相関学習における結合荷重の最適な減衰率について

赤穂 昭太郎
電子技術総合研究所

Optimal Decay Rate of Connection Weights in Covariance Learning

Shotaro Akaho
Electrotechnical Laboratory

Abstract — Associative memory cannot store items more than its memory capacity. When new items are given one after another, connection weights should be decayed so as not to exceed the memory capacity. This report presents the optimal decay rate that maximizes the number of stored items.

1. はじめに

記憶容量は記憶モデルとしてのニューラルネットワークの能力をはかる目安となるばかりでなく，ネットワークの汎化能力とも関係しており重要な問題となっている．また，連想記憶で出力パターンをスパース符号化しておくと記憶容量が非常に大きくなることが知られており注目されている [1].

しかしながら，ニューラルネットワーク自身は記憶しているパターンの個数を覚えているわけではなく，記憶容量を越えたかどうかを判定できない．従って，次々に新しいパターンを覚えていったときに，記憶容量を越えないよう自動的に古いパターンを忘れていくような機構が必要となる．

さて，n 個のニューロンからなる 2 層の自己相関型の連想記憶の結合荷重は次のように単純な Hebb 則 (相関学習) で強化される．ただしここでは離散時間で考える．

$$w_{ij}(t+1) = aw_{ij}(t) + cx_i(t)x_j(t), \qquad i \neq j, \qquad 0 < a < 1, \quad 0 < c \tag{1}$$

ここで w_{ij} は i 番目の入力から j 番目のニューロンへの結合荷重であり，$x_i(t)$ は時刻 t に提示されたパターンの第 i 成分，a と c は時定数である (以下一般性を失うことなく $c=1$ と仮定する)．(1) 式の右辺第 1 項は結合荷重の値が発散しないように便宜的に入れられた減衰項であるが，もっと積極的に解釈すれば，忘却によって学習空間の構造を抽出したり記憶容量を越えるのを防ぐ働きをしていると見ることもできる．以下ではランダムなパターンを次々と覚えていった時に，記憶容量がもっとも大きくなるように減衰率の値を決定するという問題を考える．

2. 問題設定

まず，学習パターンの各要素 $x_i(t)$ は ± 1 を確率 1/2 で独立かつランダムにとるとする．このとき，時刻 $t+1$ まで経った時に結合荷重は

$$w_{ij}(t+1) = x_i(t)x_j(t) + ax_i(t-1)x_j(t-1) + \cdots + a^m x_i(t-m)x_j(t-m) + \cdots \tag{2}$$

になっている．パターンは指数的に減衰していくため，昔に覚えたものほど想起される確率は低くなる．最も新しく覚えた m 個のパターンが想起される確率が (n が大きいとき漸近的に) 1 となるような m の最大値 M を**記憶容量**と呼ぶ．減衰が速過ぎると昔のパターンが想起できなくなるし，遅過ぎると昔のパターンの影響で新しいパターンの想起さえできなくなる．従って問題は，記憶容量 M を最大にするような減衰率 a およびそのときの M の値を求めることである．

3. 主要結果

最適減衰率および記憶容量の bound として次の定理が成り立つ.(証明は省略)

Theorem 最適な減衰率 a_{opt} は n が十分大きい時,

$$a_0 = 1 - \frac{4\log n}{n} \leq a_{\mathrm{opt}} \leq 1 - \frac{2\log n}{n} = a_1 \tag{3}$$

Keywords — 連想記憶, 記憶容量, 相関学習, スパース符号化, 減衰率; Associative memory, Memory capacity, Covariance learning, Sparse coding, Weight decay

の間にあり, 記憶容量 M_{opt} は

$$M_1 = \frac{n}{8e\log n} \le M_{\mathrm{opt}} \le \frac{n}{4e\log n} = M_0 \tag{4}$$

の間にある. 減衰率として a_1 を使った場合の記憶容量の下限が M_1 で与えられる.

この theorem で得られた記憶容量は, 通常の連想記憶の記憶容量 $\dfrac{n}{4\log n}$ より小さいが, オーダーは同じである.

