

LH-003

選択的不感化原理に基づく文脈依存的ニューロン活動のモデル化

Modeling Context-Dependent Neuronal Activity Based on Selective Desensitization Principle

末光 厚夫†
Atsuo Suemitsu

諸上 茂光††
Shigemitsu Morokami

森田 昌彦†††
Masahiko Morita

1 はじめに

視覚性記憶において中心的な役割を果たしている下側頭葉皮質において, Naya ら [1] は文脈依存的活動を示すニューロンを報告している. 彼らは, 従来よく用いられてきた 2 種類の記憶課題を試行の途中で切り替えることによって, 文脈に応じた target の変化を反映するニューロン活動を得た. この結果は, 脳における文脈依存的な連想の過程を知る上で非常に興味深いものである.

ところで, 同様な文脈依存的連想を神経回路モデルで実現しようとする, 入力刺激及び文脈とモデルの出力との多対多に起因する, 深刻な計算論的困難が生じる. この問題を解決する方法として, 「選択的不感化」という方法が森田ら [2] によって提案されている. 本研究では, 選択的不感化に相当するメカニズムが脳にも存在するという仮定の下に, 下側頭葉における文脈依存的連想のモデルを構築し, モデルの挙動と Naya らの実験データとを比較検討する.

2 PACS 課題における下側頭葉ニューロンの活動

サルの実験でよく用いられる記憶課題に, 遅延見本合わせ (DMS) 課題及び遅延対連合 (DPA) 課題がある. いずれの課題でもサルは遅延期間の後に提示された刺激が target であるかどうかを判断することが求められるが, DMS 課題では target は cue と同じであるのに対し, DPA 課題では cue と対になっている刺激が target となる.

このような記憶課題を発展させたのが Naya ら [1] が考案した PACS (Pair-Association with Color Switch) 課題である. 彼らは, 以下のような手順で実験を行った.

まず, コンピュータで 12 対の異なるフラクタル図形を生成し, 図形対の一方を green, 他方を cyan で着色した. この 24 個の図形のうち 1 つを cue として短時間提示するが, target は 2 種類の課題で異なり, cue と同じ図形が target の場合は DMS 課題, cue の対図形を target とする場合は PACS 課題である.

DMS 課題では, cue 提示後 5 秒の遅延期間 (d1) を置くが, この間ディスプレイ画面に cue と同色の刺激を表示しつづける (以下, この刺激を色信号と呼ぶ). d1 が終了すると同時に画面の色が gray に変化し, 1 秒の遅延期間 (d3) の後に test 刺激が提示される. 一方 PACS 課題では, cue 提示後 test 刺激が提示されるまでの遅延期間は d1 (2 秒), d2 (3 秒), d3 (1 秒) に分かれており, d1 では色信号は cue と同じ色だが, d2 で target の色に変化し

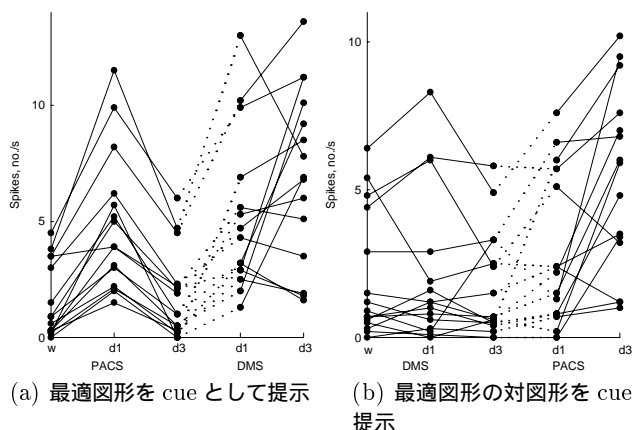


図 1: 両課題の遅延期間における各ニューロンの発火頻度 (文献 [1] より引用)

てから d3 で gray になる. いずれの課題においても, サルは提示された test 刺激が target であるかどうかを判断しなければならない.

