離散的かつ連続的な情報表現を持つ神経回路モデル
Neural network model with discrete and continuous information representation

北園 淳 (PY)[†], 大森 敏明 [†,‡], 岡田 真人 [†,‡]
Jun Kitazono(PY), Toshiaki Omori, and Masato Okada
† 東京大学新領域創成科学研究科, ‡ 理化学研究所 脳科学総合研究センター
kitazono@mns.k.u-tokyo.ac.jp

Abstract— Associative memory model has discretely distributed fixed-point attractors. In other words, discrete information representation is used in AM model. On the other hand, several types of neural networks with Mexican-hat type interaction have been studied to model continuous information representation, found in, for example, working memory and columnar activity in the visual cortex. Recently, the presence of both discrete and continuous information representation in temporal lobe is suggested. Here we propose associative memory model with Mexican-hat typelike interaction, and succeed in achieving both discrete and continuous information representation.

Keywords— **Mexican-hat type interaction, Associative memory, Localized activity, Statistical mechanics**

1 はじめに

連想記憶モデルとメキシカンハット型相互作用を持つモデルは，二大アトラクターネットワークである．全ての構造を持つアトラクターネットワークは，この二つのいずれかのモデルに属するといっても過言ではない．メキシカンハット系は初期視覚野のハイパーコラムやメモリーガイドサッケードの記憶保持中の前頭葉など，外界の距離関係を反映した中立安定で連続的なアトラクター構造を持つ．一方，連想記憶モデルは Hopfield モデルに代表されるような離散的で点状のアトラクター構造を持つ．

しかしながら，世の中には離散的でなおかつ連続的な情報表現を持つものも多い．たとえば，ある視覚物体を回転させた場合，その回転に対応する脳内情報表現の変化は連続的であろう．一方，その視覚物体を他の物体と識別するには，二つの異なった物体間での情報表現の変化は離散的でなければならない．Tamura[1] らの電気生理実験からも，側頭葉に離散的で連続的な情報表現が存在することが示唆されている．これら考察から，離散的で連続的な情報表現に対応する離散的で連続的なアトラクター構造を持つ神経回路モデルを議論する．

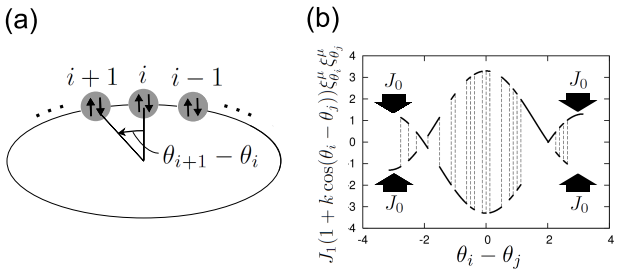

図 1: (a) N 個のスピンが円環状に配置されている．(b) $J_1(1+k\cos(\theta_i-\theta_j))\xi^\mu_{\theta_i}\xi^\mu_{\theta_j}$ は, $\theta_i-\theta_j=0$ の周辺では, スピンを記憶パターン μ に合わせるように, $\theta_i-\theta_j=\pm\pi$ の周辺では記憶パターン μ と逆向きになるように作用する．J_0 は強磁性相互作用である．

ここでは，その一例として相関学習で埋め込んだ複数個の記憶パターンを，メキシカンハット型相互作用で変調することを考える．このモデルの平衡状態は一つの記憶パターンと有限のオーバーラップをもつ．この平衡状態は，従来のメキシカンハット系と同様に，発火の重心位置に関して中立安定であり，連続的な情報表現に対応する連続的なアトラクター構造を持つ．一方，異なった記憶パターンに対応する平衡状態は，たとえ発火の重心位置が同じであっても，オーバーラップは 0 となり, 互いに直交する．これらの性質は, この系のアトラクターが離散的な情報を表現できることを意味する．

2 モデル

N 個のスピンからなる連想記憶モデルを考える．各スピンは $[-\pi,\pi)$ の周期境界条件で一次元状に等間隔に配置されている (図 1(a)). 系のエネルギーは,

$$H(\{S\}) = -\frac{1}{2}\sum_{i\neq j} J_{\theta_i\theta_j} S_{\theta_i} S_{\theta_j} \tag{1}$$

