

## 複合ニューラルネットワークによる効率的なカオス時系列予測

福永 好伸・井上 浩孝\*・成久 洋之\*\*

岡山理科大学大学院工学研究科修士課程情報工学専攻

\* 岡山理科大学大学院工学研究科博士課程システム科学専攻

\*\* 岡山理科大学工学部情報工学科

(2001年11月1日 受理)

### 1 はじめに

これまで時系列予測問題は AR モデルのような統計的な手続きを使用して解析されてきた [1]. しかし, 従来の統計的手法の予測能力は不規則に変化し続けるデータに対しては無力であり, 限界があるとされてきた. そこで近年, ニューラルネットワークやローカルモデルがカオス時系列予測に応用され, 比較的満足のいく結果が報告されている [2, 3, 4, 5].

ニューラルネットワーク [6] は時系列予測問題に対して時系列の周期性に依存した大域的な予測が期待できる. しかし, 階層型のように単純なネットワークモデルではデータの急激な変化に対応しきれないという問題点がある. それに対し, 最近傍決定則を用いたローカルモデルは短期的なカオス時系列予測に対して有効であることが知られている. これはローカルモデルの入力が過去に存在するようなパターンならば的確に再現が可能のため, 急激な変化にも対応することができるからである. しかし, カオスはデータの初期状態に鋭敏に反応するという特徴を持つため, カオス時系列の長期予測は不可能とされている.

本研究において我々はバックプロパゲーション学習を行う一般的な階層型ニューラルネットワークと最近傍決定則を使用する単純なローカルモデルを組み合わせた複合ニューラルネットワークを提案する. この手法の特徴は階層型ニューラルネットワーク及びローカルモデルで単独に予測するよりも長期的な予測が可能になる点にある. これはそれぞれ単独の予測器によって出力された予測値を各予測器の持つ出力特性で補正することにより実現されるものである. 本研究ではカオス時系列予測のベンチマーク問題である Mackey-Glass データ [7] と 1992 年のニューメキシコ州の Santa Fe で開催された時系列予測コンテスト [4], 及び 1998 年のベルギーの K. U. Leuven [5] で開催された時系列予測コンテストで使用されたカオス時系列データに適応させて本提案手法の有効性を示す.

### 2 複合ニューラルネットワーク

#### 2.1 複合ニューラルネットワークの構造

複合ニューラルネットワークの構造は図 1 のようになる. このように階層型ニューラルネットワーク ( $c_1$ ) とローカルモデル ( $c_2$ ) のそれぞれの出力を統合して, 新たな予測値としてネットワークの出力とするものである. ある時刻  $t$  の複合ニューラルネットワークの期待出力を  $x_t$  とするとネットワークの出力  $y_t$  は次式のように計算される.

$$y_t = w_{c_1} y_{c_1 t} + w_{c_2} y_{c_2 t} \quad (1)$$

式 (1) における  $w_{c_1}$  は階層型ニューラルネットワーク,  $w_{c_2}$  はローカルモデルそれぞれの寄与率を示すものである. 寄与率とは各予測器の時系列予測問題に対する適応度を表すものであり, 予測時刻以前の既知のデータ区間を利用して求める. 具体的には扱う問題に対する  $c_1, c_2$  それぞれの評価値を  $e_{c_1}, e_{c_2}$  とすると, 次のように表すことができる.

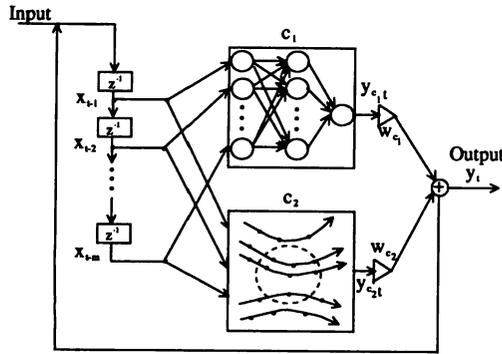


図1 複合ニューラルネットワークの構造

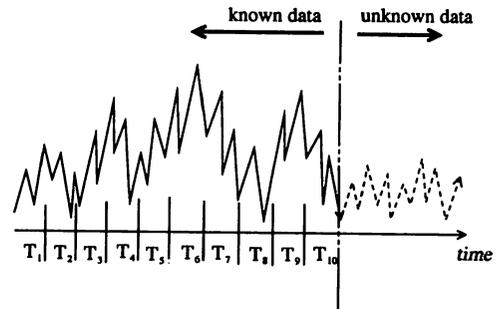


図2 10-fold cross-validation を用いたテスト予測

$$w_{c_1} = \frac{e_{c_2}}{e_{c_1} + e_{c_2}}, \quad w_{c_2} = \frac{e_{c_1}}{e_{c_1} + e_{c_2}} \quad (2)$$

