

局所近似法によるカオス時系列データの予測

伊 藤 憲 一

Predicting Chaotic Time Series based on a Local Approximation Technique

Ken-ichi ITOH

Abstract: A method is presented for predicting chaotic time series based on a local approximation technique. In the local approximation technique, a state space is reconstructed from a time series using delay coordinate embedding and then a local predictor is constructed on the basis of the motion of the nearest neighbors in the state space. The parameters, such as the embedding dimension and the number of the nearest neighbors, have a significant effect on the prediction accuracy. The method can be used as a means of choosing the parameters suitable for the prediction. The efficacy of the method is demonstrated using chaotic time series generated by the Hénon map and the Ikeda map.

Key Words: chaos, time series, prediction, embedding dimension

1. ま え が き

複雑で不規則に変動する時系列データの背後に決定論的な法則が潜んでいれば、この法則を推定することにより、統計的な方法に比べ精度の良い予測ができると考えられる。すなわち、不規則な時系列データがカオス力学系に基づくものであれば、カオスの決定論的な特徴を生かした精度の良い短期予測が可能である。このような考え方に基づくカオス時系列データの短期予測手法の研究がこれまで活発に行われてきた[1]~[5]。

カオス時系列データの予測においては、まず観測された1変数の時系列データから、埋込みの手法[6], [7]を用いて多次元空間上にアトラクタの軌道を再構成し、次にこのアトラクタの軌道を生成する力学モデルの推定を行う。この推定方法として、多次元空間上で局

所的に予測モデルを構成する方法（局所近似法）[1]~[3]、Radial basis functionsを用いる方法[4]、ニューラルネットワークによる方法[5]などがある。このうち局所近似法は、アトラクタ上のある点の将来の値をその近傍点の推移をもとに近似する方法であり、アトラクタ全体に対しモデルを構築する方法に比べて、データへの適合が容易である[8]。

局所近似法においては、埋込み次元や近傍点の数などのパラメータをいくつに設定するかが、予測精度上重要な問題である。アトラクタの再構成のためには、最小埋込み次元を求める必要があるが、観測データから最小埋込み次元を正しく推定できるかどうかという問題がある。特にデータ数が少ない時には推定が困難である[9], [10]。また、仮に最小埋込み次元を推定できたとしても、最小埋込み次

元が予測精度上最適であるという保証はない。一方、近傍点の数をいくつに設定すべきかという問題がある。近傍点の数が少なすぎると個々の近傍点の動きに解が大きく依存するため、解の安定性がなくなる。近傍点の数が多すぎると予測点から離れた点まで近傍点として選択されるため、解は予測点の局所的な動きを反映しなくなり、予測誤差が増加する。

本論文では、上記問題の解決を図るため、局所近似法において予測に適した埋込み次元と近傍点の数を決定する手法を提案する。さらに、Hénon map[11]および Ikeda map[12]から生成されるカオス時系列データを用いて予測精度に関する評価実験を行い、本提案手法の有効性を示す。

2. 局所近似法

カオス時系列データの予測手法である局所近似法の概要と問題点を述べる。

ある時系列データが ν 次元空間上のカオス力学系から生成され、1変数の時系列データ x_1, x_2, \dots, x_K が観測されたとする。この1変数の時系列データからもとの ν 次元空間上のアトラクタの軌道を再構成するため、一定の時間遅れ τ を用いて m 次元空間上に次の m 次元ベクトル X_i を作成する。

$$X_i = (x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau}),$$

$$i = 1 + (m-1)\tau, 2 + (m-1)\tau, \dots, K-1, K \quad (1)$$

ここで、 $m \geq 2\nu + 1$ であれば、この変換は埋込みであり、もとの ν 次元空間上のアトラクタの構造が m 次元空間上に保存されることが、Takensにより証明されている[6]。さらにその後Sauerらにより、アトラクタのボックスカウント次元を D_0 とすると、 $m > 2D_0$ であれば埋込みになることが示されている[7]。この m 次元空間上での X_i の動きを推定することにより予測を行う。

