

基于鲁棒回声状态网络的混沌时间序列预测研究*

李德才 韩敏[†]

(大连理工大学电子信息与电气工程学部, 大连 116023)

(2010年12月15日收到; 2011年1月16日收到修改稿)

针对回声状态网络模型易受异常点影响的问题, 提出一种基于拉普拉斯先验分布的鲁棒回声状态网络模型. 通过采用对异常点不敏感的拉普拉斯分布代替高斯分布作为模型输出的先验, 以增强网络对于异常点的抑制能力. 此外, 为解决由引入拉普拉斯分布所造成的求解困难的问题, 根据边际优化方法, 构建适当的替代函数, 使拉普拉斯先验等价转化为易于计算的高斯形式, 并通过贝叶斯方法实现模型参数的自适应估计. 仿真结果表明, 在异常点存在的情况下, 本文所提出的模型具有较好的鲁棒性, 并仍能保持较高的预测精度.

关键词: 回声状态网络, 鲁棒模型, 替代函数, 拉普拉斯分布

PACS: 89.90.+n

1. 引言

混沌是指发生在确定性系统中的貌似随机的不规则运动. 自然界中的混沌现象是普遍存在的, 诸如太阳黑子, 河流径流量, 电力系统, 气温及降雨量等都可以较好的运用混沌理论来解释^[1,2].

近些年, 国内外学者对混沌系统的分析和预测研究与日剧增. Volterra 滤波器^[3], 多层感知机神经网络, 径向基神经网络^[4], 支持向量机^[5], 模糊模型^[6], 状态空间神经网络, Elman 网络^[7]以及回声状态网络^[8]等都被广泛地应用于混沌时间序列的建模和预测. 其中回声状态网络作为一种新型的递归神经网络, 无论是模型构建, 还是学习算法都与传统的神经网络具有很大差别^[9,10]. 回声状态网络的核心组成部分是储备池, 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, 输出权值是唯一而且是全局最优的, 因此不存在传统神经网络普遍存在的局部最小问题, 从而显示出较好的预测性能.

伪逆方法是目前回声状态网络学习过程中的常规方法, 但在处理实际问题时, 易导致病态解问题, 且伴随着具有较大幅值的输出权值, 从而导致过拟合现象发生. 文献^[11]建议采用岭回归方法解

决上述问题, 但是模型中正则项系数通常采用交叉检验方法进行确定, 计算量较大. 在此基础上, 文献^[12]提出基于贝叶斯回归的回声状态网络预测模型, 实现了网络输出权值及正则化系数的自适应估计, 避免了繁琐的交叉检验过程, 降低了算法的计算复杂度. 在贝叶斯回归方法当中, 通常假设所有数据的噪声均为高斯分布. 然而, 实际观测的数据不免会包含一些异常点, 这些异常点的存在导致数据噪声不再满足高斯分布, 从而影响模型的泛化能力.

针对此问题, 本文提出一种鲁棒回声状态网络(robust echo state network, RESN)模型, 采用更具鲁棒性的拉普拉斯分布代替高斯分布对数据的噪声进行描述, 借以克服异常点对模型的影响. 此外, 在模型参数的求解过程中, 为解决引入拉普拉斯分布造成的求解困难的问题, 根据边际优化方法, 构建适当的替代函数, 使其等价转化为易于求解的高斯分布形式, 从而简化模型参数的求解过程.

2. 回声状态网络

2.1. 回声状态网络的结构

回声状态网络的方程可以表示为

* 国家自然科学基金(批准号:61074096)资助的课题.

[†] 通讯联系人. E-mail: minhan@dlut.edu.cn

$$\begin{aligned} \mathbf{x}(k+1) &= \text{sig}(\mathbf{W}_x \cdot \mathbf{x}(k) + \mathbf{W}_{in} \cdot \mathbf{u}(k) + \mathbf{b}_x) \\ y(k) &= \mathbf{w}^T \mathbf{x}(k) + b, \end{aligned} \quad (1)$$

其中 sig 表示 Sigmoid 激活函数. $\mathbf{x}(k)$, $\mathbf{u}(k)$ 和 $y(k)$ 分别为储备池的状态变量, 输入变量和输出变量. \mathbf{W}_x 和 \mathbf{W}_{in} 分别表示内部连接矩阵和输入矩阵, \mathbf{b}_x 为输入偏置, 其中 \mathbf{W}_x , \mathbf{W}_{in} 和 \mathbf{b}_x 按照特定方式随机生成, 并且在训练过程中保持不变. \mathbf{W}_x 通常保持 1%—5% 的稀疏连接, 其谱半径 (矩阵所有特征值的最大值) 则一般保持小于 1. \mathbf{w} 和 b 是需要根据训练样本确定的输出权值和输出偏置. 为了表示方便, 以下将偏置向量当作与固定值输入 1 之间的连接权值, 从而可以分别将 \mathbf{b}_x 和 b 合并到矩阵 \mathbf{W}_x 和 \mathbf{w} 中去.

对于给定的输入输出样本对 $\{(u_k, y_k), k = 1, 2, \dots\}$, 回声状态网络的建立过程可以简单描述为: 1) 选择网络的规模, 内部连接矩阵 \mathbf{W}_x 的稀疏度和谱半径, 并随机初始化 \mathbf{W}_x 和 \mathbf{W}_{in} ; 2) 根据输入样本激励系统(1)的内部状态, 并记录各个时刻状态变量的大小; 3) 根据状态变量和期望输出之间的线性回归关系, 计算输出权值 \mathbf{w} .

