

文章编号: 1001-0920(2007)03-0258-04

ESN 岭回归学习算法及混沌时间序列预测

史志伟, 韩敏

(大连理工大学 电子与信息工程学院, 辽宁 大连 116024)

摘要: ESN(回声状态网络)是一种新型的递归神经网络,可有效处理非线性系统辨识以及混沌时间序列预测问题. 针对 ESN 学习算法中可能存在的解的奇异问题,利用岭回归方法代替原有的线性回归算法. 通过贝叶斯或 Bootstrap 方法确定岭回归方法中的正则项系数,从而有效地控制输出权值的幅值,改善 ESN 的预测性能. 该方法在月太阳黑子预测问题中显示出较好的结果.

关键词: 回声状态网络; 岭回归; 混沌时间序列预测

中图分类号: TP183

文献标识码: A

Ridge regression learning in ESN for chaotic time series prediction

SHI Zhi-wei, HAN Min

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, China.
Correspondent: HAN Min, E-mail: minhan@dlut.edu.cn)

Abstract: As a new type of recurrent neural network, echo state network (ESN) is applied to nonlinear system identification and chaotic time series prediction. A technique is proposed to improve the properties of ESN solution, which performs ridge regression in the reservoir state space instead of the previous linear regression. In addition, the ridge regression parameter is systematically determined by Bayesian method or Bootstrap method. The prediction results for monthly sunspots time series show a satisfying performance.

Key words: Echo state networks; Ridge regression; Chaotic time series prediction

1 引言

近年来,基于神经网络的混沌时间序列预测技术表现出较快的发展势头,每一种新的神经网络方法都促进了混沌预测技术的丰富和改进^[1-7]. 多层前向网络和递归神经网络是两种最基本的预测工具. 一般来说,如果将预测任务当作一个静态函数逼近问题,则可通过非线性回归方法来解决,比如使用多层前向网络、支持向量机等. 另一方面,混沌时间序列本身源自动态系统,因此可使用递归神经网络来模拟动态系统. 作为动态系统的递归神经网络常常比静态的多层前向网络更有效.

作为一种新型递归神经网络,ESN(回声状态网络)^[8,9]在非线性系统辨识方面较传统的递归神经网络有较大改进. 首先,在稳定性方面,可以通过预先设定储备池权值矩阵的谱半径来保证递归网络的稳定性;其次,在网络训练方面,输出权值的确定是唯一且是全局最优的,因此没有传统神经网络普遍

存在的局部最小问题;再次,ESN 避免了传统递归神经网络求取时序偏微分的过程,因此网络的训练过程变得特别简单;最后,在混沌时间序列预测方面,ESN 方法显示出较好的预测性能.

在经典的 Mackey-Glass 84 步标杆预测问题中,由 ESN 得到 NRMSE(Normalized Root Mean Square Error) 达到了 $10^{-4.2}$ ^[7],而目前常规神经网络的预测精度仅为 $10^{-1.2} \sim 10^{-1.7}$. 与此同时,ESN 对其他混沌时间序列也有较突出的预测性能. 然而,目前 ESN 的应用仅局限于不含噪声混沌时间序列的迭代预测^[10,11]. 由于稳定性和误差积累两方面的问题,对于实际观测时间序列而言,迭代预测方法表现较差. 本文旨在研究和发展一种基于 ESN 的直接预测方法,因其不包含迭代过程,故可从根本上避免迭代预测过程中出现的问题并应用于实际.

ESN 在求解输出权值时使用线性回归算法,在处理实际问题时,易导致病态解问题,且伴随着具有

收稿日期: 2005-11-17; 修回日期: 2006-03-20.

基金项目: 国家自然科学基金项目(60674073, 60374064).

作者简介: 史志伟(1980—),男,河南洛阳人,博士生,从事神经网络、非线性系统辨识等研究; 韩敏(1959—),女,吉林延吉人,教授,博士生导师,从事神经网络、专家系统等研究.