時刻 i までの観測値 $x_1 \sim x_i$ が得られているとき、時間 p 経過後の値 x_{i+p} を予測する。観測値 $x_1 \sim x_i$ を用いて式(1)によりアトラクタの再

構成を行い、アトラクタ上で x_i に対応する点 X_i と時間 p 経過後の予測点 X_{i+p} の関係を関数 F により近似する。

$$X_{i+p} \cong F(X_i) \quad (2)$$

局所近似法では、関数 F を次の考え方で求める。アトラクタ上の X_i の時間変化は、微小時間内であればその近傍点の時間変化にほぼ等しいと考え、アトラクタ上のすべての点の中から X_i の近傍点 X_{T_h} ($h=1, 2, \dots, n$)を選択し、これらの n 個の近傍点の平均的な推移を最小2乗法により求め、関数 F を推定する。

x_{i+p} の予測のためには、 X_{i+p} の m 個の要素のうち1番目の要素 x_{i+p} を求めればよいので、次の線形近似式により推定する。

$$x_{i+p} \cong f(X_i)$$

$$= a_0 + \sum_{k=1}^m a_k x_{i-(k-1)\tau} \quad (3)$$

係数 a_0, a_1, \dots, a_m は次のようにして求める。まず X_i を除くすべての X_j について、 X_i とのユークリッド距離 r を計算する。

$$r = \left(\sum_{k=0}^{m-1} (x_{i-k\tau} - x_{j-k\tau})^2 \right)^{1/2} \quad (4)$$

次に、 r の小さい順に n 個の点を X_i の近傍点 X_{T_h} ($h=1, 2, \dots, n$)とし、次の最小2乗条件より係数 a_0, a_1, \dots, a_m を決定する。

$$\sum_{h=1}^n (x_{T_h+p} - f(X_{T_h}))^2 = \min \quad (5)$$

この係数の値を式(3)に代入することにより、 x_{i+p} を求めることができる。

最小2乗解を得るために必要な近傍点の数 n は、次式を満たす必要がある。

$$n \geq n_{\min} = m+1 \quad (6)$$

局所近似法においては、埋込み次元 m および近傍点の数 n をどのように決定するかが、予測精度の点から重要である。

埋込み次元 m は $m > 2D_0$ を満たす必要があるため、まず D_0 を推定する必要があるが、観測データから D_0 を正しく推定できるかどうかという問題がある。特にデータ数が少ない時は推定が困難である[9], [10]。また、仮に D_0 を推定できたとしても、 $m > 2D_0$ を満たす m をいくつに設定すべきかという問題がある。従来は $m > 2D_0$ を満たす最小値 m_{\min} を使用してきたが、予測精度上 m_{\min} が最適である保証はない。

近傍点の数については以下のような問題がある。すなわち、近傍点の数が少なすぎると個々の近傍点の動きに解が大きく依存するため、解の安定性がなくなる。特に実データを予測する場合、実データには通常雑音が含まれるため、近傍点の数が少ない程雑音の影響を受けやすい。一方、近傍点の数が多すぎると予測点から離れた点まで近傍点として選択されるため、解は予測点の局所的な動きを反映なくなり、予測誤差が増加する。近傍点の数は、最小2乗解を得るために必要な最小値 n_{\min} の2倍にするのが適当であるという報告[4]もあるが、どのような時系列データにも常に当てはまるとは限らない。

3. 提案手法

ここでは、局所近似法において予測に適した埋込み次元と近傍点の数を決定する手法を提案する。

本手法では、観測データを前半と後半に分け、前半のデータを用いて後半のデータの予測を行い、予測誤差を求める。埋込み次元と近傍点の数を変化させながら予測を繰り返す、予測に適した埋込み次元と近傍点の数を決定する。

具体的な方法は以下のとおりである。

(1) 観測データ L 個を、前半の L_f 個と後半の L_r 個に2分する。前半の L_f 個のデータをアトラクタ再構成用の学習データとして使用し、後半の L_r 個のデータについて p ステップ先を局所近似法により順次予測する。この L_r 個の予測値と観測値より相対誤差 E を求める。相対誤差 E の定義は4.2節に示す。

