

複雑な時系列の神経回路網モデルによる学習過程とその能力の評価

東北大通研 本堂 毅 沢田 康次

<はじめに>

我々は、日々刻々変化する現象に接するとき、無意識のうちにもその法則性を理解し、その学習によって未来の変化を予測しようとする。無意識のうちにもこの<学習>が経験によって達成される場合としては、漁師たちが身近な自然の様子から明日の天気を知ることが挙げられるだろう。このように、時系列の学習に対して理解を深めることは、科学の最も基本的な問題の1つであると言えよう。しかしながら、神経回路網モデル等を用いた時系列の学習については、連想記憶等の問題と異なり、その一般的評価法が確立されていないために議論が深まり難い側面もあると思われる。

そこで、我々は、時系列の学習過程について神経回路網モデルを用いたコンピュータシミュレーションを行い、その学習の過程と能力についてその評価法をも含めて考察を行った。

<背景>

時系列の神経回路網モデルなどによる学習については、近年、バックプロパゲーションアルゴリズムが提案され、そのリカレント型への拡張可能性がPinedaらにより指摘されたこと (ref.1)と、3層のフィードフォワード型神経回路網モデルで任意の連続関数が任意の精度で近似出来ることが舟橋らにより明らかにされ(ref.2)てから多くの研究がなされている(ref.3-6)。また、数学的に見れば、ニューラルネットワークによる学習は特定の関数空間を用いたパラメーターフィッティングであると言えるため、神経回路網モデルで提案されたアルゴリズムは有限な(したがって非線形な)出力を有するユニットをもとに構成される多様なネットワークに容易に拡張出来る。Moodyらはシグモイド型ではなく、ガウス型のユニットを用いた学習を提案している(ref.7)。

一般に学習系の構成としてはフィードフォワード型とリカレント型の2種類があるが、リカレント型でも問題が状態空間の位置 X から $dx/dt = F(x)$ なる速度を求めるベクトル場の学習になる場合、本質的にフィードフォワード型と同じである。

本研究では、高い学習能力を持つネットワークを見つけることは目的とせず、時系列の学習における一般的なダイナミクスと能力を知ることが目標とする。

<対象と方法>

学習対象とする時系列としては、低自由度無限周期の決定論的時系列であるカオスを用いる。これは自然現象との対応や解析の容易さ等から、その時系列を発生する系の自由度は低く、またそれでいて複雑な時系列を出すものが望まれるからである。

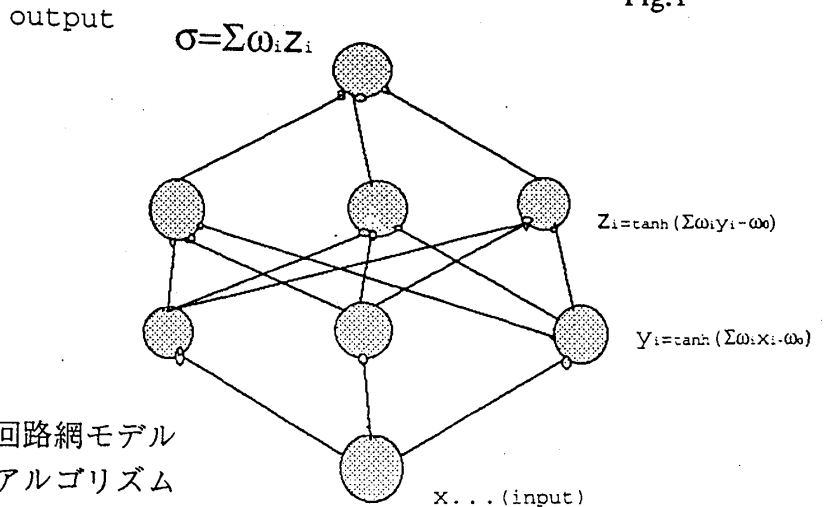
学習アルゴリズムとしては、まず最も広く使われている教師付き最急降下法である

バック・プロパゲーションを用いた。

時系列の学習は、学習信号が有限自由度の時系列である場合、一般的に入力ベクトルから出力ベクトルへの非線形写像の実現の問題となる。本研究では、この写像を4層の神経回路網モデルで学習させた。また、CPU負担が小さい等の理由で離散時間の時系列を用いた。

以下では方程式のパラメーターとリアプノフ数の対応が容易につけられるなどの理由でテント (トライアングル) マップを学習する時系列として採用する(ref.8).

