多元混沌时间序列的因子回声状态网络预测模型

许美玲[1]　　韩 敏[1]

摘　要　针对采用回声状态网络预测多元混沌时间序列时存在的病态解问题, 本文建立了因子回声状态网络模型, 通过因子分析 (Factor analysis, FA) 方法提取高维储备池状态矩阵的公因子, 去除冗余和噪声成分. 利用降维后的因子变量与期望输出之间的线性回归关系, 求解网络未知参数. 基于 Lorenz 序列和大连月平均气温 – 降雨量的仿真实验验证了本文所提模型的有效性.

关键词　多元混沌时间序列, 预测, 回声状态网络, 因子分析

引用格式　许美玲, 韩敏. 多元混沌时间序列的因子回声状态网络预测模型. 自动化学报, 2015, **41**(5): 1042−1046

DOI　10.16383/j.aas.2015.c140604

Factor Echo State Network for Multivariate Chaotic Time Series Prediction

XU Mei-Ling[1]　　HAN Min[1]

Abstract　When an echo state network is used to predict multivariate time series, there may exist ill-posed problem. This paper proposes a novel prediction model, named factor echo state network, to solve the problem. It uses a factor analysis (FA) algorithm to extract the common factors of the reservoir matrix, and to remove the redundancies and noises. Then the unknown output weights are calculated by linear regression of the output and common factors. The model is used to predict Lorenz series and monthly average temperature-rainfall time series in Dalian, and simulation results substantiate its effectiveness.

Key words　Multivariate chaotic time series, prediction, echo state network, factor analysis (FA)

Citation　Xu Mei-Ling, Han Min. Factor echo state network for multivariate chaotic time series prediction. *Acta Automatica Sinica*, 2015, **41**(5): 1042−1046

混沌是自然界确定性系统中广泛存在的一种貌似无规则、类似随机、对初始条件敏感、非周期、短期可预测但长期不可预测的现象. 随着信息采集与记录技术的发展, 从实际复杂系统收集到的具有混沌特性的时间序列规模越来越庞大, 比如气象系统中的气温、降雨量、大气压力序列等[1]. 多数情况下, 由于系统的复杂性, 无法建立精确的解析模型. 因此, 在机理分析的基础上, 如何利用数学的方法从大量的混沌时间序列中挖掘出有用的、相互关联的序列, 构成多元混沌时间序列, 共同表征复杂系统的演化规律, 是近年来的研究热点之一[2−4].

收稿日期 2014-08-19　录用日期 2014-12-04
Manuscript received August 19, 2014; accepted December 4, 2014
国家自然科学基金 (61374154), 国家重点基础研究发展计划 (973 计划) (2013CB430403) 资助
Supported by National Natural Science Foundation of China (61374154), and National Basic Research Program of China (973 Program) (2013CB430403)
本文责任编委 桂卫华
Recommended by Associate Editor GUI Wei-Hua
1. 大连理工大学电子信息与电气工程学部 大连 116023
1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023

神经网络是一种应用比较广泛的混沌时间序列预测模型[5−6], 理论上可以拟合任意非线性函数[7−9], 但结构比较复杂、收敛速度慢、易陷入局部最优. Jaeger[10] 于 2001 年提出一种改进的递归神经网络 – 回声状态网络 (Echo state network, ESN). 回声状态网络内部包含一个稀疏的、递归连接的储备池, 起到存储历史信息的回声作用[11], 储备池到输出层为线性映射. 输入权值和储备池内部连接权值随机初始化后保持不变[12], 只有输出权值通过线性回归的方法求解, 克服了一般递归神经网络收敛速度慢、易陷入局部最优的缺点, 成功应用于时间序列预测、模式识别和工业过程等领域[13−14]. 然而在使用回声状态网络预测多元混沌时间序列时, 仍存在着一些不足. 例如, 为映射数据复杂的混沌特性, 往往需要采用较大规模的储备池, 大规模储备池会产生大规模状态矩阵. 随着状态矩阵维数的增高, 可能出现不满秩现象或者近似相关现象, 导致病态解, 得到幅值较大的输出权值, 影响泛化性能[15−16].

