

雑音を含むカオス時系列データの子測

伊 藤 憲 一

Predicting Chaotic Time Series Corrupted with Measurement Noise

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 coordinates and then a local predictor is constructed on the basis of the motion of the nearest neighbors in the state space. The number of the nearest neighbors has an important effect on the prediction accuracy. We evaluate the prediction error as a function of the number of the nearest neighbors for time series of the Hénon map and the Lorenz model contaminated with Gaussian white noise and show basic guidelines for selecting the nearest neighbors.

Key Words: chaos, time series, prediction

1. ま え が き

不規則に変動する時系列データの背後に何らかの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、従来の統計的な方法に比べ精度の良い予測が可能と考えられる。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、短期的には決定論的な特徴を生かした精度の良い予測が可能となる。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、近年活発である[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、時間遅れの方法[6]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を推定する。この推定方法には、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

ラルネットワークを用いる方法[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る。アトラクタ全体に対する近似モデルを構築する場合に比べて、一般にデータへの適合が容易であり、特にデータ数が大きい場合その利点は大きい[7]。局所近似法においては、近傍点の数をいくつに決定するかが予測精度上重要な問題である。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ようになり、解の安定性が損なわれる。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ているため、近傍点の数が少ない程雑音の影響を受けやすくなる。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しなくなり、予測誤差が増加する。

本論文では、近傍点の数と予測精度の関係を明らかにするために、Hénon map[8]および Lorenz model[9]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。実験では、雑音の影響を調べるために、上記カオス時系列データにガウス分布型雑音を加えたデータを用いる。本実験により、予測精度が近傍点の数や雑音にどのような影響を受けるかを明らかにする。

2. 局所近似法

ここでは、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の概要を述べる。

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 m 次元ベクトル X_i を作成する[6]。

$$X_i = (x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau}), \\ i = 1 + (m-1)\tau, 2 + (m-1)\tau, \dots, K-1, K \quad (1)$$

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[10]。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 とするとき、 $m > 2D_0$ であれば埋込みになることが示されている[11]。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。

いま、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 後の値 x_{t+p} を予測する問題を考える。 $x_1 \sim x_t$ 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_i と時間 p 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。すなわち、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、これらの複数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} 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。

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

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、 n の値をどのように決定するかが予測精度上重要な問題である。 n の値が小さすぎると個々の近傍点の動きに解が大きく依存するようになり、最小2乗解の安定性が損なわれる。また、 n の値が大きすぎると、 X_i から離れた点まで近傍点として選択されるため、最小2乗解は X_i の局所的な動きを反映しなくなり、予測誤差が増加する。

3. 評価実験

局所近似法における近傍点の数と予測精度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[8]および Lorenz model[9]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。また、雑音の影響を調べるために、上記カオス時系列データにガウス分布型雑音を加えたデータも使用した。

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。

Lorenz modelは次の3変数微分方程式である。

$$\begin{aligned} \dot{x} &= -\sigma(x-y), \\ \dot{y} &= -y - xz + rx, \\ \dot{z} &= xy - bz \end{aligned} \quad (8)$$

式(8)において、パラメータ値 $\sigma=10$, $r=28$, $b=8/3$, 初期値 $x=0.1$, $y=0$, $z=0$, 時間刻み $\delta t=0.01$ として4次のルンゲ・クッタ法で x の時間変化を求め、これを時系列データとした。Lorenz modelについては、初期の過渡的なデータを取り除くため、最初の3000点のデータは使用せず、それ以降のデータを時系列データとして用いた。

雑音データを生成するために、上記Hénon mapおよび Lorenz modelの時系列データに対し、標準偏差の0.1%、1%、10%のガウス分布型雑音を加えた。

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。

実験で使用したパラメータ値を表1に示す。学習データ長 L は、400から3200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 $p=1$, 埋込み次元 $m=3$ とした。

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	400, 800, 1600, 3200
予測データ長	N	L と同一
予測ステップ数	p	1
埋込み次元	m	3
時間遅れ	τ	1 (Hénon map) 10 (Lorenz model)
近傍点の数	n	4 ~ 64

時間遅れ τ については、Hénon map の場合は 1, Lorenz model の場合は 10 とした。近傍点の数 n は、最小 2 乗解の算出に必要な最小値 $n_{\min} = 4$ から 64 まで変化させた。

3.4 実験結果と考察

Hénon map および Lorenz model の時系列データにガウス分布型雑音を加えたデータに対して、局所近似法による予測を行い、近傍点の数と相対誤差の関係を調べた。結果を図 1, 図 2 に示す。