で定義され, 相互作用 $J_{\theta_i\theta_j}$ は以下のようである.

$$J_{\theta_i\theta_j} = \frac{J_0}{N} + \frac{J_1}{N}\sum_{\mu=1}^{p}(1+k\cos(\theta_i-\theta_j))\xi^\mu_{\theta_i}\xi^\mu_{\theta_j} \tag{2}$$

ここで, J_0 は強磁性相互作用, J_1 は Hebb 則的な相互作用, k はそのうち, 空間局所的な相互作用の強さを表す (図 1(b)). p は埋め込んだパターン数である．また, 記

憶パターンの各成分 $\xi^\mu_{\theta_i}$ は ±1 の値を確率 $\mathrm{Prob}[\xi^\mu_{\theta_i} = \pm1] = \frac{1}{2}$ でとる.

3 解析

分配関数 Z は以下のようになる.

$$Z = \left(\frac{\beta N}{2\pi}\right)^{(1+3p)/2} J_0^{1/2} J_1^{3p/2} k^p \; e^{-\beta(J_0+(1+k)pJ_1)/2} \int\cdots\int_{-\infty}^{\infty} \mathrm{d}m_0 \prod_{\mu=1}^{p} \mathrm{d}m^\mu \mathrm{d}m_c^\mu \mathrm{d}m_s^\mu \exp\left[-\beta N\left\{\frac{J_0}{2}m_0^2 + \frac{J_1}{2}\sum_{\mu=1}^{p}\left\{{m^\mu}^2 + k\left({m_c^\mu}^2 + {m_s^\mu}^2\right)\right\} - \frac{1}{\beta N}\sum_{i=1}^{N}\log\left(2\cosh\beta\tilde{h}(\theta_i)\right)\right\}\right] \quad (3)$$

$$\tilde{h}(\theta) = J_0 m_0 + J_1 \sum_{\mu=1}^{p} \xi_\theta^\mu \left(m^\mu + k(m_c^\mu \cos\theta + m_s^\mu \sin\theta)\right) \quad (4)$$

$N \to \infty$ の熱力学極限において, 式 (4) を鞍点法によって評価すると, $Z \sim \exp(-\beta N f)$ となる. ただし f は自由エネルギーで,

$$f = \frac{J_0}{2}m_0^2 + \frac{J_1}{2}\sum_{\mu=1}^{p}\left\{{m^\mu}^2 + k\left({m_c^\mu}^2 + {m_s^\mu}^2\right)\right\} - \frac{1}{2\pi\beta}\int_{-\pi}^{\pi}\mathrm{d}\theta \log\left(2\cosh\beta\tilde{h}(\theta)\right) \quad (5)$$

と表わされる. ここで, $m_0, m^\mu, m_c^\mu, m_s^\mu$ はそれぞれ $m_0 = \frac{1}{N}\sum_i S_{\theta_i}$, $m^\mu = \frac{1}{N}\sum_i \xi^\mu_{\theta_i} S_{\theta_i}$, $m_c^\mu = \frac{1}{N}\sum_i \xi^\mu_{\theta_i} S_{\theta_i}\cos\theta_i$, $m_s^\mu = \frac{1}{N}\sum_i \xi^\mu_{\theta_i} S_{\theta_i}\sin\theta_i$ なるオーダーパラメータで, 鞍点条件より以下を満たす.

$$m_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\theta \tanh\beta\tilde{h}(\theta) \quad (6)$$

$$m^\mu = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\theta\, \xi_\theta^\mu \tanh\beta\tilde{h}(\theta) \quad (7)$$

$$m_c^\mu = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\theta\, \xi_\theta^\mu \cos\theta \tanh\beta\tilde{h}(\theta) \quad (8)$$

$$m_s^\mu = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{d}\theta\, \xi_\theta^\mu \sin\theta \tanh\beta\tilde{h}(\theta) \quad (9)$$

また, $m_L^\mu = \sqrt{(m_c^\mu)^2 + (m_s^\mu)^2}$, $\phi^\mu = \arctan(m_s^\mu/m_c^\mu)$ と定義する.