在回声状态网络的建模过程中, 常常需要舍弃一段初始的暂态过程, 假设舍掉的初始暂态长度为 ϕ . 此时回声状态网络的状态矩阵 Φ 和对应的期望输出 \mathbf{y} 可以分别表示为

$$\begin{aligned} \Phi &= [\mathbf{x}(\phi), \mathbf{x}(\phi+1), \dots, \mathbf{x}(\phi+N-1)]^T, \\ \mathbf{y} &= [y(\phi), y(\phi+1), \dots, y(\phi+N-1)]^T, \end{aligned} \quad (2)$$

其中 ϕ 一般不从 1 开始, 以去除初始暂态的影响, N 对应于训练样本的最后一个时间点. 其训练过程即是求解关于状态矩阵 Φ 和期望输出 \mathbf{y} 的线性系统, 可以得到

$$\mathbf{y} = \Phi \mathbf{w}. \quad (3)$$

2.2. 回声状态网络的学习算法

目前伪逆方法是回声状态网络的常规学习方法. 但现有结果表明, 在采用伪逆方法处理实际问题时易导致病态解, 从而产生过拟合现象. 对此可以采用岭回归代替伪逆方法求取网络的输出权值^[11]

$$\mathbf{w} = (\lambda \mathbf{I} + \Phi^T \Phi)^{-1} \Phi^T \mathbf{y}, \quad (4)$$

其中正则项系数 λ 的确定影响着回声状态网络的性能, 通常需要采用交叉检验方法对其进行有效确定. 鉴于交叉检验方法计算量较大, 可以采用贝叶斯回归对网络输出权值及正则项系数进行自适应

估计^[12], 从而避免繁琐的交叉检验过程. 然而, 由于贝叶斯回归假设所有数据的噪声均为同方差的高斯分布, 导致其对于异常点不具有鲁棒性^[13], 因此, 当观测数据中存在异常点时, 模型的泛化能力将受到影响.

3. 鲁棒回声状态网络

为解决上述问题, 本文将采用对异常点敏感性较差的拉普拉斯分布替代贝叶斯方法中的高斯分布, 实现对异常点的有效抑制. 并通过贝叶斯框架实现模型参数的自适应估计, 此外, 鉴于拉普拉斯分布的使用将增加模型求解过程的复杂性, 本文将进一步根据边际优化方法, 构造适当的替代函数, 简化模型参数的求解过程. 具体实现过程如下文所述.

3.1. 模型结构设计

假设模型输出 \mathbf{y} 由某种非线性函数所产生, 且各数据满足独立同分布. 为提高模型对异常点的抑制能力, 其先验分布可以选为拉普拉斯分布

$$p(\mathbf{y} | \mathbf{w}, \beta) = \left(\frac{\beta}{2}\right)^N \exp\{-\beta |\mathbf{y} - \Phi \mathbf{w}|\}. \quad (5)$$

对于网络的输出权值 \mathbf{w} , 在没有先验知识的情况下, 可以假设 $p(\mathbf{w})$ 服从指数型分布, 而其中最常见的是高斯分布, 因此可以定义输出权值 \mathbf{w} 的先验分布为

$$p(\mathbf{w} | \alpha) = \left(\frac{\alpha}{2\pi}\right)^{L/2} \exp\left\{-\frac{\alpha}{2} \mathbf{w}^T \mathbf{w}\right\}, \quad (6)$$

输出权值 \mathbf{w} 及超参数 α 和 β 可以分别通过最大化相应的后验分布 $p(\mathbf{w} | \mathbf{y})$ 和边际似然函数 $p(\mathbf{y} | \alpha, \beta)$ 获得

$$p(\mathbf{w} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{w}, \beta) p(\mathbf{w} | \alpha)}{p(\mathbf{y})}, \quad (7)$$

$$p(\mathbf{y} | \alpha, \beta) = \int p(\mathbf{y} | \mathbf{w}, \beta) p(\mathbf{w} | \alpha) d\mathbf{w}, \quad (8)$$

其中

$$p(\mathbf{y}) = \int p(\mathbf{y} | \mathbf{w}, \beta) p(\mathbf{w} | \alpha) d\mathbf{w}. \quad (9)$$

然而, 由于假设模型输出 \mathbf{y} 的先验分布 $p(\mathbf{y} | \mathbf{w}, \beta)$ 为拉普拉斯分布, 因此难以通过最大化(7)和(8)式得到模型参数的解析表达式, 为解决此问题可以引入边际优化方法^[14], 通过选择适当的替代函数, 将其转化为易于求解的形式.

3.2. 基于替代函数的模型转化

根据边际优化方法, 最大化目标函数 $L(\theta)$ 可以等价于迭代最大化相应的替代函数 Q , 即

$$\hat{\theta}^{(t+1)} = \operatorname{argmax}_{\theta} Q(\theta | \hat{\theta}^{(t)}). \quad (10)$$

如果替代函数满足条件: $L(\theta) - Q(\theta | \hat{\theta}^{(t)})$ 在 $\theta = \hat{\theta}^{(t)}$ 处取得最小值, 则可保证目标函数 $L(\theta)$ 在整个迭代过程中单调递增.