両課題に対する正答率が十分高くなるまで訓練した 2 頭のサルについて, 課題実行時の前部下側頭葉皮質 (TE 野) におけるニューロン活動を測定した. そのうち遅延期間中に活動を示した 15 個のニューロンについて解析したところ, 図 1 に示す結果を得た.

これは, DMS 課題及び PACS 課題における遅延期間 d1 及び d3 での発火頻度をプロットしたものであり, (a) は各ニューロンの最適図形 (cue 提示による反応が最大となる図形) を, (b) はその対図形を, それぞれ cue として提示した場合である. グラフから, d1 における発火頻度は課題間でほとんど差がないが, d3 では大きく異なる, すなわち色スイッチによってその後のニューロン活動が大きく変化することがわかる.

3 選択的不感化

著者ら [3] は, 以前の研究で, DPA 課題の学習および実行のメカニズムを説明する対連合記憶の神経回路モデル (以下, DPA モデルと略称する) を構築した. しかしながら, これを PACS 課題に適用しようとする, 文脈である色信号に応じて異なる target を連想するにはどうすればよいか, という問題を解決しなければならない.

これに対して, cue 図形を表現する細胞群とは別に文脈 (色) を表現する細胞群を用意し, 両方を合わせた活動パターンから target を連想するという方法が考えられる. これは, 文脈を扱う神経回路モデルにおいてこれまで普通に用いられてきた方法であるが, ある重大な欠点がある. それが, 文脈と想起すべき target が 1 対多に対応することに起因する計算論的問題である. これについては文献 [2] で詳細に論じられているので, ここでは省略する.

† 筑波大学大学院工学研究科, Doctoral Program in Engineering, University of Tsukuba

†† 筑波大学大学院システム情報工学研究科, Doctoral Program in Systems and information Engineering, University of Tsukuba

††† 筑波大学機能工学系, Institute of Engineering Mechanics and Systems, University of Tsukuba

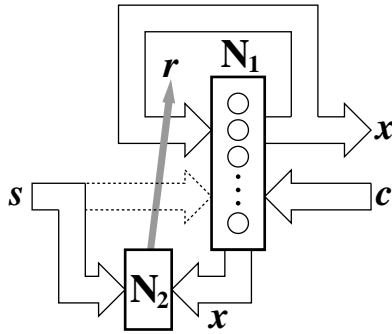


図 2: モデルのブロック図

この問題を解決する方法は、現時点で一つだけ知られている。それが選択的不感化による文脈修飾 [2] を用いる方法である。この不感化とは、細胞の感度、すなわち入力の変化に対する出力の変化を 0 にすることである。式で表すならば、 i 番目の細胞 ($i = 1, \dots, n$) の出力 x_i が

$$x_i = g_i \cdot (f(u_i) - \bar{x}_i) + \bar{x}_i \quad (1)$$

で与えられるとき、感度 $g_i = 0$ とすることである。ここで、 u_i は細胞への入力 (連続時間モデルの場合は内部ポテンシャル)、 $f(u)$ は出力関数、 \bar{x}_i はこの細胞の平均出力レベルであり、不感化された細胞は \bar{x}_i を出力することになる。一方、不感化されない細胞では $g_i = 1$ であり、通常の細胞と同じく $x_i = f(u_i)$ を出力する。そして、文脈に応じて異なる細胞群を不感化する、別の見方をすれば感度ベクトル (g_1, \dots, g_n) が文脈を表す、というのが選択的不感化による文脈修飾である。この方法を用いれば、文脈は target と直接連合されないで、上記の計算論的問題は生じない。

また、脳内で不感化を実現するメカニズムはいくつか考えられる。なかでも、細胞の出力 $x_i \geq 0$ の場合には、強い抑制を加えるという単純な方法で実現可能である。すなわち、通常の入力に加えて大きな抑制性信号を与えたとき、出力の飽和特性 (下限が 0) によって、入力に変化しても出力はほとんど変化せず $x_i \approx 0$ となる。もし細胞の平均活動レベル \bar{x}_i が 0 に近ければ、これにより近似的に不感化が実現されることになる。実際、下側頭葉では図形はスパースなパターンで表現されることが知られており、ニューロンの平均活動レベルはかなり低い。従って、下側頭葉における不感化をモデル化するにはこの方法でも十分であり、後述のモデルではこの方法を用いている。

