情報システム工学科
伊達 章

神経回路網特論
2006 年 6 月 8 日

自己組織化神経回路モデル (3)

論文紹介
P. Foldiak
Forming sparse representations by local anti-Hebbian learning,
Biological Cybernetics, vol. 64, pp. 165-170, 1990.

1. 情報表現の仕方 （発火率 p の違いに着目）
   1.1. 局所表現 （$p = 1/n$）
   $n \cdots$ ニューロン数，$p \cdots$ 活動度 $0 \leqq p \leqq 1$,（興奮しているニューロンの割合）
   別名「おばあさん細胞」表現．SOM による表現形成． ヘブ学習と競合学習で実現
   1.2. 分散表現 （完全分散 $p = 1/2$）
   1.3. スパース表現 （例： $p = \log n$）

   それぞれの利点： 容量，耐ノイズ性，$\cdots$

2. 数理モデル
   2.1. ネットワークの構造（二層の回路）
   入力 $x_1, x_2, \cdots, x_m \in \{0, 1\}$
   出力 $y_1, y_2, \cdots, y_n \in \{0, 1\}$
   2.2. ニューロン活動のダイナミックス

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = f\left(\sum_{j=1}^{m} q_{ij} x_j + \sum_{j=1}^{n} w_{ij} y_j^* - t_i\right) - y_i^*$$

$$f(u) = \frac{1}{1 + \exp(-\lambda u)}$$

   2.3. 学習アルゴリズム
   入力層の素子 x_j と出力層の素子 y_i の間：Hebb 学習 $\Delta q_{ij} = \beta y_i (x_j - q_{ij})$
   出力層内：anti-Hebb 学習 $\Delta w_{ij} = \alpha (y_i y_j - p^2) \to$ 学習の平衡状態で $\mathrm{E}[y_i y_j] = p^2$
   出力素子の閾値：$\Delta t_i = \gamma (y_i - p) \to$ 平均活動度を p にする

3. コンピュータシミュレーション その 1： 縦線，横線の学習
   3.1. 図 1： ネットワークの構造（入力層 64 個，出力層 16 個）
   3.2. 図 2： 入力信号の例 （$8 \times 8 = 64$ ピクセル）
   入力は縦線，横線の組合わせ．各縦線，横線が確率 1/8 で独立に on/off
   表現可能なパターンの総数 2^{64} 個
   そのうち入力として出現する可能性のあるパターンの総数 2^{16} （$\ll 2^{64}$）通り
   3.3. 図 3： 自己組織化が進む様子（$t = 0, 400, 800, 1200$）． $q_{ij}, i = 1, \cdots, 16$ を表示．

4. コンピュータシミュレーション その2： 文字の学習

4.1. 図1： ネットワークの構造（入力層120個，出力層16個）

4.2. 図4： 自己組織化の様子（$t = 0, 4000, 8000, 16000$）．$q_{ij}, i = 1, \cdots, 16$ を表示.

4.3. 入力パターン： Sun-3 ワークステーションのフォント.
各文字が出現する頻度を，ある英文で使われている頻度に設定.
入力層で表現可能なパターンの総数は 2^{120} 個．そのうち出現するのは 82 個.
出力層には 16 素子あるので $2^{16} = 256$ 通りのパターンを 1 対 1 で表現できる.
（実際には 256 パターンのうち，48 パターンが使われるようになった）.

4.4. 表1： 学習後，各文字を入力すると，どういう出力パターン（16 ビット）が出力されるようになったかを示した一覧表．多対 1 の写像になっていることに注意.
例：k, B, R, F を入力すると，どの場合も出力は 0000010000010000 になる.

4.5. 表2：各表現（入力パターン出力パターン）の性質

4.5.1. 入力
$A_j \cdots j$ 番目の文字の出現頻度．$j = 1, \cdots, 82$.
$$E(A) = -\sum_{j=1}^{82} A_j \log_2 A_j = 4.34 \text{ bit}$$
j 番目の文字の入力パターン $(a_{1j}, a_{2j}, \cdots, a_{120j})$
$q_i = \sum_{j=1}^{82} a_{ij} A_j \cdots$ 入力パターンの i 番目の要素の平均値（1 になる確率）
$$e(A, a) = -\sum_{i=1}^{120} [q_i \log q_i + (1 - q_i) \log(1 - q_i)] = 24.14 \text{ bit}$$

4.5.2. 出力
82 通りの入力に対して得られる出力 $\boldsymbol{b}_k = (b_{1k}, b_{2k}, \cdots, b_{16k})$，$k = 48$
$B_k \cdots \boldsymbol{b}_k$ を観測する確率
$$E(B) = -\sum_k B_k \log_2 B_k = 4.22 \text{ bit}$$
各文字パターンを 16 ビットの列に変換したところ，出力パターンの総数は 82 個より小さく，48 個になっていた（$k = 48$）．そのため，もとのエントロピーより小さくなっている.
$p_i = \sum_{k=1}^{48} b_{ik} B_k \cdots$ 出力層の素子の i 番目の要素が 1 になる確率
$$e(B, b) = -\sum_{i=1}^{16} [p_i \log p_i + (1 - p_i) \log(1 - p_i)] = 5.86 \text{ bit}$$
スパース性：16 ビットのうち，1 の個数は 0 〜 4 になっている.
※出現頻度が高い文字ほど 1 の数が少ないパターンに符合化されている.

4.5.3. 各ビットのエントロピーの和は何を意味するか.
$H(X), H(Y) \leqq H(X, Y) \leqq H(X) + H(Y)$ の関係を思い出す.

$H(p_i) \leqq H(p_i, p_2, \cdots, p_{16}) \leqq H(p_1) + H(p_2) + \cdots + H(p_{16})$
16 個のビットが互いに独立な場合に $=$ が成立する.

$$H(p_i, p_2, \cdots, p_{16}) = 4.22 \text{ [bits]}$$
$$H(p_1) + H(p_2) + \cdots + H(p_{16}) = \sum_{i=1}^{16} H(p_i) = 5.86 \text{ [bits]}$$

冗長性： $[e(B,b) - E(B)]/E[B] = (5.86 - 4.22)/4.22 = 0.388$
この意味：すべてのビットが独立だと仮定すると 5.86 ビットの情報を得られる．実際には 4.22 ビットしか得られなかった．これをどうして著者らは冗長性と定義したか．→ 4.22 ビットの情報を表現するのに 5.86 ビット使用，と考える．入力パターンについても同様に考える：
冗長性： $[e(A,a) - E(A)]/E[A] = (24.14 - 4.34)/4.34 = 4.562$

5. その他

5.1. パラメータ $\alpha, \beta, \gamma, \lambda$ などは，どう決定したか.

5.2. 速修 確率論と情報理論

- 情報の量をいかに定めるか
- エントロピーとは
- 確率変数の依存性，確率変数が互いに独立であるとは