

カオス時系列データの予測のための埋込み次元決定方法

伊 藤 憲 一

Determining embedding dimension for predicting chaotic time series

Ken-ichi ITOH

Abstract: In a local approximation technique for predicting chaotic time series, a state space is reconstructed from a time series using delay coordinate embedding and then a local predictor is constructed in the state space. The embedding dimension of the state space has a significant effect on the prediction accuracy. We evaluate the prediction error as a function of the embedding dimension for chaotic time series generated by the Hénon map and the Ikeda map and show basic guidelines for determining the embedding dimension.

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

1. ま え が き

時間と共に不規則に変動する時系列データがカオスの性質を有していれば、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。この考え方にに基づき、カオス時系列データの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を行う場合、まず1変数の時系列データから、埋込みの手法を用いて多次元空間上にアトラクタの軌道を再構成する[6], [7]。次に、このアトラクタの軌道を生成する力学モデルを推定する。この推定方法には、多次元空間の中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。局所近似法は、アトラクタ上のある点の予測値をその近傍点だけの推移をもとに推定する方法であり、アトラクタ全体に対する

モデルの構築方法に比べて、データへの適合が容易である[8]。

局所近似法により予測を行う場合、埋込み次元をいくつに設定するかは、予測精度上重要な問題である。埋込み次元がアトラクタの再構成に必要な最小の埋込み次元と同等以上であれば、もとのアトラクタの構造が埋込んだ多次元空間上に保存されるため、局所近似法による予測は可能と考えられる。通常は、最小の埋込み次元を用いて予測が行われるが、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。しかしこれまで、予測精度の観点から埋込み次元をいくつに設定すべきかについて、十分な検討はなされていなかった。

本論文では、埋込み次元と予測精度の関係を明らかにするために、Hénon map[9]およびIkeda map[10]から生成されるカオス時系列データを用いて評価実験を行う。本実験によ

り、埋込み次元に応じて予測精度がどのように変化するのかを明らかにし、予測に適した埋込み次元に関する考察を行う。

2. 局所近似法

カオス時系列データの予測手法の一つである局所近似法について、概要を述べる。

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み出され、1変数の時系列データ x_1, x_2, \dots, x_K が観測されたものとする。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に、一定の時間遅れ τ を用いて m 次元空間上に次のような 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 t までの観測値 $x_1 \sim x_t$ が得られているとき、時間 p 経過後の値 x_{t+p} を予測することとする。観測値 $x_1 \sim x_t$ 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_t に対応する点 X_t と時間 p 経過後の予測点 X_{t+p} の関係を関数 F により近似する。

$$X_{t+p} \cong F(X_t) \quad (2)$$

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_t 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_t 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_{t+p} を予測するためには、 X_{t+p} の m 個の要素のうち1番目の要素 x_{t+p} のみを求めればよいので、次の線形近似式により推定する。

$$x_{t+p} \cong f(X_t)$$

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

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

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

次に、 r の小さい順に n 個の点を X_t 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_{t+p} を求めることができる。

最小2乗解を得るためには、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。

3. 評価実験

局所近似法における埋込み次元および近傍点の数と予測精度との関係を明らかにするために、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を述べる。

3.1 予測対象データ

Hénon map[9]および Ikeda map[10]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。

3.2 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。

3.3 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 については、Hénon mapの $D_0=1.26$, Ikeda mapの $D_0=1.32$ であり、 $m > 2D_0$ が必要なことから、 m の最小値 m_{\min} は3となる。埋込み次元と相対誤差の関係を調べるため、 $m=3$ から開始して $m=10$ まで変化させた。また、近傍点の数 n については、最小値 $n_{\min} = m+1$ から開始して、 n_{\min} の2倍、3倍、4倍、5倍と変化させた。

3.4 実験結果と考察

まず、近傍点の数が予測精度にどのような影響を与えるかについて調べた。Hénon mapにおける近傍点の数と相対誤差の関係を図1に、Ikeda mapにおける近傍点の数と相対誤差の関係を図2に示す。

図1より以下のことがいえる。 $n = n_{\min}$ の場合、一般に相対誤差は最小とはならず、 $n = 2n_{\min}$ で相対誤差は最小に近くなる。さらに n が大きくなると相対誤差は増加する傾向となる。