(2) 予測において、埋込み次元 m は2から順に増加させる。近傍点の数 n は、 $n = 2n_{\min} = 2(m+1)$ とする。埋込み次元 m を2から順に増加させながら各々の m に対する相対誤差 E を求め、 E が最初に増加する m の直前の値を m_{fix} とする。

(3) 次に、埋込み次元 m を m_{fix} とし、近傍点の数を $n = 2n_{\min} = 2(m_{\text{fix}} + 1)$ から順に1つずつ減少させながら各々の n に対する相対誤差 E を求め、 E が最初に増加する n の直前の値を n_d とする。

(4) さらに、埋込み次元 m を m_{fix} とし、近傍点の数を $n = 2n_{\min} = 2(m_{\text{fix}} + 1)$ から順に1つずつ増加させながら各々の n に対する相対誤差 E を求め、 E が最初に増加する n の直前の値を n_u とする。

(5) 近傍点の数 n が n_d の時の相対誤差と n_u の時の相対誤差を比較し、相対誤差が小さい方の n を n_{fix} とする。

(6) 以上により求めた埋込み次元 m_{fix} と近傍点の数 n_{fix} を用いて予測を行う。

提案手法の流れ図を図1に示す。

本手法では、予測に適した埋込み次元と近傍点の数は、観測データの数が半分になっても大幅に変化しないと考え、観測データのみを用いて予測誤差の評価を行い、埋込み次元と近傍点の数を決定する。