较大幅值的输出权值. 文献[12]建议在状态变量上加入噪声以改变矩阵的性质, 从而改善解的性能. 然而噪声幅值的确定目前主要靠经验, 缺乏可靠的确定方法. 此外可以证明, 当训练样本个数小于输出权值维数时, 无论噪声的幅值取多少, ESN 的解必为奇异的. 本文使用岭回归方法代替常用的线性回归方法, 岭回归方法借助于正则项消除原线性回归方法中出现的病态问题, 从而提升回声状态网络解的性质. 而且, 即使训练样本个数小于输出权值维数, 岭回归方法也能可靠地工作. 另外, 本文通过贝叶斯和 Bootstrap 方法确定正则项系数的大小, 从而进一步提高了该方法的可操作性.

2 基于回声状态网络的直接预测方法

ESN 不仅是一种新的网络结构, 也是一种新型学习机制. ESN 首先假设非线性系统的动态特性由一个大规模的“储备池”产生. “储备池”包含大量随机生成且稀疏连接的神经元, “储备池”蕴含了系统的运行状态, 并具有记忆功能; 在外部输入作用下, 构成“输入-状态-输出”驱动系统. ESN 网络的方程可以写成

$$x(k+1) = \text{tansig}(W_x \cdot x(k) + W_{in} \cdot u(k) + v(k)), \quad (1)$$

$$y(k) = w \cdot x(k) + b. \quad (2)$$

其中: tansig 为双曲正切函数; $x(k)$ 为 k 时刻“储备池”的状态变量, 通常其维数较高, 取 300 ~ 1 000 之间; $u(k)$ 为 k 时刻系统输入向量; $v(k)$ 为可加随机噪声向量; $y(k)$ 为 ESN 网络的输出; W_x 为随机产生且稀疏连接(通常保持 1% ~ 5% 连接)的高维方阵, “储备池”一经产生, 其连接权值保持不变; W_{in} 和 w 分别为 ESN 的输入权值矩阵和输出权值向量, W_{in} 也随机产生且保持不变, 唯一需要调整的是输出权值 w . 在下面的讨论中使用 $x(k)$ 代替 $[x(k)^T \ 1]^T$, 从而可以使用 w 代替 $[w^T \ b]^T$ 进行计算.

对于给定非线性系统输入输出对 $\{u(k), y_d(k), k = 1, 2, \dots\}$, 利用 ESN 辨识该系统的过程为: 首先初始化储备池中的权值 W_x 和 W_{in} ; 其次, 使用输入 $u(k)$ 激励系统(1), 求得 ESN 的各个时刻状态响应; 输出权值 w 可通过 ESN 的状态变量与期望输出之间的线性回归得到.

给定混沌时间序列 $\{y_d(k), k = 1, 2, \dots\}$, 假设其嵌入维数为 m , 延迟时间为 τ , 那么可以得到重构的延迟坐标向量为

$$d(k) = [y_d(k) \ y_d(k-\tau) \ \dots \ y_d(k-(m-1)\tau)]^T. \quad (3)$$

根据 Takens 定理, 存在一个映射 $F(\cdot)$ 满足

$$d(k+1) = F(d(k)), \quad (4)$$

假设 $y_d(k)$ 通过函数 $g(\cdot)$ 观测重构的状态向量 $d(k)$, 即

$$y_d(k) = g(d(k)). \quad (5)$$

另外有以下等式成立:

$$\begin{aligned} d(k) &= d(k), \\ d(k+1) &= F(d(k)), \\ &\vdots \\ d(k+h) &= F(d(k+h-1)), \\ y_d(k+h) &= g(d(k+h)). \end{aligned} \quad (6)$$

进一步假设 $F_0(\cdot) \equiv 0$, 则式(6)可表示成如下的状态空间表达式:

$$\begin{aligned} \begin{bmatrix} d(k) \\ d(k+1) \\ \vdots \\ d(k+h) \end{bmatrix} &= \begin{bmatrix} F_0(d(k-1)) \\ F(d(k)) \\ \vdots \\ F(d(k+h-1)) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot d(k), \\ y_d(k+h) &= [0 \ 0 \ \dots \ 1] \begin{bmatrix} d(k) \\ d(k+1) \\ \vdots \\ g(d(k+h)) \end{bmatrix}. \end{aligned} \quad (7)$$

其中: h 为预先设定预测时域, $[d(k)^T \ d(k+1)^T \ \dots \ d(k+h)^T]^T$ 为预测时域 h 对应的增广状态变量. 注意, 状态空间表达式(7)并不是输入输出 $\{d(k), y_d(k+h)\}$ 的唯一实现.