为构造合适的替代函数, 首先考虑如下不等式:

$$-|\varepsilon| \geq -\frac{1}{2} \left(\sum_{i=1}^N \frac{\varepsilon_i^2}{|\hat{\varepsilon}_i^{(t)}|} + \sum_{i=1}^N |\hat{\varepsilon}_i^{(t)}| \right), \quad (11)$$

其中 $\varepsilon = \mathbf{y} - \mathbf{A}\mathbf{w}$, 且不等式左右两边当且仅当 $\varepsilon = \hat{\varepsilon}^{(t)}$ 时相等.

进而可以得到

$$\begin{aligned} p(\mathbf{y} | \mathbf{w}, \beta) &= \left(\frac{\beta}{2}\right)^N \exp\{-\beta|\mathbf{y} - \mathbf{A}\mathbf{w}|\} \geq \left(\frac{\beta}{2}\right)^N \\ &\times \exp\left\{-\frac{\beta}{2} \left(\sum_i \frac{\varepsilon_i^2}{|\hat{\varepsilon}_i^{(t)}|} + \sum_i |\hat{\varepsilon}_i^{(t)}| \right)\right\} \\ &= \prod_{i=1}^N \left(\frac{\beta}{2\pi|\hat{\varepsilon}_i^{(t)}|}\right)^{\frac{1}{2}} \exp\left\{-\frac{\beta}{2|\hat{\varepsilon}_i^{(t)}|} \varepsilon_i^2\right\} \\ &\times \prod_{i=1}^N \left(\frac{\pi\beta|\hat{\varepsilon}_i^{(t)}|}{2}\right)^{1/2} \exp\left\{-\frac{\beta|\hat{\varepsilon}_i^{(t)}|}{2}\right\} \\ &= p(\mathbf{y} | \mathbf{w}, \beta, \hat{\varepsilon}^{(t)}). \end{aligned} \quad (12)$$

综上所述, 基于(10)和(11)式可得 $p(\mathbf{w} | \mathbf{y})$ 和 $p(\mathbf{y} | \alpha, \beta)$ 的替代函数分别为

$$p(\mathbf{w} | \mathbf{y}, \hat{\varepsilon}^{(t)}) = \frac{p(\mathbf{y} | \mathbf{w}, \beta, \hat{\varepsilon}^{(t)})p(\mathbf{w} | \alpha)}{p(\mathbf{y})}, \quad (13)$$

$$p(\mathbf{y} | \alpha, \beta, \hat{\varepsilon}^{(t)}) = \int p(\mathbf{y} | \mathbf{w}, \beta, \hat{\varepsilon}^{(t)})p(\mathbf{w} | \alpha) d\mathbf{w}. \quad (14)$$

由(12)式容易看出, 通过替代函数可以将输出变量 \mathbf{y} 的先验分布转化为易于求解的高斯分布, 从而简化模型参数的求解过程.

3.3. 参数求解

将(6)和(12)式代入(13)式中, 可得

$$\begin{aligned} p(\mathbf{w} | \mathbf{y}, \hat{\varepsilon}^{(t)}) &= \\ &\frac{\exp(M(\mathbf{w})) \left(\frac{\beta}{2}\right)^N \left(\frac{\alpha}{2\pi}\right)^{L/2} \prod_{i=1}^N \exp\left\{-\frac{\beta|\hat{\varepsilon}_i^{(t)}|}{2}\right\}}{2p(\mathbf{y})}, \end{aligned} \quad (15)$$

其中

$$M(\mathbf{w}) = -\frac{\beta}{2}(\mathbf{y} - \Phi\mathbf{w})^T \Lambda^{-1}(\mathbf{y} - \Phi\mathbf{w}) - \frac{\alpha}{2}\mathbf{w}^T \mathbf{w}, \quad (16)$$

$$\Lambda = \operatorname{diag}\{|\hat{\varepsilon}_1^{(t)}|, |\hat{\varepsilon}_2^{(t)}|, \dots, |\hat{\varepsilon}_N^{(t)}|\}. \quad (17)$$

进而可得 $p(\mathbf{w} | \mathbf{y}, \hat{\varepsilon}^{(t)})$ 的对数似然函数为

$$\begin{aligned} \ln p(\mathbf{w} | \mathbf{y}, \hat{\varepsilon}^{(t)}) &= -\frac{\beta}{2}(\mathbf{y} - \Phi\mathbf{w})^T \Lambda^{-1}(\mathbf{y} - \Phi\mathbf{w}) \\ &- \frac{\alpha}{2}\mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \frac{\beta|\hat{\varepsilon}_i^{(t)}|}{2} \\ &+ N \log \frac{\beta}{2} + \frac{L}{2} \log \left(\frac{\alpha}{2\pi}\right) \\ &- \log p(\mathbf{y}). \end{aligned} \quad (18)$$

固定参数 α 和 β , 最大化 $\ln p(\mathbf{w} | \mathbf{y}, \hat{\varepsilon}^{(t)})$ 可得到输出权值 \mathbf{w} 为

$$\mathbf{w} = \beta \Sigma \Phi^T \Lambda^{-1} \mathbf{y}, \quad (19)$$

其中

$$\Sigma = (\mathbf{A} + \beta \Phi^T \Lambda^{-1} \Phi)^{-1}, \quad (20)$$

$$\mathbf{A} = \operatorname{diag}\{\alpha_1, \alpha_2, \dots, \alpha_L\}. \quad (21)$$