本研究では予測器の評価基準に誤差を利用するため、評価値が小さいほど寄与率が高くなるように式を設定している。また、各予測器の寄与率の間には次の関係が成り立つ。

$$w_{c_1} + w_{c_2} = 1.0 \quad (3)$$

評価値の求め方には種々の方法があるが、本研究では NMSE(Normalized Mean Squared Error) を利用した。NMSE は評価を行う区間を  $T$  とすると以下の式で表すことができる。

$$e_{c_1} = \frac{\sum_{t \in T} (y_{c_1 t} - x_t)^2}{\sum_{t \in T} (x_t - E[x])^2}, \quad e_{c_2} = \frac{\sum_{t \in T} (y_{c_2 t} - x_t)^2}{\sum_{t \in T} (x_t - E[x])^2} \quad (4)$$

式(4)の  $y_{c_1 t}$ ,  $y_{c_2 t}$  は各予測器における出力であり、これらを求めるために既知のデータを利用してテスト予測を行う。本研究で用いたのは 10-fold cross-validation という手法である。これは図2のように与えられたデータを10等分し、それぞれの区間において評価するというものである。例えば、 $T_1$  をテスト予測する場合には  $T_2$  から  $T_{10}$  までの区間を利用し、 $T_2$  をテスト予測する場合には  $T_1$  及び  $T_3$  から  $T_{10}$  までを利用して行う。しかし、この手法では階層型ニューラルネットワークは各区間ごとにあらかじめネットワークを学習させておく必要があるため、非常に処理の効率が悪いため、我々は処理の効率化を図るために一度に全既知区間を対象として学習を行った。そして、学習済みの区間に対してテスト予測を実施するという有利性を軽減させるためにテスト予測時にガウス雑音(平均0, 標準偏差0.1)を入力パターンに付加して予測させた。

## 2.2 階層型ニューラルネットワーク

式(1)における  $y_{c_1 t}$  を計算するために、階層型ニューラルネットワークを使用した。階層型ニューラルネットワークの構造には入力層、中間層、出力層からなる3層構造を用い、バックプロパゲーション学習を行った。これは具体的には以下のように行われる。

いま、図3のように時刻  $t-1$  の入力パターン  $x_{t-1} \in R^m$  がネットワークに入力されるとする。このとき、入力層から中間層への重みを  $w_{ij}^{(1)}$  とすると中間層の内部状態  $s_j^{(1)}$  は次のように表される。

$$s_j^{(1)} = \sum_i w_{ij}^{(1)} x_i \quad (5)$$

同様に、中間層から出力層への重みを  $w_j^{(2)}$  とすると、出力層の内部状態  $s^{(2)}$  とネットワークの出力  $y_{c_1 t}$  は

$$s^{(2)} = \sum_j w_j^{(2)} f(s_j^{(1)}) \tag{6}$$

$$y_{c1t} = f(s^{(2)}) \tag{7}$$

のようになる。また、各ニューロンの入出力関数  $f(x)$  には次式で表されるシグモイド関数を使用した。

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{8}$$

バックプロパゲーション学習による重みの修正式は以下のようなになる。

$$w^{new} = w^{old} + \Delta w \tag{9}$$

中間層から出力層への重みを修正するときの  $\Delta w$  は

$$\Delta w_j^{(2)} = \eta f'(s^{(2)}) e_t f(s_j^{(1)}) \tag{10}$$

となり、入力層から中間層では次のようになる。

$$\Delta w_{ij}^{(1)} = \eta f'(s_j^{(1)}) (f'(s^{(2)}) e_t w_j^{(2)}) f(x_i) \tag{11}$$

