

# 基于信息准则的局域预测法邻近点的选取方法<sup>\*</sup>

孟庆芳<sup>†</sup> 彭玉华 曲怀敬 韩 民

(山东大学信息科学与工程学院, 济南 250100)

(2007 年 5 月 23 日收到, 2007 年 7 月 1 日收到修改稿)

基于信息准则, 提出了选取局域预测法中邻近点个数的定量方法. 并用此方法分析 Lorenz 模型生成的混沌时间序列和 Santa Fe 时间序列竞赛的激光数据(Data A). 实验结果表明用该方法选取邻近点的局域预测法的一步和多步预测性能较好, 在满足预测精度较高的条件下, 计算量较小.

关键词: 邻近点, 邻域, 局域预测, 信息准则

PACC: 0545

## 1. 引 言

近 20 年来, 许多预测混沌时间序列的方法被提出, 它们可被分为两类, 全局预测法和局域预测法<sup>[1]</sup>. 全局预测法利用全部的去信息来预测未来值, 用全部已知数据来拟合动力方程. 例如用支持向量机建立的全局预测模型, 它通过全部的输入-输出对支持向量机进行训练. 但当新的数据加入已知样本时, 全局预测模型须重新估计其参数, 因此全局预测法计算量较大. 并且因为吸引子的结构非常复杂, 所以拟合全局动力方程的难度也往往较大. 而局域预测法仅利用部分的过去信息来预测未来值, 局域动力方程较容易拟合, 且计算量较小. Farmer 和 Sidorowich 早已证明, 在相同的嵌入维数下, 局域预测法的效果比全局预测法更好<sup>[2]</sup>.

正是因为如此, 局域预测法受到了学术界的重视和关注, 成为预测混沌时间序列最基本的方法之一. 文献[2]和文献[3]曾在这方面作过开创性的研究, 堪称经典. 时至今日, 国内外对局域预测法的研究仍然在继续进行, 国内近些年来在这方面的研究如文献[4—7]等所示.

在建立局域预测模型前, 首先必须确定三个参数, 即嵌入维数、延迟时间与邻近点个数. 关于嵌入维数和延迟时间的选取, 人们已经作了大量的研究, 有不少定量的方法可供应用, 而邻近点的个数却大

多仅凭经验选取, 或者简单地取为  $m+1$  ( $m$  为嵌入维数)<sup>[3]</sup>, 难以见到关于邻近点个数选取的定量研究. 其实, 邻近点的个数既与计算量密切相关, 也影响到局域模型的预测精度, 同样是十分重要的一个参数. 本文基于信息准则, 提出了确定局域预测法中邻近点个数的定量方法. 对 Lorenz 模型生成的混沌时间序列和 Santa Fe 时间序列竞赛的激光数据(Data A)的计算结果表明, 这种方法是切实可行的.

## 2. 用信息准则选取局域预测法邻近点

相空间重构理论是非线性时间序列分析的基础<sup>[1]</sup>. 假设观测到的混沌时间序列为  $\{x(t), t=1, 2, \dots, N\}$ , 由延迟坐标相空间重构法<sup>[8]</sup>可得延迟矢量为

$$X(n) = [x(n), x(n-\tau), \dots, x(n-(m-1)\tau)]^T, \\ n = (m-1)\tau + 1, (m-1)\tau + 2, \dots, N. \quad (1)$$

其中  $m$  为嵌入维数,  $\tau$  为延迟时间. 由嵌入定理知, 在重构的相空间中, 存在一个光滑映射  $f: R^m \rightarrow R$ , 使得

$$x_p(n+T) = f(X(n)). \quad (2)$$

根据对动力方程的拟合方式, 混沌时间序列预测可分为全局法与局域法. 其中, 全局法用全部已知数据在整个重构的混沌吸引子上拟合动力方程; 而局域法则是通过分段逼近的形式来拟合函数  $f$ , 即每次利用相空间中的几个局域邻近点来逼近函数

<sup>\*</sup> 教育部留学启动基金(批准号 2004.176.4); 山东省自然科学基金(批准号 2004G01, 2004ZRC03016)资助的课题.

<sup>†</sup> E-mail: qfmeng@sohu.com

$f$  的一个局域子集,更能体现混沌系统的动态性,具有比全局法更好的预测性能.