对状态矩阵降维[17] 可以解决这一问题. 因子分析 (Factor analysis, FA) 是一种应用广泛的降维方法[18−19]. 假设彼此关联的多个变量存在着一些不能直接观测到的、但影响可观测变量变化的隐变量, 称为公共因子. 然后以这些公共因子为基底分解原变量, 即用有限个不可观测的隐变量来解释原始变量之间的相关关系. 本文将因子分析方法与回声状态网络结合, 建立一种新的多元混沌时间序列预测模型 — 因子回声状态网络 (Factor echo state network, FESN). 所提模型首先采用储备池映射混沌时间序列的动力学特性, 得到高维储备池状态变量, 然后从储备池状态变量间的依赖关系出发, 采用因子分析方法将大量的具有复杂关联关系的储备池变量归结为少量综合因子, 因子的个数通过 HQ 准则确定[20], 超参数采用 EM (Expectation maximization) 算法求解[18, 21]. 最后利用公因子求取输出权值, 以解决病态解问题, 提高泛化性能.

1　回声状态网络及其病态解问题

回声状态网络包括输入层、内部储备池和输出层三层. 其核心部分储备池包含成百上千个隐层节点, 具有强大的非线性映射能力[13]. 假设 t 时刻 l 元输入向量为 $\boldsymbol{u}_{in}(t)=[u_1(t),u_2(t),\cdots,u_l(t)]^{\rm T}$, l 元输出向量为 $\boldsymbol{y}(t)$, 有 $\boldsymbol{y}(t)=\boldsymbol{u}_{in}(t+\eta)$, 通过改变 η 的值实现对不同时域的预测.

回声状态网络的状态方程和输出方程为

$$\begin{aligned}\boldsymbol{x}(t)&=\tanh\left(W_{in}\boldsymbol{u}_{in}(t)+W_x\boldsymbol{x}(t-1)\right)\\ \boldsymbol{y}(t)&=W_{out}^{\rm T}\boldsymbol{x}(t)\end{aligned}\tag{1}$$

其中, $W_{in}\in\mathbf{R}^{p\times l}$ 为输入连接矩阵, 元素值在区间 $[-\gamma,\ \gamma]$ 内随机均匀分布[12], γ 称为输入变换系数. $W_x\in\mathbf{R}^{p\times p}$ 为储备池内部连接矩阵, 元素在 $[-1,\ 1]$ 之间随机赋值[12], 同时要求稀疏度在 1%～10% 之间. 为保证网络稳定, 设定 W_x 的谱半径小于 1[13]. $W_{out}\in\mathbf{R}^{p\times l}$ 为储备池输出连接矩阵, 是唯一由训练求解的参数. $\boldsymbol{x}(t)=[x_1(t),x_2(t),\cdots,x_p(t)]^{\rm T}$ 为 t 时刻 p 元储备池内部状态向量, 由当前输入 $\boldsymbol{u}_{in}(t)$ 和上一时刻状态 $\boldsymbol{x}(t-1)$ 激发产生. 为消除储备池状态随机初始化的影响, 从 $Init+1$ 时刻开始收集储备池状态 $X=[\boldsymbol{x}(Init+1),\cdots,\boldsymbol{x}(Init+n)]^{\rm T}\in\mathbf{R}^{n\times p}$, 对应地, $Y=[\boldsymbol{y}(Init+1),\cdots,\boldsymbol{y}(Init+n)]^{\rm T}\in\mathbf{R}^{n\times l}$. n 为训练数据长度, 一般大于储备池规模 p.

使用最小二乘法求解输出权值, 得到

$$W_{out}=X^{\dagger}Y=\left(X^{\rm T}X\right)^{-1}X^{\rm T}Y\tag{2}$$

其中, $X^{\dagger}$ 是 X 的伪逆.

