

ニューロ・フロー・ダイナミクスによる学習とその特徴

早大理工 龍野正実、相澤洋二

1. 序

生物の神経系におけるシナプスの可塑性は、基本的には Hebb 則によって説明できると考えられている。また、情報はニューロンの発火の時空パターンやニューロン群の同期現象に担われていると推測されている。そしてこれらの仮説を確かめるために多くの実験的、理論的研究が行われ、優れた成果が数多く報告されている。しかしモデルを用いた理論的研究に関しては、実際の生物の神経回路網とコンピュータ上に実現されるニューラル・ネットワーク・モデルとの間の大きな隔たりのために、モデル研究から得られた結果を生物の神経系の理解にどのように役立てたらよいか大きな問題となっている。

我々は、現在特に興味を持っている学習のメカニズムについて基本的データがまだ十分ではないと考え、実際の神経系をマクロスコピックにとらえた単純なモデルを用いて学習の性質を詳しく調べてみた。その際、生物の神経系が常に外界から情報を受け取る「流れ」の中にあることに注目して、「流れ」を意識したモデル化を試みた。すなわち、「結合の変化は二つのニューロンの発火状態に依存する」という標準的な Hebb 則に加えて、「結合の変化はニューロン間を伝わるフローにも依存する」という視点に立ったモデルをつくり、その学習の様子を詳しく調べた。

2. モデル化の方法

ニューロン間の結合の可塑性の生物学的要因としては、

- 1 回の活動電位がシナプス終末に流入させるカルシウムイオンの増加
- シナプス後レセプターの活性効率の変化
- レセプター数の変化
- シナプスの数や位置の変化

などが考えられるが、これらの変化はニューロンの発火状態に加えて「ニューロン間の樹上突起や軸索を伝わる電気インパルスなどに依存する」と考える。ここで電気インパルスを含むニューロン間の入出力を「フロー」と呼び、フローを主要な変数にとったニューラル・ネットワーク・モデルを用いて「フローの寄与も考慮した」学習則の性質を詳しく調べた。今回は Hopfield モデル（時間離散、状態変数離散、非同期）を基本とした状態変数連続の 1 層ニューラル・ネットワーク・モデルを構成し、シミュレーションではニューロン数を 100、記憶させるパターン数を 10 とした。

3. フロー・ダイナミクスとフローの寄与を考慮した学習

今回考えたフロー・ダイナミクスは以下のように与えられる。

$$X_{ki}(t+1) = w_{ki}(t) f_T \left(\sum_{j=1}^N X_{ij}(t) - h_i(t) \right)$$

ここで、

$X_{ki}(t)$: i 番目のニューロンから k 番目のニューロンへの時刻 t におけるフロー。

$w_{ki}(t)$: i 番目のニューロンから k 番目のニューロンへの時刻 t における結合の強さ

$h_i(t)$: i 番目のニューロンの時刻 t における閾値

$f_T(x) = \frac{1}{1+\exp(-x/T)}$: ニューロンの応答関数

である。

また学習方程式は、Hebb 則を用い減衰を考慮した学習則

$$w_{ij}(t+1) = Aw_{ij}(t) + C\xi_i\xi_j$$

を

$$w_{ij}(t+1) = Aw_{ij}(t) + BX_{ij}(t+1) + C\xi_i\xi_j$$

のように拡張しフローの寄与を付け加える。ただし、 $A < 1.0$ 、 B 、 C はそれぞれ定数である。具体的なモデルは以下の 4 種類である。

- モデル 1 (結合の強さが学習項、フローの項と同じ速さで変化し、単純に減衰する場合)

$$w_{ki}(t+1) = Aw_{ki}(t) + BX_{ki}(t+1) + C\xi_k(t+1)\xi_i(t+1)$$

- モデル 2 (結合の強さが学習項、フローの項と同じ速さで変化し、平均場的に結合の大きさを調整する場合)

$$w_{ki}(t+1) = w_{ki}(t) + b \left(1 - a \frac{\sum_{i,j=1}^N |w_{ij}(t)|}{N^2} \right) w_{ki}(t) + BX_{ki}(t+1) + C\xi_k(t+1)\xi_i(t+1)$$

- モデル 3 (結合の強さが学習項、フローの項それぞれの平均と共に変化し、単純に減衰する場合)

$$w_{ki}(t+1) = Aw_{ki}(t) + \frac{B}{p} \sum_{\mu=1}^p X_{ki}(t+\mu) + \frac{C}{p} \sum_{\mu=1}^p \xi_k(t+\mu)\xi_i(t+\mu)$$

