

文章编号: 1001-0920(2011)10-1469-04

回声状态网络 LM 算法及混沌时间序列预测

韩 敏, 穆大芸

(大连理工大学 电子信息与电气工程学部, 辽宁 大连 116023)

摘 要: 回声状态网络 (ESN) 学习算法中可能存在解的奇异问题, 在时间序列预测时易导致病态解问题, 且伴随着具有较大幅值的输出权值, 尤其是当训练样本个数小于输出权值维数时, ESN 的解必为奇异的. 鉴于此, 考虑使用 LM(Levenberg Marquardt) 算法代替常用的线性回归方法, 自适应选择 LM 参数, 从而有效地控制输出权值的幅值, 提高 ESN 的预测性能. 通过 Lorenz 混沌时间序列进行预测研究, 对大连月平均气温实际数据进行仿真研究, 取得了较好的预测效果.

关键词: 回声状态网络; 储备池; LM 算法; 时间序列预测

中图分类号: TP183

文献标识码: A

LM algorithm in echo state network for chaotic time series prediction

HAN Min, MU Da-yun

(Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023, China. Correspondent: HAN Min, E-mail: minhan@dlut.edu.cn)

Abstract: The problem of singular solution in echo state network(ESN) learning algorithm is existed, which is easy to cause ill issue. Especially when training samples are less than the dimensions of the output, the solution of the ESN must be singular. Therefore, Levenberg Marquardt(LM) algorithm is used to learn ESN instead of linear regression method, which can effectively control the amplitude of the output weight result in improved predictive performance. The presented model is tested on the Lorenz time series and applied to some real life time series such as Dalian monthly average temprture time series. Actual simulation results show that the predictive model has higher predictive accuracy, and is of great practicality and effectiveness.

Key words: echo state network; reservoir; Levenberg Marquardt algorithm; time series prediction

1 引 言

随着非线性混沌动力学理论和应用的发展, 人们更加深刻地认识到时间序列的复杂性, 对于复杂性混沌时间序列的预测已成为混沌信息处理领域中一个重要的分支^[1-2], 并广泛应用于自动控制等领域. 近年来, 国内外学者利用各种神经网络对混沌时间序列进行预测建模作了研究, 并取得了成功应用^[3-8], 但因神经网络存在过拟合、训练过程中易陷入局部极小、网络结构的选择过分依赖于经验等缺陷^[9], 直接影响了时间序列建模的可靠性.

回声状态网络 (ESN) 是一种新型的递归神经网络, 在非线性混沌系统预测建模方面与传统的递归神经网络相比, 有了较大改进^[8-10]. 首先, 通过预设 ESN 网络内部连接权值矩阵的谱半径来保证储备池

内部递归网络的稳定性; 其次, 利用线性回归算法训练唯一要求计算的输出权值. 学习算法本身属于凸优化, 因此避免了陷入局部极小值的问题.

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, 且得到较大幅值的输出权值. 尤其是当训练样本个数小于输出权值维数时, ESN 的解必为奇异的. 本文使用 LM (Levenberg Marquardt) 算法代替常用的线性回归方法, LM 方法消除原线性回归方法中出现的病态问题, 从而提升了回声状态网络解的性质, 简称这种预测模型为 RLM-ESN. 而且, 即使训练样本个数小于输出权值维数, RLM-ESN 也能有效地用于预测.

2 基于回声状态网络的直接预测方法

ESN 首先假设非线性系统的动态特性由一个大

收稿日期: 2010-06-03; 修回日期: 2010-11-01.

基金项目: 国家自然科学基金项目(61074096); 国家科技支撑计划项目(2006BAB14B05).

作者简介: 韩敏(1959—), 女, 教授, 博士生导师, 从事复杂工业系统建模与控制、智能技术及优化算法等研究; 穆大芸(1985—), 女, 硕士生, 从事时间序列建模与预测的研究.