また、出力誤差  $e_t$  は

$$e_t = x_t - y_{c1t} \tag{12}$$

である。なお、 $\eta$  は学習率とする。

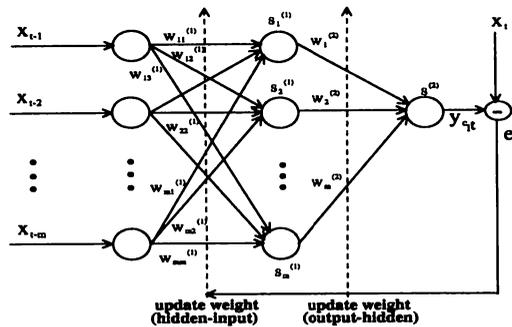


図3 3層型ニューラルネットワークの構造

### 2.3 ローカルモデル

式(1)における  $y_{c2t}$  はローカルモデルで計算した。本研究におけるローカルモデルでの予測は次の二つのステップで行うことができる。

まず、最近傍決定則により入力パターンの最近傍点を探し出す。最近傍決定則は一般にパターン認識問題などに用いられる手法で、ある特徴空間内において最も近い点を同一のクラスとして認識するというものである。本研究では  $k$  個の近傍点を探す  $k$ -NN( $k$ -nearest neighbor)を使用した。図4は  $k=2$  として、最近傍決定則を時系列予測問題に利用する場合の具体例である。このように、予測時の入力パターン  $x_{t-1}$  を時刻  $t-2$  から  $t-n$  までの各入力パターンのデータ間の距離とを比較していくと、距離の近い順に  $x_{t-a-1}$  と  $x_{t-b-1}$  の二つのパターンを検索することができる。そして、実際に出力するのは予測値であるため、最近傍パターンの次点の値である  $x_{t-a}$  と  $x_{t-b}$  を出力する。

次にこの  $k$  個の近傍点を使って予測値を求めるが、本研究では問題に応じて、式(13)のように近傍点を単純に平均するモデルと式(14)のように近傍点とその直前のデータとの差を利用するモデルを用いた。

$$y_{c_2t} = \frac{1}{k} \sum_{i=1}^k x_{t-n_i} \quad (13)$$

$$y_{c_2t} = y_{c_2t-1} + \frac{\sum_{i=1}^k (x_{t-n_i} - x_{t-n_i-1})(k-i+1)}{\sum_{i=1}^k i} \quad (14)$$

また、式 (14) のモデルでは  $k$  個のデータを単純に平均するのではなく、予測時の入力パターンに距離に近いパターンほど、合成に占める割合が高くなるように設定した。

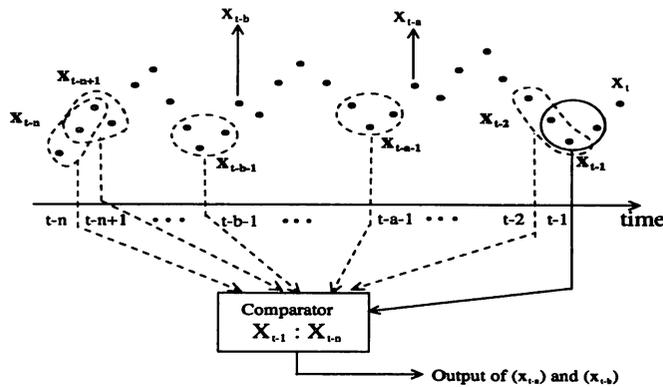


図 4  $k$ -NN の時系列予測問題への適用

### 3 時系列予測問題への適用

複合ニューラルネットワークの性能を評価するために、Mackey-Glass [7] のデータ\*1と 1992 年のニューメキシコ州の Santa Fe で開催された時系列予測コンテスト [4] のデータ、及び 1998 年にベルギーの K. U. Leuven で開かれたコンテスト [5] のデータ\*2で予測を行った。また、評価基準は NMSE, 平均自乗誤差 (Mean Squared Error: MSE), そして自乗誤差を時刻ごとに累積させた累積自乗誤差 (Accumulated Squared Error: ASE) とする。