4 結果とまとめ

鞍点条件 (6)-(9) を用いて分岐解析を行った結果, $J_0 : J_1 : kJ_1 \approx 1 : 1 : 2.5$ 付近に, $(m_0 \lessgtr 0,\ m^\mu \gtrless 0,\ m_L^\mu > 0,\ m^\nu = m_L^\nu = 0 (\nu \neq \mu))$ を満たす相 (以下 LR:Localized Retrieval 相と呼ぶ) が存在することがわかった. この LR 相において, 平均場近似 $< S_{\theta_i} >= \tanh\beta\tilde{h}(\theta_i)$ を用いてスピンの期待値を計算すると, 図 2 のようなスピン配位が得られる. これは, ある特定の i 番目のスピン S_{θ_i} の周囲 ($\theta_i - \theta_j \approx 0$ 付近) では記憶パターンに合わせる作用が優位に働き, S_{θ_i} から離れた位置 ($\theta_i - \theta_j \approx \pm\pi$ 付近) では強磁性相互作用が優位に働き, 上手くバランスのとれた状態であるといえる (図 1(b) 参照).

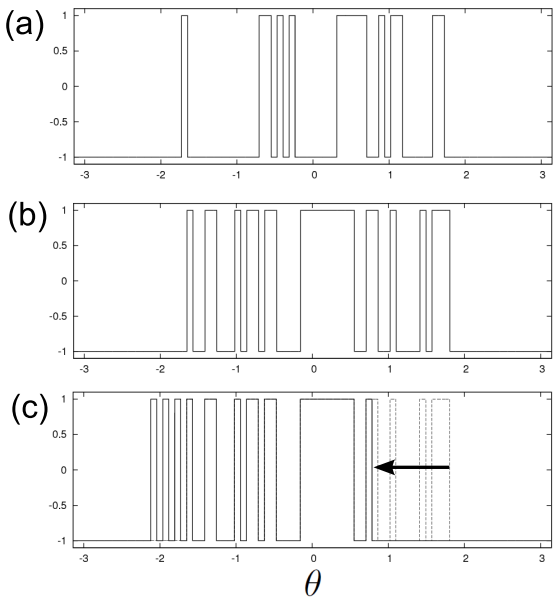

図 2: LR 相における局在想起状態のスピン配位の期待値. (a)(b)(c) はそれぞれ, (a) パターン 1 が $\theta = 0$ の周囲でのみ, (b) パターン 2 が $\theta = 0$ の周囲でのみ, (c) パターン 2 が $\theta = -\pi/4$ の周囲でのみ, 想起されている状態である. (c) の点線は, (b) におけるスピン配位の期待値. また, パターン 1 と 2 は互いに直交している.

図 2(a) では, パターン 1 が $\theta = \phi^1 = 0$ の周囲でのみ局在的に想起されている. 図 2(b) では, パターン 2 が $\theta = \phi^2 = 0$ の周囲でのみ, 図 2(c) では, パターン 2 が $\theta = \phi^2 = -\pi/4$ の周囲でのみ, 局在的に想起されており, 図 2(c) は図 2(b) に比べて, 発火の重心位置 ϕ^2 が $-\pi/4$ だけずれている.

ある特定のパターン μ について, 任意の発火の重心位置 ϕ^μ に対し自由エネルギーは同じ値をとり, ϕ^μ に関して連続的にアトラクターが分布している (これをラインアトラクターと呼ぶ). この意味で本モデルは, 連続的な情報を表現している. また, 異なるパターン同士は互いに直交しており, それぞれのパターンに対応するラインアトラクターが離散的に分布している. この意味で本モデルは, 離散的な情報を表現している.

以上のように, 今回提案したモデルは, 側頭葉での存在が示唆されている離散的かつ連続的な情報表現を持つといえる.

参考文献

[1] H. Tamura, H. Kaneko, K. Kawasaki, I. Fujita (2004) Journal of Neurophysiology, **91**, 2782–2796

[2] K. Wada, K. Kurata, M.Okada (2004) Neural Networks, **17** 1039–1049