Fig.1



<モデル>

2層の中間層をもつアナログ型神経回路網モデル学習則は、バックプロパゲーションアルゴリズム学習は逐次的に行う。(1単位時間ごとに行う)

学習系の入出力関係 : $x_n \rightarrow x_{n+1}$ が
 $x_{n+1} = f(x_n)$: 教師系
 に近づくように、1写像毎学習が進む

教師系としての時系列は、テントマップから生成される時系列を用いる。従って、離散的時系列を扱う。

シミュレーションにあたっては、系のシナプス荷重値の初期状態等に依存しない一般的性質を議論するため、必要に応じてアンサンブル平均を用いる

2乗誤差 E

$$E = (\sigma - f(x_n))^2 / 2$$

σ : 学習系の出力

$f(x_n)$: 教師信号 (正しい時間発展 x_{n+1})

学習方程式

シナプス荷重を ω , 学習信号と系の出力の2乗誤差をEとすると

$$\frac{d\vec{\omega}}{dt} = -\epsilon \frac{\partial E}{\partial \vec{\omega}}$$

テントマップ

$$x_{n+1} = f(x_n) = \begin{cases} r x_n & (x < 0.5) \\ r(1 - x_n) & (x \geq 0.5) \end{cases}$$

以下にシミュレーション例を示す。

Fig.2 Map

'learn.3'