#### 3.1 Mackey-Glass データ

次式で与えられる Mackey-Glass データはカオス時系列予測では多くの研究者が評価に使用しており、ベンチマーク的な問題である。

$$\frac{dx}{dt} = \frac{ax_{t-\tau}}{1 + (x_{t-\tau})^{10}} - bx_t \quad (15)$$

本研究で使用したデータは  $a=0.2, b=0.1, \tau=17$  として発生させたものであり、これによりカオス時系列を発生させることが知られている。

Mackey-Glass データは学習及び評価用のデータとして 2000 ポイントを使用し、それ以降の 200 ポイントを予測させた。予測時の各種パラメータは表 1 のように設定した。階層型ニューラルネットワークの入力層と中間

\*1 <http://www.ece.pdx.edu/~mcnames/DataSets> より入手可能

\*2 <http://www.esat.kuleuven.ac.be/sista/workshop> より入手可能

層にはそれぞれバイアス項を付加してある。ここで、本研究における学習回数とは各回数ごとに学習区間からランダムに一点のみを選出し、学習させることを示している。また、ローカルモデルの予測値は近傍点を平均するモデル（式(13)）により求めた。

まず、学習データを使用してテスト予測を行った結果、各予測器の寄与率は階層型ニューラルネットワークが0.109、ローカルモデルが0.891となった。よって、この問題においてはローカルモデルの方が圧倒的に寄与率が高く、この時系列に適応し易い予測器といえる。表2は各予測器における平均と最良の予測精度を比較したものである。これを見ると、寄与率の高いローカルモデルの方が実際の予測においても階層型ニューラルネットワークに比べて精度が良いことがわかる。図5は複合ニューラルネットワークが最良の予測精度を出力したときの各予測器の出力波形である。それぞれの予測器を比較すると、階層型ニューラルネットワークが予測を進めていくにつれて、次第に飽和状態に近づいているのに対し、ローカルモデルはうまく波形の軌道を捕えている。そして、複合ニューラルネットワークはこれよりも更に長期にわたって波形の軌道を捕えており、非常に精度の高い予測法であることがわかる。また、図6は図5で使用した各予測器の出力データ、及び複合ニューラルネットワークの平均の出力における予測の累積自乗誤差を表したものである。図6より階層型ニューラルネットワークは20ポイント目、ローカルモデルは120ポイント目あたりから誤差が大きく増加し始めていることがわかる。それに対し、複合ニューラルネットワークは平均でも160ポイント目まではほとんど誤差の増加が見られず、他の予測器のような急激な誤差の増加は見られない。

表1 Mackey-Glass 予測時のパラメータ

パラメータ	値
・階層型ニューラルネットワーク	
入力層のニューロン数	20+1
中間層のニューロン数	10+1
学習率	0.01
学習回数	500000
・ローカルモデル	
入力次元	20
k の数 (k-NN)	3

表2 Mackey-Glass データの予測精度

	NMSE	
	average	best
feed-forward network	0.7664	0.6797
local model	0.1975	0.1975
hybrid model	0.0404	0.0099

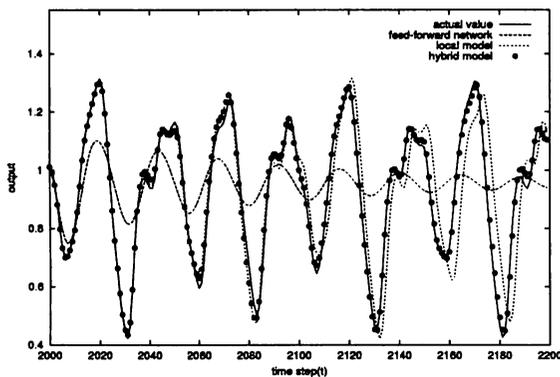


図5 Mackey-Glass データの予測波形

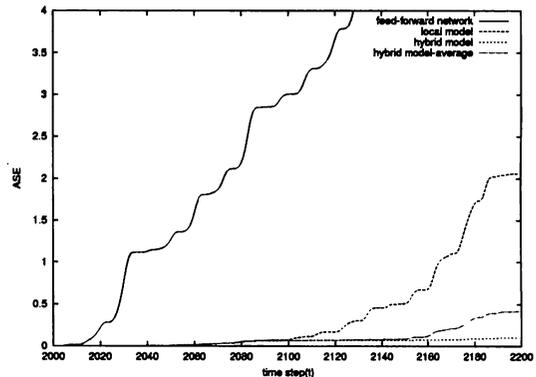


図6 予測の累積自乗誤差

### 3.2 Sante Fe コンテストデータ

図7の Santa Fe コンテストデータは1992年の NATO Advanced Research Workshop (Santa Fe: ニューメキシコ州) において開催された時系列予測コンテストで使用されたデータであり、最も有名な時系列予測コンテ

