最近の技術から

神経回路網モデルによる系列の連想*

矢内浩文

玉川大学工学部情報通信工学科　〒194 町田市玉川学園 6-1-1

1. 神経回路網モデルと系列の連想

人間の情報処理を納得しようという時にも，抽象的あるいは実利的な情報処理装置を考察しようという時にも，系列情報の処理は本質的に重要である．あらゆる情報処理は何らかの記号の流れ（系列）だからである．人間の場合でいえば，音声の生成や認識は当然のこと，他のどの情報処理においても，リズムや時系列の処理が付きまとう．

神経回路網モデルを用いた情報処理としては，動的過程が本質的でないモデル（パーセプトロン，誤差逆伝播法による層状回路網モデルなど）や，たとえ動的な動作が本質的でも結果的に達成される平衡状態にのみ意味のあるモデル（自己連想モデル）が取り上げられることが多い．系列の処理の重要性については疑問の余地はないので，系列を扱うモデルも古くから議論されてはいる[1]が，系列情報の本質的な意味や役割が不明確なために，理解はそれほど深まってはいない．

ここでは，神経回路網モデルを用いた系列連想についての基本的な事実を述べて，今後のより本質に迫った議論への一助となることを期待したい．

2. 系列連想のモデル

最も単純な形の系列連想モデルを定式化し，その性質について述べる．以下では神経回路網モデルで一般に用いられているように，ニューロン（回路網を構成するしきい値素子），シナプス（ニューロンの結合部），シナプス荷重（ニューロン間の結合効率）などの用語を用いる．

ニューロンがN個あり，それらがシナプスを介して相互に結合している．離散時刻 $t\,(=0,1,2,\cdots)$ での第 i ニューロンの状態を $x_i(t)$，第 j ニューロンから第 i ニューロンへのシナプス荷重を w_{ij} とする．連想過程においては，ニューロンの状態は次の力学に従って更新される：

$$x_i(t+1)=f\Big(\sum_{j\neq 1}^{N} w_{ij}x_j(t)+bx_i(t)-h\Big).\qquad i,j=1,2,\cdots,N \quad (1)$$

$$f(u)=\begin{cases}1, & u\geq 0,\\ -1, & u<0.\end{cases} \quad (2)$$

(1)式で，b は自己結合の大きさ，h はしきい値である．自己結合は w_{ii} と書いてもよいのだが，適応（学習）で得られる自己結合と最適な自己結合は一般には異なるので，それを明確にするために b を用いた．また，b や h はニューロンごとに異なっていてよいのだが，以下に述べる範囲ではすべて等しいのが最適である．

シナプス荷重は次の規則で適応する（相関型）：

$$w_{ij}(\tau+1)\leftarrow w_{ij}(\tau)+\alpha x_i(\tau)x_j(\tau-1) \quad (3)$$

α は正の定数，τ は適応過程での離散時刻である．連想過程と適応過程では時間きざみが異なっていてよいので，ここでは τ を用いた．なお，ここでは，適応と連想は分離して別々に行うとしている．パーセプトロン型[2]の規則などではそれらが分離していないことが本質的に重要であるが，この相関型の場合には分離している方が望ましい．

さて，記憶したい系列 $\boldsymbol{s}^1\rightarrow\boldsymbol{s}^2\rightarrow\cdots\rightarrow\boldsymbol{s}^m\rightarrow\cdots\rightarrow\boldsymbol{s}^M$ に対して，$w_{ij}(0)=0$ の条件で(3)式に従って適応すれば，シナプス荷重は

$$w_{ij}=\alpha\sum_{m=1}^{M} s_i^{m+1}s_j^m \quad (4)$$

となる．s_j^m は $\boldsymbol{s}^m$ の第 j 成分で，1または -1 をとる．

(4)式のシナプス荷重を用いて(1)式の力学を調べてみよう．$t=0$ で系列の第 μ パターンが与えられたとすると，(1)式は $x_i(1)=f(s_i^{\mu+1}+n_i)$ の形に書ける．ここに，n_i は他のパターンからの干渉や $t=0$ での状態，しきい値に由来する項である．n_i が $s_i^{\mu+1}$ と同符号か，または十分に小さければ，$t=1$ で回路網は第 $\mu+1$ パターンを再現（連想）する．

神経回路網モデルの特徴は，情報が分散かつ多重に蓄えられ，必要な時にそれを分離して再現できるところにある．したがって，記憶したいパターン全体がわからな

* この研究の大部分は東北大学電気通信研究所および東京大学工学部計数工学科で行われた．

い限り n_i は規定できず，連想の性質を厳密に議論することはできない．ただいえるのは，互いに他と直交しているパターンからなる系列ならば，$b=h=0$ のときに連想は必ずうまくゆくということである．

しかしながら，Nが十分に大きく系列パターンが確率的に生成されるならば，場合によっては，統計的な意味で厳密な解析をすることができ，理解を深めるのに役立つ．

3. 統計的手法による連想過程の解析結果

系列連想の代表的な場合について結果を列挙する．

3.1 系列パターンの成分が確率的にかつ独立に生成される場合

時刻 t でのニューロンの状態を並べたベクトルを $\boldsymbol{x}(t)$ と書き，t で連想すべきパターン $\boldsymbol{s}^t$ との類似度 l_t を次のように定義する：

$$l_t=(1/N)\boldsymbol{x}(t)\cdot\boldsymbol{s}^t=(1/N)\sum_{i=1}^{N}x_i(t)s_i^t. \qquad (5)$$