对于参数 α 和 β , 可以根据证据函数方法确定. 通过对输出权值 \mathbf{w} 积分, 可得边际似然函数 $p(\mathbf{y} | \alpha, \beta, \hat{\varepsilon}^{(t)})$ 为

$$\begin{aligned} p(\mathbf{y} | \alpha, \beta, \hat{\varepsilon}^{(t)}) &= \int p(\mathbf{y} | \mathbf{w}, \beta, \hat{\varepsilon}^{(t)}) p(\mathbf{w} | \alpha) d\mathbf{w} \\ &= \frac{1}{(2\pi)^{m/2} |\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}\right\} \\ &\times \prod_{i=1}^N \left(\frac{\pi\beta|\hat{\varepsilon}_i^{(t)}|}{2}\right)^{1/2} \\ &\times \exp\left\{-\frac{\beta|\hat{\varepsilon}_i^{(t)}|}{2}\right\}, \end{aligned} \quad (22)$$

其中

$$\mathbf{C} = \beta^{-1} \mathbf{A} + \Phi \mathbf{A}^{-1} \Phi^T. \quad (23)$$

为了便于计算, 一般不直接最大化边际似然函数 $p(\mathbf{y} | \alpha, \beta, \hat{\varepsilon}^{(t)})$, 而是等价于最大化其对数形式. 对 $p(\mathbf{y} | \alpha, \beta, \hat{\varepsilon}^{(t)})$ 取负对数, 并去除与超参数 α 和 β 无关的项后, 可得

$$\begin{aligned} \ln p(\mathbf{y} | \alpha, \beta, \hat{\varepsilon}^{(t)}) &= \frac{1}{2} \log |\Sigma| + \frac{1}{2} \log |\beta \Lambda^{-1}| \\ &+ \frac{1}{2} \log |\mathbf{A}| - \beta(\mathbf{y} - \Phi\mathbf{w})^T \\ &\times \Lambda^{-1}(\mathbf{y} - \Phi\mathbf{w}) \\ &- \mathbf{w}^T \mathbf{A} \mathbf{w} + \frac{N}{2} \log \beta \\ &- \frac{\beta}{2} \sum_{i=1}^N |\hat{\varepsilon}_i^{(t)}| + \text{const}. \end{aligned} \quad (24)$$

将 $\ln p(y | \alpha, \beta, \hat{\epsilon}^{(t)})$ 分别对 α 和 β 求偏导, 并令导数为零, 可以得到超参数 α 和 β 的估计式分别为

$$\alpha_j = \frac{1}{w_j^2 + \sum_{jj}} = \frac{\gamma_j}{w_j^2}, \quad (25)$$

$$\beta = \frac{2N}{(y - \Phi w)^T \Lambda^{-1} (y - \Phi w) + \text{tr}(\Sigma \Phi^T \Lambda \Phi) + |\hat{\epsilon}^{(t)}|}, \quad (26)$$

其中

$$\gamma_j = 1 - \alpha_j \sum_{jj}. \quad (27)$$

3.4. 鲁棒回声状态网络 (RESN) 的实现步骤

综上所述, 可以将基于拉普拉斯先验分布的鲁棒回声状态网络的实现步骤表示如下:

1) 选择网络的规模, 内部连接矩阵 W_x 的稀疏度和谱半径, 并随机初始化 W_x 和 W_{in} .

2) 根据输入样本激励系统(1)的内部状态, 并记录各个时刻状态变量的大小.

3) 根据状态变量和期望输出之间的线性回归关系, 计算输出权值 w . 为简化计算过程, 根据(10)–(14)式将 $p(w | y)$ 和 $p(y | \alpha, \beta)$ 转化为对应的替代函数 $p(w | y, \hat{\epsilon}^{(t)})$ 和 $p(y | \alpha, \beta, \hat{\epsilon}^{(t)})$.

4) 令 $t=0$, 采用贝叶斯方法优化替代函数, 并初始化参数 $\alpha^{(t)}$ 和 $\beta^{(t)}$.

5) 分别根据(19)和(27)式计算变量 $w^{(t)}$ 及 $\gamma^{(t)}$, 并据此估计模型的训练误差 $\epsilon^{(t)}$.

6) 由(25)和(26)式更新 $\alpha^{(t+1)}$ 和 $\beta^{(t+1)}$, 部分 α_j 在迭代过程中会趋于正无穷 (通常判断其若大于某一阈值如 10^{12}), 其对应的输出权值将趋于零, 从而实现模型结构的稀疏化^[15].

7) $t = t + 1$. 若超参数 α 和 β 收敛 (如参数 α 和 β 的变化量小于某一给定阈值或达到最大迭代次数). 则停止迭代, 否则转至步骤(5).

3.5. 鲁棒回声状态网络 (RESN) 的稳定性分析

稳定性问题是递归神经网络应用中的一个基本问题, 如果预测器的稳定性无法保证, 就无法进行预测. 相对与传统的递归神经网络, ESN 的一个突出优点是可以在网络训练之前确定网络的稳定性. 比如当矩阵 W_x 的谱半径小于 1 ($\rho(W_x) < 1$) 时, ESN 是渐进稳定的. 这个条件可以很容易从线性系统的状态空间表达式得到

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

$$y(k) = w^T x(k). \quad (28)$$

上述系统渐进稳定的充分必要条件是 W_x 的特征值必须位于单位圆内. 设储备池内部的激活函数为常用的双曲正切函数, 对任意 k 恒有下式成立:

$$\begin{aligned} & \| \text{sig}(W_x \cdot x(k) + W_{in} \cdot u(k)) \| \\ & \leq \| W_x \cdot x(k) + W_{in} \cdot u(k) \|. \end{aligned} \quad (29)$$

当要求 $\rho(W_x) < 1$ 时, 系统(1)将是渐进稳定的. W_x 是在 ESN 网络训练之前确定的, 因此可以在设计阶段对由(28)式描述的系统施加稳定性约束. 这种稳定性特性不会由于 w 的训练结果而改变, 这一点是储备池方法特有的. 虽然 $\rho(W_x) < 1$ 仅仅是个充分条件, 系统(1)稳定时, 该条件不一定满足. 但不过无论如何, 储备池网络的稳定性都是可以在训练之前分析的.