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, 蕴含了系统的运行状态, 并具有记忆功能. 在给定网络输入的情况下, 构成“输入-状态-输出”驱动系统, ESN 网络表达式为^[11]

$$\begin{aligned} x(k+1) &= \tanh(W_x x(k) + W_{in} u(k)), \\ y(k) &= w x(k) + b. \end{aligned} \quad (1)$$

其中: \tanh 为双曲正切函数; $x(k)$, $u(k)$, $y(k)$ 分别为 k 时刻储备池的状态变量、系统输入量、网络输出, 通常储备池维数较高, 取 300~1 000 之间; W_x 为随机产生并且稀疏连接的高维方阵 (稀疏度在 0.01~0.05 之间), 储备池一经产生, 其内部连接权值不变; W_{in} 和 w 分别为 ESN 的输入权值矩阵和输出权值向量, W_{in} 也随机产生而且保持不变, 唯一需要调整的是输出权值 w . 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 非线性系统建模的过程为: 首先初始化储备池中的权值 W_x 和 W_{in} ; 然后, 使用输入 $u(k)$ 激励系统 (1), 求得 ESN 各个时刻状态响应; 输出权值 w 可以通过 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 $\{y_d(k), k = 1, 2, \dots\}$, 假设嵌入维数为 m , 延迟时间为 τ , 则可以得到重构的延迟坐标向量为

$$d(k) = [y_d(k) y_d(k - \tau) \cdots y_d(k - (m - 1)\tau)]^T. \quad (2)$$

根据 Takens 定理, 存在一个映射 $F(\cdot)$ 满足

$$d(k+1) = F(d(k)). \quad (3)$$

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

$$y_d(k) = g(d(k)), \quad (4)$$

则另有

$$\begin{aligned} d(k) &= d(k), \\ d(k+1) &= F(d(k)), \\ d(k+2) &= F(d(k+1)), \\ &\vdots \\ d(k+h) &= F(d(k+h-1)), \\ x(k+h) &= g(d(k+h)). \end{aligned} \quad (5)$$

进一步假设 $F_0(\cdot) \equiv 0$, 式 (5) 可以表示为如下状态空间表达式:

$$\begin{bmatrix} d(k) \\ d(k+1) \\ d(k+2) \\ \cdots \\ d(k+h) \end{bmatrix} = \begin{bmatrix} F_0(d(k-1)) \\ F(d(k)) \\ F(d(k+1)) \\ \cdots \\ F(d(k+h-1)) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} d(k),$$

$$y_d(k+h) = [0 \ 0 \ \cdots \ 1] \begin{bmatrix} d(k) \\ d(k+1) \\ d(k+2) \\ \cdots \\ g(d(k+h)) \end{bmatrix}. \quad (6)$$

其中: h 为预先设定预测时域, $[d(k)^T \ d(k+1)^T \ d(k+2)^T \ \cdots \ d(k+h)^T]^T$ 为预测时域 h 对应的增广状态变量. 状态空间表达式 (6) 并不是输入输出 $\{d(k), y_d(k+h)\}$ 的唯一实现.

基于 ESN 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 (6) 表示的输入输出关系 $\{d(k), y_d(k+h), k = 1, 2, \dots\}$.

注意到, 当 $d(k)$ 缩减成 $y_d(k)$, 而预测时域仅为 $h = 1$ 时, 上述预测方法即为迭代预测方法中的单步预测器问题^[7]. 迭代方法正是依靠在单步预测器的迭代操作而形成多步预测.

3 回声状态网络 LM 学习方法

3.1 RLM-ESN 时间序列预测模型

LM 算法是高斯-牛顿法和梯度下降法的结合, 既有高斯牛顿法的快速收敛性, 也有梯度下降法全局搜索特性^[12-14]. LM 算法具有高斯牛顿法的局部快速收敛特性, 并克服了牛顿法不能有效处理奇异和非正定矩阵的缺点, 在远离解时进行修正, 沿误差曲面进行搜索, 继承了梯度下降法的全局搜索特性, 精确度较高. 由于 LM 算法利用了近似二阶导数信息, 比梯度下降法快得多; 同时, LM 算法不需要过多的调整参数, 在许多参数优化的场合得到了广泛的应用.