図 1 より、Hénon map については以下のことがいえる。雑音の割合が 0% (オリジナルデータ) の場合、相対誤差 E は近傍点の数 n にかなり依存し、 $n = 10$ 前後が最小となる。 n がさらに大きくなると相対誤差は増大し、逆に $n = n_{\min}$ の近辺でも相対誤差は増大する。また、学習データ長 L が大きい程相対誤差は減少し、 $L = 400$ と $L = 3200$ とでは 2 桁程度の差となる。

雑音の割合が増加するにつれて、相対誤差は増加する。同時に、 $n = n_{\min}$ 近辺を除き、相対誤差は近傍点の数 n および学習データ長 L にだんだん依存しなくなる。雑音の割合が 10% になると、 $n = n_{\min}$ 近辺以外では相対誤差は n および L にほとんど依存しない。一方、 $n = n_{\min}$ 近辺では相対誤差は極めて大きい。

図 2 より、Lorenz model については以下のことがいえる。雑音の割合が 0% (オリジナルデータ) の場合、相対誤差 E は近傍点の数 n にそれ程依存しないが、 $n = n_{\min}$ 近辺では相対誤差は大きくなる。また、Hénon map の場合と同様、相対誤差は学習データ長 L に大きく依存し、 $L = 400$ と $L = 3200$ で 2 桁程度の差となる。

雑音の割合が増加してくると、Hénon map の場合と同様、相対誤差は増加するとともに、 $n = n_{\min}$ 近辺を除き、相対誤差は近傍点の数 n および学習データ長 L にだんだん依存しなくなる。 $n = n_{\min}$ 近辺の相対誤差は極めて大きい。

以上の結果より、局所近似法を用いた時系列データの予測においては、予測精度向上の観点から次の点が重要と考えられる。

(1) 近傍点の数は、最小 2 乗解の算出に必要な最小値の数倍程度にする必要がある。特に、実データには通常雑音が含まれているため、近傍点の数を大きくしても予測誤差はそれ程増加しないものと推定できる。逆に近傍点の数を最小値近辺にするのは、避ける必要がある。

(2) 学習データ長は、可能な限り大きくするのがよい。ただし、雑音の割合が増加するにつれて、学習データ長を大きくする効果は減少する。

4. むすび

本論文では、Hénon map および Lorenz model から生成されるカオス時系列データにガウス分布型雑音を加えたデータを用いて、局所近似法の予測精度に関する評価実験を行った。

この結果、予測精度が近傍点の数や雑音にどのような影響を受けるかを明らかにし、近傍点の選択方法などに関する指針を示した。

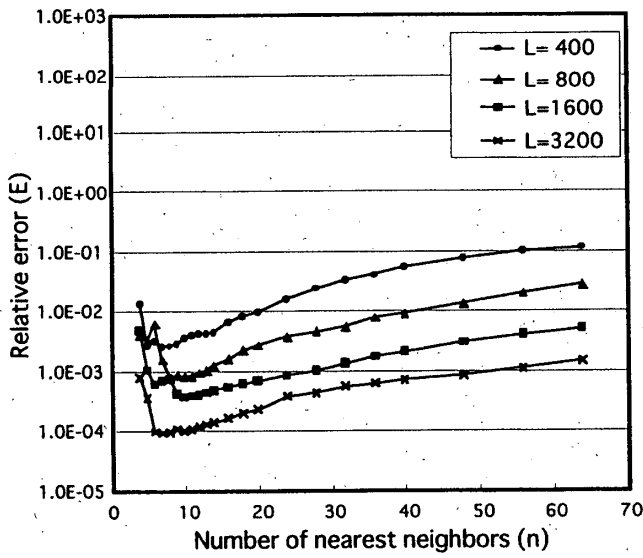
今後、実データに対して、局所近似法の評価実験を行っていきたいと考える。

文献

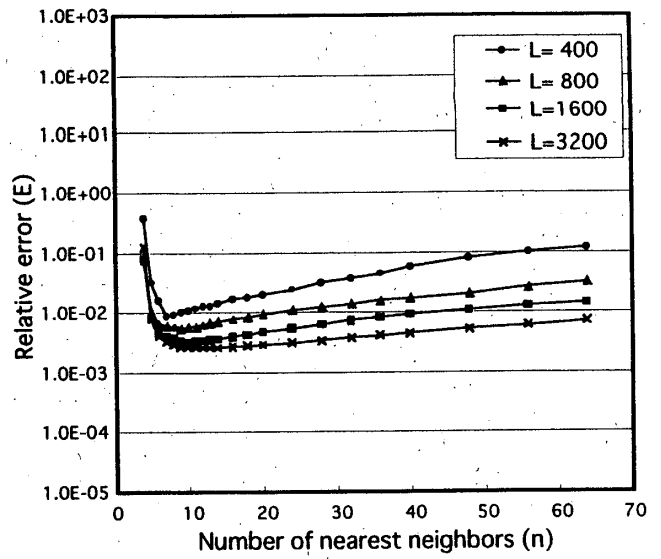
- [1] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.*, vol.59, no.8, pp.845-848, Aug. 1987.
- [2] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, vol.344, no.19, pp.734-741, April 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] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Phys. Rev. Lett.*, vol.45, no.9, pp.712-716, Sept. 1980.
- [7] 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.