局域预测的首要步骤是在重构的相空间中选取当前延迟矢量  $X(n)$  的邻近点. 根据重构轨迹,计算当前延迟矢量  $X(n)$  与前面的  $n-1$  个延迟矢量  $X(i), i=1, 2, 3, \dots, n-1$  的距离

$$d(i) = \|X(i) - X(n)\|_2, \quad (3)$$

可得到  $k$  个最近的邻近点  $X(n_i), i=1, 2, \dots, k$ . 局域线性预测模型为

$$x_p(n+T) = a_n X(n) + b_n. \quad (4)$$

其中  $a_n$  为  $1 \times m$  维行向量,  $b_n$  为常数,  $T$  为预测步长. 局域线性预测法利用当前延迟矢量  $X(n)$  的邻近点  $X(n_i), i=1, 2, \dots, k$  来估计参数  $a_n$  和  $b_n$  的值. 参数  $a_n$  和  $b_n$  的值可通过如下最小化公式来得到:

$$\min \sum_{i=1}^k |x(n_i+T) - a_n X(n_i) - b_n|^2. \quad (5)$$

实现多步预测的方法通常有两种,即直接法和迭代法. 直接法直接建模,实现对未来值  $x_p(n+T)$  的预测,即对(4)式模型预测步长  $T > 1$ ,直接实现多步预测. 迭代法每次仅做一步预测,即对(4)式模型预测步长  $T=1$ ,并把预测值  $x_p(n+1)$  加入原始数据,再预测  $x_p(n+2)$ ,依此类推,实现多步预测. 直接法的问题在于预测步长  $T > 1$  时模型比预测步长  $T=1$  时模型更复杂且更难于建模. 迭代法的问题在于在迭代多步预测过程中上一步预测误差将积累到下一步中. 大部分研究者发现迭代多步预测比直接多步预测更精确. 本文用迭代法实现多步预测.

局域线性预测法用线性映射来近似当前延迟矢量  $X(n)$  与未来值  $x_p(n+1)$  的演化方程,并用  $X(n)$  的邻近点来近似该线性映射. 文献[6,7]用支持向量回归模型代替局域线性模型,并用邻近点训练该局域支持向量回归模型,提出局域支持向量机预测法. 即把  $X(n_i), i=1, 2, \dots, k$  作为支持向量机预测模型中的输入值,  $x(n_i+1), i=1, 2, \dots, k$  作为输出值训练支持向量机,得到对应的支持向量及权重系数(即  $f$  一个子集的逼近);然后利用训练得到的局域支持向量机,计算得到对未来值的预测值  $x_p(n+1)$ .

选取邻近点的方法通常有两种,即固定邻近点个数法和固定邻域半径法. 固定邻域半径法是指选取落在以当前延迟矢量  $X(n)$  为中心,半径取固定值的超球内的点为邻近点,即对邻近点  $X(n_i), i=$

$1, 2, \dots, k$  满足下列条件:

$$\|X(n_i) - X(n)\|_2 \leq r, i=1, 2, \dots, k. \quad (6)$$

固定邻域半径法对不同的点选取的邻近点的个数将不同.

局域预测模型由三个参数决定,即嵌入维数、延迟时间与邻近点个数. 邻近点个数影响局域模型的预测精度和计算量. 对于局域线性预测法,如果邻近点的个数  $k$  取的太小或者邻域的半径  $r$  取的太小,邻近点将过少,那么将不能充分利用历史信息,并且将增加噪声的影响. 文献[2,3]提出邻近点个数应大于或等于  $m+1$ . 如果邻近点的个数  $k$  取的太大或者邻域的半径  $r$  取的太大,邻近点将过多,那么局域线性模型的线性假设条件将不满足,将降低模型预测精度,并且,邻近点过多时,邻近点将包括与基准点所在轨道相距较远的轨道上的那些点,由于混沌系统本身所存在的指数发散的性质,那些点加入到模型中以后,将降低模型预测精度. 所以,对于局域线性预测法,在满足预测精度较高的条件下,邻近点不宜过多.

对于局域支持向量机预测法,如果邻近点的个数  $k$  取的太小或者邻域的半径  $r$  取的太小(这也意味着邻近点过少),则将不能发挥局域支持向量回归模型非线性拟合的优势,不能提高模型的预测精度. 文献[7]就提出局域支持向量机预测法的邻近点个数必须大于一定的数值. 同时,局域支持向量机预测法不受局域线性这一假设条件的限制,因此其邻近点个数和邻域半径也都完全可以比局域线性预测法更大. 但是,另一方面,如果邻近点的个数  $k$  取得太大或者邻域的半径  $r$  取得太大,不仅局域支持向量机预测法的计算量将大大增大,计算时间大大增加,而且有可能出现“过拟合(overfitting)”的现象,反而降低预测的精度. 所以,对于局域支持向量机预测法,在满足预测精度较高的条件下,邻近点也不宜过多.

