

自己組織化神経回路モデル (4)

論文紹介:

B.A. Olshausen, D.J. Field

Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature*, vol. 381, pp.607-609, 1996.

1. Q

なぜ視覚一次野の神経細胞は短い線分の傾きに反応選択性をもつのか。
どうすれば画像を効率よく表現できるか(どのような基底を選べばいいか)。

2. 数理モデル

2.1. 入力と出力: 16×16 の画像

入力 $I(0,0), I(0,1), \dots, I(15,14), I(15,15) \in \{0, \dots, 255\}$

出力 $\hat{I}(0,0), \hat{I}(0,1), \dots, \hat{I}(15,14), \hat{I}(15,15) \in \{0, \dots, 255\}$

2.2. 基底画像 (192 枚): 16×16 の画像 (学習により形成)

$\vec{\phi}_1 = (\phi_1(0,0), \phi_1(0,1), \dots, \phi_1(15,15))$

$\vec{\phi}_2 = (\phi_2(0,0), \phi_2(0,1), \dots, \phi_2(15,15))$

\vdots

$\vec{\phi}_{192} = (\phi_{192}(0,0), \phi_{192}(0,1), \dots, \phi_{192}(15,15))$

2.3. 仮定: 入力は基底の線形結合で表現できる

$$I(x,y) = \sum_{i=1}^{192} a_i \phi_i(x,y), \quad x=0, \dots, 15, \quad y=0, \dots, 15 \quad (1)$$

基底の組 $\{\phi_i(x,y)\}$ を決めれば, 各画像をどう表現するか決まる。

$\{a_i\}$ は, 表現する画像に依存して変化する値。

→ どのような基底の組を使い画像を表現するのがいいだろうか

2.4. コスト関数 (この関数を最小化する最適化問題)

$$E = -[\text{preserve information}] - \lambda [\text{sparseness of } a_i] \quad (2)$$

$$\frac{1}{2} \sum_{x,y} \left[I(x,y) - \sum_{i=1}^{192} a_i \phi_i(x,y) \right]^2 + \lambda \sum_i S\left(\frac{a_i}{\sigma}\right) \quad (3)$$

これを最小にするような $\{a_i\}, \{\vec{\phi}_i\}, i=1, 2, \dots$ を求めたい。

どの入力画像に対しても同じ基底の組 $\{\vec{\phi}_i\}$ を使い表現。

2.4.1. 基準 1: 再現性

基底 $\{\vec{\phi}_i\}$ を使い, できる限り精度よくもとの画像 $I(x,y)$ を再構成したい。

2.4.2. 基準 2: スパース性

基底の重み付け和により入力を再構成するが, その際, できる限り重みが少ない枚数の基底を使い表現したい。

$a_i, i=1, 2, \dots$ のほとんどは, 値 0 になってほしい。

2.4.3. 基準3：独立性（ここでは直接的には使わない）

各 $a_i, a_j, i \neq j$ 間の相関をなくしたい。同時エントロピー $H(a_1, a_2, \dots, a_n)$ は保ったまま、個々のエントロピーの和 $\sum_i H(a_i)$ を小さくしていく。このような a_1, a_2, \dots を見つける。数学的性質：例えば a_1, a_2 間に依存性があれば、 a_1 の値がわかれば a_2 の値がある程度、推測できる。これは $H(a_1, a_2) < \sum_i H(a_i)$ ということ。確率変数 a_1, a_2, \dots, a_n 間に依存性があれば、 $H(a_1, a_2, \dots, a_n) < \sum_i H(a_i)$ となる。これを a minimum-entropy code と H.B. Barlow は名付けた。