- モデル 4 (結合の強さが学習項、フローの項それぞれの平均と共に変化し、平均場的に結合の大きさを調整する場合)

$$w_{ki}(t+1) = w_{ki}(t) + b \left(1 - a \frac{\sum_{i,j=1}^N |w_{ij}(t)|}{N^2} \right) w_{ki}(t) + \frac{B}{p} \sum_{\mu=1}^p X_{ki}(t+\mu) + \frac{C}{p} \sum_{\mu=1}^p \xi_k(t+\mu)\xi_i(t+\mu)$$

4. 学習の特徴

数値計算により明らかになったフローの寄与を考慮した学習の特徴をまとめると、次のようになる。

1. フローの寄与が無い場合にはアトラクターの Basin Size 分布がベキ的となり多数の Spurious State が存在するのに対して、フロー学習ではごく少数の記憶パターンのみがドミナントとなり、Spurious State が激減する。このことはどの学習モデルでも共通に見られる性質である。
2. モデル1、モデル2のように記憶パターンを順番に覚えさせる場合には、フローの寄与が大きくなるにつれて最初に入力した記憶パターンが最もよく学習される初頭効果 (Primacy Effect) が現れる。さらに、初頭効果と新近効果 (Recency Effect、最後の方で学習したことをよく記憶する効果) がともに現れる領域も存在する。これは Hebb 則と減衰項の場合に新近効果だけが見られることと大きく異なっている。
3. モデル3、モデル4では、最初に大きな Basin を持つ記憶パターンがフローによってより大きな Basin を持つ様子が観測された。しかしなぜ特定の記憶パターンだけが記憶されるのかは分かっていない。

なお、モデル2の場合の想起の様子を Fig.1 に示す。この図は学習を終えた直後の状態で結合の強さを固定し、初期点をランダムに1,000個ふったときの各記憶パターンへの想起数を示している。横軸がフローの寄与の程度を示す B の大きさ、縦軸が想起数である。フローの寄与が大きくなるに従って最初に入力した記憶パターンの Basin が大きくなっていくのが分かる。また $0.07 \leq B$ では、最初に入力したパターンと最後の方で学習したパターンの両方をよく記憶していることが分かる。

5. 学習終了後のアップデート (スリープ)

学習方程式 $w_{ij}(t+1) = Aw_{ij}(t) + BX_{ij}(t+1) + C\xi_i\xi_j$ において、右辺第3項の学習項がない場合のフロー学習をスリープと呼ぶ。ここでは外部からの入力遮断された状態においてネットワークが自分自身で結合の状態を変化させるとき、記憶がどのように変化に変化するのかに注目する。

スリープ中の結合の強さの変化は、例えばモデル1の場合には

$$\begin{aligned} w_{ki}(t+1) &= Aw_{ki}(t) + BX_{ki}(t+1) \\ &= Aw_{ki}(t) + Bw_{ki}(t)f_T\left(\sum_{j=1}^N X_{ij}(t) - h_i(t)\right) \\ &= w_{ki}(t)\left[A + Bf_T\left(\sum_{j=1}^N X_{ij}(t) - h_i(t)\right)\right] \end{aligned}$$

より

$$w_{ki}(t+1) \geq w_{ki}(t); \quad \text{if } f_T(\sum_{j=1}^N X_{ij}(t) - h_i(t)) \geq \frac{1-A}{B}$$

$$w_{ki}(t+1) \leq w_{ki}(t); \quad \text{if } f_T(\sum_{j=1}^N X_{ij}(t) - h_i(t)) \leq \frac{1-A}{B}$$

にしたがって変化することが分かる。他のモデルの場合も同様に評価できる。

6. スリープの効果

スリープの効果の特徴をまとめると次のようになる。

1. スリープによって、弱いフローでも特定の入力パターンをよりよく学習するようになるという結果が得られた。この結果はどの学習モデルでも共通に見られる特徴である。
2. モデル1、モデル2では最初に入力した記憶パターンが弱いフローでもドミナントになる。すなわち、初頭効果がさらに強められる。ここで心理学における仮説：「初頭効果は長期記憶に関係しており、新近効果は短期記憶に関係している。また睡眠中には短期記憶から長期記憶への移行が行われている」を考えあわせると、モデルにおいて初頭効果、すなわち長期記憶がスリープ中にさらに強まるという結果はかなり興味深いと思われる。
3. モデル3、モデル4では最初に大きな Basin をもつ記憶パターンが弱いフローでもドミナントになる。

なお、モデル2におけるスリープ100回後の結合を用いた想起の様子を Fig.2 に示す。図から明らかなように、フローの寄与が弱い場合でも最初に入力したパターンの Basin が大きくなっていることが分かる。