人们早就发现,在时间序列建模问题中,当模型参数取的太多时,虽然模型拟合误差较小,但是模型外推预测的误差却反而较大,这就是过拟合的情况. 为了避免出现过拟合的情况, Akaike 提出了确定 ARMA 模型阶数的 AIC 信息准则<sup>[9]</sup>,即对于 ARMA  $(p, q)$  模型,使下列 AIC 值达到最小:

$$AIC(p, q) = \ln \hat{\sigma}^2 + (p + q + 1) \frac{2}{N'}, \quad (7)$$

其中  $\hat{\sigma}^2$  为拟合误差的方差,  $N'$  为拟合数据个数.

AIC 准则得到了普遍的认同和广泛的应用. 但

是,人们也发现了 AIC 准则仍然存在一些局限性.有人证明这一准则确定的模型阶数在理论上并不是相容估计,即当样本长度趋于无穷大时,用 AIC 准则定出的模型阶数估计值并不能依概率收敛到真值.此外,在实际问题中,AIC 准则所要求的那些前提条件也不一定成立.因此人们又对 AIC 准则作了一些改进,提出了其他一些准则.

Akaike 在贝叶斯原理的基础上提出了 AIC 准则的贝叶斯改进形式,即 BIC 信息准则<sup>[10]</sup>.对于 ARMA( $p, q$ )模型,使下列 BIC 值达到最小:

$$BIC(p, q) = \ln \hat{\sigma}^2 + (p + q + 1) \frac{\ln N'}{N'}. \quad (8)$$

它所确定的阶数是真阶的相容估计.显然,BIC 准则第二项的系数比 AIC 准则的大.

Hannan 和 Quinn 提出了另一种定阶准则<sup>[11]</sup>,即

$$\Phi(K) = \ln \hat{\sigma}^2 + K \frac{D \ln \ln N'}{N'}, \quad (9)$$

其中  $K = p + q + 1$ ,为独立参数个数; $D$  为常数,且  $D > 2$ , $D$  值越大,第二项的权重越大,则阶数的估计值越偏低.该准则所确定的阶数也是真阶的相容估计.Hannan-Quinn 准则第二项的系数也比 AIC 准则的大.用 BIC 准则和 Hannan-Quinn 准则得到的阶数比用 AIC 准则得到的阶数低.

由于一步预测误差不能反映模型迭代多步预测过程中误差积累作用,本文用多步预测误差来表征模型的短期可预测性能.把训练样本分为两部分.把训练样本第二部分的数据均匀地分为  $l$  段,并对每段数据进行多步预测,预测步长为  $n_a$ .其中,为了充分利用已知数据,对训练样本第二部分某段数据进行多步预测时,使用训练样本除去该段数据的所有剩余数据寻找邻近点.用归一化均方误差来表征模型的短期可预测性能,其定义为

$$\epsilon^2 = \frac{\frac{1}{l \cdot n_a} \sum_{i=1}^l \sum_{j=1}^{n_a} |x[\nu(i) + j - 1] - x_p[\nu(i) + j - 1]|^2}{(1/N) \sum_{i=1}^N [x(i) - \bar{x}]^2}, \quad (10)$$

其中  $\nu(i)$  为每段数据的起始数据的序号,  $\bar{x} = (1/N) \sum_{i=1}^N x(i)$ .

本文分别用 BIC 信息准则和 Hannan-Quinn 信息准则,确定局域线性预测法和局域支持向量机预测法的邻近点个数和邻域半径.

在采用 BIC 信息准则时,通过对以下两式极小化来选取局域线性预测法和局域支持向量机预测法的邻近点个数和邻域半径.

对于固定邻近点个数法

$$\alpha(k) = \ln \epsilon(k)^2 + (k + 1) \frac{\ln N'}{N'}. \quad (11a)$$

对于固定邻域半径法

$$\alpha(r) = \ln \epsilon(r)^2 + (\bar{k} + 1) \frac{\ln N'}{N'}. \quad (11b)$$

其中  $\bar{k}$  为当  $r$  取某固定值时邻近点个数的平均值,拟合数据个数  $N' = l \cdot n_a$ .

在采用 Hannan-Quinn 信息准则时,通过对以下两式极小化来选取局域线性预测法和局域支持向量机预测法的邻近点个数和邻域半径.