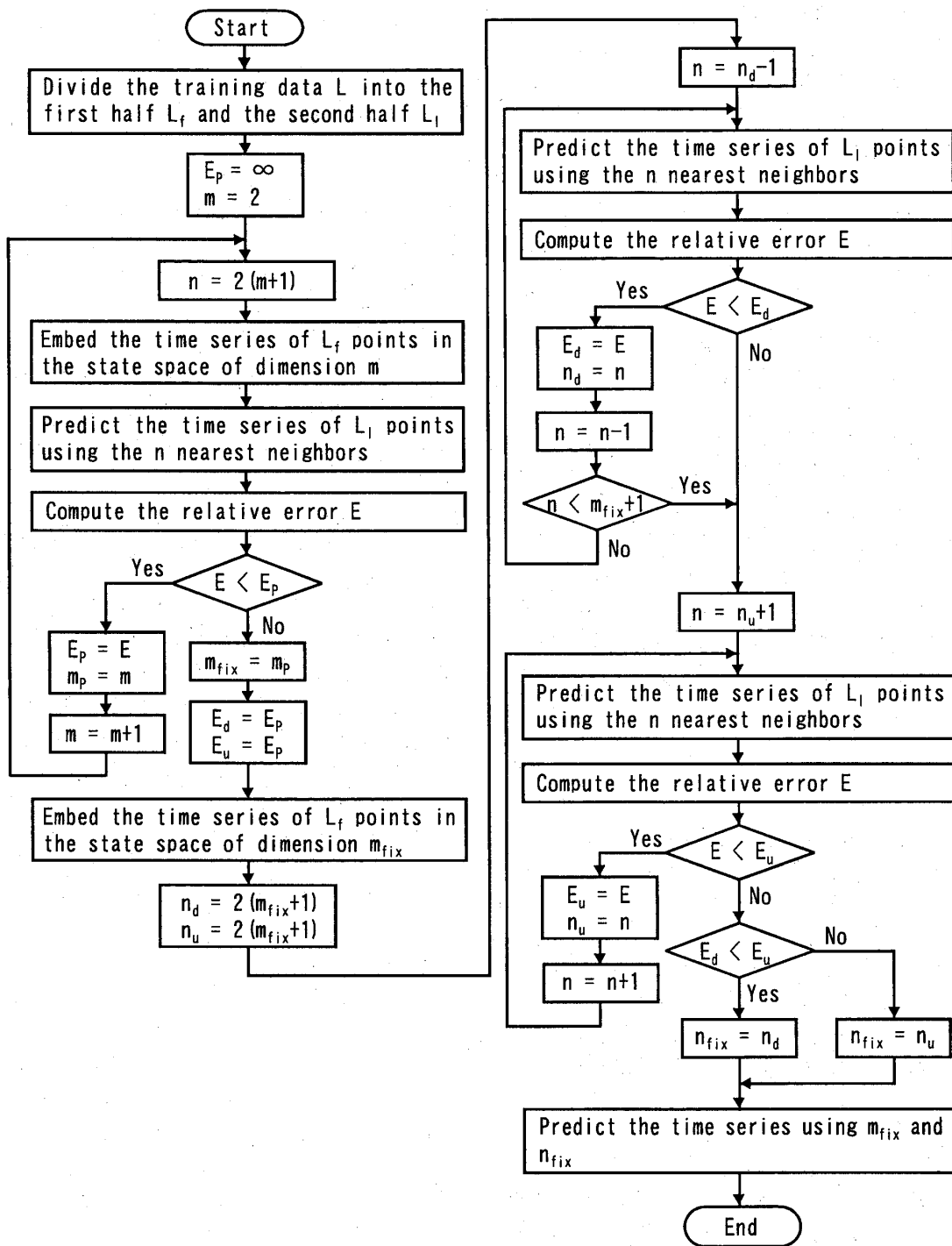


図1 提案手法の流れ図
Fig. 1 Flow chart of the proposed method.

4. 評価実験

提案手法の有効性を確認するために、評価実験を行った。本章では、実験内容、実験結果および考察について述べる。

4.1 予測対象データ

Hénon map[11]および Ikeda map[12]から生成されるカオス時系列データを、予測対象データとした。

Hénon mapは、次の2次元写像である。

$$\begin{aligned} x_{n+1} &= y_n + 1 - Ax_n^2, \\ y_{n+1} &= Bx_n \end{aligned} \quad (7)$$

式(7)において、パラメータ値 $A=1.4$, $B=0.3$, 初期値 $x_0=0.3$, $y_0=0.3$ としたときの x を時系列データとした。

Ikeda mapは、次の2次元写像である。

$$\begin{aligned} x_{n+1} &= 1 + \mu(x_n \cos t - y_n \sin t), \\ y_{n+1} &= \mu(x_n \sin t + y_n \cos t) \end{aligned} \quad (8)$$

ここで、 $t=0.4-6.0/(1+x_n^2+y_n^2)$ である。式(8)において、パラメータ値 $\mu=0.7$, 初期値 $x_0=0.3$, $y_0=0.3$ としたときの x を時系列データとした。

4.2 予測値の評価指標

N ステップ分の観測値 x_i ($i=1,2,\dots,N$) とその予測値 y_i ($i=1,2,\dots,N$) が与えられたとき、次式の相対誤差 E を用いて予測精度の評価を行う[1]。

$$E = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}} \quad (9)$$

ここで、 $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ である。

$E=0$ のとき、予測は完全であることを示す。 $E=1$ のとき、単に平均値を予測したにすぎないことを示す。

4.3 実験方法

予測対象データの先頭 L 個のデータをアトラクタ再構成用の学習データとして使用し、引き続き N 個のデータについて p ステップ先を順次予測する。すなわち、過去の観測値 x_1, x_2, \dots, x_L および予測点の観測値 x_i をもとに p ステップ先の予測値 y_{i+p} を求め、観測値 x_{i+p} との誤差を求める。これを $i=L+1$ から $i=L+N$ まで N 回行い、この N 回分の予測値と観測値より相対誤差 E を求める。