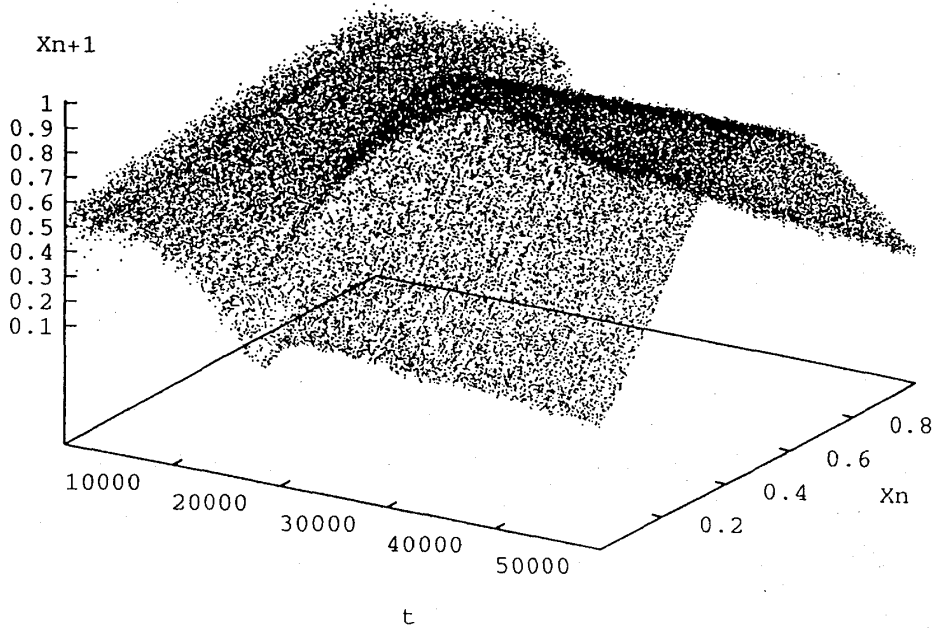


Fig.3

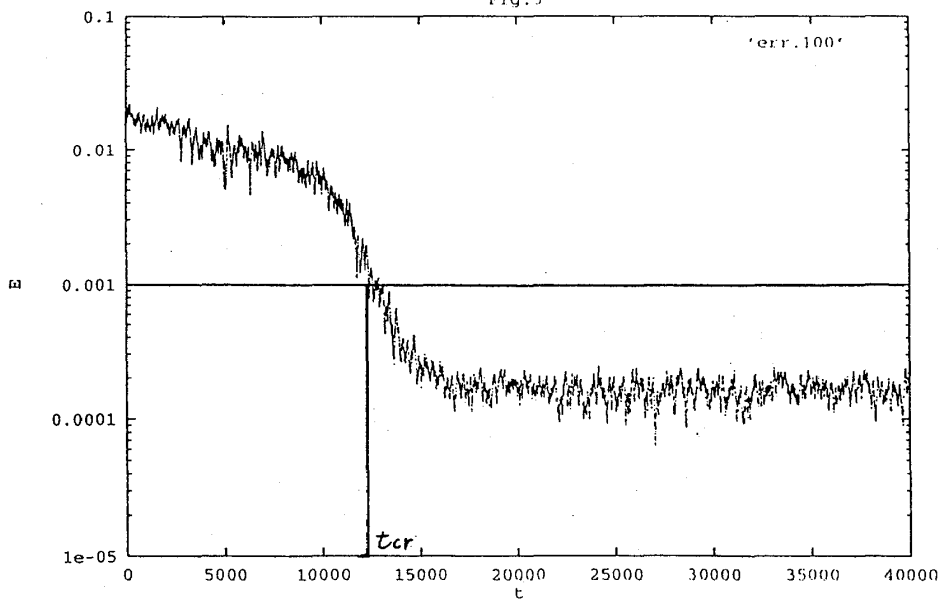


図2, 図3に見られるように, バックプロパゲーションアルゴリズムによるテン
トマップの学習では, 学習が急激に進む段階がある. そこで, このエラーの2乗誤差E
が $E=10^{-3}$ になる時間を t_{cr} 定義する. (図3参照)

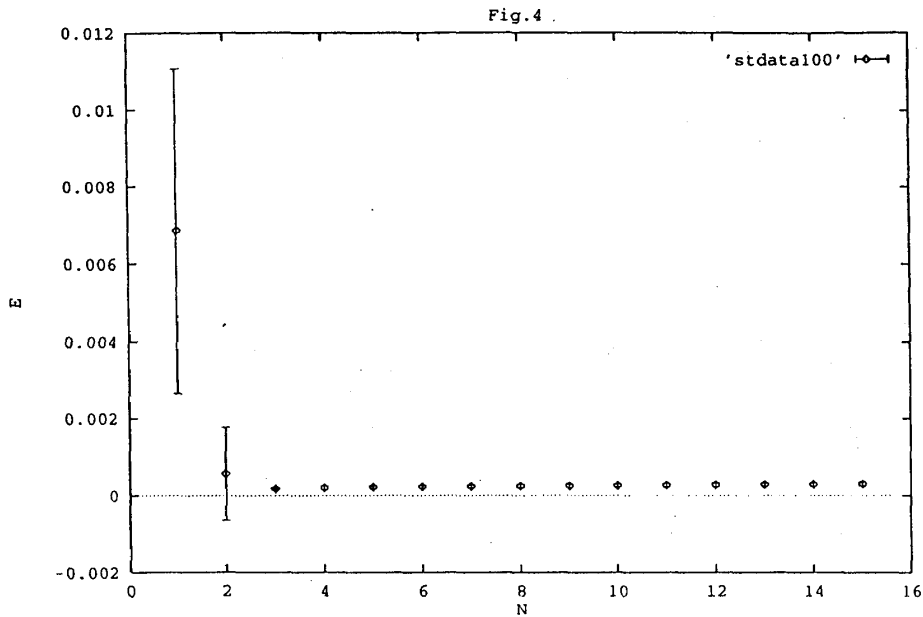


図4は、中間層の素子数 N と時間 $t = 40,000$ における二乗誤差 E との関係である。100コの独立な初期値を持つ系の集合のアンサンブル平均と標準偏差を示す。 $N = 3$ が最小になっている点特徴的である。