对于固定邻近点个数法

$$\alpha(k) = \ln \epsilon(k)^2 + (k + 1) \frac{D \ln \ln N'}{N'}. \quad (12a)$$

对于固定邻域半径法

$$\alpha(r) = \ln \epsilon(r)^2 + (\bar{k} + 1) \frac{D \ln \ln N'}{N'}. \quad (12b)$$

$D$  为常数,且  $D > 2$ .

### 3. 实验结果及讨论

本文以 Lorenz 模型生成的混沌时间序列和 Santa Fe 时间序列竞赛的激光数据(Data A)为实验数据,验证本文提出的选取局域预测法邻近点方法的有效性.

Lorenz 模型为

$$\begin{aligned} \dot{x} &= \sigma(y - z), \\ \dot{y} &= x(R - z) - y, \\ \dot{z} &= xy - bz, \end{aligned} \quad (13)$$

其中  $\sigma = 10, R = 28, b = 8/3$ ,用四阶 Runge-Kutta 算法求解获得数据,采样间隔  $\delta = 0.01$ ,丢弃过渡数据,取一组  $x$  变量生成的数据作为仿真数据,序列总长度为 2500,前 2000 个点作为训练样本,后 500 个点

作为测试样本. 对该序列相空间重构时, 嵌入维数取  $m = 3$ , 延迟时间取  $\tau = 14$ . 局域支持向量回归模型的核函数取为径向基核函数

$$K(x, y) = \exp(-\|x - y\|_2^2 / 2\sigma^2),$$

参数取为  $C = 10^6$ ,  $\epsilon = 0.00001$ ,  $\sigma = 2$ .

将该序列训练样本分为两部分, 前 1000 个点作为第一部分, 后 1000 个点作为第二部分. 用本文提出的方法选取局域线性预测法和局域支持向量机预测法的邻近点个数和邻域半径, 参数  $l = 200$ ,  $n_a = 3$ ,  $D = 8$ , 仿真结果如图 1—图 4 所示.

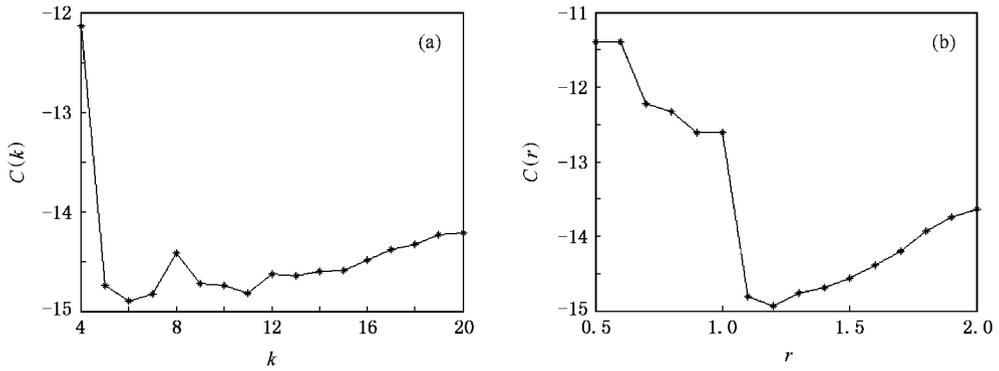


图 1 对 Lorenz 系统  $x$  分量混沌序列, 局域线性预测法的基于 BIC 准则的邻近点个数和邻域半径 (a)  $C(k) \sim k$ ; (b)  $C(r) \sim r$

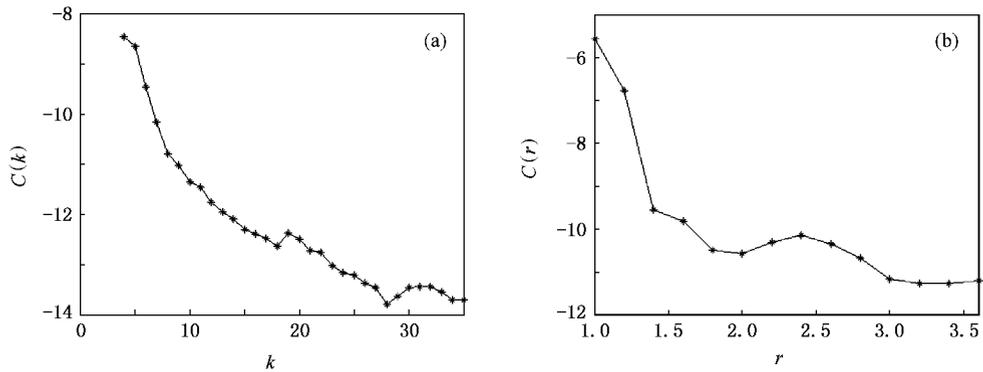


图 2 对 Lorenz 系统  $x$  分量混沌序列, 局域支持向量机预测法的基于 BIC 准则的邻近点个数和邻域半径 (a)  $C(k) \sim k$ ; (b)  $C(r) \sim r$