基于回声状态网络的混沌时间序列直接预测方法借助于储备池机制, 通过求解合适的输出权值 w , 使得回声状态网络的输入输出关系 $\{u(k), y(k), k = 1, 2, \dots\}$ 逼近由状态空间表达式(7)表示的输入输出关系 $\{d(k), y_d(k+h), k = 1, 2, \dots\}$.

应该注意到, 当 $d(k)$ 缩减成 $y_d(k)$, 而预测时域仅为 $h = 1$ 时, 上述模型辨识问题就是迭代预测方法中的单步预测器辨识问题^[7]. 迭代方法正是依靠单步预测器的迭代操作而形成的多步预测.

3 回声状态岭回归方法

回声状态网络的岭回归学习算法包括两个组成部分: 首先, 使用岭回归方法代替原有的线性回归以及“加噪声项”方法; 其次, 通过有效的方法确定岭回归中的正则项系数. 本文通过贝叶斯后验概率最大化方法和 Bootstrap 方法确定合适的正则项系数.

3.1 回声状态岭回归

储备池中的状态变量与期望输出之间是线性关系, 因此, 输出权值 w 的确定可使用如下基本的线性回归算法:

$$\hat{w} = (X^T X)^{-1} X^T y_d. \quad (8)$$

其中

$X = [x(p)^T, x(p+1)^T, \dots, x(p+T_n-1)^T]^T$,
 $y_d = [y_d(p) \ y_d(p+1) \ \dots \ y_d(p+T_n-1)]^T$,
 T_n 为训练样本个数. 注意, 这里的 $x(p)$ 和 $x(p+1)$
 并不一定是储备池的两个相继的状态变量, 而是被
 选作训练样本索引, 只要保证 $x(p)$ 和 $y_d(p)$ 之间的
 对应关系即可. 也可以通过求取 $X^T X$ 伪逆方法获得
 w 的估计. 另外根据文献[9]的建议, 训练样本的选
 取应尽量避免 ESN 初始暂态的影响.

在实际操作中, 常常会出现病态解的问题, $X^T X$
 接近奇异, 结果使得输出权值 \hat{w} 的幅值较大^[12], 导
 致预测结果恶化. 由于 $X^T X$ 接近奇异, 误差估计非
 常保守, 即 $\|w - \hat{w}\| \leq \| (X^T X)^{-1} X^T \| \| X^T \hat{w} -$
 $y_d \|$. 如果在储备池的状态变量上加入小的噪
 声^[12], 则可改变矩阵 $X^T X$ 的性质, 进而改善解的性
 能. 噪声的大小严重影响着模型的质量, 然而其大小
 的确定主要靠经验. 通过岭回归方法, 求取输出权值
 的方法变成

$$\hat{w} = (X^T X + C^{-1} I)^{-1} X^T y_d, \quad (9)$$

其中 $C^{-1} \in R^+$, 它等价于最小化如下的代价函
 数^[13]:

$$L(w, b) = C \sum_{j=1}^{T_n} (\langle wx_j \rangle + b - y_{dj})^2 + \|w\|^2. \quad (10)$$

C 的作用是在 $X^T X$ 之上施加了一个“噪声项”, 从而
 可代替式(1)中的 $v(k)$. 如果 C 选择合适, 便可有效
 地平衡误差项和模型的复杂性, 提高回声状态网络
 的性能.