ストとして知られている。この時系列を予測するときのパラメータは表3のように設定した。また、学習データは1000ポイントとし、それ以降の100ポイントを予測させた。ローカルモデルの予測値は近傍点とその直前のデータとの差を利用するモデル(式(14))により求めた。

Santa Fe データでは各予測器の寄与率は階層型ニューラルネットワークが0.747、ローカルモデルは0.253になった。表4は各予測器の精度を比較したものである。これを見ると、この時系列ではMackey-Glassのように大幅な精度の向上がないことがわかる。そこで、複合ニューラルネットワークによる長期予測性能を検証するために1000ポイント先までを予測させた。図8は二つの予測器のうち、精度の高いモデルで1000ポイント予測させたときの真値と予測値との波形であり、図9は同様に複合ニューラルネットワークの波形である。これを見ると、明らかに複合ニューラルネットワークの方が安定した長期予測が出来ていることがわかる。表5は複合ニューラルネットワークをコンテストの上位者の精度[8, 9]と比較を行ったものである。この表における予測はどの予測区間に対しても、学習区間は0~1000ポイント目までを利用し、予測を始める地点だけを変更して100ポイントの予測を行ったものである。表5を見ると、区間(c)では本提案手法が最も優れており、区間(d)でも同程度の精度が得られていることがわかる。図10は表5の各予測区間における複合ニューラルネットワークの予測波形である。この図より表5において精度の高い(a), (c), (d)の区間における予測波形はうまく波形の周期を捕えていることがわかる。また、表5においてあまり精度の高くない(b)と(e)の区間の予測波形を見ると、波形の軌道自体は真値の波形にうまく適合していることがわかる。

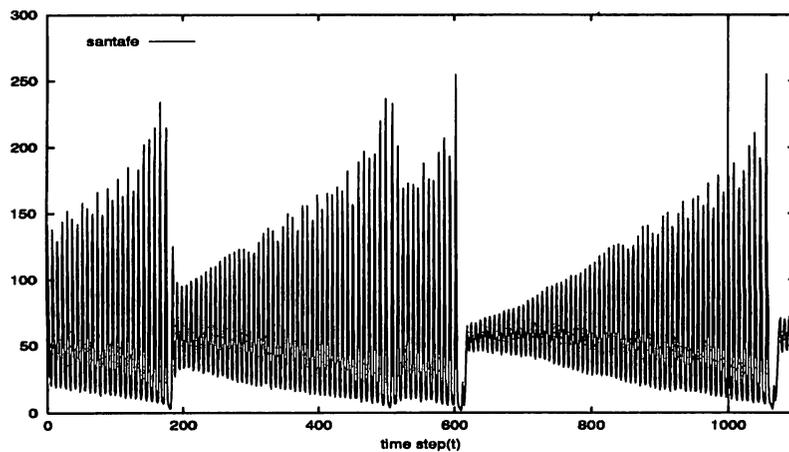


図7 Santa Fe コンテストデータ

パラメータ	値
・階層型ニューラルネットワーク	
入力層のニューロン数	5+1
中間層のニューロン数	5+1
学習率	1.0
学習回数	500000
・ローカルモデル	
入力次元	40
$k$ の数 ( $k$ -NN)	3

	NMSE	
	average	best
feed-forward network	1.2524	0.6270
local model	0.2377	0.2377
hybrid model	0.2070	0.1866