実験では、表1に示すパラメータ値を使用した。学習データ長 L は、100から10000まで変化させた。予測データ長 N は L と同一とした。予測ステップ数 $p=1$, 時間遅れ $\tau=1$ とした。埋込み次元 m および近傍点の数 n については、提案手法と従来手法の比較を行うため、次のようにした。すなわち、提案手法では、3章に示した方法により求めた m_{fix} と n_{fix} を使用した。従来手法では、埋込み次元は $m > 2D_0$ を満たす最小値 m_{min} を使用し、近傍点の数は $n = 2n_{min} = 2(m_{min} + 1)$ とした。Hénon mapの $D_0=1.26$, Ikeda mapの $D_0=1.32$ であり、両時系列データとも、埋込み次元 $m=3$, 近傍点の数 $n=8$ となる。また、局所近似法による予測精度の限界を示すため、埋込み次元 m と近傍点の数 n を変化させて予測を行い、相対誤差の最小値および相対誤差が最小となる m , n の値を求めた。

表1 パラメータ値
Table 1 Values of parameters.

種類	記号	値
学習データ長	L	100, 500, 1000, 5000, 10000
予測データ長	N	L と同一
予測ステップ数	p	1
時間遅れ	τ	1

4.4 実験結果と考察

学習データ長 L を100から10000まで変化させ、提案手法および従来手法による予測を行い、相対誤差を求めた。同時に、局所近似法において相対誤差が最小となる限界値についても求めた。

Hénon mapにおける学習データ長と相対誤差の関係を図2に示す。また、各々の予測において実際に使用した埋込み次元と近傍点の数を表2に示す。図2より、提案手法では、ほぼ限界値に近い相対誤差となっている。一方、従来手法では、学習データ長が大きくなる程相対誤差は限界値から乖離する。すなわち、提案手法は、ほぼ限界に近い予測精度を実現でき、特に学習データ長が大きくなる程従来手法に比べてより有効であるといえる。

表2からわかるように、従来手法の問題点の1つは埋込み次元の不足である。従来手法では常に最小埋込み次元である3を用いているが、限界値のケースからわかるように、学習データ長が大きい場合は埋込み次元をさらに増加させることにより予測精度が向上する。提案手法では、学習データ長が5000以上で埋込み次元が4となっており、埋込み次元の不足という従来手法の問題点のある程度解決している。また、近傍点の数 n について、従来手法では常に $n = 2n_{\min} = 2(m_{\min} + 1)$ を用いているが、限界値のケースからわかるように、学習データ長が大きい場合は近傍点の数をさらに増加させることが予測精度上望ましい。この点についても、提案手法ではある程度対応している。

Ikeda mapにおける学習データ長と相対誤差の関係を図3に示す。また、各々の予測において実際に使用した埋込み次元と近傍点の数を表3に示す。図3より、提案手法では、ほぼ限界値と等しい相対誤差となっている。一方、従来手法では、学習データ長が大きくなる程相対誤差は限界値から大きく乖離し、その度合いは、Hénon mapの場合より極めて大きい。その原因の1つは、表3からわかるように埋込み次元の不足と考えられる。Ikeda mapでは、学習データ長が大きい場合はもちろ

ん、小さい場合も埋込み次元が最小埋込み次元3では不足していることがわかる。提案手法では、学習データ長が100でも埋込み次元が4となっており、埋込み次元の不足という従来手法の問題点のある程度解決している。近傍点の数については、Hénon mapの場合と同様、常に $n = 2n_{\min} = 2(m_{\min} + 1)$ を用いるのが最適とはいえ、限界値のケースからわかるように、学習データ長が大きい場合は近傍点の数をさらに増加させることが予測精度上望ましい。この点についても、提案手法ではある程度対応している。

以上の結果をまとめると次のとおりである。

(1) 予測に適した埋込み次元は、常に最小埋込み次元であるとは限らない。特に学習データ長が大きくなる程、埋込み次元をさらに増加させることが、予測精度向上の点から必要である。

(2) 予測に適した近傍点の数は、最小2乗解を得るために必要な最小値 n_{\min} の2倍が適当であるとはいえない。学習データ長が大きくなる程、近傍点の数をさらに増加させることが、予測精度向上の点から必要である。