ESN 在求解输出权值时使用线性回归算法, 处理实际问题, 易导致病态解问题, 且伴随着具有较大幅值的输出权值. 因此考虑使用 LM 算法代替常用的线性回归方法, 自适应选择 LM 参数, 从而有效地控制输出权值的幅值, 提高 ESN 的预测性能.

ESN 的状态空间表达式为式 (1) 和 (2) 所示, 其中只需求取网络的输出权值 w^T , 因此可以设 $Y = f(x_i, w)$. 设计 RLM-ESN 网络的过程如下: 利用 LM 算法求储备池输出权值 w 可以理解为选择合适的参数使得

$$S(w) = \sum_{i=1}^m [y_i - f(x_i, w)]^2 \quad (7)$$

变为最小.

LM 算法也是一个迭代的过程, 首先初始化参数向量 w , 大多情况下进行一般设置, 仅在初始值设置得接近最终值时, 算法才会收敛, 因此, 算法的收敛性

也是一个研究热点.

在每一次迭代中, 参数向量 w 会被新的估计值 $w + \delta$ 所代替, 即 $w_{t+1} = w_t + \delta$, 因此 $f(x_i, w + \delta) \approx f(x_i, w) + J_i \delta$. 其中: 误差 $e(w) = y - f(w)$, 误差函数

$$V(w) = \sum_{i=1}^N e_i^2(w), \text{ 雅克比矩阵为}$$

$$J_i = \Delta f(x_i, w)/w = \begin{bmatrix} \frac{\Delta e_1(w)}{\Delta w_1} & \frac{\Delta e_1(w)}{\Delta w_2} & \cdots & \frac{\Delta e_1(w)}{\Delta w_n} \\ \frac{\Delta e_2(w)}{\Delta w_1} & \frac{\Delta e_2(w)}{\Delta w_2} & \cdots & \frac{\Delta e_2(w)}{\Delta w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\Delta e_N(w)}{\Delta w_1} & \frac{\Delta e_N(w)}{\Delta w_2} & \cdots & \frac{\Delta e_N(w)}{\Delta w_n} \end{bmatrix}. \quad (8)$$

因为

$$S(w + \delta) \approx \sum_{i=1}^m (y_i - f(x_i, w) - J_i \delta)^2, \quad (9)$$

写成向量的形式为

$$S(w + \delta) \approx \|y_i - f(x_i, w) - J_i \delta\|. \quad (10)$$

令式 (10) 为 0, 有

$$(J^T J) \delta = J^T [y - f(w)], \quad (11)$$

变形为

$$\delta = (J^T J)^{-1} J^T [y - f(w)]. \quad (12)$$

当矩阵 $J^T J$ 为奇异阵, 即矩阵非正定时, δ 的解不会存在.

LM 方法的贡献是引进了 LM 参数 λ , 即式 (12) 改写为

$$\delta = (J^T J + \lambda I)^{-1} J^T [y - f(w)]. \quad (13)$$

其中: $\lambda > 0$ 为常数, I 为单位矩阵.

从式 (13) 可以看出, 如果 λ 很大, 则 LM 算法近似于梯度下降法; 如果 λ 为 0, 则是高斯牛顿法. 因为利用二阶导数信息, LM 算法比梯度法快得多, 且 $[J^T(x)J(x) + \mu I]$ 是正定的, 所以式 (14) 的解总存在, 从而 LM 算法优于高斯牛顿算法, 且速度比梯度下降法快得多. 对于高斯-牛顿法, $J^T J$ 是否满秩也是一个潜在的问题. 实际操作中, λ 是一个试探性的参数, 对于给定的 λ , 如果求得的 δ 能使误差函数 $V(w)$ 降低, 则用因子 β 除以 λ ; 若误差函数 $V(w)$ 增加, 则 λ 乘以因子 β . 在仿真中, 选取初始值 $\lambda_0 = 0.01$, $\beta = 10$. 采用 LM 算法时, 为了使收敛速度更快, 增加了学习率 α , 并取 α 为 0.4.