若 X 的秩为 r, 对 X 进行奇异值分解:

$$X = U\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^{\mathrm{T}} = \sum_{i=1}^{p} \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^{\mathrm{T}} \tag{3}$$

其中, $U = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n] \in \mathbf{R}^{n\times n}$ 和 $V = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p] \in \mathbf{R}^{p\times p}$ 是酉阵, $\Sigma = \mathrm{diag}\{\sigma_1, \cdots, \sigma_r\} \in \mathbf{R}^{r\times r}$ 是对角阵. 奇异值 σ_i, $i = 1, 2, \cdots, r$ 为 $X^{\mathrm{T}}X$ 的特征值, 有如下关系: $\sigma_1 \geq \cdots \geq \sigma_r > 0$. 将式 (3) 代入式 (2) 得

$$W_{out} = X^{\dagger}Y = \sum_{i=1}^{r} \frac{\boldsymbol{v}_i \boldsymbol{u}_i^{\mathrm{T}}}{\sigma_i} Y \tag{4}$$

由上式可以看出, 输出权值幅值与奇异值大小成反比.

当数据中含有噪声时, 相近幅值的输入 $\boldsymbol{u}(t)$ 映射到高维储备池中, 会产生近似相关的状态 $\boldsymbol{x}(t)$, 即 X 的列近似相关, 导致 $X^{\mathrm{T}}X$ 的列向量或行向量近似共线, $X^{\mathrm{T}}X$ 接近奇异, 最小非零奇异值 σ_r 很小, 趋于零. 设实际观测的时间序列含有噪声 δY, 即 $\tilde{Y} = Y + \delta Y$, 相应地, $\hat{W}_{out} = W_{out} + \delta W_{out}$, 有

$$\frac{\|\delta W_{out}\|}{\|W_{out}\|} \leq cond_2(X)\frac{\|\delta Y\|}{\|Y\|} = \frac{\sigma_1}{\sigma_r}\frac{\|\delta Y\|}{\|Y\|} \tag{5}$$

可以看出, 由于矩阵 X 的最小非零奇异值 σ_r 接近于零, 故 σ_1/σ_r 较大, W_{out} 的计算值受扰动信号 δY 的影响较大, 泛化能力弱.

2　因子回声状态网络

为解决病态解问题, 本文提出因子回声状态网络模型, 对储备池进行因子分析操作, 提取储备池状态变量的公因子, 把高维状态变量转化为少数元变量, 去除近似、冗余及噪声成分, 得到一个 "良态" 的储备池状态矩阵. 在此基础上, 进行求解, 避免病态解问题. 模型结构如图 1 所示.

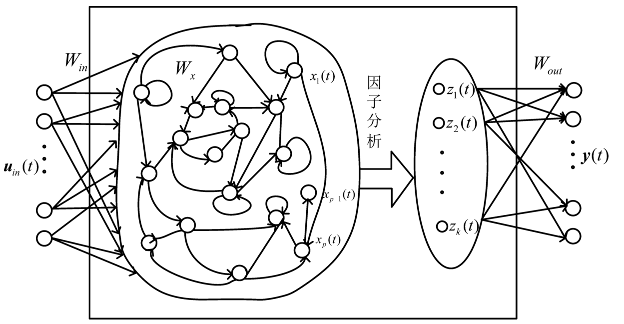

图 1　FESN 结构示意图

Fig. 1　The structure of FESN

2.1　储备池因子分析

采用因子分析方法, 将储备池状态向量 $\boldsymbol{x}(t) \in \mathbf{R}^{p\times 1}$ 用 k 维向量 $\boldsymbol{z}(t) \in \mathbf{R}^{k\times 1}$ 来近似, k 远小于 p, 一般描述为

$$\boldsymbol{x}(t) = \Lambda \boldsymbol{z}(t) + \boldsymbol{\varepsilon}(t) \tag{6}$$