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
埋込み次元	m	3~10
近傍点の数	n	$\alpha \cdot n_{\min}$ $\left(\begin{array}{l} n_{\min} = m+1, \\ \alpha = 1,2,3,4,5 \end{array} \right)$

図2の場合、埋込み次元 m および学習データ長 L にかかわらず、 $n = 2n_{\min}$ で相対誤差は最小となる。

以上、図1および図2の結果から、局所近似法においては、近傍点の数 n は $n = 2n_{\min}$ 程度が適当であるといえる。 $n = n_{\min}$ の場合は個々の近傍点の動きに解が大きく依存するため、最小2乗解の安定性が損なわれ、相対誤差が増加したと考えられる。また、 n が $n = 2n_{\min}$ よりも大きくなると、予測点から離れた点まで近傍点として選択されるため、最小2乗解が予測点の局所的な動きを反映なくなり、相対誤差が増加したと考えられる。 $n = 2n_{\min}$ が適当であるという本実験の結果は、従来の報告[4]とも一致する。

次に、 $n = 2n_{\min}$ という設定のもとで、予測精度上最適な埋込み次元を調べた。Hénon map について、埋込み次元と相対誤差の関係を図3に、学習データ長と相対誤差の関係を図4に示す。また、Ikeda map について、埋込み次元と相対誤差の関係を図5に、学習データ長と相対誤差の関係を図6に示す。

図3、図4より以下のことがいえる。Hénon map の場合、学習データ長 L が小さい時(1000以下)は埋込み次元 m を $m = m_{\min}$ とすることで特に問題はないが、 L が大きい時(5000以上)は m を大きくすることにより相対誤差は低下し、 $m = 6$ で相対誤差は最小となる。さらに m を大きくすると相対誤差は増加する。すなわち、学習データ長 L が大きい時は、予測精度上最適な埋込み次元は、 m_{\min} ではないことがわかる。

図5、図6より以下のことがいえる。Ikeda map の場合、学習データ長 L が小さい時(1000以下)は $m = 4$ で、 L が大きい時(5000以上)は $m = 6$ で相対誤差は最小となる。さらに m を大きくすると相対誤差は増加する。 $m = m_{\min}$ とするのは予測精度上最適ではない。

以上、図3～図6の結果から、予測精度上最適な埋込み次元は学習データ長 L により変化し、 $m = m_{\min}$ が常に最適とはいえないこと、学習データ長 L が大きい程最適埋込み次

元は m_{\min} よりも大きくなるのがわかる。学習データ長 L と最適埋込み次元の関係は、時系列データの種類にも依存すると考えられ、一意に最適埋込み次元を決定することはできないが、少なくとも埋込み次元 $m = m_{\min}$ が予測精度上最適ではないことを認識しておくことは重要である。Hénon map および Ikeda map の場合は、埋込み次元 m を $m = m_{\min} + 1$ 程度とするのが、学習データ長にあまり影響されない現実的な解であるといえる。

4. むすび

本論文では、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、局所近似法における埋込み次元および近傍点の数と予測精度との関係を調べた。

この結果、以下の点を明らかにした。

(1) 近傍点の数は、最小2乗解の算出に必要な最小値の2倍程度とするのが好ましい。

(2) 予測精度上最適な埋込み次元は、埋込みに必要な最小値とは限らない。学習データ長が大きい程、最適埋込み次元は最小値よりも大きくなる。Hénon map および Ikeda map の場合は、埋込み次元を最小値 + 1 程度にするのが適当である。

今後、実データに対して、予測精度上最適な埋込み次元に関する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] M. Hénon, "A two-dimensional mapping with a strange attractor," *Commun. Math. Phys.*, vol.50, pp.69-77, 1976.
- [10] 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.