2.4.4. ここで $S(x)$ というのは、スパース性を促進する、例えば以下のようなもの。

- $S(x) = -e^{-x^2}$
- $S(x) = \log(1 + x^2)$

図2：コスト関数の模式図。各 a_i はなるべく0になろうとする。

- $S(x) = |x|$

2.5. 2つのダイナミクス。

2.5.1. 速いダイナミクス $\{a_i\}$ ：ニューロン活動に対応。

$I(x, y)$ を基底 $\{\vec{\phi}_i\}$ 重みつき線形和で表現。その時の重みが $\{a_i\}$ 。

$\{a_i\}$ の求め方：

目的関数 E を a_i で微分してみる。

$$\begin{aligned} -E &= -\frac{1}{2} \sum_{x,y} \left[I(x,y)^2 - 2I(x,y) \sum_{i=1}^n a_i \phi_i(x,y) + (a_1 \phi_1 + \dots + a_n \phi_n)^2 \right] - \lambda \sum_i S\left(\frac{a_i}{\sigma}\right) \\ &= \frac{1}{2} \left[-\sum_{x,y} I(x,y)^2 + 2 \sum_{x,y} I(x,y) \sum_{i=1}^n a_i \phi_i(x,y) - \sum_{x,y} (a_1 \phi_1 + \dots + a_n \phi_n)^2 \right] - \lambda \sum_i S\left(\frac{a_i}{\sigma}\right) \\ -\frac{\partial E}{\partial a_i} &= \sum_{x,y} I(x,y) \phi_i(x,y) - \sum_{x,y} \sum_j (a_j \phi_j \phi_i) - \frac{\lambda}{\sigma} S'\left(\frac{a_i}{\sigma}\right) \end{aligned}$$

したがって、 E を減らす方向に a_i を動かすには $\Delta a_i = -\tau \frac{\partial E}{\partial a_i}$ とすればよい。

2.5.2. 遅いダイナミクス $\{\vec{\phi}_i\}$ ：学習（結合係数の変化に対応）。

$$-\frac{\partial E}{\partial \phi_i(x,y)} = I(x,y) a_i - (a_1 \phi_1 \dots a_n \phi_n) a_i = a_i \left[I(x,y) - \hat{I}(x,y) \right]$$

したがって、 E を減らす方向に $\phi_i(x,y)$ を動かすには $\Delta \phi_i(x,y) = -\eta \frac{\partial E}{\partial \phi_i(x,y)}$ とすればよい。

3. コンピュータシミュレーション その1（図1）

3.1. 入力：大量の 8×8 データ（自然画像のパッチ）

3.2. 図1：入力データに対して主成分分析をおこなった結果。

64個すべての基底（固有ベクトル）を表示。

4. コンピュータシミュレーション その2（図3）

4.1. 入力：大量の人工的に作成したデータ

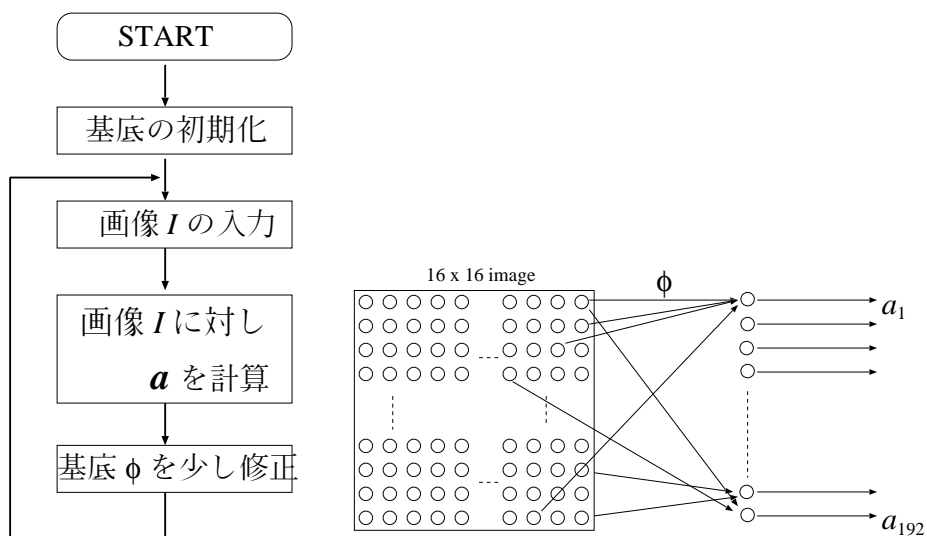
- a. 各ピクセルは独立でランダムな値（大きな値は滅多にとらないようモデル化）.
 - b. フーリエ変換した領域で a のようなもの
 - c. a, b では互いの基底は直交していたが, c では, 互いに直交していない Gabor 関数（基底）の重ねあわせで画像をランダムに生成
- 4.2. 図3：この論文で提案しているアルゴリズムが確かに動くことを, 人工的に作った画像で確認. あらかじめ, ある基底集合の重みづけで画像を生成. 3種類の基底を試しており, それが a, b, c に対応. もとの基底表現が自己組織的に獲得できたことに注目.