其中, $\Lambda \in \mathbf{R}^{p\times k}$ 为因子载荷矩阵. k 个因子表示数据的 k 个投影方向. $\boldsymbol{z}(t)$ 服从正态分布, 即 $\boldsymbol{z}(t) \sim \mathrm{N}(\mathbf{0}, I)$, I 是 $k\times k$ 单位阵. 随机变量 $\boldsymbol{\varepsilon}(t) \in \mathbf{R}^{p\times 1}$ 表示每一个 $\boldsymbol{x}(t)$ 独立的噪声, 服从正态分布, 即 $\boldsymbol{\varepsilon}(t) \sim \mathrm{N}(\mathbf{0}, \Psi)$, $\Psi \in \mathbf{R}^{p\times p}$ 是对角阵. $\boldsymbol{x}(t)$ 服从均值为零、方差为 $\Lambda\Lambda^{\mathrm{T}} + \Psi$ 的正态分布. 若给定 Λ 和 Ψ, 因子的期望通过以下线性投影得到

$$\mathrm{E}\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right] = \beta \boldsymbol{x}(t) \tag{7}$$

其中, $\beta \equiv \Lambda^{\mathrm{T}}(\Psi + \Lambda\Lambda^{\mathrm{T}})^{-1} \in \mathbf{R}^{k\times p}$. 二次矩[18] 为

$$\begin{aligned} &\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right] = \\ &\quad Var\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right] + \mathrm{E}\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right]\mathrm{E}[\boldsymbol{z}(t)|\boldsymbol{x}(t)]^{\mathrm{T}} = \\ &\quad I - \beta\Lambda + \beta\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\beta^{\mathrm{T}} \end{aligned} \tag{8}$$

2.2　参数求解

因子分析中超参数 Λ 和 Ψ 采用 EM 算法[18, 21] 迭代估计. 对数似然函数为

$$\begin{aligned} Q = &\mathrm{E}\left\{\log\prod_{t=1}^{n}(2\pi)^{\frac{p}{2}}|\Psi|^{-\frac{1}{2}} \times \right.\\ &\left.\exp\left[-\frac{1}{2}(\boldsymbol{x}(t) - \Lambda\boldsymbol{z}(t))^{\mathrm{T}}\Psi^{-1}\left(\boldsymbol{x}(t) - \Lambda\boldsymbol{z}(t)\right)\right]\right\} = \\ &-\frac{n}{2}\log|\Psi| - \sum_{t=1}^{n}\left[\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}(t)\Psi^{-1}\boldsymbol{x}(t) - \right.\\ &\left.\boldsymbol{x}^{\mathrm{T}}(t)\Psi^{-1}\Lambda\mathrm{E}[\boldsymbol{z}(t)|\boldsymbol{x}(t)]\right] - \\ &\sum_{t=1}^{n}\frac{1}{2}\mathrm{tr}\left(\Lambda^{\mathrm{T}}\Psi^{-1}\Lambda\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right]\right) + c \end{aligned} \tag{9}$$

其中, c 是常数, 独立于参数, $\mathrm{tr}(\cdot)$ 表示矩阵的迹, $|\cdot|$ 表示行列式.

E 步:

给定矩阵 Λ 和 Ψ, 对于每一个数据点 $\boldsymbol{x}(t)$, 按照式 (7) 和式 (8) 计算 $\mathrm{E}\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right]$ 和 $\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right]$.

M 步:

将似然函数式 (9) 对因子载荷矩阵 Λ 求偏导:

$$\begin{aligned} \frac{\partial Q}{\partial \Lambda} = &-\sum_{t=1}^{n}\Psi^{-1}\boldsymbol{x}(t)\mathrm{E}[\boldsymbol{z}(t)|\boldsymbol{x}(t)]^{\mathrm{T}} + \\ &\sum_{t=1}^{n}\Psi^{-1}\Lambda^{new}\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right] = 0 \end{aligned}$$

化简得到

$$\Lambda^{new} = \sum_{t=1}^{n}\boldsymbol{x}(t)\mathrm{E}[\boldsymbol{z}(t)|\boldsymbol{x}(t)]^{\mathrm{T}}\left(\sum_{t=1}^{n}\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right]\right)^{-1}$$