4. 実験結果

ランダムにパターンを発生させ, 減衰率を変えて記憶容量の変化を見るというシミュレーションを, 素子数 $n = 100, 200, \ldots, 1100$ のそれぞれについて 6 回ずつ行なった.

- Fig.1 横軸: 減衰率を $1 - d\dfrac{\log n}{n}$ と表した時の d の値. 縦軸: 容量 (正しく想起されたパターンの数). 見やすくするため n =100(o), 500(+), 900(X) の場合だけをプロットしてある. Theorem では $2 \le d \le 4$ の間に容量が最大になる値が存在することを主張しているが, それによく符合している.
- Fig.2 横軸: 素子数 n. 縦軸: 容量を $M_0 = n/(4e\log n)$ で割った値. また, 図中の数字は 6 回の実験のうちで同じ容量になったものの頻度を表す. すべての n に対して, 容量が Theorem で保証された値 (1/2 以上 1 以下) を上回っているが, その理由としては, (1)Theorem では記憶容量のための条件が厳しい (2)n が比較的小さくて漸近性がでていない, などが考えられる.

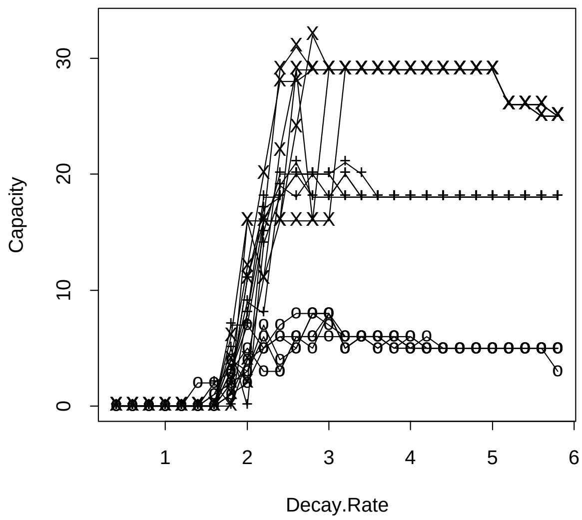

Figure 1: 減衰率と容量の関係

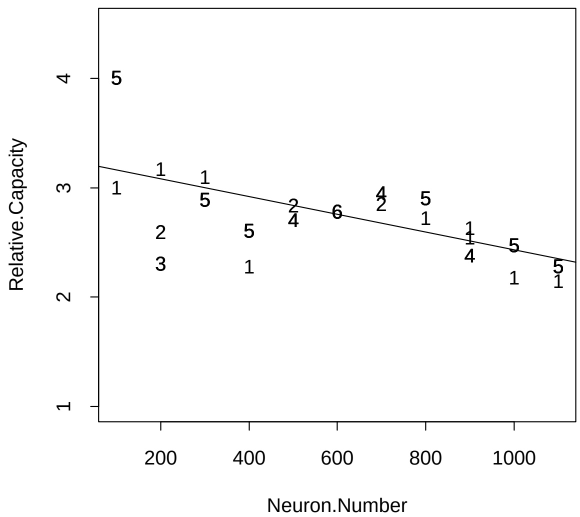

Figure 2: 素子数と容量の関係

5. 考察

パターンをスパース符号化しておけば飛躍的に記憶容量が増えることが知られている. 本報告ではパターンの各要素が確率 1/2 でランダムに発生する場合のみを述べたが, スパースな場合についても theorem と同様の結果を導くことができて, やはり通常の記憶容量と同じオーダーの記憶容量を達成する減衰項を求めることができる. スパースな場合についてのシミュレーションをおこなったり, より tight な bound を計算する問題は今後の課題として残されている.

また, ニューラルネットをより一般に帰納的な学習システムととらえた場合には, 記憶容量だけではなく汎化能力も考慮に入れて最適な減衰率を求める必要があり, 興味深い問題である.

References

[1] S. Amari: Characteristics of sparsely encoded associative memory. *Neural Networks*, Vol. 2, No. 6, pp. 451–457, 1989.

[2] 甘利 俊一: 神経回路網の数理. 産業図書, 1978.