3.2 评价指标

为了分析数值实验结果的优劣, 采用均方根误差 E_{RMSE} 作为性能指标, 对预测结果进行比较, 定量说明各个模型的预测效果. 设 $y(k)$ 为预测变量的实际观测值, $\hat{y}(k)$ 为该变量的网络预测值, N 对应训练的

样本数, 则均方根误差定义为

$$E_{\text{RMSE}} = \left[\frac{1}{N-1} \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \right]^{1/2}. \quad (14)$$

均方根误差反映了预测值相对于观测值的偏离程度, 取值大于或等于零, 预测无误差时等于零. 均方根误差值越小, 表明预测效果越好.

4 仿真实例

4.1 Lorenz 数据仿真实验

将设计的 RLM-ESN 模型与文献 [14] 中的方法进行比较, 并采用 Lorenz 数据进行仿真研究, 主要对 X 时间序列进行仿真研究, 利用相空间参数选择方法确定延迟时间为 18, 嵌入维数为 6.

Lorenz 混沌方程如式 (1), 当取 $a = 10$, $b = 8/3$, $c = 28$, $x(0) = y(0) = z(0) = 1.0$ 时, 系统产生混沌. 利用四阶 Runge-Kutta 法迭代产生 1000 组时间序列, 前 800 组数据用于训练, 剩余 200 组用来预测 $x(t+1)$, 有

$$\begin{cases} \dot{x} = a(y - x), \\ \dot{y} = (c - z)x - y, \\ \dot{z} = xy - bz. \end{cases} \quad (15)$$

储备池模型各参数确定如表 1 所示. 对 $x(t+1)$ 预测的仿真结果如图 1 所示. 仿真结果与文献 [3] 中相关方法进行比较, 主要从训练时间与测试时间之和及测试误差方面进行了比较, 结果如表 2 所示.

由表 2 可见, 在训练时间与测试时间的总时间上, 所提出的方法与其他方法时间相当, 但在测试精度上

表 1 ESN 的参数

名称	取值
储备池模型	200 × 200
内部权值矩阵 W_x 稀疏度	0.01
内部权值矩阵 W_x 谱半径	1
输入权值	[-1, 1]

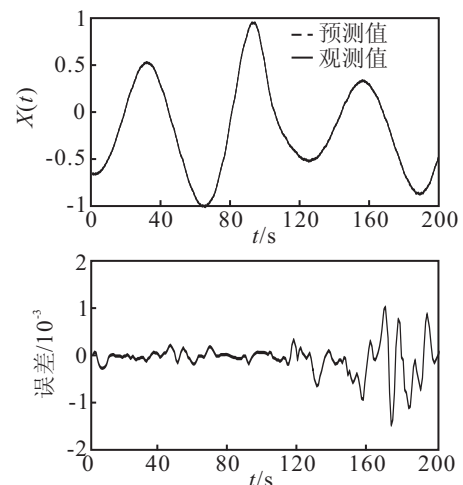


图 1 Lorenz 时间序列预测

表 2 Lorenz 时间序列预测结果对比

Model	RLM-ESN	RLM-RNN ^[14]	RNN-EKF ^[15]
Ermse/ 10^{-4}	2.549	3.600	12.100
CPU/(t/s)	2.09	2.12	2.16

要高于 RLM-RNN 等 3 种方法. 从理论上分析, ESN 网络只有一个输出权值需要利用 LM 方法求取, 而 RNN 及多层感知机等传统网络除了网络输出权值需要求取, 还有多个内部复杂的连接权值需要利用 LM 算法迭代求取, 因此网络学习和训练速度较快. ESN 作为一个新型递归神经网络, 能够得到较高的时间序列预测“精度”, 并且利用 LM 算法求取输出权值, 解决了伪逆方法求取过程中存在的病态问题. 由上述分析可以得出本文所提出的 RLM-ESN 模型在混沌时间序列中的有效性.