将似然函数式 (9) 对 Ψ^{-1} 求偏导

$$\begin{aligned} \frac{\partial Q}{\partial \Psi^{-1}} = &\frac{n}{2}\Psi^{new} - \\ &\sum_{t=1}^{n}\left\{\frac{1}{2}\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t) - \Lambda^{new}\mathrm{E}\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right]\boldsymbol{x}^{\mathrm{T}}(t)\right\} - \\ &\frac{1}{2}\sum_{t=1}^{n}\Lambda^{new}\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right](\Lambda^{new})^{\mathrm{T}} = 0 \end{aligned}$$

化简得到

$$\Psi^{new} = \frac{1}{n}\mathrm{diag}\left\{\sum_{t=1}^{n}\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t) - \Lambda^{new}\mathrm{E}\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right]\boldsymbol{x}^{\mathrm{T}}(t)\right\}$$

EM 算法迭代估计参数 Ψ^{new} 和 Λ^{new}, 有 $Q(\Psi^{new},\Lambda^{new}) \geq Q(\Psi,\Lambda)$[22], 可保证收敛.

公因子个数 k 采用 HQ 准则[20] 确定:

$$HQ(k) = -2 \times h \times Q + [2p(k+1) - k(k-1)]\frac{\ln(\ln(n))}{n} \tag{10}$$

其中, Q 为式 (9) 所示的对数似然函数. 第 1 项是关于未知参数的对数似然项, 第 2 项是关于公因子个数的惩罚项, h 是平衡两项主次关系的常数. 该准则确定的阶数是真阶的相容估计.

2.3 算法分析与描述

从储备池状态矩阵 $X \in \mathbf{R}^{n\times p}$ 中提取出本征成分矩阵 $Z = [\boldsymbol{z}(1), \boldsymbol{z}(2), \cdots, \boldsymbol{z}(n)]^{\mathrm{T}} \in \mathbf{R}^{n\times k}$, $k \ll p$, 滤除了噪声干扰成分. 矩阵 Z 的奇异值分解表示如下:

$$Z = \widehat{U}\left[\begin{array}{c}\widehat{\Sigma}\\ 0\end{array}\right]\widehat{V}^{\mathrm{T}} = \sum_{i=1}^{k}\widehat{\boldsymbol{u}}_i\widehat{\sigma}_i\widehat{\boldsymbol{v}}_i^{\mathrm{T}} \tag{11}$$

其中, $\widehat{U} = [\widehat{\boldsymbol{u}}_1, \cdots, \widehat{\boldsymbol{u}}_n] \in \mathbf{R}^{n\times n}$ 和 $\widehat{V} = [\widehat{\boldsymbol{v}}_1, \cdots, \widehat{\boldsymbol{v}}_k] \in \mathbf{R}^{k\times k}$, $\widehat{\Sigma} = \mathrm{diag}\{\widehat{\sigma}_1, \cdots, \widehat{\sigma}_k\} \in \mathbf{R}^{k\times k}$, 有 $\widehat{\sigma}_1 \geq \cdots \geq \widehat{\sigma}_k > 0$. 进而, 输出权值计算公式为

$$\widehat{W}_{out} = Z^{\dagger}Y = (Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}Y = \sum_{i=1}^{k}\frac{\widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{u}}_i^{\mathrm{T}}}{\widehat{\sigma}_i}Y \tag{12}$$

当观测序列存在噪声 δY 时, 可以推导出

$$\frac{\left\|\delta\widehat{W}_{out}\right\|}{\left\|\widehat{W}_{out}\right\|} \leq cond_2(Z)\frac{\|\delta Y\|}{\|Y\|} = \frac{\widehat{\sigma}_1}{\widehat{\sigma}_k}\frac{\|\delta Y\|}{\|Y\|} \tag{13}$$

经因子分析处理后得到的矩阵 Z 去除了 X 中近似共线的成分, 为良态矩阵, 最小奇异值 $\widehat{\sigma}_k$ 明显大于零, 故 $\widehat{\sigma}_1\Big/\widehat{\sigma}_k$ 值相对 σ_1/σ_k 的计算值受扰动信号 δY 的影响较小, FESN 具有较好的泛化能力.