7. 考察と今後の課題

10組の記憶パターンの学習を n 回繰り返した場合の結合の強さは、例えばモデル1の場合には近似的に、

$$w_{ki}(10n) = C(A + \frac{B}{2})^9 [(A + \frac{B}{2})^{10(n-1)} + (A + \frac{B}{2})^{10(n-2)} + \dots + 1] \xi_k^1 \xi_i^1$$

$$+ C(A + \frac{B}{2})^8 [(A + \frac{B}{2})^{10(n-1)} + (A + \frac{B}{2})^{10(n-2)} + \dots + 1] \xi_k^2 \xi_i^2$$

$$+ \dots$$

$$+ C(A + \frac{B}{2})^{10(n-1)} + (A + \frac{B}{2})^{10(n-2)} + \dots + 1 \xi_k^{10} \xi_i^{10}$$

と書くことができる。上式よりフローの項が無い Hebb 則 ($B = 0$) では後から入力されるパターンがより強く記憶されることが分かる。また、フローが加わると $A + B/2 > 1$ となりうるため逆に初頭効果が生じることも理解できる。

さらに l 回スリープさせると結合の強さは近似的に、

$$\begin{aligned}
 w_{ki}(10n+l) &= C\left(A + \frac{B}{2}\right)^{l+9} \left[\left(A + \frac{B}{2}\right)^{10(n-1)} + \left(A + \frac{B}{2}\right)^{10(n-2)} + \dots + 1 \right] \xi_k^1 \xi_i^1 \\
 &+ C\left(A + \frac{B}{2}\right)^{l+8} \left[\left(A + \frac{B}{2}\right)^{10(n-1)} + \left(A + \frac{B}{2}\right)^{10(n-2)} + \dots + 1 \right] \xi_k^2 \xi_i^2 \\
 &+ \dots \\
 &+ C\left(A + \frac{B}{2}\right)^l \left[\left(A + \frac{B}{2}\right)^{10(n-1)} + \left(A + \frac{B}{2}\right)^{10(n-2)} + \dots + 1 \right] \xi_k^{10} \xi_i^{10}
 \end{aligned}$$

となる。Hebb 則のようにフローのない場合にはただ減衰するだけであるが、フローが加わっている場合には初頭効果をさらに強める効果があることが分かる。またその他のモデルについても同様に解析することが可能である。

この研究によってフローの寄与を考慮した学習則の興味深い特徴が明らかになりつつあるが、これらのモデルと実際の脳神経系が行っている記憶情報処理の間には依然として大きな開きがある。例えば脳の実質核は 3～4 段階にわたって投射する複雑な 3 次元構造物であり、ここで取り上げたような簡単なモデルとは全く違った処理をしている可能性が高い。また、常に外部からの情報の「流れ」の中にあって記憶を想起するモデルもまだできていない。しかしこのような簡単なモデルにもかかわらず、フローを考慮することで Spurious State の激減や初頭効果、新近効果が見られることは、非常に興味深いと思われる。現在は実際の海馬により近いモデルを構築し、その学習の基本的性質を明らかにする努力をしているところである。