脳は何らかの方法で文脈依存的連想の際の計算論的問題を解決しているに違いない。そして、現時点でその唯一の候補が選択的不感化による文脈修飾である。もちろん他にも存在するかもしれないが、少なくともこれが最も簡単であることは確実と思われる。更に、生物学的な根拠も少なからず存在する。従って、脳も同じ原理に基づいているという可能性は十分に考えられる。

4 モデル

以上の背景の下に、PACS 課題を遂行する神経回路モデルを構築した。本モデルは、基本的に DPA モデルに色信号による選択的不感化を導入したものである。DPA モデルと異なる点を中心に説明するので、共通部分に関する詳しい説明は文献 [3] を参照されたい。

図 2 は、モデル全体の構造を示したブロック図である。N₁ と N₂ の二つの回路網からなり、前者が TE 野に

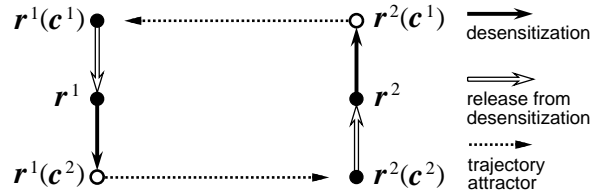


図 3: N₁ の状態空間中に形成するアトラクタの模式図

相当する。N₂ は、図形刺激に相当する刺激パターン s とともに N₁ の出力 x を受けて学習信号 r を生成し、それを N₁ に送る。本来 s は N₁ にも入力されるべきであるが、単純化のためこの直接入力 (図の破線) は省略し、N₂ を経由して外部からの入力刺激を受けるようにした。以上は DPA モデルと同じであるが、このほかに色信号 $c = (c_1, \dots, c_n)$ が N₁ に対して入力され、これによって N₁ の一部が不感化される。

回路網 N₁ は、興奮性細胞 C_i^+ と抑制性細胞 C_i^- からなるユニットが相互に結合した構造をしている。 C_i^+ は N₂ の出力 r_i 及び他のユニットからの再帰的な入力を受けて x_i を出力し、それがユニットの出力となる。 C_i^- は他のユニットからの入力に応じて y_i を出力し、 C_i^+ を feedforward に抑制する働きがある。また、ユニットの不感化は、色信号の入力を受けて C_i^- の活動が高まり C_i^+ が強く抑制されることによって実現する。

以上を式で表すと、

$$y_i = f \left(\sum_{j=1}^n w_{ij}^- x_j + \zeta c_i - \theta \right) \quad (2)$$

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^n w_{ij}^+ x_j - w_i^* y_i + \lambda r_i \quad (3)$$

$$x_i = f(u_i) \quad (4)$$

となる。ここで、 w_{ij}^+ と w_{ij}^- はそれぞれ j 番目のユニットから C_i^+ および C_i^- へのシナプス荷重、 w_i^* は C_i^- から C_i^+ への抑制性シナプスの効率、 u_i は C_i^+ の平均膜電位、 λ は N₂ の出力 r の入力強度を表す。また、 τ, ζ, θ は正の定数、 $f(u)$ は 0 から 1 の値をとるシグモイド関数である。

回路網 N₂ は、N₁ のユニットと同数の n 個の細胞からなっている。 i 番目の細胞 C_i は、刺激パターン $s = (s_1, \dots, s_m)$ および N₁ の出力パターン x をそれぞれランダムな値をとるシナプス荷重を介して受け、学習信号 r_i を出力する。また、各細胞は他の細胞への側抑制と自分自身への興奮性結合をもつ ([3] 参照)。