综上, 因子回声状态网络步骤如算法 1 所示.

算法 1.

1) 初始化储备池参数.
2) 初始化 Λ 和 Ψ.
3) 计算储备池状态 $\boldsymbol{x}(t)$, $t = 1, 2, \cdots, n$.
4) 进行因子分析
   For kk = 1 : p
   重复
   E 步: 计算 $\mathrm{E}\left[\boldsymbol{z}(t)|\boldsymbol{x}(t)\right]$ 和 $\mathrm{E}\left[\boldsymbol{z}(t)\boldsymbol{z}^{\mathrm{T}}(t)|\boldsymbol{x}(t)\right]$.
   M 步: 计算 Ψ^{new} 和 Λ^{new}.
   直到算法收敛 (Ψ^{new} 和 Λ^{new} 的变化量小于某一给定阈值或达到最大迭代次数).
   计算 HQ(kk).
   End for
   选择最小的 HQ(kk), 令 $k = kk$, 记录对应的 Ψ 和 Λ, 计算变换阵 β 和变换后的储备池状态变量 $\boldsymbol{z}(t)$, $t = 1, 2, \cdots, n$.
5) 按式 (12) 计算输出权值.
6) 采用测试集样本进行测试.

3 仿真实验

为验证本文所提模型的有效性, 将其应用于 Lorenz 混沌时间序列和大连月平均气温 – 降雨量序列的预测仿真中. 选取均方根误差 (Root mean square error, RMSE)、标准化均方根误差 (Normalized root mean square error, NRMSE)、平均对称绝对百分率误差 (Symmetric mean absolute percentage error, SMAPE) 和相关系数 (Correlation coefficient, CR) 作为衡量模型预测性能好坏的评价指标. 仿真结果为 50 次实验的平均值.

3.1 含噪声 Lorenz 序列

Lorenz 系统方程为

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(-x+y),\ \frac{\mathrm{d}y}{\mathrm{d}t} = bx - y - xz,\ \frac{\mathrm{d}z}{\mathrm{d}t} = xy - cz$$

当选取参数 $a = 10$, $b = 28$, $c = 8/3$ 时[20, 23], 系统具有混沌特性. 设定初始状态为 $x(0) = 12$, $y(0) = 2$, $z(0) = 9$, 取步长为 0.02, 选用四阶 Runge-Kutta 方法产生 3 000 组样本, 舍弃前 500 组初始样本, 采集后 2 500 组数据用于仿真实验. 本文将 $[x(t), y(t)]^{\mathrm{T}}$ 作为输入, $[x(t+1), y(t+1)]^{\mathrm{T}}$ 作为输出. 为说明本文方法的抗噪声干扰能力, 输入数据归一化后, 附加均值为零、方差为 10 % 的高斯白噪声. 前 100 组数据用于消除储备池网络初始暂态的影响, 剩余的前 80 % 数据用于训练, 后 20 % 数据用于测试. FESN 模型的储备池规模设置为 300, 谱半径为 0.98, 稀疏度为 0.05, 输入变换系数为 0.1.

图 2 给出了当 h 取 0.035 时不同维数 k 下的 HQ 值. 可以看出, 当维数为 54 时, HQ 达到局部极小值. 在此条件下, 计算 FESN 单步预测 Lorenz-$x(t)$ 的误差, 如表 1 所示. 为说明多元时间序列预测的有效性, 本文也做了仅采用 Lorenz $x(t)$ 序列预测 $x(t+1)$ 的 FESN 单元对比实验. 为说明因子分析方法提取储备池状态矩阵本征成分的有效性, 还做了一般 ESN[11] 模型在储备池规模为 300 维和 54 维条件下的仿真实验, 以及其他降维方法[17, 24] 与 ESN 结合的仿真实验. 同时, 将本文所提模型仿真结果与其他 ESN 改进模型 SVESM[15] 和 ESGP[16] 进行比较.

