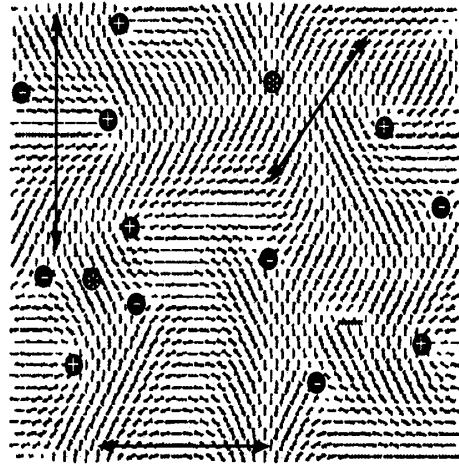


図 2: Orientation preference map. ‘+’ は  $+1/2$  pinwheel singularity, ‘-’ は  $-1/2$  pinwheel singularity, ‘\*’ は saddle point を各々表わしている. また, 細い直線が linear zone, 及び太い直線が fracture を表わしている.

に於てこの平均場方程式を与えるハミルトニアン形式を与えた [4, 5, 6]. そのハミルトニアンは以下のように与えられる.

$$\begin{aligned} H = & -\frac{1}{2} \sum_i \sum_{\mu \neq \nu} \sum_j J_{\mu\nu}^{(z)} \sigma_{i\mu} \sigma_{i\nu} \\ & - \frac{1}{2} \sum_{i \neq j} \sum_j \sum_\mu J_{ij}^{(xy)} \sigma_{i\mu} \sigma_{j\mu} \\ & - \frac{1}{2} \sum_{i \neq j} \sum_{\mu \neq \nu} \sum_j \sum_\nu J_{ij\mu\nu}^{(xyz)} \sigma_{i\mu} \sigma_{j\nu} \\ & - \sum_i \sum_\mu (h_{i\mu} + h_{i\mu}^{\text{self}} + h_{i\mu}^{\leq \sigma}) \sigma_{i\mu} \quad (6) \end{aligned}$$

交換相互作用  $J_{ij}^{(xy)}$ ,  $J_{\mu\nu}^{(z)}$  と  $J_{ij\mu\nu}^{(xyz)}$  は次のように定義される.

$$J_{ij}^{(xy)} = \frac{1}{2} w_{ij} \quad (7a)$$

$$J_{\mu\nu}^{(z)} = \frac{1}{2P} (c_1 v_{\mu\nu} - c_2 X_0 X_0) \quad (7b)$$

$$J_{ij\mu\nu}^{(xyz)} = \frac{1}{2P} (c'_1 v_{\mu\nu} - c'_2 X_0 X_0) e_{ij} \quad (7c)$$

また, 局所場  $h_{i\mu}$  と  $h_{i\mu}^{\text{self}}$  は次のように与えられる.

$$\begin{aligned} h_{i\mu} = & \sum_{j \neq i} J_{ij}^{(xy)} + \sum_{\nu \neq \mu} J_{\mu\nu}^{(z)} \\ & + \sum_{j \neq i} \sum_{\nu \neq \mu} J_{ij\mu\nu}^{(xyz)} - u_{th} \quad (8a) \end{aligned}$$

$$h_{i\mu}^{\text{self}} = J_{ii}^{(xy)} + J_{\mu\mu}^{(z)} \quad (8b)$$

このハミルトニアンは次の拘束条件 (自己無撞着性の条件) を満す時, 方程式 (3) に対応する平均

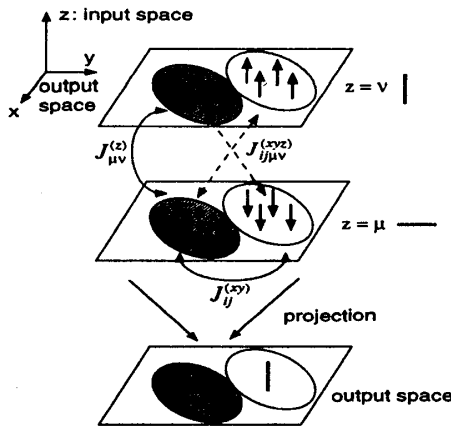


図 3: 3 次元 Ising スピン空間での交換相互作用の模式図。

場方程式を与える。

$$h_{i\mu}^{<\sigma>} = h_{i\mu}^{\text{self}} < \sigma_{i\mu} > \quad (9)$$

ここで、 $\langle \sigma_{i\mu} \rangle$  は Ising スピン  $i\mu$  の熱平均である。この自己無撞着性の条件が成立すると、局所場  $h_{i\mu}^{<\sigma>}$  はその位置のスピンの熱平均に依存する自己場とみなすことができる。この自己場が、我々の自己組織化モデルの特徴の一つを表わしている。

次に、このハミルトニアン形式を用いて、クラスター学習と抑制性ニューロンプールの役割について考察する。図 3 に於て、一番下の output space と記された平面がモデルの出力層に相当し、その中に水平な方位と垂直な方位の線分に反応するニューロン集団が隣接して存在している (fracture の存在)。その面の上が、3 次元 Ising スピン空間を表わし、 $z = \mu$  (水平方位の入力) 及び  $z = \nu$  (垂直方位の入力) 平面が表わされている。スピン空間でその熱平均が正のスピンは、出力層で発火しているニューロンに対応する。ここでの議論は簡単化のため温度を  $T \approx 0$  と仮定する。つまり、スピンの熱平均値は近似的に  $+1$  又は  $-1$  の値を取る。

