

自己組織化神経回路モデル (3)

論文紹介 (テーマ: 入力信号の規則性を反映した情報表現の獲得)

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)

- 局所表現 ($p = 1/n$)
 n ... ニューロン数, p ... 活動度 $0 \leq p \leq 1$ (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.
- 分散表現 (完全分散 $p = 1/2$)
- スパース表現 (例: $p \propto \frac{\log n}{n}$)

それぞれの利点: 容量, 耐ノイズ性, ...

2. 数理モデル

- ネットワークの構造 (二層の回路)
入力 $x_1, x_2, \dots, x_m \in \{0, 1\}$
出力 $y_1^*, y_2^*, \dots, y_n^* \in \{0, 1\}$
- ニューロン活動のダイナミクス

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

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

- 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 $\Delta w_{ij} = \alpha (y_i y_j - p^2)$ 学習の平衡状態で $E[y_i y_j] = p^2$
出力素子の閾値: $\Delta t_i = \gamma (y_i - p)$ 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

- 図1: ネットワークの構造 (入力層 64 個, 出力層 16 個)
- 図2: 入力信号の例 ($8 \times 8 = 64$ ピクセル)
入力は縦線, 横線の組み合わせ. 各縦線, 横線が確率 $1/8$ で独立に出現
入力層の 64 個の素子で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
- 図3: 自己組織化が進む様子 ($t = 0, 400, 800, 1200$). $q_{ij}, i = 1, \dots, 16$ 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

- 図1：ネットワークの構造（入力層 120 個，出力層 16 個）
- 図4：自己組織化の様子（ $t = 0, 4000, 8000, 16000$ ）． $q_{ij}, i = 1, \dots, 16$ を表示．
- 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）．
- 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．
- 表2：各表現（入力パターン出力パターン）の性質

－ 入力

$$A_j \cdots j \text{ 番目の文字の出現頻度 } \cdot j = 1, \dots, 82. \longrightarrow \sum_{j=1}^{82} A_j = 1$$

$$E(A) = - \sum_{j=1}^{82} A_j \log_2 A_j = 4.34 \text{ bit}$$

j 番目の文字の入力パターン $(a_{1j}, a_{2j}, \dots, a_{120j})$

$$q_i = E[a_i] = \sum_{j=1}^{82} a_{ij} A_j \cdots i \text{ 番目の入力素子が値 1 をとる確率 } \longrightarrow \sum_{i=1}^{120} q_i = 1$$

$$e(A, a) = - \sum_{i=1}^{120} [q_i \log q_i + (1 - q_i) \log(1 - q_i)] = 24.14 \text{ bit}$$

－ 出力

82 通りの入力に対して得られる出力 $b_k = (b_{1k}, b_{2k}, \dots, b_{16k}), k = 48$

$B_k \cdots b_k$ を観測する確率

$$E(B) = - \sum_k B_k \log_2 B_k = 4.22 \text{ bit}$$

82 通りある各文字パターンを 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（4.34 ビット）より小さくなっている．

$$p_i = \sum_{k=1}^{48} b_{ik} B_k \cdots \text{出力層の素子の } i \text{ 番目の要素が 1 になる確率}$$

$$e(B, b) = - \sum_{i=1}^{16} [p_i \log p_i + (1 - p_i) \log(1 - p_i)] = 5.86 \text{ bit}$$

スパース性：16 個の出力素子のうち，出力値 1 をとる個数は 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．

－ 各ビットのエントロピーの和は何を意味するか．

$$H(X), H(Y) \leq H(X, Y) \leq H(X) + H(Y) \text{ の関係を思い出す．}$$

$$H(p_i) \leq H(p_i, p_2, \dots, p_{16}) \leq H(p_1) + H(p_2) + \dots + H(p_{16})$$

16 個のビットが互いに独立な場合に = が成立する .

$$H(p_i, p_2, \dots, p_{16}) = 4.22 \text{ [bits]}$$

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

$$\text{冗長性: } [e(B, b) - E(B)]/E[B] = (5.86 - 4.22)/4.22 \approx 0.389$$

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 .

実際には出力素子どうしに相関が残っており 4.22 ビットしか得られなかった .

4.22 ビットの情報を表現するのに 5.86 ビット使用, したと考える .

入力パターンについても同様に考える :

$$\text{冗長性: } [e(A, a) - E(A)]/E[A] = (24.14 - 4.34)/4.34 \approx 4.562$$

5. 演習課題

結果を予想できるものは, 予想してから実験をしてみる .

- 5.1. 例 1 では, 出力層の素子数が都合よく 16 個, あらかじめ用意されていた .
出力層の素子数が 8 個なら, どういう表現が形成されるか . 出力層の素子数が 30 個なら, どういう表現が形成されるか . いろいろ変化させてみる .
- 5.2. 例 1 では, どの縦棒も横棒も確率 1/8 で出現したが, 図 2 に示しているようなパターンのうち, 頻繁に出現するパターンがいくつかある場合, どのような表現が形成されるか .
- 5.3. 入力信号として白黒 (0,1) でなく, グレースケール (256 値の画像) の信号を入力すると, どういう表現が形成されるか .
- 5.4. 入力信号 x に対する内部表現 y が得られた後に, 入力信号を入力するのをやめると, y には, 直前に入力した x の値に依存した値になるか, それとも 0 になるか . 0 になる場合, 直前に入力した x の値に依存した値を保持するためには, どうしたらいいか .
- 5.5. 入力の空間的な規則性は獲得できたが, 時間的な規則性も獲得するのは, どうしたらよいか .
- 5.6. 入力空間と回路が獲得した表現との相関を測る指標としては, 論文で書かれているもの以外に, どのようなものがあるか .
- 5.7. パラメータ $\alpha, \beta, \gamma, \lambda$ などは, どう決定すればいいか .
- 5.8. ほか, 自由にいろいろ試してみる .

6. 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