$\Pr\{s_i^m=1\}=1/2$ であるようなパターン系列に対して，負荷率を $r=M/N$ とし，(1)式で $b=h=0$（最適値）とすれば，類似度は次の力学に従う：

$$\begin{cases} l_{t+1}=\phi\left(\dfrac{l_t}{\sigma_t}\right), & (6)\\ \sigma_{t+1}=\sqrt{r+4\gamma\left(\dfrac{l_t}{\sigma_t}\right)^2}, \quad \sigma_0=\sqrt{r}. & (7)\end{cases}$$

ただし

$$\gamma(u)=(1/\sqrt{2\pi})\exp(-u^2/2),\quad \phi(u)=\int_{-u}^{u}\gamma(s)\,\mathrm{d}s$$

である．

これらの式は，例えば，時刻 0 で $\boldsymbol{x}(0)$ と $\boldsymbol{s}^\mu$ の類似度が l_0 である時に，時刻 t では $\boldsymbol{x}(t)$ と $\boldsymbol{s}^{\mu+t}$ の類似度 l_t がどうなるかを，統計的な意味で厳密に与える．それによれば，連想過程には次の性質がある：

- r が 0.27 以下ならば系列はほぼ正しく連想できる．
- 連想の誤りは r と共に増大する．$r<N/(4\log N)$ ならば 0%，$r=0.27$ で約 8% のニューロンが誤った状態をとる．

なお，系列が複数あったり，周期的な系列がある場合にも全く同じ式が成り立つ．ただし，その場合には M は系列の長さの総和であり，周期的な系列については長さが 3 以上でなければならない．また，相関型のシナプス荷重を一般化した共分散型を用いれば，系列パターンの成分が 1 をとる確率が 1/2 から離れるほどより多くのパターンを連想することができるようになる（確率が 1/2 の場合には相関型と共分散型のシナプス荷重は一致する）．

3.2 系列の連続するパターンに相関がある場合[3]

一般の時刻に対する厳密な解析はできないが，相関の大きさを

$$q=1-2\Pr\{s_i^{m+1}\neq s_i^m\} \qquad (8)$$

とすれば，$b\approx-\alpha Nq$，$h=0$ のときが最適で，$|q|$ の増加と共に連想できる系列の数が小さくなることがわかっている．(4)式をそのまま自己結合にも適用すると $w_{ii}=\alpha Nrq$ となって最適値とはほど遠いことに注意．

3.3 層状回路網による系列連想

系列を時間方向に埋め込むのではなく，層を追っての空間方向に埋め込んだ場合をみてみよう．系列が K 本あり，第 k 系列の第 m パターンが第 m 層に割り当てられ（$k=1,2,\cdots,K$），引き続く層のみとのシナプス荷重が(4)式と同様に定められている．同一系列中の連続するパターンには(8)式と同様の相関があるとする．この場合には，(5)式と同様の類似度について，q の任意の値に対してかなりよい近似で式を導くことができ，相関のある系列の連想過程を理解するのに役立つ．時刻 0 で第 1 層に初期状態を与え，その後は時刻 t で第 $t+1$ 層の連想をすると，類似度の従う力学は，(6)式と次式を組み合わせたものとなる：

$$\sigma_{t+1}=\sqrt{r+4\gamma\left(\frac{l_t}{\sigma_t}\right)^2+4\frac{|q|rl_t}{\sigma_t}\gamma\left(\frac{l_t}{\sigma_t}\right)\phi\left(\frac{l_t}{\sigma_t}\right)},$$
$$\sigma_0=\sqrt{r}. \qquad (9)$$

ただし，ここでは $r=K/N$ である．$q=0$ のとき(6)と(9)式による記述は厳密で，その時(9)式は(7)式と一致する．一般に $|q|$ が小さい系列ほど連想がうまくゆくことがわかる．特に，$q=1$ ならば第 2 層以降には第 1 層のパターンがそのまま割り当てられ，自己連想モデルになる．つまり，$q=1$ ならば，連想がうまくゆくための r の限界は約 0.15 である．

4. 補　　足

以上述べたモデルそのままでは，動作が同期的でないと系列の連想ができないが，非同期的動作でも系列を連想するモデルがある程度解析されている[4]．

文　　献

1) S. Amari: "Learning patterns and pattern sequences by self-organizing nets of threshold elements," IEEE Trans. Comput., **C-21** (1972) 1197–1206.
2) 甘利俊一：神経回路網の数理（産業図書，1978）．
3) 矢内浩文，沢田康次："系列を連想する神経回路網モデルの性質"，信学論 D-II, **J 73-D-II** (1990) 1192–1197.
4) D. Kleinfeld: "Sequential state generation by model neural networks," Proc. Natl. Acad. Sci. USA, **83** (1986) 9469–9473.

(1991 年 8 月 2 日受理)