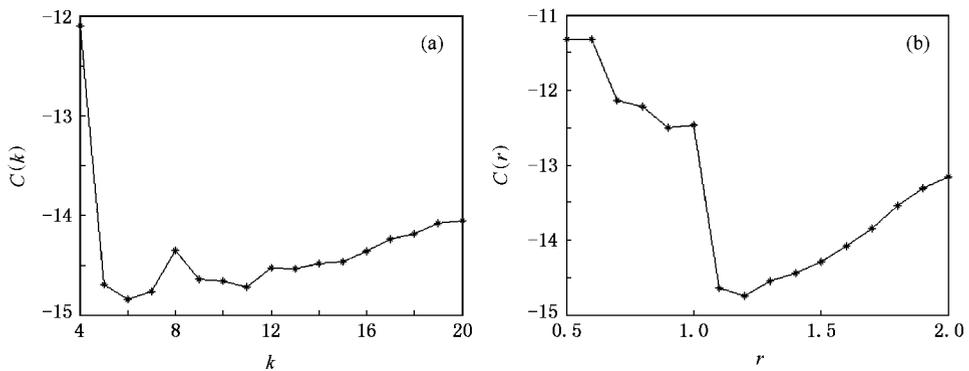


图 3 对 Lorenz 系统  $x$  分量混沌序列, 局域线性预测法的基于 Hannan-Quinn 准则的邻近点个数和邻域半径 (a)  $C(k) \sim k$ ; (b)  $C(r) \sim r$

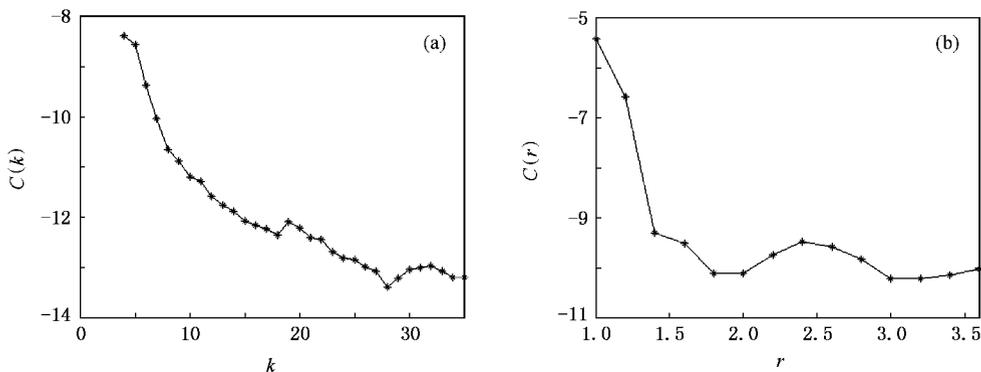


图4 对 Lorenz 系统  $x$  分量混沌序列,局域支持向量机预测法的基于 Hannan-Quinn 准则的邻近点个数和邻域半径  
(a)  $C(k) \sim k$ ; (b)  $C(r) \sim r$

为了分析用本文提出的方法选取邻近点的局域预测法的预测性能,对该序列测试样本进行 1 至 10 步预测,用归一化均方误差(NMSE)作为评测标准,该单步预测的归一化均方误差定义为

$$NMSE(T) = \frac{(1/L) \sum_{i=1}^L |x(N+i+T) - x_p(N+i+T)|^2}{(1/N) \sum_{i=1}^N [x(i) - \bar{x}]^2} \quad (14)$$

其中  $T=1, 2, \dots, 10$ , 为预测步长,  $L$  为外推预测数据长度. 仿真结果如表 1 所示.

对该组 Lorenz 模型生成的仿真数据,从图 1 和图 3 可见,用 BIC 准则和 Hannan-Quinn 准则选取的局域线性预测法的邻近点个数和邻域半径分别为  $k=6, r=1.2$ ;从图 2 和图 4 可见,用 BIC 准则和 Hannan-Quinn 准则选取的局域支持向量机预测法的邻近点个数和邻域半径分别为  $k=28, r=3.2$ . 可见局域支持向量机预测法的邻近点个数和邻域半径大于局域线性预测法的邻近点个数和邻域半径.