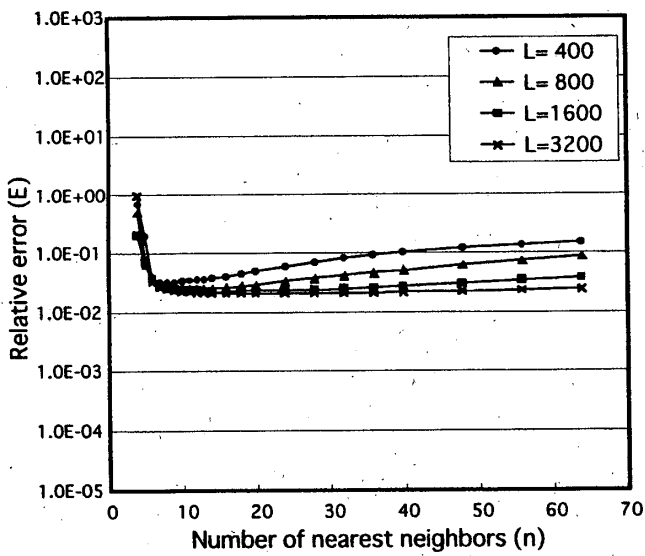
- [8] M. Hénon, "A two-dimensional mapping with a strange attractor," *Commun. Math. Phys.*, vol.50, pp.69-77, 1976.
- [9] E. N. Lorenz, "Deterministic non-periodic flow," *J. Atmos. Sci.*, vol.20, pp.130-141, 1963.
- [10] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, eds. D. A. Rand and L. S. Young, *Lecture Notes in Mathematics*, vol.898, pp.366-381, Springer, Berlin, 1981.
- [11] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Stat. Phys.*, vol.65, no.3,4, pp.579-616, 1991.