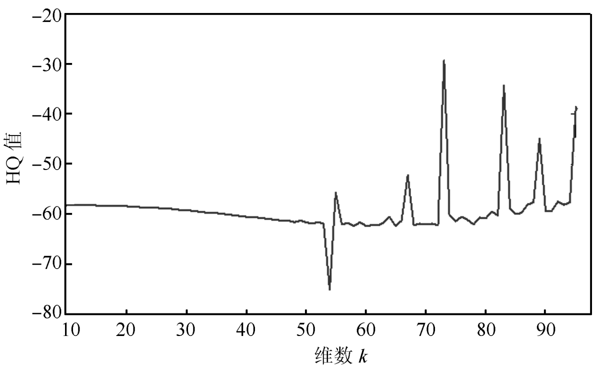

图 2 不同维数 k 下的 HQ 值 (Lorenz 序列)

Fig. 2 HQ values for different model orders k (Lorenz series)

从表 1 中可以看出, 当 ESN 的储备池规模设置为 300 时, 误差较大. 储备池状态矩阵最大奇异值为 88.7317, 最小非零奇异值为 4.8277×10^{-4}, 条件数为 1.8379×10^5, 病态特性严重. 而本文所提 FESN 模型经因子分析降维后, 储备池状态矩阵的最大奇异值为 54.7142, 最小非零奇异值为 0.0079, 条件数为 6.9258×10^3. 可见奇异值幅值分布集中, 条件数减小, 解决了 ESN 的病态解问题, 提高了预测精度. 如果一开始将 ESN 的储备池规模设置为降维后的维数 54, 则 ESN 无法映射混沌时间序列多个变量间复杂的非线性关系, 预测误差明显大于 ESN 改进模型.

表 1　Lorenz-$x(t)$ 序列仿真结果
Table 1　Prediction results for Lorenz-$x(t)$ time series

方法	RMSE	NRMSE	SMAPE (%)	CR
SVESM[15]	0.5696	0.0617	52.32	0.9981
ESGP[16]	0.6466	0.0855	95.09	0.9968
PCAESN[17]	0.5335	0.0684	37.81	0.9977
LDA[24]+ESN	0.6653	0.0853	41.02	0.9965
SNE[24]+ESN	0.6695	0.0858	47.33	0.9967
GPLVM[24]+ESN	0.5335	0.0684	37.81	0.9977
LLE[24]+ESN	0.5570	0.0714	93.21	0.9977
Isomap[24]+ESN	0.5563	0.0713	33.20	0.9975
LTSA[24]+ESN	0.5513	0.0707	44.62	0.9975
ESN[11](300)	4.1628	0.5467	89.81	0.9384
ESN[11](54)	0.8976	0.1179	96.94	0.9974
FESN(单元)	0.7791	0.0998	61.99	0.9950
FESN(多元)	0.3927	0.0503	20.86	0.9988

从表 1 中还可以看出, 大部分降维方法 +ESN 模型得到了较小的均方根误差和标准化均方根误差, 但本文提出的 FESN 模型误差最小. 因为 HQ 准则很好地确定了储备池的本征维数, 准确地构造出储备池状态矩阵的本征成分, 使得求解的输出权值幅值较小, 提高了泛化能力. 值得注意的是, 采用多变量作为输入的预测误差明显小于单变量输入的预测误差, 这与 Lorenz 系统方程是相符的, $x(t)$ 和 $y(t)$ 之间具有较强的耦合关系, 综合应用多个变量共同进行预测, 可以提高预测精度. 图 3 给出了 FESN 单步预测 Lorenz-$x(t)$ 的预测曲线及误差曲线. 从图 3 中也可以看出, 本文模型较精确地拟合了含噪声 Lorenz-$x(t)$ 序列.