此外, 从式(8)可以看出, 当训练样本个数 T_n
 小于输出权值的维数时, 无论噪声取何值, 矩阵
 $X^T X$ 必为奇异矩阵. 但使用岭回归方法就可以完全
 避免这种情况的发生, 因为岭回归能直接改善矩阵
 $X^T X$ 的性质.

3.2 正则项系数的确定

由前面所述, 正则项系数 C 的确定影响着回声
 状态网络的建模质量, 因此需要有效地确定正则项
 系数. 储备池一经初始化后便保持不变, 那么确定 C
 的过程就是模型的选择过程. 下面将通过贝叶斯后
 验概率最大化方法和 Bootstrap 重采样方法来确定
 正则项系数.

3.2.1 贝叶斯后验概率最大化方法

Van 等人^[14] 已将贝叶斯方法应用于最小二乘
 支持向量机分类器的设计过程, 分别以 3 个层次确
 定模型参数、正则项系数和核参数. 具体来说, 第 1
 个层次是给定正则项系数和核参数推理模型参数;
 第 2 个层次是给定核参数推理正则项系数; 第 3 个
 层次是核参数的贝叶斯推理过程. 在回声状态岭回

归方法中, 由于需要确定正则项系数, 需要在固定储
 备池的条件下推理合适的正则项系数. 正则项系数
 的后验概率最大化等价于最小化标量指标函数^[14]

$$\min_C J(C) = \sum_{i=1}^{T_n-1} \log(\lambda_{G,i} + C^{-1}) +$$

$$(N-1) \log[E_w(w_{MP} +$$

$$CE_D(w_{MP}, b_{MP})]. \quad (11)$$

其中: T_n 为训练样本个数, $\lambda_{G,i}$ 是 $M\Omega M$ 的非零特征
 值 ($\Omega = X^T X$, M 为相应的幂等矩阵), $E_w = \frac{1}{2} w^T w$,

$E_D = \frac{1}{2} \sum_{i=1}^N (y_{di} - [w^T x_i + b])^2$, w_{MP} 和 b_{MP} 是给定
 正则项系数 C 后得到的最优模型参数.

3.2.2 Bootstrap 方法

在神经网络设计方面, 重采样技术, 如交叉检验
 和 Bootstrap^[15,16], 是两种重要的模型选择方法. 其
 中, Bootstrap 方法的基本思想是: 已知随机样本 χ
 $= \{X_1, X_2, \dots, X_n\}$ 符合未知分布 F , 假设感兴趣的
 某个随机变量为 $R(\chi, F)$, 现希望通过 n 个 χ 的观测
 值 $\chi = \{x_1, x_2, \dots, x_n\}$ 来估计随机变量 R 的分布特
 性, 比如均值和置信区间. 方法是根据样本观测值构
 造经验概率分布 \hat{F} , 使得每一个样本出现的概率均
 为 $1/n$. 按照分布 \hat{F} , 生成 B 个随机子样 $\chi^* = \{X_1^*,$
 $X_2^*, \dots, X_n^*\}$, 最后根据随机子样估计出随机变量
 $R(\chi, F)$ 的分布特性.

对某个正则项系数 C , 使用 Bootstrap 方法估计
 出模型的泛化误差, 即感兴趣的随机变量为 ESN 模
 型的泛化误差. 在固定储备池的情况下, 使用
 Bootstrap 方法估计泛化误差的均值和置信区间. 泛
 化误差的样本方差计算公式为

$$S^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta})^2. \quad (12)$$

其中: $\hat{\theta} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*$ 为样本均值; $\hat{\theta}_b^*$ 为使用第 b 个
 Bootstrap 子样估计出来的泛化误差, 泛化误差的计
 算是在每个 Bootstrap 样本之外的样本上进行的. 由
 于真实的样本方差未知, 泛化误差置信度为 $1-\alpha$ 的
 置信区间为

$$\left(\hat{\theta} \pm \frac{S}{\sqrt{B}} t_{\alpha/2}(B-1) \right), \quad (13)$$

其中 $t_{\alpha/2}(B-1)$ 可通过 t 分布得到.