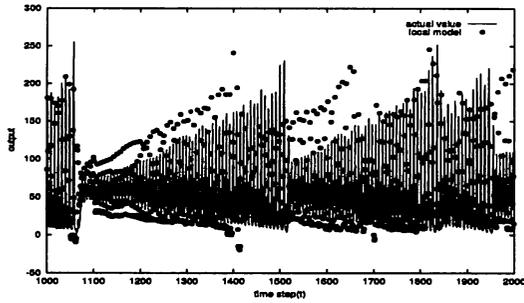


図 8 ローカルモデルによる Santa Fe データの予測波形

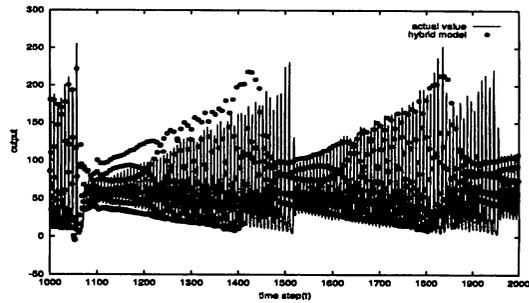
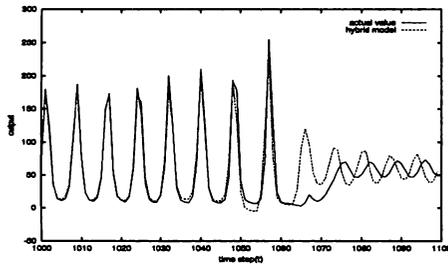


図 9 複合ニューラルネットワークによる Santa Fe データの予測波形

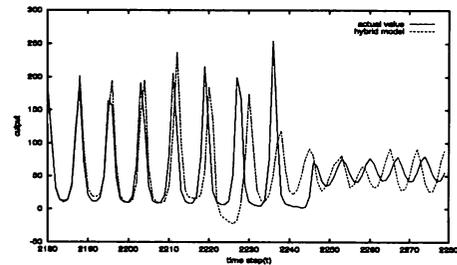
表 5 Sante Fe データにおけるコンテスト上位者との予測精度 (NMSE) の比較

starting point	hybrid model	T. Sauer [8]	E. A. Wan [9]
(a)1000	0.187	0.077	0.027
(b)2180	0.980	0.174	0.065
(c)3870	0.079	0.183	0.487
(d)4000	0.013	0.006	0.023
(e)5180	1.129	0.111	0.160

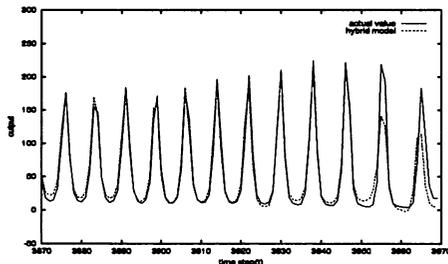
(a)



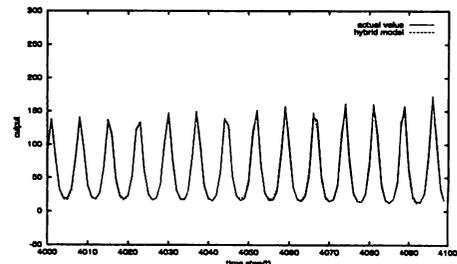
(b)



(c)



(d)



(e)

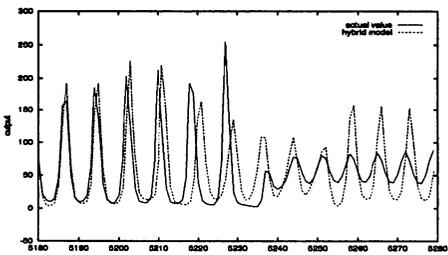


図 10 複合ニューラルネットワークによる Santa Fe データの予測波形

### 3.3 K. U. Leuven コンテストデータ

K. U. Leuven コンテストデータは 1998 年の International Workshop(K. U. Leuven : ベルギー)において開催された時系列予測コンテストで使用されたデータである(図 11)。この時系列データを予測するときの各種パラメータを表 6 のように設定した。また、学習データは 2000 ポイントとし、それ以降の 200 ポイントを予測させた。ローカルモデルの予測値は近傍点を平均するモデル(式(13))により求めた。