另一方面, 在对输出权值 w 的训练过程中, 虽然为提高模型对异常点的抑制能力, 选择模型输出 y 的先验分布为拉普拉斯分布, 但通过选择适当的替代函数, 可以将其转化为易于计算的高斯形式. 这种形式下的超参数求解过程可以等价于 Tipping 所提出的相关向量机方法^[15], 其中超参数可以通过最大化似然函数, 经由有限步收敛到其最优值.

综上所述, 通过选择适当的储备池谱半径及替代函数形式, 可以有效保证所提的鲁棒回声状态网络方法的稳定性以及超参数在求解过程中的收敛性.

4. 仿真实例

对 Lorenz 方程的 $x(t)$ 序列以及黄河年径流序列两组实验数据进行仿真, 以验证所提算法的有效性.

4.1. Lorenz 方程 $x(t)$ 混沌时间序列预测仿真

Lorenz 方程组为

$$\begin{aligned} \frac{dx}{dt} &= a(-x + y), \\ \frac{dy}{dt} &= bx - y - xz, \\ \frac{dz}{dt} &= xy - cz, \end{aligned} \quad (30)$$

当参数 $a = 10, b = 28, c = 8/3$, 初始值 $x(0) = 12, y(0) = 2, z(0) = 9$ 时, 方程表现出混沌特性.

储备池的参数设置见表 1, 其中储备池的大小取为 200×200 , 内部连接矩阵 W_x 的稀疏度和谱半

径分别取为 0.95 和 0.02. 取其中的 $x(t)$ 时间序列作为仿真对象, 利用四阶 Runge-Kutta 法产生混沌时间序列, 步长为 0.02. 首先对 $x(t)$ 序列进行相空间重构, 延迟时间和嵌入维数分别选择为 $\tau_x = 8$, $m_x = 6$.

$$d(t) = \{x(t), x(t-8), \dots, x(t-5 \times 8)\}.$$

用延迟坐标向量 $d(t)$ 作为输入激励系统 (1), 得到 2200 个相继的状态变量 $x(t)$, 同时找到 t 时刻的期望输出 $x(t+1)$, 舍掉长度为 200 的初始暂态过程, 得到 2000 组“状态-输出”对. 选取其中前 1500 组作为训练样本, 其余 500 组作为测试样本.

表 1 用于 Lorenz 序列预测的储备池参数设置

储备池规模	谱半径	稀疏度/%
200 × 200	0.95	2

为体现实验条件的真实性, 在训练样本的输出目标上叠加均值为零的高斯噪声, 噪声水平取 20%. 进一步, 为分析异常点对建模质量的影响, 额外加入 6 个幅值为 50 的异常点进行干扰, 得到的 Lorenz 混沌时间序列如图 1 所示.

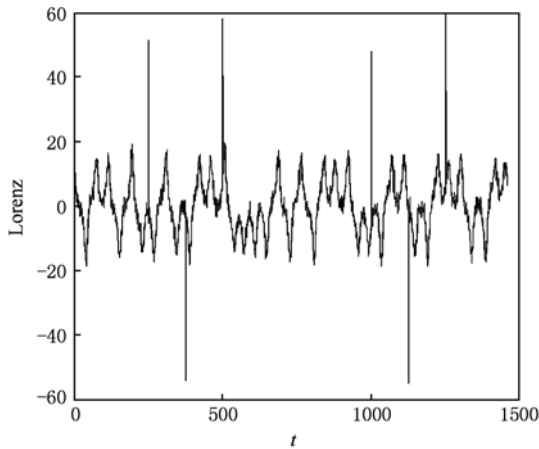


图 1 包含噪声和异常点的 Lorenz 混沌时间序列

采用本文所提出的鲁棒回声状态网络对包含噪声及异常点的 Lorenz 混沌时间序列进行仿真, 得到的预测及误差曲线如图 2 所示. 从图 2 中可以看出, 由于采用拉普拉斯分布替换了对异常点较为敏感的高斯分布作为模型输出的先验, 增强了网络对异常点的抑制能力, 即使在训练数据包含噪声及异常点的情况下, 仍能得到可靠的预测结果.

为进一步验证所提算法的有效性, 分别对基于岭回归方法, 贝叶斯回归方法以及 Huber 鲁棒损失函数的回声状态网络模型进行仿真, 并同本文所提

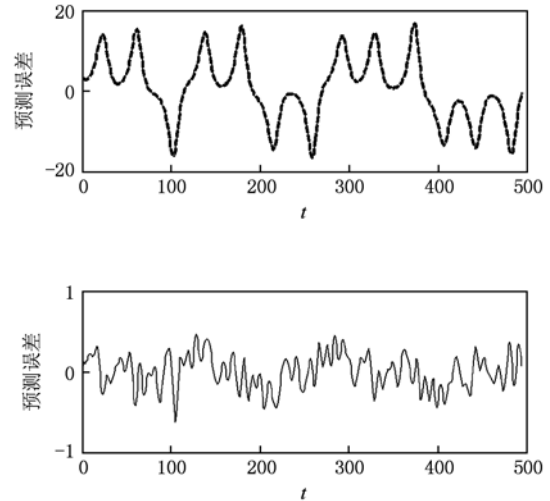


图 2 鲁棒回声状态网络对 Lorenz 混沌时间序列的预测及误差曲线

出的鲁棒回声网络方法进行比较. 相应的仿真结果如表 2 所示. 其中 Huber 损失函数为一种常用的鲁棒损失函数, 能够有效抑制奇异点对模型的影响^[16]. 其定义式可以表示为

$$L_{\text{Huber}}(\varepsilon) = \begin{cases} \frac{1}{2}\varepsilon^2, & |\varepsilon| < \mu, \\ \mu|\varepsilon| - \frac{\mu^2}{2}, & \text{其他}, \end{cases} \quad (31)$$