(3) 従来手法では、埋込み次元を最小埋込み次元に、近傍点の数を n_{\min} の2倍に固定して予測を行うため、学習データ長が大きくなる程局所近似法の限界に比べ予測精度が低下する。

(4) 提案手法では、観測データのみを用いて事前に予測誤差の評価を行い、予測に適した埋込み次元と近傍点の数を決定するため、学習データ長にかかわらず局所近似法の限界に近い予測精度を実現できる。

今回のHénon mapとIkeda mapを用いた実験結果から、予測に適した埋込み次元と近傍点の数は、観測データの数が半分になっても大幅に変化しないという提案手法の考え方が概ね妥当であることを確認した。実データにおいては、従来手法で必要となる最小埋込み次元などのパラメータが不明確な場合が多いと想定され、このような実データの予測にも、提案手法は有効であると考えられる。

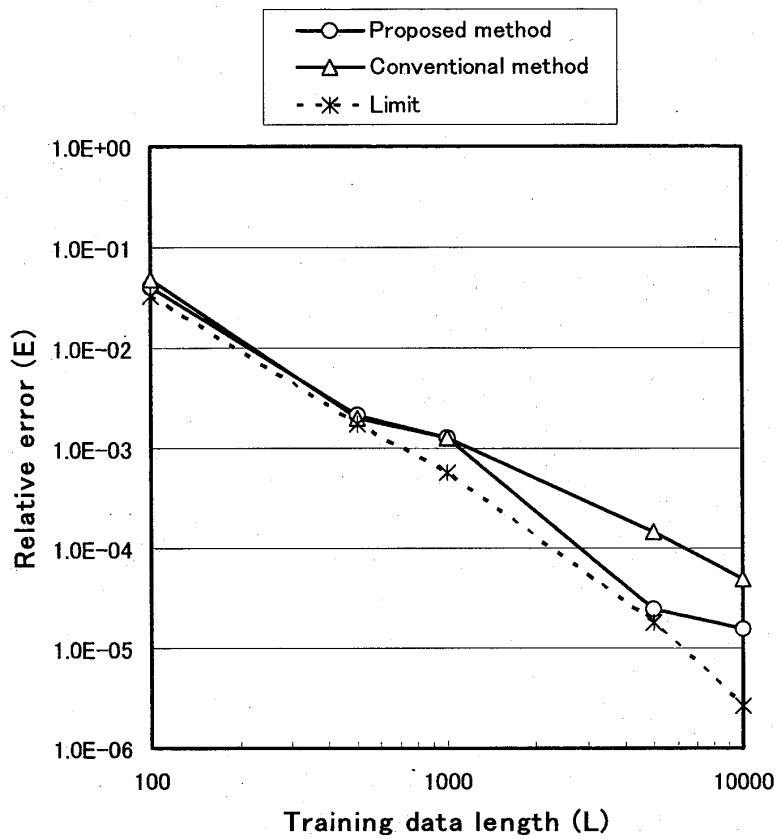


図2 Hénon mapにおける学習データ長と相対誤差の関係
 Fig. 2 Relative error vs. training data length for the Hénon map.

表2 Hénon mapにおける埋込み次元と近傍点の数
 Table 2 Embedding dimension and number of nearest neighbors for the Hénon map.

学習データ長	埋込み次元			近傍点の数		
	提案手法	従来手法	限界	提案手法	従来手法	限界
100	3	3	3	7	8	6
500	3	3	3	6	8	4
1000	3	3	3	8	8	9
5000	4	3	6	10	8	13
10000	4	3	6	10	8	13