寄与率は階層型ニューラルネットワークが 0.312, ローカルモデルが 0.688 となった。図 12 は複合ニューラルネットワークが最良の結果を出力したときの各予測器の出力波形である。階層型ニューラルネットワークは予測波形の軌道を捕えることができず、60 ステップ目あたりからはほぼ一定の値しか出力していない。また、ローカルモデルは 80 ステップ目まではうまく出力されているが、そこまでが限界であり、それ以降は一定の周期を出力するだけとなっている。それに対し、複合ニューラルネットワークは 80 ステップ目から 100 ステップ目にかけて若干のずれがあるものの、その後は再び波形の軌道を捕えていることがわかる。図 13 は K. U. Leuven データにおけるそれぞれの予測器の累積自乗誤差を表したものである。図 13 から階層型ニューラルネットワーク、ローカルモデルともに階段状に誤差が増大していることがわかる。複合ニューラルネットワークは 80 ステップ目から 100 ステップ目にかけて誤差が大きく増えており、一時はローカルモデルの累積自乗誤差を超えているものの、その後はほとんど誤差の増加がなく、最終的な累積自乗誤差はローカルモデルの三分の一程度に抑えられている。表 7 は本提案手法に使用している予測器、及びコンテストの優勝者と準優勝者の精度 [10] を比較したものである。これを見ると、階層型ニューラルネットワーク及びローカルモデルを単独でコンテストの参加者と比較すると、かなり精度が劣っている。しかし、この二つを組み合わせた複合ニューラルネットワークでは優勝者の予測精度には及ばないが、準優勝者よりは優れていることがわかる。

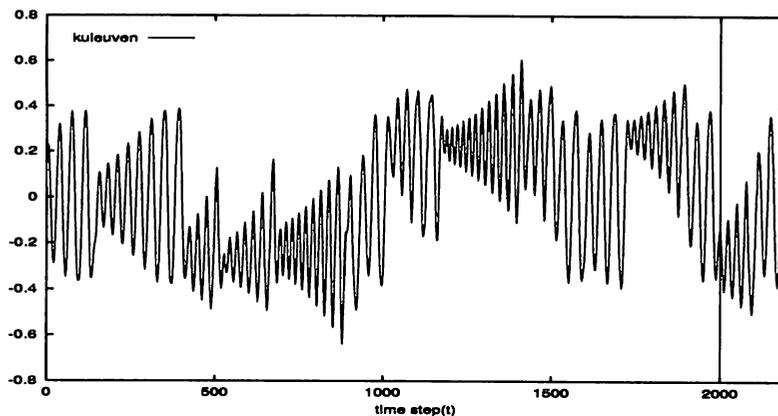


図 11 K. U. Leuven コンテストデータ

表 6 K. U. Leuven 予測時のパラメータ

パラメータ	値
・階層型ニューラルネットワーク	
入力層のニューロン数	10+1
中間層のニューロン数	10+1
学習率	0.01
学習回数	500000
・ローカルモデル	
入力次元	20
k の数 (k-NN)	3

表 7 K. U. Leuven データの予測精度

	NMSE(MSE)	
	average	best
feed-forward network	2.5084 (0.1436)	1.1076 (0.0634)
local model	1.3197 (0.0756)	1.3197 (0.0756)
hybrid model	0.4605 (0.0264)	0.3040 (0.0174)
nearest trajectory	—	(0.0018)
local learning model	—	(0.0475)