学習は、文献 [3] と同じであるが、N₁ の不感化されていないユニットのシナプス荷重のみ修正を行う。学習の目的は、N₁ の状態空間中に図 3 に模式的に示すようなアトラクタを形成することである。これは、刺激パターン (s^1, s^2) の対を学習した場合である。この図で、 r^1, r^2 はそれぞれ s^1, s^2 のコードであり、いずれも点アトラクタである。また、 $r^\mu(c^\nu)$ は色信号 c^ν による修飾を受けた r^μ を表し、 $r^1(c^2)$ から $r^2(c^2)$ へ、および $r^2(c^1)$ から $r^1(c^1)$ への破線の矢印は軌道アトラクタを表す。

課題の実行は、基本的に N₁ だけで行われる。但し、モデルの単純化のため、刺激パターン s^μ からそのコード r^μ への変換は N₂ を用いて行う。以下では s^1 を cue とした場合を例にとり、学習後のモデルが課題を実行する過

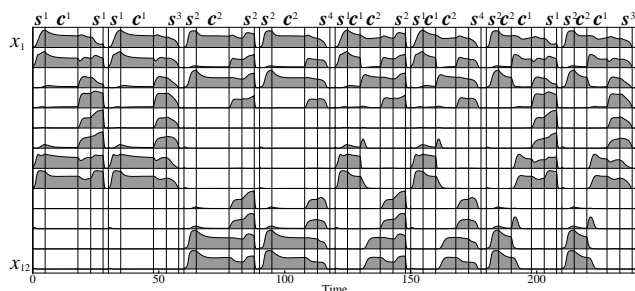


図 4: 学習後のモデルの挙動

程を説明する。

まず, s^1 を cue 入力し, 同時に (またはその後) 色信号 c^1 を与えると, $x \simeq r^1(c^1)$ となる. この状態は点アトラクタであるから, cue 入力が終わっても同じ状態が維持される. これが遅延期間 $d1$ に相当する.

DMS 課題の場合, $d1$ が終わると $d3$ となるので $c = O \equiv (0, \dots, 0)$ とする. これは, $d3$ 以降ではサルにとって図形の形だけが問題であって, 色信号は無視してよいからである. このとき x は点アトラクタ r^1 に引き込まれる. この状態で test パターンとして s^1 が再度入力されたならば, s^1 をコードするユニットが非常に強く反応するから, target を容易に識別することができる.

一方 PACS 課題の場合, $d1$ が終わるとき c が c^2 に切り替わる. x はすぐに $r^1(c^2)$ に近い状態に変化するから, $d2$ では軌道アトラクタに沿った状態遷移が生じ, x は $r^2(c^2)$ に近い状態 $r^2(c^2)$ に達する. $d3$ で $c = O$ に変化した直後の状態は安定ではないが, r^2 が点アトラクタであることから, x は r^2 に近づいていくが, 少なくとも $r^2(c^2)$ 付近に留まる. いずれにせよ x は r^2 に比較的近いから, その状態で test パターンとして s^2 を入力したならば, 他の test パターンを入力したときよりもユニットの反応が強いと考えられる.

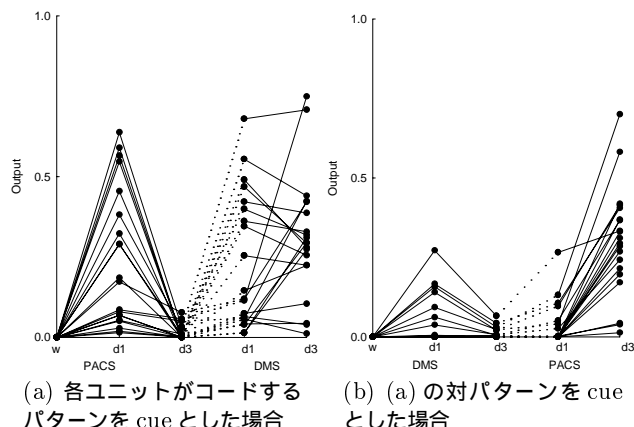
5 シミュレーション

本モデルの動作を確認するために, 計算機シミュレーションを行った. モデルのサイズ (N_1 のユニット数) は $n = 1000$, 学習する刺激パターンの次元は $m = 1000$ と設定し, 12 組のパターンを順次入力して, 20 回ずつ学習を行った.