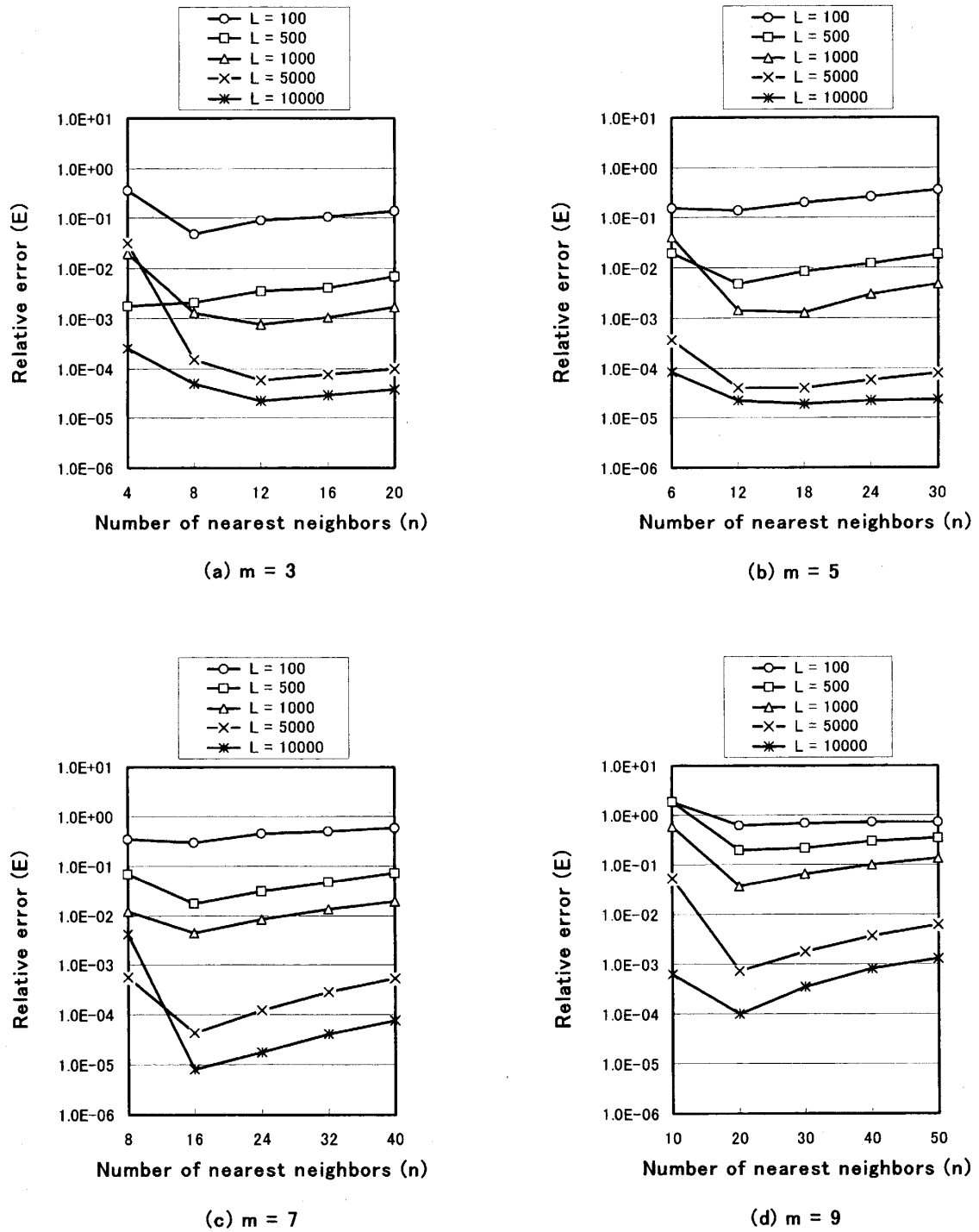


図1 Hénon mapにおける近傍点の数と相対誤差の関係
 Fig. 1 Relative error vs. number of nearest neighbors for the Hénon map.