从表 1 可见,用本文提出的方法选取邻近点的局域线性预测法和局域支持向量机预测法都能够有效地预测混沌时间序列,一步和多步预测精度高,归一化均方误差(NMSE)很小. 可见,用本文提出的方法选取邻近点的局域线性预测法和局域支持向量机预测法的一步和多步预测性能较好,在预测精度较高的条件下,计算量较小.

Santa Fe 时间序列竞赛的激光数据(Data A)的总长度为 1000,前 900 个点作为训练样本,后 100 个点作为测试样本. 对该序列相空间重构时,嵌入维数  $m=8^{[12]}$ ,延迟时间  $\tau=1$ . 把该序列训练样本分



表 1 对 Lorenz 系统  $x$  分量混沌序列,局域预测法的归一化均方误差(NMSE)

| 预测步长 | 局域线性预测法                 |                         | 局域支持向量机预测法              |                         |
|------|-------------------------|-------------------------|-------------------------|-------------------------|
|      | $k=6$                   | $r=1.2$                 | $k=28$                  | $r=3.2$                 |
| 1    | $6.0107 \times 10^{-8}$ | $1.5407 \times 10^{-7}$ | $7.4017 \times 10^{-9}$ | $3.2605 \times 10^{-9}$ |
| 5    | $2.1664 \times 10^{-6}$ | $3.9678 \times 10^{-6}$ | $8.6975 \times 10^{-7}$ | $1.5109 \times 10^{-6}$ |
| 10   | $1.5941 \times 10^{-5}$ | $2.0649 \times 10^{-5}$ | $8.0121 \times 10^{-6}$ | $1.2180 \times 10^{-5}$ |

为两部分,前 700 个点作为第一部分,后 200 个点作为第二部分. 用本文提出的方法选取局域线性预测法和局域支持向量机预测法的邻近点个数和邻域半径,参数  $l=100, n_a=2, D=3$ ,仿真结果如图 5—图 8 所示. 其中局域支持向量回归模型的核函数取为径向基核函数,参数取为  $C=10^6, \epsilon=0.01, \sigma=150$ . 为了分析用本文提出的方法选取邻近点的局域预测法的预测性能,对该序列测试样本进行 1 至 10 步预测,仿真结果如表 2 所示.

对 Santa Fe 激光数据,从图 5 和图 7 可见,用 BIC 准则和 Hannan-Quinn 准则选取的局域线性预测法的邻近点个数和邻域半径分别为  $k=16, r=29$ ;从图 6 和图 8 可见,用 BIC 准则和 Hannan-Quinn 准则选取的局域支持向量机预测法的邻近点个数和邻域半径分别为  $k=26, r=38$ .

从表 2 可见,用本文提出的方法选取邻近点的局域线性预测法和局域支持向量机预测法都能够有效地预测 Santa Fe 激光数据,一步和多步预测精度高,归一化均方误差(NMSE)很小. 可见,用本文提出的方法选取邻近点的局域线性预测法和局域支持向量机预测法的一步和多步预测性能较好,在预测精度较高的条件下,计算量较小.

