

予測符号化モデルの多層化に関する研究

A study on multilayered predictive coding model

広瀬 裕樹 (PY)[†], 渡部 修[†]

Hiroki Hirose(PY) and Osamu Watanabe

[†] 室蘭工業大学工学部情報工学科

hirose@cortex.csse.muroran-it.ac.jp

Abstract— Predictive coding is a hierarchical model with feedback connections as well as feedforward connections. This model shows endstopping and other extra-classical RF effects in the early visual cortex. We investigated whether a multilayered predictive coding model explains characteristics of higher level areas in the visual cortex.

Keywords— Neural Network, Predictive Coding, Area IT, Pattern Recognition

1 はじめに

大脳皮質の領野間には、低次領野から高次領野へのフィードフォワード結合だけでなく、逆に高次領野から低次領野へのフィードバック結合が存在する。このような階層的構造をもとに提唱された情報処理モデルとして、予測符号化モデルがある。

Raoら [1] は、2階層の予測符号化モデルによるシミュレーションを行った。彼らは、予測符号化モデルに様々な自然画像を学習させた。その結果、図1に示す予測器は、V1野のニューロンが持つ受容野特性に類似する、線分方位を表す基底画像を獲得した。また、端点を含む線分やポップアウトを生じさせる刺激において、V1野の文脈的修飾と同様に、それらに強く応答した。

以上の結果より、予測符号化モデルは、初期視覚における視覚系の情報処理をよく表していると言える。もし視覚系の高次領野が、低次領野と同様の構造の積み重ねで構成されているのであれば、予測符号化モデルを多層化することによって高次パターン認識が可能になることが期待できる。そこで、本研究では予測符号化モデルを多層化し、受容野の特性や時間応答について、高次視覚領野の特性と比較する。

2 予測符号化モデル

予測符号化モデルの構造を図1に示す。上位モジュールは下位モジュールの神経活動を予測し、予測値をフィードバック結合によって下位モジュールに送る。下位モジュールは上位モジュールからの予測値と実際の神経活動を比較し、誤差信号をフィードフォワード結合によって上位モジュールに送る。

各モジュールは、複数のニューロンからなる神経活動

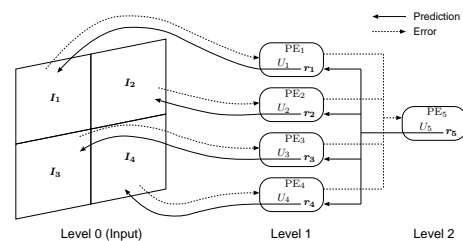


図1: 予測符号化モデルの構造。PE: Predictive Estimator.

を持ち、これをベクトル r で表す。各モジュールは神経活動 r を、シナプス結合 U を介して伝搬することにより予測値を生成する。これを視覚系の問題に当てはめると、神経活動は入力画像の持つ隠れ状態に対応し、シナプス結合は隠れ状態から可視状態への変換行列に対応すると考えられる。したがって、予測誤差を最小化するように r, U を学習することで、隠れ状態を推定することができる。以下に目的関数を示す。

$$E = \frac{1}{\sigma^2} (\mathbf{I} - f(U\mathbf{r}))^T (\mathbf{I} - f(U\mathbf{r})) + \frac{1}{\sigma_{id}^2} (\mathbf{r} - \bar{\mathbf{r}})^T (\mathbf{r} - \bar{\mathbf{r}}) \quad (1)$$

上位モジュールは複数の下位モジュールの組み合わせを予測するため、入力画像のより大域的な状態を表すことができる。 $\bar{\mathbf{r}}$ が上位モジュールからの予測を表す。

3 計算機シミュレーション

本研究では、3階層のモデルによるシミュレーションを行った。27×27ピクセルの人工画像を学習し、学習後の受容野の特性や、想起する際の情報量の変化を調べた。

3.1 入力画像

IT野には、星形や縦縞の三角形など、中程度に複雑な図形に選択的に応答するニューロンが存在する [2]。視覚系は、学習によって様々な入力画像に共通する中程度に複雑なパターンを獲得し、これらを組み合わせることで物体像を表現していると考えられている。

また、サルにヒトの顔、サルの顔、単純図形を含んだ画像を呈示し、IT野ニューロンの発火との相互情報量

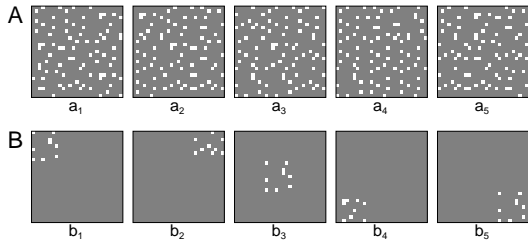


図 2: 入力画像 . A, B からそれぞれ任意の一枚を選んで重ね合わせた 25 枚を入力画像とした .