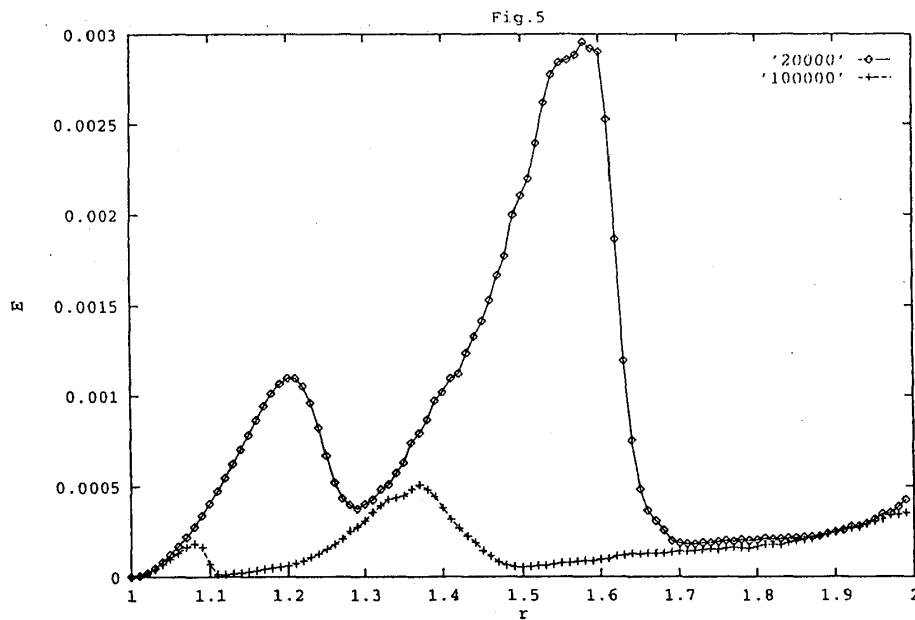


図5は、 $t = 20,000$ と $t = 100,000$ におけるテントマップのパラメータ r と二乗誤差 E の関係である。時間の経過と共に、ピーク値は小さくなくても、その形はあまり変わっていない。

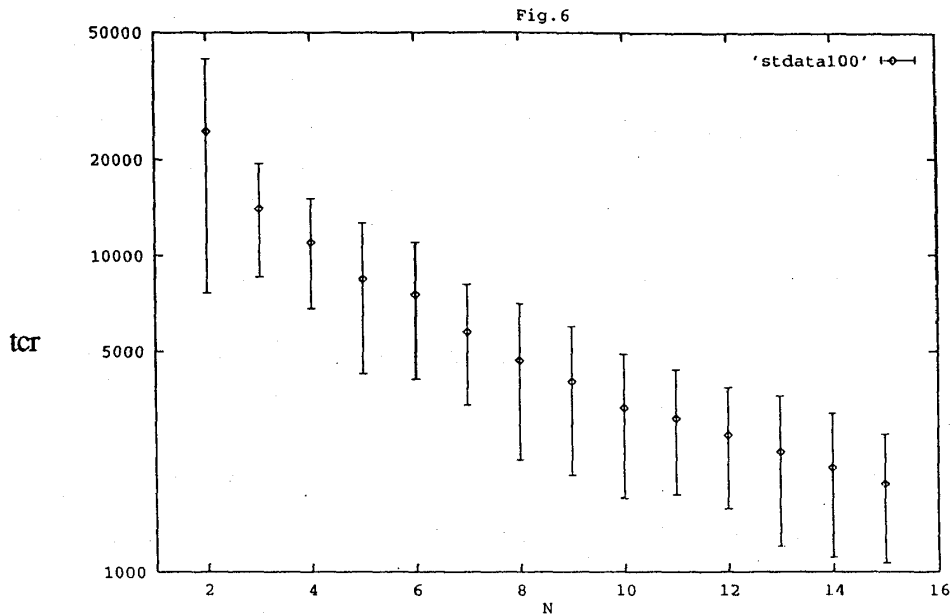


図6, 中間層の素子数Nと t_{cr} の関係である.