5. コンピュータシミュレーション その3（図4）

5.1. 入力：16 × 16-pixel の画像パッチ. もとの画像は 512 × 512 の自然画像 10 枚.

5.2. 図4:

- a. 学習により獲得した基底 $\phi_i(x, y)$, $i = 1, \dots, 192$ を表示. 様々な傾き, 位置, 空間周波数の基底が得られた.
- b. 各場所のスポット入力（1ピクセルだけ on?）に対する再構成時の a_i の値を表示. $16 \times 16 = 256$, それぞれの場所で試す.
- c. どんな空間周波数の基底が得られたか, その分布.
- d. 自然画像のパッチを表現したときの a_i の確率分布（ヒストグラム）.
 実線：学習済の基底を使った場合. $H(a_i) = 4.0$
 点線：ランダムに生成した基底. $H(a_i) = 4.6$
 確かにエントロピーが下がっている
 （ $\sum_i H(a_i)$ をできるだけ小さくするのが目的だった）.



P.Földiák (1990) の論文との関連：基本的に同一の思想をもった研究

6. 同じ点：

各々の入力に対し，自己組織的に（教師なし学習により）内部表現を形成する．

2つのダイナミックス：

1. 各入力を表現する（ニューロン活動のダイナミックス）速いダイナミックス．
2. 表現するための基底（結合係数）を学習により獲得する遅いダイナミックス．

7. 若干異なる点：

モデル

Földiák (1990) は神経回路モデル．

Olshausen & Field (1996) は神経回路モデルではないが，神経回路的な解釈はできる．

再構成（生成）性

Földiák (1990) では再構成は考えていない．再構成してみると，どういう画像が得られるかは，示していない（見てみたい）．

Olshausen & Field (1996) は，各入力を基底の線型重み和で再構成．

入力画像

Földiák (1990) では文字が描かれた画像の 0,1 パターン．

Olshausen & Field (1996) では自然画像の濃淡値．視覚一次野のニューロンの反応選択性と比較したいこともあり，画像全体ではなく一部分（ 16×16 のパッチなど）．

8. 異なる点：

線型・非線型モデル

Földiák (1990) は非線型．

Olshausen & Field (1996) は各基底の線型の重み和で画像を表現．

基底を使った表現の仕方

Földiák (1990) の場合，入力を表現する際，各基底に対するの重みづけはほぼ 0,1（ニューロンの活動）．

Olshausen & Field (1996) では a_i はスパースとは言ってもアナログ値（このアナログ値を実際の神経活動でどう表現しているのかは，よくわからない）

エントロピーを使った評価方法

Földiák (1990) では，入力に与える各アルファベット・記号の出現頻度をある英文と同じにして，エントロピーによる評価をおこなった．符合化の前後で情報量はほぼ落ちていないが，各ビットのエントロピー（ a_i に対応）の和は低く押さえていること，各 a_i が独立に近くなっていることを示した．

Olshausen & Field (1996) は入力が自然画像の一部分で，しかも濃淡値であり，可能な画像の数は膨大なものになることから，Földiák (1990) のような解析はおこなえない．そこで，どの a_i も，その確率分布が同じ形をするはずであると考え，すべての $a_i, i = 1, 2, \dots$ を同一視して，その確率分布のエントロピーを計算した．その結果，エントロピーが確かに小さくなっていることを示した．