を調べたところ, 表情などの詳細な情報は, 大まかな分類に対して遅れて表現されることが報告されている [3] .

予測符号化モデルにおいて同様の学習や物体像の表現が行われることを示すために, 図 2 の A, B の組み合わせによって入力画像を構成した .

3.2 パラメータ

スパースネスの拘束条件として, r, U の事前分布を $g(r) = \sum_i \alpha \log(1 + r_i^2)$, $h(U) = \sum_{i,j} \beta \log(1 + u_{ij}^2)$ と仮定し, 式 (1) に加えた . 係数 α, β によってスパースネスの拘束条件の強さを設定することができる . ここでは, スパースネスが強い場合と弱い場合の二通りのシミュレーションを行った .

4 結果

学習後に最上位層のニューロンが抽出した画像を図 3 に示す . スパースネスが強い場合, 第 3 層ニューロンは, 図 2 の a_1, b_5 など, 特定のパターンに選択性を持つ受容野を獲得した . これらのニューロンは, 学習によって, 入力画像に共通する中程度に複雑なパターンを獲得したと言える . また, スパースネスが弱い場合, 各パターンを混合したような基底画像が得られた .

図 4 に, 学習後に各画像を入力した時の情報量の変化を示す . スパースネスが弱い場合, 小分類に関する情報量は, 大分類に関する情報量に対して遅れて上昇した . これは, サルの IT 野ニューロンの特性と類似している . また, スパースネスが強い場合, 二つの情報量は同時に上昇した .

以上の結果より, スパースネスの拘束を強くすると IT 野の受容野特性をよく表わし, 弱くすると IT 野の情報量の時間変化をよく表わした . したがって, IT 野ではこれらの中間の表現を行っていると考えられる .

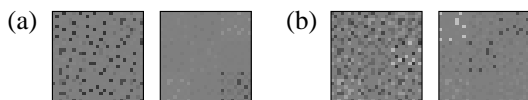


図 3: 第 3 層ニューロンが抽出した基底画像の例 . (a) スパースネスが強い場合 . (b) スパースネスが弱い場合 .

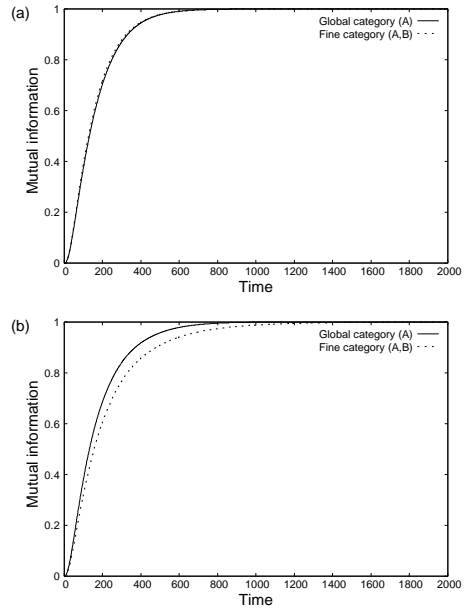


図 4: 想起時における入力画像に関する情報量 (最大値を正規化したもの) . (a) スパースネスが強い場合 . (b) スパースネスが弱い場合 .

5 今後の課題

高次視覚野が解いている問題に, 3 次元形状の認識がある . 視覚系に入力される情報は, 網膜に投影された 2 次元画像であり, 視覚系はそこから 3 次元構造を推定しなければならない .

顔を認識する場合, 正面だけでなく斜めや真横から見た画像からも, 個人を特定する必要がある . 視覚系はこの問題に対応するため, 主要な角度から見たいくつかの顔画像を保持し, これらの組み合わせで中間の角度から見た顔を表現していると考えられる [4] .

このような処理を予測符号化モデルで検証するためには, 様々な角度から見た物体像を学習する必要がある . しかし, 今回用いたモデルでは, 位置, 大きさのずれを許容しないため, 同様の物体でも角度が変われば全く別の物体として認識される . したがって今後は, 位置のずれなどに対応する必要がある .

参考文献

- [1] Rao, R.H. and Ballard, D.H. (1999) "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects." *Nature Neuroscience*, **2**, 79–87.
- [2] Kobatake, E. and Tanaka, K. (1994) "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex." *J. Neurophysiol.*, **71**, 856–867.
- [3] Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999) "Global and fine information coded by single neurons in the temporal visual cortex." *Nature*, **400**, 869–873.
- [4] Wang, G., Tanaka, K. and Tanifuji, M. (1996) "Optical imaging of functional organization in the monkey inferotemporal cortex." *Science*, **272**, 1665–1668.