4.2 大连降雨及气温数据仿真

为了验证 RLM-ESN 预测模型对于混沌数据建模的有效性, 采用大连气温时间序列数据对所提出的预测建模进行仿真实验. 采用 1951~2001 年大连月平均气温数据, 共 612 个数据点, 其中 412 组数据作为训练样本, 其余 200 组用于测试预测模型. 图 2 是对大连 200 个月的气温预测结果, 与文献 [7,14] 预测方法比较结果如表 3 所示. 可以看出本文所提出的预测模型 RLM-ESN 对大连气温预测时的均方根误差比文献 [7,14] 要小, 且本文方法在实际时间序列预测研究中同样具有有效性.

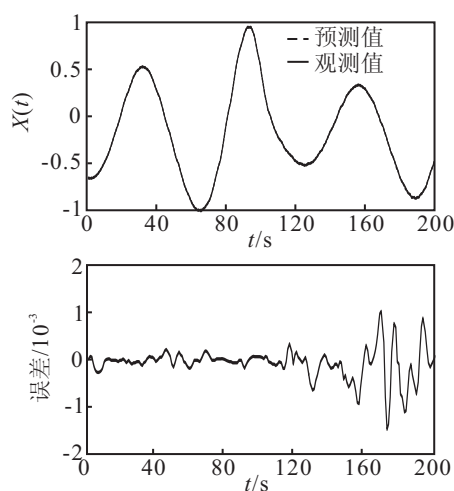


图 2 大连气温预测及误差

表 3 不同模型的预测效果

Model	RLM-ESN	ESN ^[7]	RLM-RNN ^[14]
Ermse	1.3157	2.1725	1.9790

5 结 论

为了解决 ESN 学习算法中可能存在解的奇异问题、时间序列预测时易导致病态解问题, 尤其是当训

练样本个数小于输出权值维数时, ESN 的解必为奇异的问题, 本文使用 LM 算法代替常用的线性回归方法, 提出 RLM-ESN 预测模型, 自适应选择 LM 参数, 从而有效地控制输出权值的幅值, 提高了 ESN 的预测性能. 通过 Lorenz 混沌时间序列进行预测研究, 并对大连月平均气温实际数据进行仿真研究, 取得了较好的预测效果.

参考文献(References)

- [1] Gurupdes S. Pandher. Forecasting multivariate time series with linear restrictions using constrained structural state-space models[J]. J of Forecasting, 2002, 21(4): 281-300.
- [2] 崔万照, 朱长纯. 混沌时间序列的支持向量机预测[J]. 物理学报, 2004, 53(10): 3304-3310.
(Cui W Z, Zhu C C. Prediction of the chaotic time series using support vector machines[J]. Acta Physica Sinica, 2004, 53(10): 3304-3310.)
- [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. Reading MA: Addison-Wesley, 1993: 195-217.
- [4] William W W H. Nonlinear multivariate and time series analysis by neural network methods[J]. Review Geophysics, 2004, 42(1): 1003-1008.
- [5] Vesanto J. Using the SOM and local models in time-series prediction[C]. Proc of WSOM'97, Workshop on Self-Organizing Maps. Espoo, 1997: 209-214.
- [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.
- [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]. German National Research Center for Information Technology, 2002.
- [10] Shi Z W, Han M. Support vector echo state machine for chaotic time series prediction[J]. IEEE Trans on Neural Networks, 2007, 18(2): 359-372.

(下转第1478页)