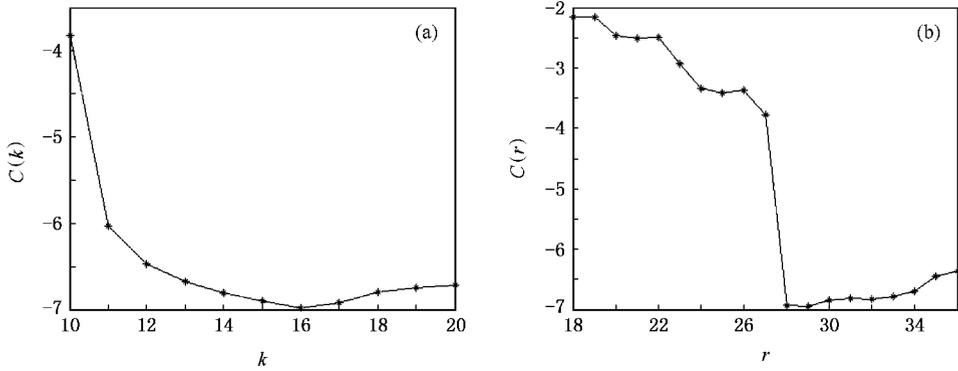


图 5 对 Santa Fe 激光数据, 局域线性预测法的基于 BIC 准则的邻近点个数和邻域半径 (a)  $C(k)-k$ ; (b)  $C(r)-r$

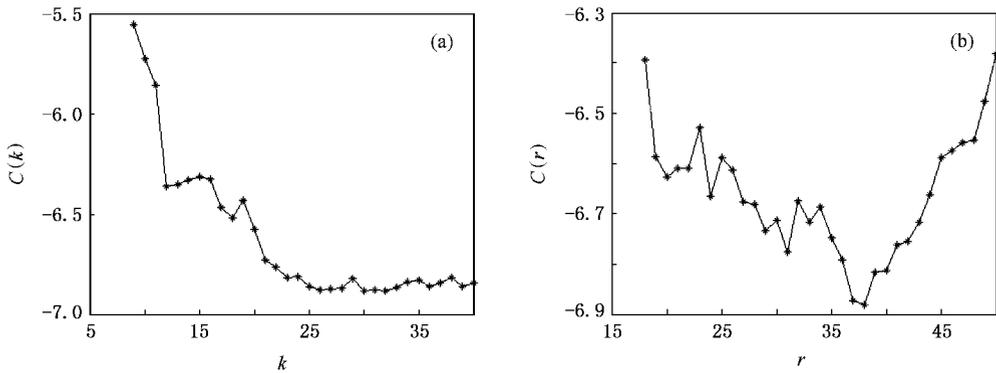


图 6 对 Santa Fe 激光数据, 局域支持向量机预测法的基于 BIC 准则的邻近点个数和邻域半径 (a)  $C(k)-k$ ; (b)  $C(r)-r$

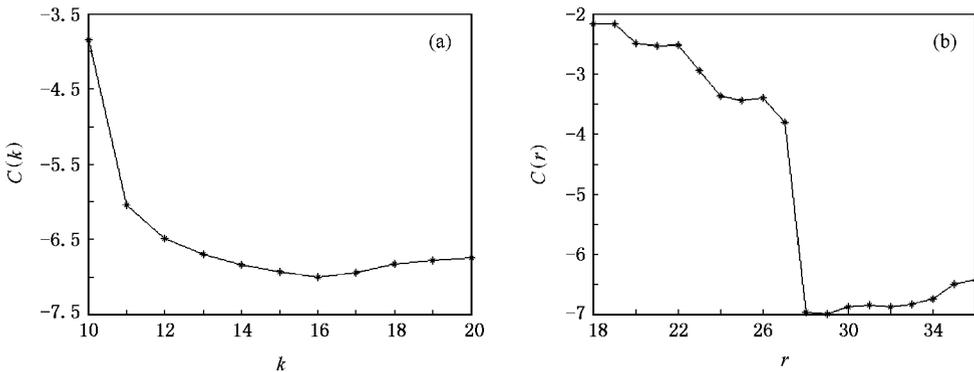


图 7 对 Santa Fe 激光数据, 局域线性预测法的基于 Hannan-Quinn 准则的邻近点个数和邻域半径 (a)  $C(k)-k$ ; (b)  $C(r)-r$

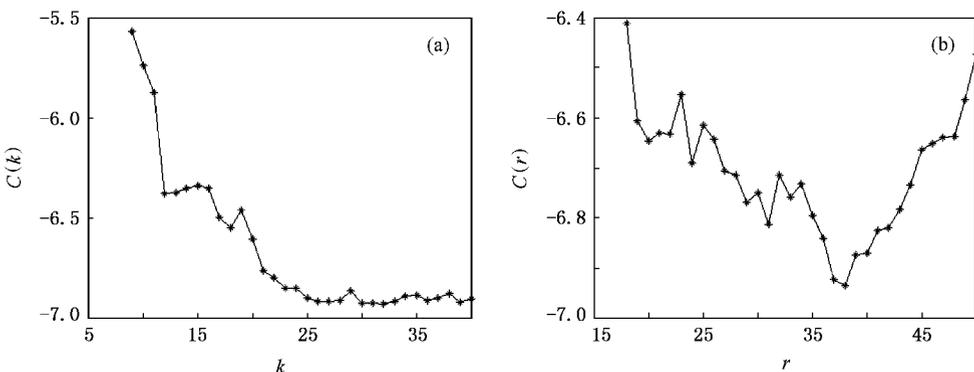


图 8 对 Santa Fe 激光数据, 局域支持向量机预测法的基于 Hannan-Quinn 准则的邻近点个数和邻域半径 (a)  $C(k)-k$ ; (b)  $C(r)-r$

