

スパースコーディングされた時系列パターンの想起のダイナミクス

北野勝則 青柳富誌生

京都大学 工学研究科 数理工学専攻

Hopfield モデルに代表される数理モデルとしてのニューラルネットワークは、これまでさまざまな点で現実的な改良をされてきたが、ネットワークで処理される情報（連想記憶モデルの場合パターンと呼ぶ）もその重要な一つとして取り上げられてきたのは、主として次に挙げる2つの理由と考えられる。まず第一に、解析の容易さのためこれまでに扱われてきたような一様ランダムに生成されるパターン、つまり“foreground”と“background”のビットが等しい確率で生成されるようなパターンは、情報処理の観点から現実的でない、第二に、生理学の知見により、現実の神経系の活動度は非常に低いと考えられていることである。こうした点を考慮に入れたモデルとして、スパースコーディングされたパターンを処理するネットワークが提案され、研究されてきた。このネットワークが注目を集めていたのは、処理するパターンの活動度が低くなるにつれ、ネットワークに記憶できるパターン数（記憶容量）飛躍的に増大し、ネットワークの素子数に対するアクティブなニューロンの割合を f とすると、この記憶容量は、 $f \rightarrow 0$ の極限で $1/f |\log f|$ の漸近特性を持って発散し、この漸近形は活動度 f のパターンを保持するネットワークの最適オーダーであることが調べられている。しかしながら、この場合、1パターン当たりの情報量自体は減少していることを考えると、むしろ注目すべきは上述の最適オーダーが Hebb 則で学習されたネットワークについて得られていることであろう。このように、平衡状態の性質はかなり理解が進んでいるが、ダイナミカルな性質にはまだ不明な点が多いと言わざるを得ない。特に、連想記憶モデルの場合、初期に与えられたノイズなパターンが動的に修正されていくのを考えると、動的な特性である引き込み領域なども議論することは、ネットワークの特徴を正しく理解する上で必要不可欠である。近年、想起過程ダイナミクスを扱ういくつか

の理論が提案されてきているが、特に統計神経力学による手法は、想起成功の場合長時間の振舞いを記述でき、実践的である。しかしながら、スパースコーディングにおいては、クロストークノイズの扱いの難しさから、自己想起モデルに対する解析結果は計算機実験と若干の定量的な隔たりが見られている。そこで、本研究では、相互想起モデルの1つである、系列想起モデルについて、この理論を適用してみる。

まず、 N 個のニューロンで構成されるネットワークが次の離散時間同期ダイナミクスに従うとする。

$$\begin{aligned} S_i(t+1) &= \Theta(h_i(t)) \\ h_i(t) &= \sum_j J_{ij} S_j(t) - \theta \end{aligned}$$

$S_i(t), h_i(t)$ はそれぞれ、時刻 t における i 番目のニューロンの状態、内部ポテンシャルであり、また、 θ, J_{ij} は閾値、シナプス結合を表している。記憶するパターンは

$$P(\xi_i^\mu) = a\delta(\xi_i^\mu) + (1-a)\delta(\xi_i^\mu)$$

で生成され、 ξ_i^μ はパターン μ において i 番目のニューロンがとるべき値である。このとき1パターンの活動度は、 $\frac{1}{N} \sum_j \xi_j^\mu = a$ となるが、 $a = 1/2$ をランダムパターン、 $a \neq 1/2$ をバイアスパターンと呼び、特に $a \rightarrow 0$ の極限をスパースコーディングという。このようなパターンの系列 $\xi^1 \rightarrow \xi^2 \rightarrow \dots$ を保持するためには

$$J_{ij} = \frac{1}{AN} \sum_\mu (\xi_i^{\mu+1} - a)(\xi_j^\mu - a) - \frac{g}{aN}$$

と定めれば良いと知られている。ここで、 $A = a(1-a)$ は規格化因子。右辺第二項は一様な抑制結合を表し、ネットワークの活動レベルをコントロールする効果を持つ。この系の想起過程は巨視的な状態を表す次のパラメータ

$$m(t) = \frac{1}{AN} \sum_j (\xi_j^t - a) S_j(t)$$

$$x(t) = \frac{1}{aN} \sum_j S_j(t)$$

によって記述される。 $m(t)$ は注目しているパターン ξ^p と系の類似度, $x(t)$ は系の活動度を表している。このパラメータを用いると内部ポテンシャルは