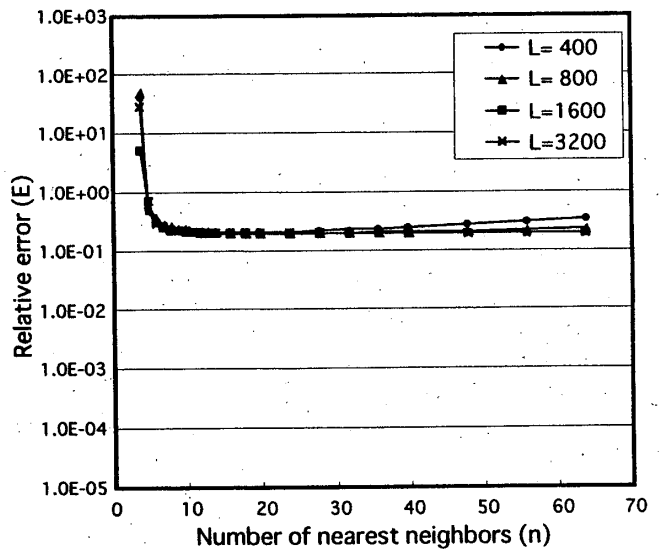
(a) Noise rate = 0 %



(b) Noise rate = 0.1 %

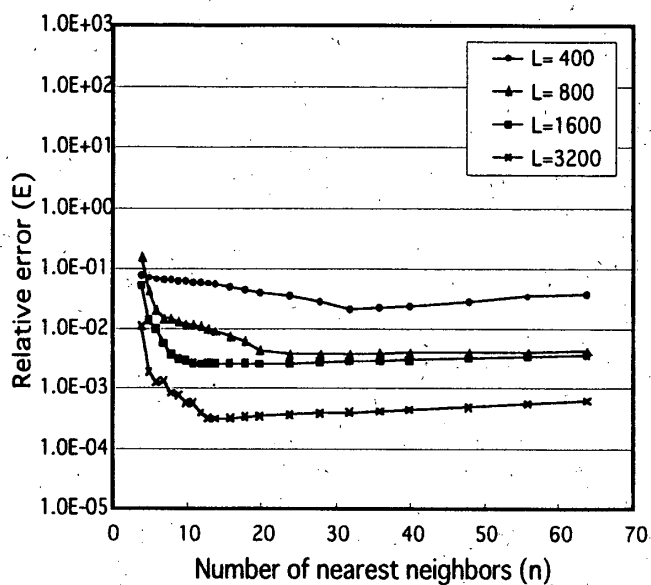


(c) Noise rate = 1 %

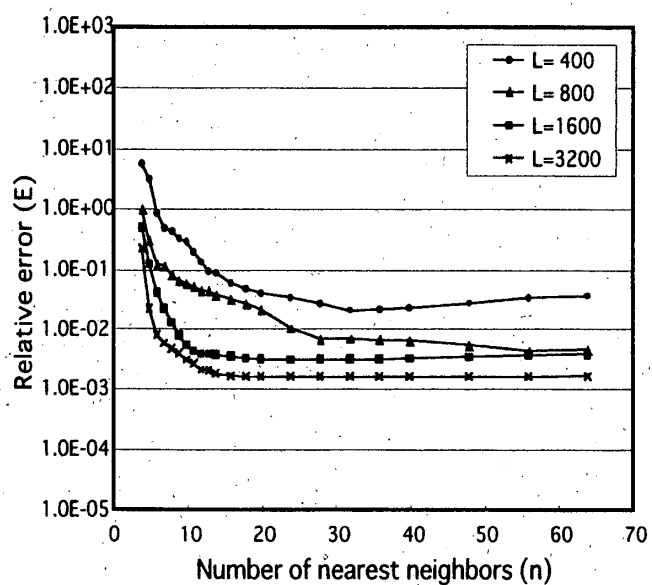


(d) Noise rate = 10 %

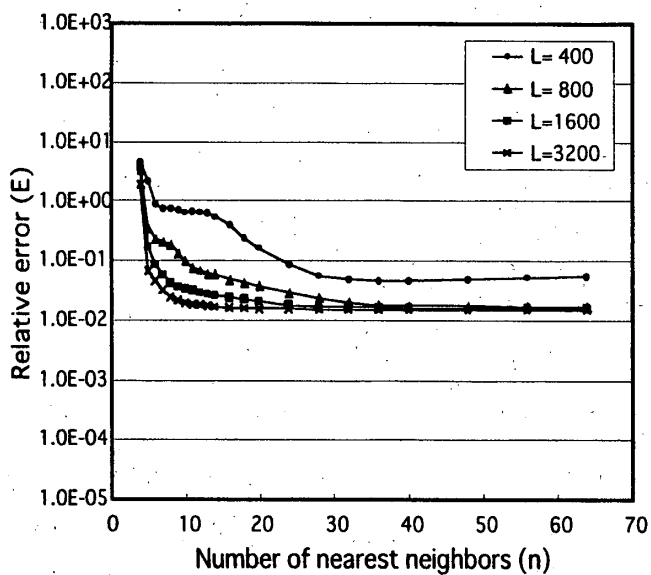
図1 Hénon mapにおける近傍点の数と相対誤差の関係
 Fig. 1 Relative error vs. number of nearest neighbors for the Hénon map.



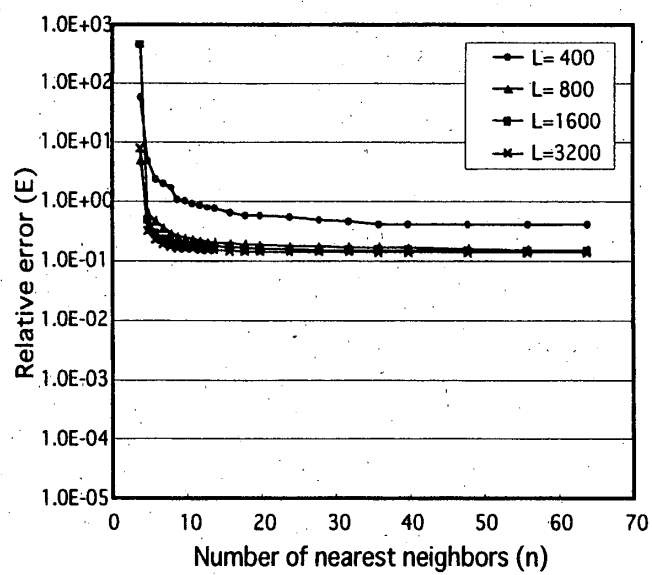
(a) Noise rate = 0 %



(b) Noise rate = 0.1 %



(c) Noise rate = 1 %



(d) Noise rate = 10 %

図2 Lorenz model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 Lorenz model.