表 2 对 Santa Fe 激光数据,局域预测法的归一化均方误差(NMSE)

| 预测步长 | 局域线性预测法  |          | 局域支持向量机预测法              |                         |
|------|----------|----------|-------------------------|-------------------------|
|      | $k = 16$ | $r = 29$ | $k = 26$                | $r = 38$                |
| 1    | 0.0030   | 0.0020   | $6.0333 \times 10^{-4}$ | $6.7478 \times 10^{-4}$ |
| 5    | 0.0025   | 0.0034   | $1.0878 \times 10^{-3}$ | $9.9174 \times 10^{-4}$ |
| 10   | 0.0070   | 0.0049   | $1.1726 \times 10^{-3}$ | $2.1776 \times 10^{-3}$ |

## 4. 结 论

邻近点个数是局域预测法的重要参数之一,它决定局域模型的预测精度和计算量.本文基于信息准则,提出了一种确定局域预测法中邻近点个数的定量方法,并用 Lorenz 模型生成的混沌时间序列和 Santa Fe 时间序列竞赛的激光数据(Data A)验证该

方法的有效性.实验结果表明用该方法选取邻近点的局域线性预测法和局域支持向量机预测法的一步和多步预测性能较好,在预测精度较高的条件下,计算量较小.

Akaike 等的信息准则是在传统的线性时间序列分析的研究中提出的,用于客观确定 ARMA 模型的参数的个数.我们这里遇到的则是非线性时间序列分析问题,情况有很大的不同,但信息准则的基本思想和原则应当同样适用.而且,由于混沌系统本身所存在的指数发散的性质,邻近点过多时,反而会降低预测效果,这更是一般的建模问题所没有的现象.本文只是这方面的一个初步尝试,信息准则在非线性的时间序列分析中的应用,还有许多问题有待进一步研究.

- [ 1 ] Kantz H , Schreiber T 2003 *Nonlinear Time Series Analysis* ( Second Edition )( Cambridge : Cambridge University Press )
- [ 2 ] Farmer J D , Sidorowich J J 1987 *Phys. Rev. Lett.* **59** 845
- [ 3 ] Sugihara G , May R M 1990 *Nature* **344** 734
- [ 4 ] Gan J C , Xiao X C 2003 *Acta Phys. Sin.* **52** 1096 ( in Chinese )  
[ 甘建超、肖先赐 2003 物理学报 **52** 1096 ]
- [ 5 ] Meng Q F , Peng Y H , Sun J 2007 *Chin. Phys.* **16** 3221
- [ 6 ] Li H C , Zhang J S 2005 *Chin. Phys. Lett.* **22** 2776

- [ 7 ] Zhang J S , Dang J L , Li H C 2007 *Acta Phys. Sin.* **56** 67 ( in Chinese ) [ 张家树、党建亮、李恒超 2007 物理学报 **56** 67 ]
- [ 8 ] Takens F 1981 *Dynamical Systems and Turbulence* ( Berlin : Springer ) p366
- [ 9 ] Akaike H 1974 *IEEE Trans. Auto. Contr.* **19** 716
- [ 10 ] Akaike H 1978 *Ann. Inst. Statist. Math.* **30** 9
- [ 11 ] Hannan E J , Quinn B G J 1979 *Roy. Statist. Soc.* **41** 190
- [ 12 ] Meng Q F , Peng Y H , Xue P J 2007 *Chin. Phys.* **16** 1253

# The neighbor point selection method for local prediction based on information criterion<sup>\*</sup>

Meng Qing-Fang Peng Yu-Hua Qu Huai-Jing Han Min

( School of Information Science and Engineering , Shandong University , Jinan 250100 , China )

( Received 23 May 2007 ; revised manuscript received 1 July 2007 )

## Abstract

The number of nearest neighbor points is an important parameter for the local prediction method , which has an important impact on the prediction accuracy and computation complexity of the local model. Based on the information criterion , the neighbor point selection method for the local prediction method is proposed in this paper. We illustrate this method by analyzing chaotic time series from Lorenz model and the experimental laser data- Santa Fe Data A. Simulation results show that using the proposed method to select neighbor points , the one-step and multi-step prediction accuracy of the local prediction method is good , and the computation complexity is reduced.

**Keywords** : nearest neighbor points , neighborhood , local prediction , information criterion

**PACC** : 0545

---

<sup>\*</sup> Project supported by the Scientific Research Starting Foundation for Returned Overseas Chinese Scholars , Ministry of Education , China ( Grant No. 2004-176.4 ) and the Natural Science Foundation of Shandong Province , China ( Grant Nos. 2004G01 , 2004ZRC03016 ).