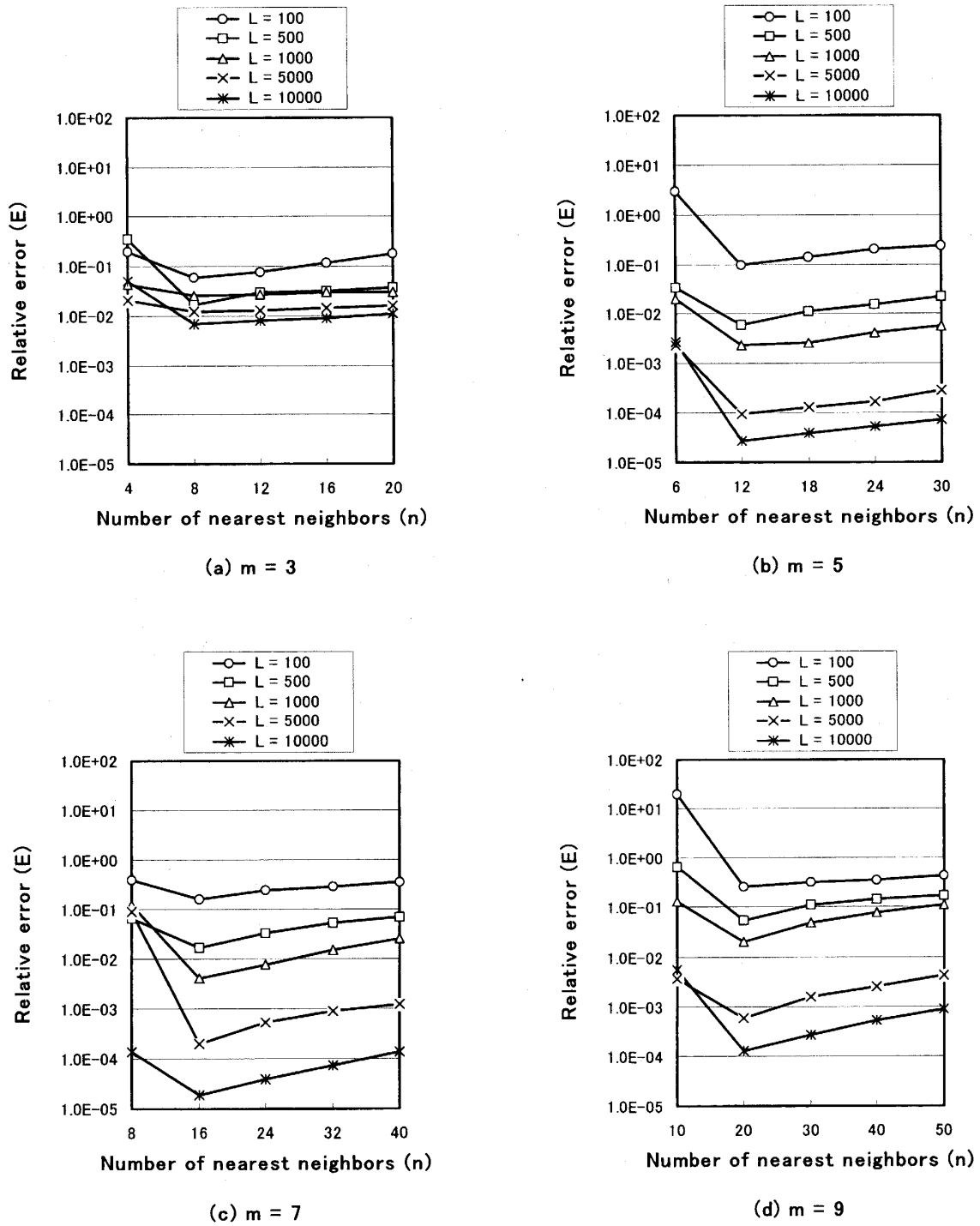


図2 Ikeda mapにおける近傍点の数と相対誤差の関係
 Fig. 2 Relative error vs. number of nearest neighbors for the Ikeda map.

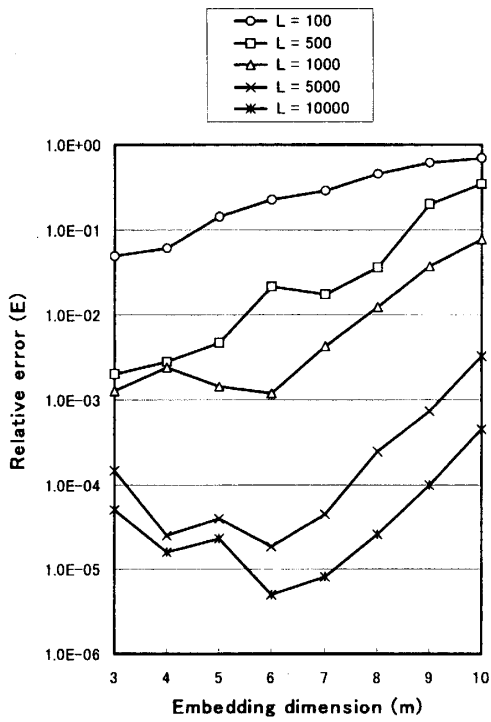


図3 Hénon mapにおける埋込み次元と相対誤差の関係
 Fig. 3 Relative error vs. embedding dimension for the Hénon map.