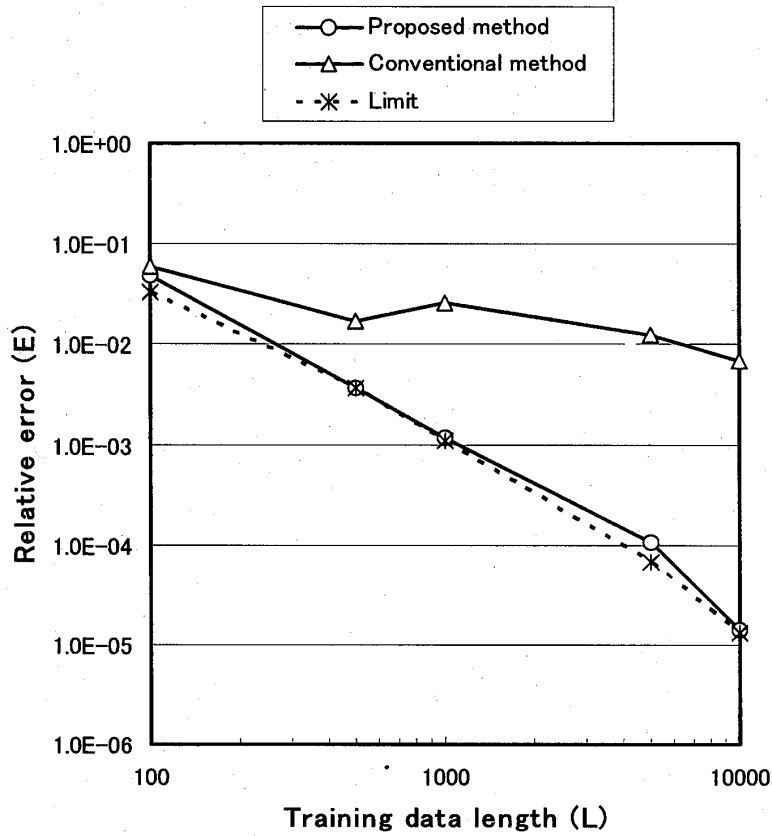


図3 Ikeda mapにおける学習データ長と相対誤差の関係
 Fig. 3 Relative error vs. training data length for the Ikeda map.

表3 Ikeda mapにおける埋込み次元と近傍点の数
 Table 3 Embedding dimension and number of nearest neighbors for the Ikeda map.

学習データ長	埋込み次元			近傍点の数		
	提案手法	従来手法	限界	提案手法	従来手法	限界
100	4	3	5	9	8	8
500	4	3	4	8	8	8
1000	4	3	4	8	8	6
5000	5	3	6	11	8	14
10000	6	3	6	14	8	13

5. むすび

本論文では、局所近似法において予測に適した埋込み次元と近傍点の数を決定する手法を提案した。さらに、Hénon mapおよびIkeda mapから生成されるカオス時系列データを用いて予測精度に関する評価実験を行い、提案手法の有効性を示した。

今後、実データに対して、提案手法の評価実験を行っていきたいと考える。

文献

- [1] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.*, vol.59, no.8, pp.845-848, Aug. 1987.
- [2] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, vol.344, no.19, pp.734-741, April 1990.
- [3] J. Jiménez, J. A. Moreno, and G. J. Ruggeri, "Forecasting on chaotic time series: A local optimal linear-reconstruction method," *Phys. Rev. A*, vol.45, no.6, pp.3553-3558, March 1992.
- [4] M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D*, vol.35, pp.335-356, 1989.
- [5] 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.
- [6] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, eds. D. A. Rand and L. S. Young, *Lecture Notes in Mathematics*, vol.898, pp.366-381, Springer, Berlin, 1981.
- [7] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Stat. Phys.*, vol.65, no.3,4, pp.579-616, 1991.
- [8] J. D. Farmer and J. J. Sidrowich, "Exploiting chaos to predict the future and reduce noise," in *Evolution, Learning and Cognition*, ed. Y. C. Lee, pp.277-330, World Scientific, Singapore, 1988.
- [9] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Phys. Rev. A*, vol.34, no.3, pp.2427-2432, September 1986.
- [10] J. P. Eckmann and D. Ruelle, "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," *Physica D*, vol.56, pp.185-187, 1992.
- [11] M. Hénon, "A two-dimensional mapping with a strange attractor," *Commun. Math. Phys.*, vol.50, pp.69-77, 1976.
- [12] K. Ikeda, "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system," *Opt. Commun.*, vol.30, no.2, pp.257-261, 1979.