4 仿真例子

将 Wolf 月太阳黑子经平滑和去噪处理^[17], 得
 到 1753 ~ 2000 年共 2 976 个月的数据. 选择 ESN 的
 储备池维数为 600, 并保持约 2% 连接, 连接权值 W_x
 的谱半径约为 0.71, 储备池初始状态随机选择. 大

量的仿真结果显示,基于 ESN 的迭代预测方法不能恰当地对经过平滑和去噪处理后的月太阳黑子进行预测,故以下采用直接方法进行预测.

选择月太阳黑子嵌入维数 $m = 3, \tau = 10$ 进行相空间重构,得到延迟坐标向量 $\{d(k), k = 1, 2, \dots\}$. 令 ESN 的输入 $u(k) = d(k), k = 1, 2, \dots$, 按照式 (1) 计算得到回声状态变量 $\{x(k), k = 1, 2, \dots\}$. 设定预测时域为 $h = 10$ 个月,即 ESN 的期望输出序列为 $\{y_d(k + 10), k = 1, 2, \dots\}$. 随机选择 400 组模式对 $\{x(p), y_d(p + 10), p = 1, 2, \dots, T_n\}$ 组成矩阵 X 和 y_d .

在贝叶斯方法中,只需使用无约束优化算法优化式(11)中的系数 C 即可. 为了与 Bootstrap 方法进行对比,图 1 分别给出了 C 取不同值时 ($10^5, 10^{5.5}, 10^6, \dots, 10^{11}$) 代价函数 $J(C)$ 的取值情况. 从 $J(C)$ 对 C 的导数信息上可以看出, $J(C)$ 在 $C = 10^{7.5}$ 与 $10^{8.0}$ 之间取得极小值,也就是说,通过贝叶斯方法估计出的最优的 C 应位于 $10^{7.5}$ 与 $10^{8.0}$ 之间.

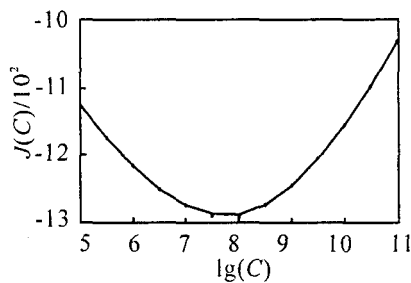


图 1 使用贝叶斯方法估计系数 C

在 Bootstrap 方法当中,取随机子样个数 $B = 25$. 当 C 取不同值时 ($10^5, 10^{5.5}, 10^6, \dots, 10^{11}$), 分别计算泛化误差估计和其 95% 置信区间. 图 2 中给出了由 Bootstrap 方法估计出来的泛化误差和 95% 置信区间. 由 Bootstrap 计算出的泛化误差均值的估计在 10^8 处左右取得最小值. 另一方面,从 Bootstrap 方法的计算结果来看,当 C 取值较小时,泛化误差具有较大的置信区间;当 C 取 $10^{7.5}$ 左右时,泛化误差的置信区间取得最小值;继续增大 C 时,泛化误差的置信区间逐渐增大,即随着 C 的进一步增大,对泛化误差的估计越来越不稳定. 因此,应选择 $C = 10^{-7.5}$

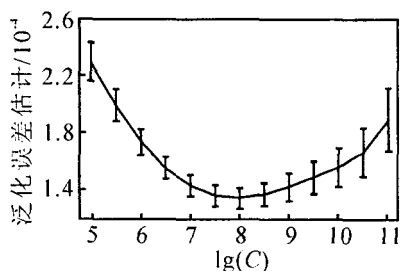


图 2 使用 Bootstrap 方法估计的泛化误差及置信区间

$\sim 10^8$ 进行计算. 可以看出,由两种方法估计出的 C 的范围基本吻合.