$$h_i(t) = (\xi_i^{p+1} - a)m^p(t) - gx(t) - h + z(t)$$

と分離でき、最後の項は注目パターン以外からのクロストークノイズを表している。この理論の本質は $z(t)$ を平均 0, 分散 $\sigma(t)^2$ のガウスノイズで近似することである。一般には $z(t)$ の時間相関が存在し、自己想起モデルではこの時間相関が想起に大きな影響を与えるが、相互想起モデルの場合、この時間相関の寄与は無視できる。よって、系の想起過程を記述する方程式は

$$m(t+1) = 1 - \frac{1}{2} \{ \text{erfc}(\phi_1) + \text{erfc}(\phi_0) \}$$

$$x(t+1) = 1 - \frac{1}{2} \left\{ \text{erfc}(\phi_1) - \frac{1-a}{a} \text{erfc}(\phi_0) \right\}$$

$$U(t) = \frac{1}{\sqrt{2\pi}\sigma(t)} \left\{ ae^{-\phi_1^2} + (1-a)e^{-\phi_0^2} \right\}$$

$$\sigma(t+1)^2 = \alpha ax(t+1) + U(t)^2 \sigma(t)^2,$$

ここで、

$$\begin{aligned} \text{erfc}(u) &= \frac{2}{\sqrt{\pi}} \int_u^\infty e^{-u^2} du \\ \phi_1 &= \frac{(1-a)m(t) - gx(t) - h}{\sqrt{2}\sigma(t)} \\ \phi_0 &= \frac{am(t) + gx(t) + h}{\sqrt{2}\sigma(t)}, \end{aligned}$$

である。得られた理論的結果と計算機実験を比較したのが図 1 である。図の 3 本の曲線は上から、平衡活動率、平衡オーバーラップ、引き込み領域を表す。これより、得られた方程式は、系の振舞いの良い予測を与え得る。さらに、閾値 h のみの場合と（図 1 上）、抑制結合 g を導入した場合（図 1 下）を比較すると、記憶容量にはその差はあまり見られないが、その効果は引き込み領域の大きさに現れており、抑制結合は引き込み領域を広げる効果があることが確認される。また、導出した方程式により、このモデルの場合も $a \rightarrow 0$ で $\alpha_c \sim \frac{1}{a|\log a|}$ となることも確認されている。

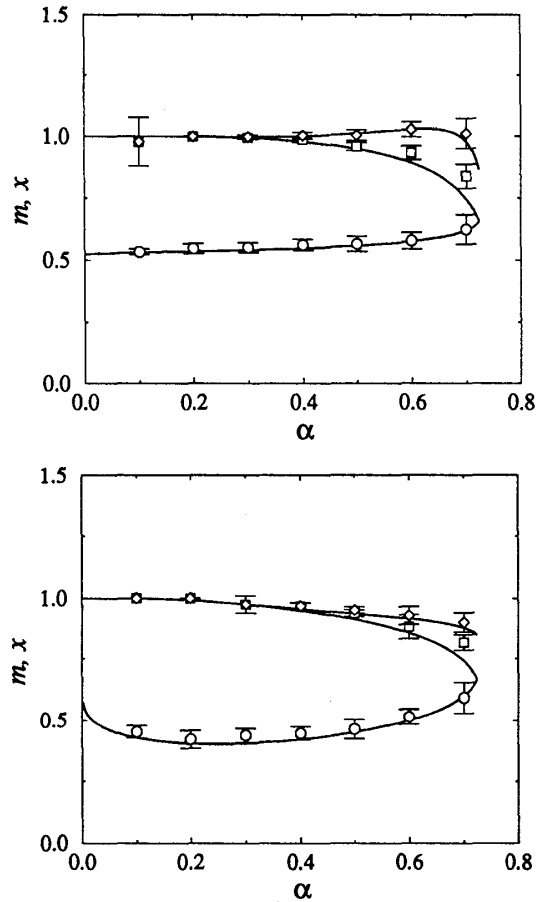


図 1: 理論解析と計算機実験の比較

参考文献

- E. Gardner, *J. Phys. A*, **21**, 257, (1988)
- M.V. Tsodyks and M.V. Feigelman, *Europhys. Lett.*, **6**, 101, (1988)
- C.J.P. Vicente and D.J. Amit, *J. Phys. A*, **22**, 559, (1989)
- J. Buhmann, R. Divko and K. Shulten, *Phys. Rev. A*, **39**, 2689, (1989)
- S. Amari and K. Maginu, *Neural Networks*, **1**, 63, (1988)
- M. Okada, *Neural Networks*, **8**, 833, (1995)