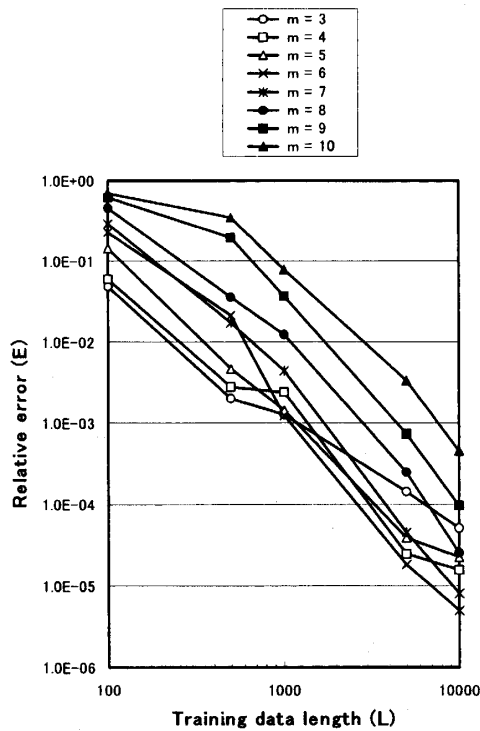


図4 Hénon mapにおける学習データ長と相対誤差の関係
 Fig. 4 Relative error vs. training data length for the Hénon map.

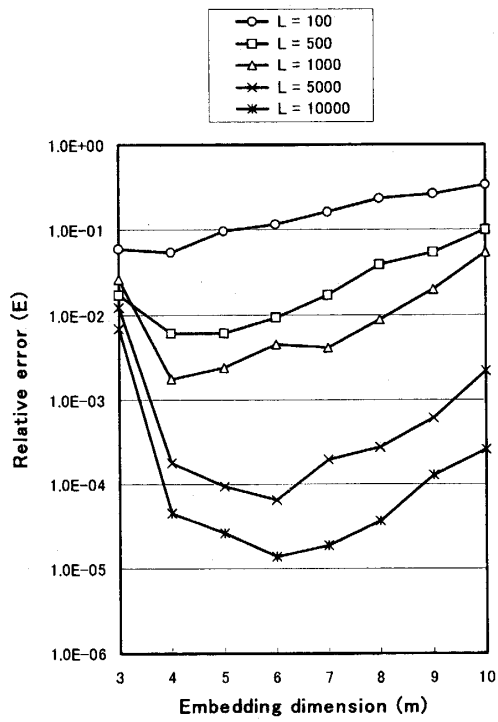


図5 Ikeda mapにおける埋込み次元と相対誤差の関係

Fig. 5 Relative error vs. embedding dimension for the Ikeda map.

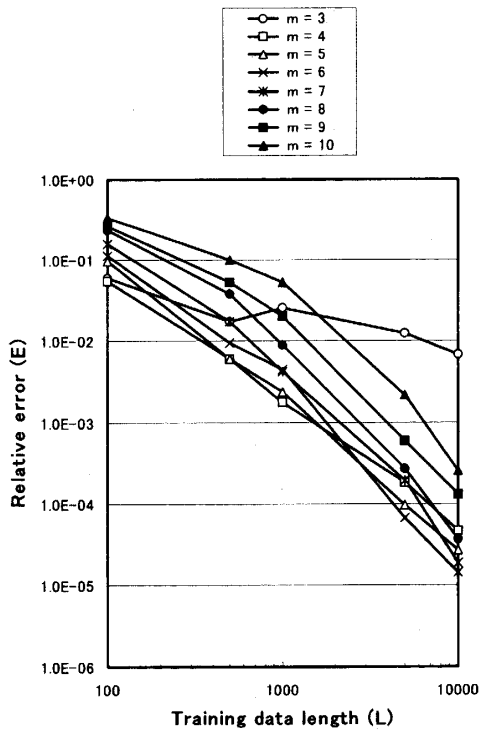


図6 Ikeda mapにおける学習データ長と相対誤差の関係

Fig. 6 Relative error vs. training data length for the Ikeda map.