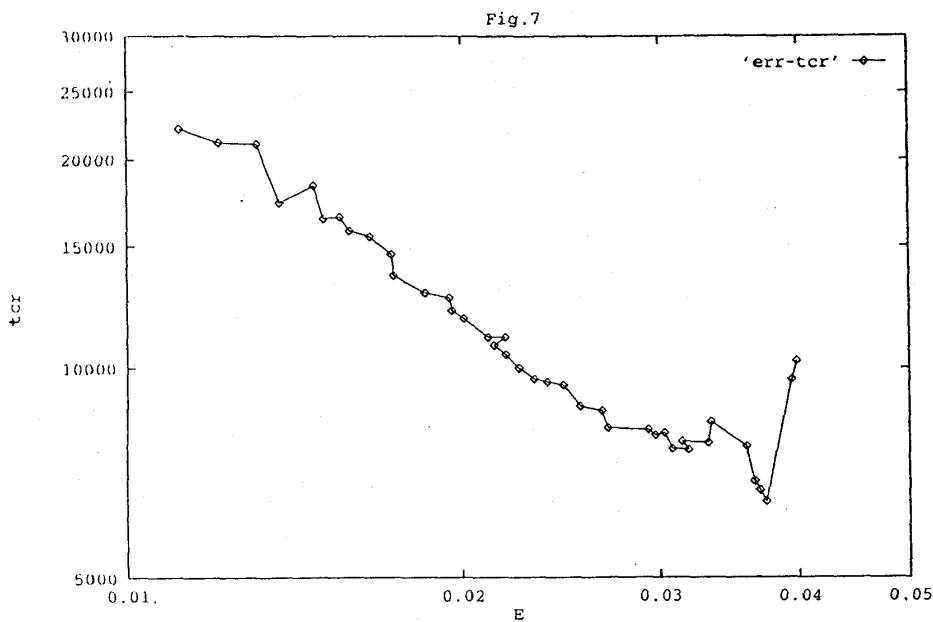


図7, 図5の2山の構造の右側部分 ($r=1.6$ から 2.0) では収束が速くなっていることを説明するためにそれぞれのパラメータを持つテントマップの学習における収束時間 t_{cr} を学習の初期の段階で ($t \ll t_{cr}$) の二乗誤差に対してプロットしたものである。ほぼ, $t_{cr} \propto 1/E$ であることが解かる。これは, 図2に見られるように, このモデルに於ては系は初期段階においては定数でフィッティングを行うことになるが, そのため, 教師系のテントマップの構造が鋭い (r が大きい) ほど初期段階での誤差も大きい, r の大きいものほど学習が早く進むので, 図7の関係になる。

これは, 初期にError が大きいものほど学習が早く進むという, 一見意外な結果となっている。

<考察>

1) 学習過程においては、急激に学習が進む段階がありその時間 (t_{cr}) は素子数 (N) の増加とともに減少する傾向がみられる。この点に関しては排他的 OR (XOR) 問題に関して Yves Chauvin (ref.9) により指摘された関係

$$t_{cr} = 2.80 - 3.3 \log_2 N \quad t_{cr} : \text{学習に要する時間}$$

$$N : \text{隠れ層のニューロン数}$$

があるが、我々のシミュレーション (6 図) では対数的減少傾向を示しているとは言えない。

2) カオスの強さを示すパラメーター (r) と系の学習過程における学習達成度 (E) との間には一般的に奇妙な関係 (2 山の構造) が見られ、初期の学習系と教師系との誤差が大きいほどかえって学習は速く進むことが解かった。