在基本的线性回归算法中,虽然也可使用 Bootstrap 方法确定噪声项 v 的大小,但可以看到,随着 B 的增加该方法产生的计算量也迅速增加. 相对而言,贝叶斯方法的计算量要小得多,因为它仅需要处理一个单变量优化问题. 除此之外,正则项系数与岭回归方法直接相关,因而避免了加入噪声项的经验性.

需要指出的是,在本文给出的仿真例子中,因为样本个数 ($T_n = 400$) 小于输出权值的维数 (600), 所以无论噪声 v 的幅值多大,矩阵 $X^T X$ 必为奇异矩阵,使用原方法给出的预测结果没有任何意义,实验结果也说明了这一点. 岭回归方法直接改变了矩阵 $X^T X$ 的性质,因此解的性能良好,而且通过贝叶斯或者 Bootstrap 方法可以很好地确定正则项系数的大小.

5 结 论

本文在回声状态网络的基础上,使用了岭回归方法代替原有的线性回归方法,并且使用贝叶斯方法或 Bootstrap 方法确定最优的正则项系数. 该方法避免了以往方法加入噪声项的经验性,提高了回声状态网络解的性能和建模质量. 将该方法应用于月太阳黑子预测问题,得到了较为满意的结果.

参考文献 (References)

- [1] Lapades A, Farbar R. How neural nets work [C]. Advances in Neural Information Processing Systems. Denver, 1987:442-456.
- [2] Casdagli M. Nonlinear prediction of chaotic time series [J]. Physica D, 1989, 35(3):335-356.
- [3] Wan E A. Time series prediction by using a connectionist network with internal delay lines[C]. Proc of the NATO Advanced Research Workshop on Comparative Time Series Analysis. New Mexico, 1993: 195-217.
- [4] Vesanto J. Using the SOM and local models in time-series prediction [C]. Proc of Workshop on Self-organizing Maps. Espoo, 1997: 209-214.
- [5] Haykin S, Principe J. Making sense of a complex world [J]. IEEE Signal Processing Magazine, 1998, 15(3): 66-68.
- [6] Han M, Xi J H, Xu S G, et al. Prediction of chaotic time series based on the recurrent predictor neural network[J]. IEEE Trans on Signal Processing, 2004, 52(12): 3409-3416.

(下转第 267 页)

- [7] Jaeger H, Haas H, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication[J]. *Science*, 2004, 304(5667): 78-80.
- [8] Jaeger H. Adaptive nonlinear system identification with echo state networks [C]. *Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2003, 15: 593-600.
- [9] Jaeger H. Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and “echo state network” approach[R]. Bremen, German National Research Center for Information Technology, 2002.
- [10] Prokhorov D. Echo state networks: Appeal and challenges [C]. *Proc of Int Joint Conf on Neural Networks*. Montreal, 2005: 1463-1466
- [11] Xi J H, Shi Z W, Han M. Analyzing the state space property of echo state networks for chaotic system prediction [C]. *Proc of Int Joint Conf on Neural Networks*. Montreal, 2005: 1412-1417.
- [12] Jaeger H. Reservoir riddles: Suggestions for echo state network research [C]. *Proc of Int Joint Conf on Neural Networks*. Montreal, 2005: 1460-1462.
- [13] Tikhonov A N, Arsenin V Y. *Solution of Ill — Posed problems*[M]. Washington DC, 1977.
- [14] Van Gestel T, Suykens J A K, Lanckriet G, et al. Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and kernel fisher discriminant analysis[J]. *Neural Computation*, 2002, 14(5):1115-1147.
- [15] Kallel R, Cottrell M, Vigneron V. Bootstrap for neural model selection[J]. *Neurocomputing*, 2002, 48 (1-4):175-183.
- [16] Anguita D, Boni A, Ridella S. Evaluating the generalization ability of support vector machines through the bootstrap[J]. *Neural Processing Letters*, 2000, 11(1): 51-58.
- [17] Jinno K, Xu S G, Kawamura A, et al. Prediction of sunspots using reconstructed chaotic system equations [J]. *J of Geophysical Research — Space Physics*, 1995, 100(A8): 14773-14781.