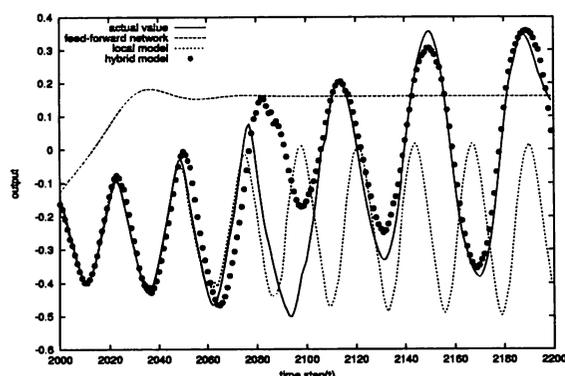


図 12 K. U. Leuven データの予測波形

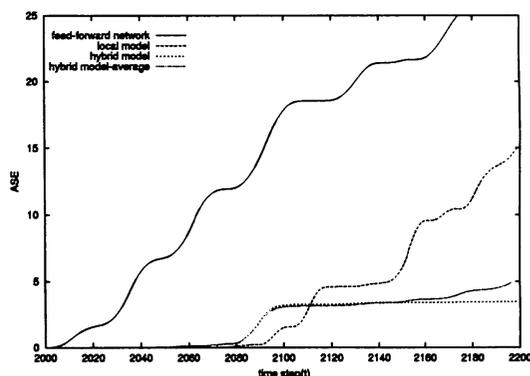


図 13 予測の累積自乗誤差

#### 4 まとめ

本研究では、カオス時系列の効率的な予測を目的として、階層型ニューラルネットワークとローカルモデルにより構成された複合ニューラルネットワークを提案した。数値実験の結果、単独の予測器で行うよりも長期的な予測が可能であることを実証した。

今後、様々なニューラルネットワークやローカルモデルを組み合わせ、時系列予測問題へのさらなる有効性を考えることやパターン認識問題などへの応用も考える。

#### 参考文献

- [1] P. J. Brockwell, and R. A. Davis, Introduction to Time Series and Forecasting, Springer-Verlag, New York, 1996.
- [2] A. Lapedes, and R. Farber, "Nonlinear signal processing using neural networks: Prediction and system modeling," Los Alamos National Laboratory Report, no.LA-UR-87-2662, 1987.
- [3] M. Casdagli, "Nonlinear prediction of chaotic time series," Physica D, vol.35, pp.335–356, 1989.
- [4] A. S. Weigend, and N. A. Gershenfeld, eds., Time Series Prediction: Forecasting the Future and Understanding the Past, Santa Fe Institute Studies in the Sciences of Complexity, Perseus Books, NM, 1994.
- [5] J. Suykens, and J. Vandewalle, eds., Nonlinear Modeling: Advanced Black-box Techniques, Kluwer Academic Publishers, Boston, 1998.
- [6] S. Haykin, Neural Networks: A comprehensive foundation, 2nd ed., Prentice-Hall, Upper Saddle River, 1999.
- [7] M. C. Mackey, and L. Glass, "Oscillation and chaos in physiologist control systems," Science, vol.197, pp.287–289, 1977.
- [8] T. Sauer, "Time series prediction by using delay coordinate embedding," in Time Series Prediction: Forecasting the Future and Understanding the Past, A. S. Weigend, and N. A. Gershenfeld, eds., pp.175–193, Perseus Books, NM, 1994.
- [9] E. A. Wan, "Time series prediction by using a connectionist network with internal delay lines," in Time Series Prediction: Forecasting the Future and Understanding the Past, A. S. Weigend, and N. A. Gershenfeld, eds., pp.195–217, Perseus Books, NM, 1994.
- [10] J. Suykens, and J. Vandewalle, "The K. U. Leuven competition data: a challenge for advanced neural networks techniques," Proc. European Symposium on Artificial Neural Network, Bruges, Belgium, pp.299–304, Apr. 2000.

## Efficient Hybrid Neural Network for Chaotic Time Series Prediction

Yoshinobu FUKUNAGA, Hirotaka INOUE\* and Hiroyuki NARIHISA\*\*

Graduate School of Engineering,  
\*Doctoral Program in System Science,  
Graduate School of Engineering,  
\*\*Department of Information & Computer Engineering,  
Faculty of Engineering,  
Okayama University of Science,  
Ridaicho 1-1, Okayama 700-0005, Japan  
(Received November 1, 2001)

Neural networks have been used for time series prediction as an effective model. However, conventional neural networks are inferior to the local models which are based on the nearest neighbor. In this paper, we propose an efficient hybrid neural network which is constructed by a traditional feed-forward neural network, which is learned by using the backpropagation and a local model such as a nearest neighbor. The feed-forward neural network is employed for the global approximation, and the local model works for the local approximation. The effectiveness of our proposed method is evaluated on the standard benchmarks: the Mackey-Glass data and K.U. Leuven competition data. Experimental results show that the hybrid neural network can more precisely predict long-term behavior than the other known predictors.