学習終了後, すべての cue および test パターンの組み合わせで各課題を実行させた. そのうち 8 試行におけるモデルの挙動を図 4 に示す. この図は, s^1, s^2 のいずれかをコードするユニットの中からなるべく異なるタイプの活動を示すものを 12 個選び, 出力値の時間変化をプロットしたものである. 最初の 4 試行が DMS 課題, 残りの 4 試行が PACS 課題である.

このグラフから, $d3$ に強く活動しているユニットは, target パターンを test 入力したとき, 出力をさらに増加させるのに対して, それ以外のパターンを test 入力したときには出力が減少することがわかる. このような出力値の分布の違いから target を識別できることがすべての試行において確かめられた.

最後に, s^1 または s^2 をコードするユニットからランダムに 20 個を選び, 図 1 と同様にグラフ化した結果を図 5 に示す. (a) は s^1 と s^2 のうち各ユニットがコードしている刺激パターン (両方の場合には cue 反応が大きい方) を cue とした場合, (b) はその対パターンを cue とした場合であり, 各ユニットの $d1$ 及び $d3$ の最後における値がプ



(a) 各ユニットがコードするパターンを cue とした場合 (b) (a) の対パターンを cue とした場合

図 5: 両課題の遅延時間における各ユニットの出力値

ロットされている.

それぞれのグラフを図 1 の (a) および (b) と比較すると, 活動レベル (刺激提示前の発火頻度) の違いを除けば, 両者はかなりよく一致することがわかる. この図で最も注目すべき点は, (a) の DMS 課題において, 一部のユニットの出力が $d3$ で急増していることである. このような出力の増加は, $d1$ で不感化されていたユニットが $d3$ において不感化から解放されることによって生じている. 一方, 図 1(a) からわかるように, 同様の活動を示すニューロンがサルの TE 野にも存在する. DMS 課題において, 遅延期間 $d1$ に比べて $d3$ の活動が 3 ~ 4 倍も増加するというのは, 他にほとんど例を見ない現象である. しかも, そのようなニューロンが 15 個中の 4 個を占めているから, 何かの偶然やノイズが原因とは考えられない. 従って, 図 1 と図 5 のシミュレーション結果との一致は, この現象に初めて説明を与えると同時に, 選択的不感化が TE 野で生じていることを示す極めて有力な証拠である.

6 おわりに

本研究の結果は, 下側頭葉において選択的不感化が用いられていることを強く示唆する. そして, 同じ原理が前頭前野をはじめとする脳の様々な領域で普遍的に使われている可能性は十分にあると思われる.

このことは, 脳の情報処理機構の解明に重要であると同時に, 脳のように文脈の変化に柔軟に対応可能な情報処理システムの実現にとっても大きな意味をもつ. すなわち, 既に選択的不感化を用いた人工神経回路網 [2] によって従来型の記号処理の限界を打破できる可能性が指摘されているが, 脳が同じ原理に基づいているならば, その可能性は更に高まると言える.

参考文献

- [1] Y. Naya, K. Sakai, and Y. Miyashita, "Activity of primate inferotemporal neurons related to a sought target in pair-association task," Proc. Natl. Acad. Sci. USA, vol.93, pp.2664-2669, 1996.
- [2] 森田昌彦, 松沢浩平, 諸上茂光, "非単調神経素子の選択的不感化を用いた文脈依存的連想モデル," 信学論 (D-II), vol.J85-D-II, no.10, pp.1602-1612, 2002.
- [3] 末光厚夫, 森田昌彦, "下側頭葉における対連合記憶の神経回路モデル," 信学論 (D-II), vol.J84-D-II, no.4, pp.718-727, 2001.