自己相関  $v_{\mu\mu}$  は異なるパターン間の相関  $v_{\mu\nu}$  ( $\mu \neq \nu$ ) よりはるかに大きな値をとるので適当な抑制性ニューロンプールからの寄与 ( $c_2$  及び  $c'_2$  の値) に対して、自己場  $h_{i\mu}^{\text{self}}$  は大きな正の値となる。一方、局所場  $h_{i\mu}$  (式 (8) (a)) に於ては、閾値  $u_{th}$  が支配的であり、負の値となる。ある位置のスピンの平均的に  $+1$  ならば、ハミルトニアン (6) に於て局所場の項は  $2h_{i\mu}^{\text{self}} + h_{i\mu} (> 0)$  となり、上向きの状態を保持する。一方、スピンの平均的に  $-1$  ならば、局所場の項は  $h_{i\mu} (< 0)$  のみとなり、下向きの状態を保持する。

側方結合  $w_{ij}$  に由来する交換相互作用  $J_{ij}^{(xy)}$  の寄与により、 $J_{ij}^{(xy)} > 0$  であるような範囲には上向きのスピんクラスターが  $z$  軸に垂直な平面内に形成される。このクラスターの大きさは、 $J_{ij}^{(xy)} < 0$  の大きさに依存する。

交換相互作用は  $z$  軸に平行な方向、つまりニューロンの出力層で同じ位置  $i$  で異なる入力パターンに対応するスピン間に作用する。抑制性ニューロンプールからの寄与によって  $J_{\mu\nu}^{(z)}$  は類似の入力パターン間 ( $\mu$  と  $\nu$  が似かよったパターン) では正の値、大きく異なる入力パターン間では負の値を取る。類似の入力パターン間 ( $J_{\mu\nu}^{(z)} > 0$ ) ではスピンは同じ向き、大きく異なる入力パターン間 ( $J_{\mu\nu}^{(z)} < 0$ ) ではスピンは反対向きとなり、大きく異なる方位に対しては反応を示さなくなる。逆に、もし抑制性ニューロンプールからの寄与が無ければ ( $c_2, c'_2 = 0$ )、交換相互作用  $J_{\mu\nu}^{(z)}$  は常に正の値を取り、全てのスピンが同じ向きとなる。つまり、方位選択性を獲得するためには、抑制性ニューロンプールからの抑制性の寄与が重要である。

しかし、上に述べた各相互作用は全てパターン方向 ( $z$  軸方向) またはニューロンの位置方向 ( $xy$  平面内) の間で作用し、fracture の形成を防ぐ効果を持っていない。一方、交換相互作用  $J_{ij\mu\nu}^{(xyz)}$  は、 $xy$  平面内に於ては隣接し ( $e_{ij} > 0, i \neq j$ ) かつ大きく異なるパターン ( $J_{\mu\nu} < 0, \mu \neq \nu$ ) に対応する位置のスピン間では、 $J_{ij\mu\nu}^{(xyz)} < 0$  となる。このようなスピン  $\sigma_{i\mu}$  と  $\sigma_{j\nu}$  には  $J_{ij\mu\nu}^{(xyz)} < 0$  として働くので、同じ向きになることを防ぐことができ、結果的に fracture が形成されにくくなる。このように、クラスター学習則は近隣のニューロンの状態を学習に積極的に取り入れることで、滑らかな orientation preference map の自己組織化に重要な寄与を与える。

## 5 おわりに

本報告では、orientation preference map の自己組織化におけるクラスター学習則と抑制性ニューロンプールの役割を統計力学的な立場から明かにした。

## 参考文献

- [1] D. H. Hubel and T. N. Wiesel: *J. Physiol. (Lond.)*, **160**, pp.106-154, 1962.
- [2] G. G. Blasdel and G. Salama: *Nature*, **321**, pp.579-585, 1986.
- [3] T. Bonhoeffer & A. Grinvald: *Nature*, **353**, pp. 429-431, 1991.
- [4] S. Inawashiro, Y. Tamori, S. Miyake & J. Kuroiwa: *J. Phys. A: Gen. Math.*, **29**, pp. 7389-7399, 1996.
- [5] J. Kuroiwa, S. Miyake, S. Inawashiro & H. Aso: *WORLD CONGRESS ON NEURAL NETWORKS* (San Diego, U.S.A, INNS Press), pp. 543-546, 1996.
- [6] J. kuroiwa, S. Inawashiro S. Miyake & H. Aso: *Progress in Connectionist-Based Information Systems* (N. Kasabov, et al., Eds., Springer), pp. 103-107, 1997.
- [7] E. Erwin, K. Obermayer & K. Schulten: *Neural Comp.*, **7**, pp. 425-468, 1995.
- [8] A. Takeuchi & S. Amari: *Biol. Cybern.*, **35**, pp. 63-72, 1979.