これを見るため 2 山の構造の 1 つだけ (右側) を取り出すと
学習初期の 2 乗誤差 E と t_{cr} の間に

$$t_{cr} \propto E \quad (\text{図 7})$$

の関係が成り立つ。

一般に神経回路網モデルの場合、任意の学習の結果はその荷重空間での解は一意ではない。よって、荷重空間に条件を満足する解が (テントマップの r に依存してきまる初期の 2 乗誤差によらず) 一様な密度で分布しているならば、

$$d\omega/dt \propto \partial E/\partial \omega \propto E \quad \text{より} \quad t_{cr} - E \text{ 曲線を説明できる。}$$

よってパラメーター空間での解の密度分布の評価が問題となる。

また、2 山の構造が生じる明確な理由等はその先の問題であり学習のポテンシャルの構造などからさらに検討が必要である。

3) また、時系列に対する系の学習能力の一般的評価法を確立するためには、さまざまなモデルに対してこれらの方法を試し、普遍性を検討することが必要であり、評価法の確立とともに、学習系の複雑さと時系列の複雑さとの間で、学習可能性がどのように決まってくるのかを理解しようと考えている。

4) 一般に神経回路網モデルの能力は素子数とともに増大すると考えやすいが、現実の学習過程に於てはある素子数 N で十分であり、さらに ε の選び方も重要である。

AIC 規準導入の議論から AR モデル等の最尤法の学習では未知データに対する予測精度は学習時のパラメータの数が多すぎずサンプル数が多いほど、予測の精度は上がることが知られている。最尤法であてはめた、回帰モデルによって予測をしたときの平均 2 乗予測誤差を見積る量として以下の量がある (ref.10)

$$FPE = \frac{n+m}{n-m} \hat{\sigma}^2$$

n : サンプル数, モデルをあてはめたときのデータの数 (ここでは ε に対応する)

m : 説明変数の数 (ここではニューロン数 N に対応する)

$\hat{\sigma}$: 残差誤差分散の最尤推定値

逐次的学習に於て、 ϵ が最尤推定における実効サンプル数に対応するものを決定しているならば、系の達成しうる能力 (予測, 汎化能力) は ϵ とニューロン数 N によって決まってくると考えられる。これらと今回のシミュレーションとの対応については今度の課題である。

今回の学習では問題は $x_{n+1} = f(x_n)$ の f の関数系を求めることに帰着するため、 $[0, 1]$ の一様乱数を学習入力に用いても結果はあまり変わらない感触である。しかし、より高次のヘノン・マップ等の場合には、2次元空間内で時系列が取る軌道が一様分布ではなく、局所的なアトラクター上になるため、単なるマップの学習ではない、時系列としての特徴が顕著になるものと思われる。

<参考文献>

- 1) F.J.Pineda : Generalization of back-propagation to recurrent neural networks , Phys Rev Lett. 59 2229-2232 , 1987
- 2) K. Funahashi : On the Approximation Realization of Continuous Mapping by Neural Networks, Neural Networks. vol.2, No.3, 1989
- 3) K.Doya & S.Yoshizawa: Memorizing oscillatory patterns in the analog neural network, proceedings of the 3rd IJCNN I, 27-32, 1989
- 4) 村上由彦, 佐藤雅昭: リカレントネットによる非線形ダイナミクスの近似とカオスの学習, 電子情報通信学会技術報告, NC 90-81, p. 77, 1990
- 5) 松葉育雄: カオスと予測, 数理科学 6月号, p. 64, 1992
- 6) J.D.Farmer & J.J.Sidorowich : Predicting Chaotic Time Series , Phys Rev Lett. 59 845-848, 1987
- 7) Moody : Fast Learning in Network of Locally -tuned Processing Unit
Neural Comput. (USA) vol.1, no.2, 281-94, Summer 1989
- 8) H.G.Shuster : Deterministic Chaos , Physik-Verlag, Germany, 1984
- 9) D.E.Rumelhart, J.L.McClelland and the PDP Research Group : Parallel Distributed Processing , vol I , chap 8 , MIT press, 1986
- 10) 坂本, 石黒, 北川 : 情報量統計学, 共立出版, 1983