其中 μ 为 Huber 损失函数的参数, 在本例中通过交叉检验方法确定其取值为 0.6.

表 2 Lorenz 混沌时间序列的测试结果比较 (RMSE)

方法	岭回归方法 ^[11]	Huber 损失函数 ^[16]	贝叶斯回归方法 ^[12]	鲁棒回声状态网络
测试误差	0.3714	0.2644	0.4192	0.2238
训练时间	95.4688	34.3751	2.0547	1.3578

从表 2 中可以看出, 基于岭回归和贝叶斯方法的回声状态网络模型具有相似的测试结果 (文中采用均方根误差 E_{RMSE} 对模型的预测结果进行度量), 由于在两种方法中分别采用对异常点不具有鲁棒性的二次损失函数和高斯先验分布作为建模基础, 因此, 对于包含较大异常点的观测数据, 模型的预测性能无法得到保证. 而在本文所提出的鲁棒回声状态网络中, 由于采用对异常点不敏感的拉普拉斯分布代替高斯分布作为模型输出的先验, 增强了网络对于异常点的抑制能力, 从而其预测精度明显优于前述两种方法. 而且在本文所提出的方法中, 由于采用贝叶斯方法对模型参数进行自适应估计, 所

以无需诸如岭回归方法中繁琐的正则项系数的确定过程.

此外, 鉴于 Huber 损失函数对异常点的鲁棒性, 使其所对应的网络模型在处理包含异常点的观测数据时仍能保证较高的预测精度. 然而对于 Huber 损失函数, 需要根据参数 μ 对训练样本进行划分, 以决定采用二次损失函数或者鲁棒损失函数进行建模 (如 (31) 式所示). 因此, 参数 μ 的取值对模型的性能具有较大影响. 但是对于参数 μ , 通常采用交叉检验方法从一组候选参数中选出最优值, 从而增加了模型的计算复杂度. 虽然交叉检验过程可以通过贝叶斯方法进行简化, 但是同正则化参数 λ 的确定过程不同, 参数 μ 无法通过最大化似然函数进行自适应估计, 而需要对应于每一个候选参数, 通过贝叶斯方法进行训练, 进而根据训练误差选择出最优参数值, 从而增加了模型的计算复杂度. 反观本文所提出的鲁棒回声状态网络模型, 在其学习过程中, 各样本数据对模型的重要性可以通过网络的训练误差进行反映, 并在迭代过程中不断调整, 避免了额外设定参数对样本数据进行划分的步骤, 简化了模型的求解过程.

4.2. 黄河年径流时间序列预测

黄河作为中国第二大河流, 对沿岸居民的产生和生活影响较大. 近年来, 国内外学者采用混沌理论对其径流序列进行分析和预测, 并取得了一定成果^[17,18]. 本例中采用河南省三门峡水文站 1700 年至 1997 年观测到的年黄河径流时间序列作为研究对象, 借以分析本文所提出的鲁棒回声状态网络的预测性能. 在建立预测模型之前, 首先对此径流序列进行简单的去噪处理, 以减弱序列的随机性, 并增强其确定性.

储备池的大小取为 200×200 , 内部连接矩阵 W_x 的稀疏度和谱半径分别取为 0.06 和 0.02, 具体的参数设置如表 3 所示. 分别由互信息及伪最近邻方法计算得到延迟时间 τ 和嵌入维数 m 为 1 和 6^[19,20], 对径流序列进行相空间重构, 得到延迟坐标向量 $d(t)$ 为

$$d(t) = \{x(t), x(t-1), \dots, x(t-5)\},$$

用延迟坐标向量 $d(t)$ 作为输入激励系统 (1), 得到 292 个相继的状态变量 $x(t)$, 同时找到 t 时刻的期望输出 $x(t+\eta)$, 其中 η 为预测步长, 本例中期望的预测输出分别为 1 ($\eta=1$) 和 2 ($\eta=2$) 年以后的

径流量.

表 3 用于 Lorenz 序列预测的储备池参数设置

储备池规模	谱半径	稀疏度/%
200 × 200	0.06	2

舍掉长度为 20 的初始暂态过程, 得到 252 组“状态-输出”对. 选取其中前 232 组作为训练样本, 其余 20 组数据作为测试样本.

为说明本文所提算法的有效性, 分别对基于岭回归方法, 贝叶斯回归方法以及 Huber 鲁棒损失函数的回声状态网络模型进行仿真, 并同本文所提出的鲁棒回声网络方法进行比较. 相应的仿真结果如表 4 所示.