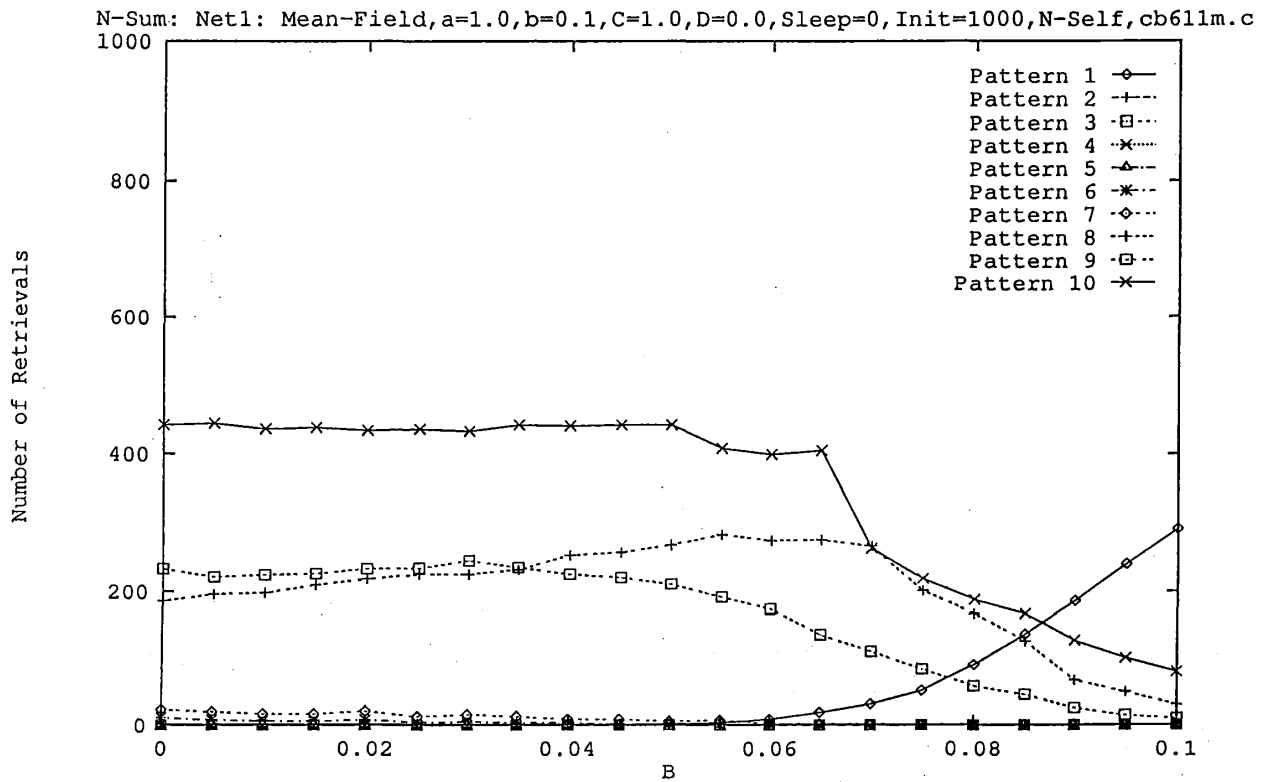


Fig.1 学習直後の想起の様子 (モデル 2)

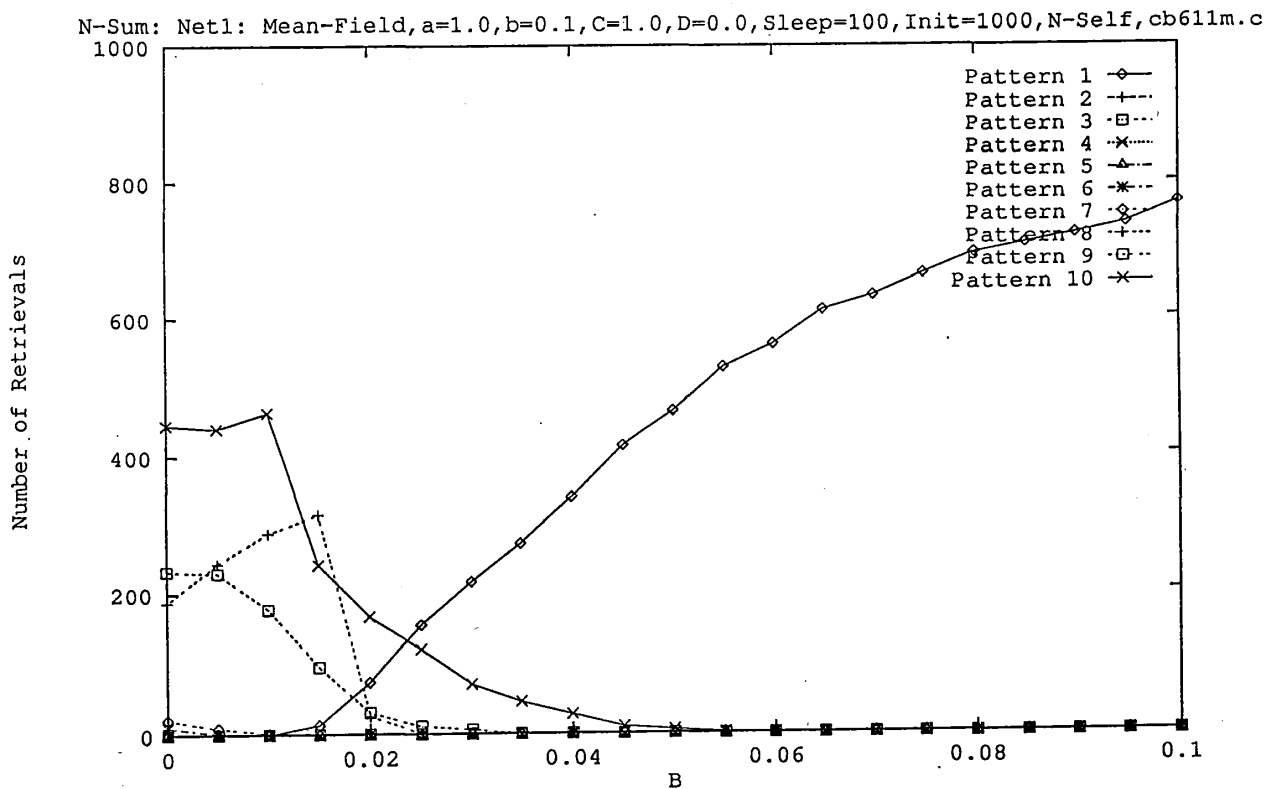


Fig.2 スリープ 100 回後の想起の様子 (モデル 2)