表 4 黄河年径流序列的测试结果比较 (RMSE)

方法	岭回归方法 ^[11]	Huber 损失函数 ^[16]	贝叶斯回归方法 ^[12]	鲁棒回声状态网络
1 年 ($\eta=1$)	25.8096	22.9267	25.4373	21.5177
2 年 ($\eta=2$)	59.5344	55.0271	58.9046	47.2391
训练时间	86.5000	18.8292	1.3156	0.9375

对比表 2 和表 4 中的仿真结果可以看出, 4 种方法在 Lorenz 以及黄河年径流序列上的预测精度具有相似的趋势. 以 1 年期的预测结果为例, 对于黄河年径流时间序列, 虽然在建立预测模型之前进行了简单的去噪处理, 但是序列中仍包含较多的噪声及异常点. 而鉴于基于岭回归和贝叶斯方法的回声状态网络模型的鲁棒性较差, 因此其预测精度受到了较大的影响. 同其余两种鲁棒方法相比, 所得到的结果较差.

此外, 在计算时间方面, 对于岭回归方法, 正则项系数 λ 通常需要通过交叉检验方法进行有效确定. 而鉴于留 1 法交叉检验能够无偏估计模型的泛化能力, 本例中采用该方法对正则项系数进行估计. 其中 λ 按照 2 的指数增长顺序选取: $\lambda = [2^{-10}, 2^{-9}, \dots, 2^9, 2^{10}]$. 但是交叉检验方法计算量较大, 因此, 其训练时间要明显多于其余三种方法. 同岭回归方法采用交叉检验确定正则项系数不同, 在贝叶斯回归方法中, 通过最大化边际似然函数可以实现输出权值及正则项系数的进行自适应估计, 从而避免繁琐的交叉检验过程, 有效缩短了模型的训练时间.

图 3 是基于 Huber 损失函数的回声状态网络的预测误差随参数 μ 的变化曲线. 从图 3 中可以看

出,当参数 μ 从0.01到0.3的变化过程中,所对应的网络模型的预测误差先降后升.模型在 $\mu = 0.05$ 处取得了最优的预测结果,并表现出较强的鲁棒性,但是当参数 μ 取值较小时,模型的泛化性能迅速恶化,而当参数 μ 由0.05逐渐增加时,模型的预测误差也随之增大,在 $\mu = 0.3$ 处模型的预测精度已同基于岭回归和贝叶斯方法的回声状态网络模型相差无几.由此可以看出,过大或者过小的参数 μ 都无法取得理想的预测结果,其取值需要在训练过程中进行有效确定.

对于 Huber 损失函数中的参数 μ ,通常需要采用交叉检验方法从一组候选参数中选出最优值,模型的计算复杂度较大.虽然交叉检验过程可以通过贝叶斯方法进行简化,但是同正则化参数 λ 的确定过程不同,参数 μ 无法通过最大化似然函数进行自适应估计,而需要对应于每一个候选参数,通过贝叶斯方法进行训练,进而根据训练误差选择出最优参数值.在本例中,候选参数 μ 分别选取: $\mu = [0.0125, 0.025, 0.050, 0.075, \dots, 0.275, 0.300]$,共13个候选值.同岭回归方法相比,由于采用贝叶斯方法代替了计算复杂度较大的交叉检验,因此计算时间明显少于前者,但是对于候选集中的每一个参数 μ ,均需要采用贝叶斯方法进行训练.因此,计算时间又多于贝叶斯回归方法.

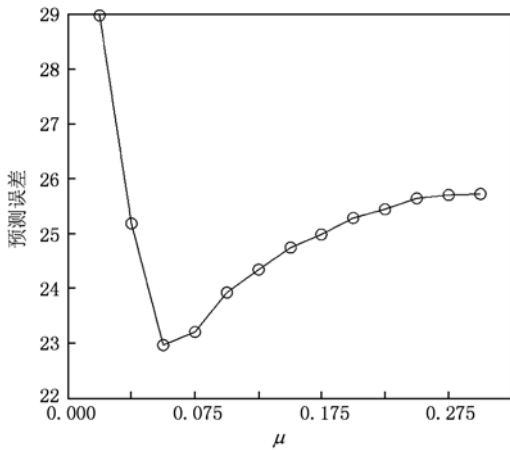


图3 基于 Huber 损失函数的回声状态网络的预测误差随参数 μ 的变化情况

对于本文所提出的鲁棒回声状态网络,拉普拉斯分布的使用提高了其对于异常点的抑制能力,因此,模型的预测精度较高.除此之外,模型参数可通过最大化边际似然函数进行自适应估计,而且无

需额外设置参数对模型的预测性能进行控制(同基于 Huber 损失函数的方法相对),也是本文所提方法区别于岭回归及 Huber 损失函数方法的一个主要方面.同贝叶斯回归方法相比较,虽然两者实现过程相似,但是对于本文所提出的方法,在其对模型训练的过程中,可以根据超参数的变化情况实现模型结构的简化(具体见3.4小节),而贝叶斯回归方法在整个训练过程中,模型都保持初始规模不变.所以,相比于贝叶斯回归方法,本文所提出的鲁棒回声状态网络模型的计算时间更短.图4为黄河年径流时间序列在测试样本上的预测结果,从图中可以看出,预测曲线较真实的反应了年径流序列的走势.

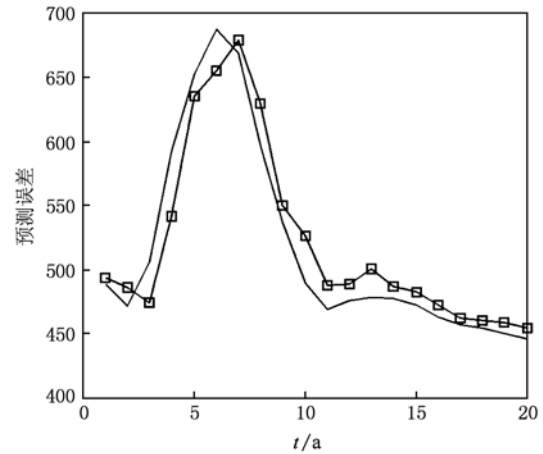


图4 鲁棒回声状态网络对黄河年径流时间序列的预测曲线

5. 结 论

针对回声状态网络模型易受异常点影响的问题,本文提出一种基于拉普拉斯先验分布的鲁棒回声网络模型,以提高模型对于异常点的抑制能力.并通过对 Lorenz 方程的 $x(t)$ 序列以及黄河年径流序列的实例仿真验证了所提方法的有效性.仿真结果表明,拉普拉斯分布的引入有效地提高了回声状态网络的鲁棒性,在异常点存在的情况下,模型仍能保持较高的预测精度.另一方面,通过引入边际优化方法,简化了基于拉普拉斯分布的模型的求解形式,并通过贝叶斯方法实现了模型参数的自适应估计,从而避免了计算量较大的交叉检验过程.

- [1] Chen S M, Hwang J R 2000 *IEEE Trans. Systems, Man and Cybernetics-Part B* **30** 263
- [2] Dhanya C T, Kumar D N 2010 *Advances in Water Resources* **33** 327
- [3] Du J, Cao Y J, Liu Z J, Xu L Z, Jiang Q Y, Guo C X, Lu J G 2009 *Acta Phys. Sin.* **58** 5997 (in Chinese) [杜杰、曹一家、刘志坚、徐立中、江全元、郭创新、陆金桂 2009 物理学报 **58** 5997]
- [4] Leung H, Lo T, Wang S C 2001 *IEEE Trans. Neural Network* **12** 1163
- [5] Cai J W, Hu S S, Tao H F 2007 *Acta Phys. Sin.* **56** 6820 (in Chinese) [蔡俊伟、胡寿松、陶洪峰 2007 物理学报 **56** 6820]
- [6] Ma Q L, Zheng Q L, Peng H, Tan J W 2009 *Acta Phys. Sin.* **58** 1410 (in Chinese) [马千里、郑启伦、彭宏、覃姜维 2009 物理学报 **58** 1410]
- [7] Farsa M A, Zolfaghari S 2010 *Neurocomputing* **73** 2540
- [8] Song Q S, Feng Z R, Li R H 2009 *Acta Phys. Sin.* **58** 5057 (in Chinese) [宋青松、冯祖仁、李人厚 2009 物理学报 **58** 5057]
- [9] Jaeger H, Haas H 2004 *Science* **304** 78
- [10] Shi Z W, Han M 2007 *IEEE Trans. Neural Networks* **18** 359
- [11] Shi Z W, Han M 2007 *Control and Decision* **22** 258 (in Chinese) [史志伟、韩敏 2007 控制与决策 **22** 258]
- [12] Han M, Mu D Y 2010 *Control and Decision* **25** 531 (in Chinese) [韩敏、穆大芸 2010 控制与决策 **25** 531]
- [13] Ting J A, Dsouza A, Schaal S 2007 *ICRA* **2007** 2489
- [14] Zhong M J 2006 *Neurocomputing* **69** 2351
- [15] Tipping M E 2001 *Journal of Machine Learning Research* **1** 211
- [16] Hong X, Chen S 2005 *IEEE Trans. Systems, Man and Cybernetics-Part B* **35** 155
- [17] Ding T, Zhou H C, Huang J H 2004 2010 *Journal of Hydraulic Engineering* **12** 15 (in Chinese) [丁涛、周惠成、黄健辉 2004 水力学报 **12** 15]
- [18] Islam M N, Sivakumar B 2002 *Advances in Water Resources* **25** 179
- [19] Fraser A M, Swinney H L 1986 *Phys. Rev. A* **33** 1134
- [20] Kennel M B, Brown R, Abarbanel H D I 1992 *Phys. Rev. A* **45** 3403

Chaotic time series prediction based on robust echo state network*

Li De-Cai Han Min[†]

(Faculty of Electronic Information and Electrical Engineering Dalian University of Technology, Dalian 116023, China)

(Received 15 December 2010; revised manuscript received 16 January 2011)

Abstract

Focusing on the problem that the echo state network is easily influenced by outliers, in this paper we propose a robust model based on the Laplace prior distribution. This is achieved by replacing the Gaussian distribution with the Laplace distribution as the prior of the model output, the Laplace prior is less sensitive to the outliers and can enhance the capability of the model to restrict outliers. Furthermore, to solve the problem arising from the introduction of the Laplace distribution, which makes the solving process of the method difficult, the bound optimization algorithm is employed and a suitable surrogate function is established. Based on the bound optimization algorithm, the Laplace prior can be equivalently transformed into the form of Gaussian prior, which is easily computed, and it can also be used to estimate the model parameters adaptively. Simulation results illustrate that the proposed method can be robust when outliers exist, while remaining acceptable prediction accuracy.

Keywords: echo state network, robust model, surrogate function, Laplace distribution

PACS: 89.90.+n

* Project supported by the National Natural Science Foundation of China (Grant No. 61074096).

[†] Corresponding author. E-mail: minhan@dlut.edu.cn