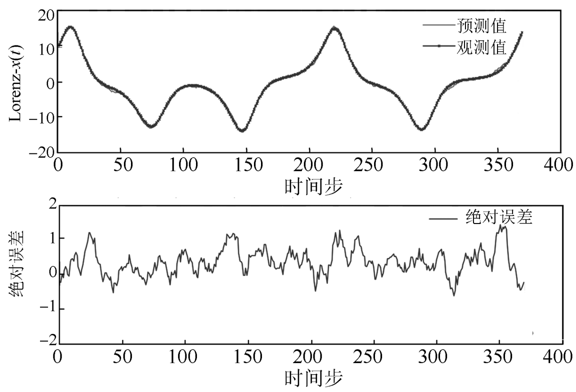

图 3　FESN 预测曲线及误差曲线 (Lorenz-$x(t)$)
Fig. 3　Prediction and error curves for Lorenz-$x(t)$ obtained by FESN

3.2　大连月平均气温 – 降雨量序列

大连月平均气温 – 降雨量序列是一组二元混沌时间序列. 本文采集 1951 年到 2001 年的大连月平均气温 – 降雨量共 612 个样本进行仿真实验, 50 个样本用于消除初始暂态的影响, 剩余样本的前 75 % 用于训练, 后 25 % 用于测试. 大连月平均气温 – 降雨量作为输入, 下一个月的大连月平均气温 – 降雨量作为输出. FESN 模型储备池规模设置为 250, 谱半径为 0.98, 稀疏度为 0.1, 输入变换系数为 0.5.

图 4 给出了当 h 取 0.5 时不同维数 k 下的 HQ 值. 当维数为 75 时, HQ 值达到局部极小值. 在此条件下, 计算 FESN 单步预测大连月平均气温的误差. 为验证本文所提模型的有效性, 也做了一般 ESN[11] 模型及其改进模型的仿真实验, 同时做了 FESN 单元预测的仿真实验, 相应的仿真结果如表 2 所示. 图 5 给出了 FESN 预测大连月平均气温的预测曲线及误差曲线.

表 2　大连月平均气温预测结果
Table 2　Prediction results for monthly average temperature series in Dalian

方法	RMSE	NRMSE	SMAPE (%)	CR
SVESM[15]	1.9829	0.2008	66.02	0.9864
ESGP[16]	1.8967	0.1882	57.00	0.9885
PCAESN[17]	1.6118	0.1607	54.30	0.9909
LDA[24]+ESN	1.6491	0.1665	76.08	0.9906
SNE[24]+ESN	1.6456	0.1661	57.06	0.9904
GPLVM[24]+ESN	1.7052	0.1721	59.34	0.9892
LLE[24]+ESN	1.7357	0.1752	100.9	0.9893
Isomap[24]+ESN	1.8555	0.1873	57.56	0.9906
LTSA[24]+ESN	2.1981	0.2219	62.16	0.9818
ESN[11](300)	2.7261	0.3073	163.8	0.9531
ESN[11](54)	1.8224	0.2054	118.4	0.9841
FESN(单元)	1.7893	0.1912	51.97	0.9830
FESN(多元)	1.4295	0.1443	42.87	0.9911

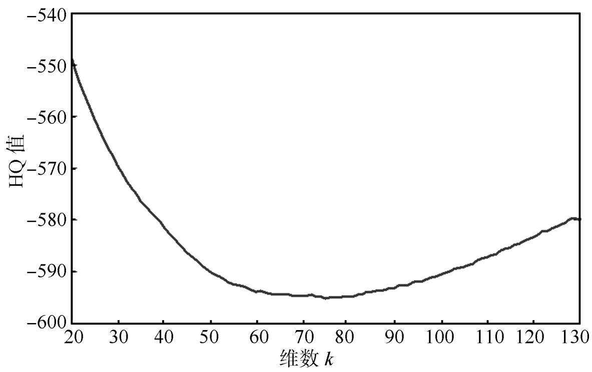

图 4　不同维数 k 下的 HQ 值 (大连月平均气温 – 降雨量序列)
Fig. 4　HQ values for different model orders k (Monthly average temperature-rainfall series in Dalian)

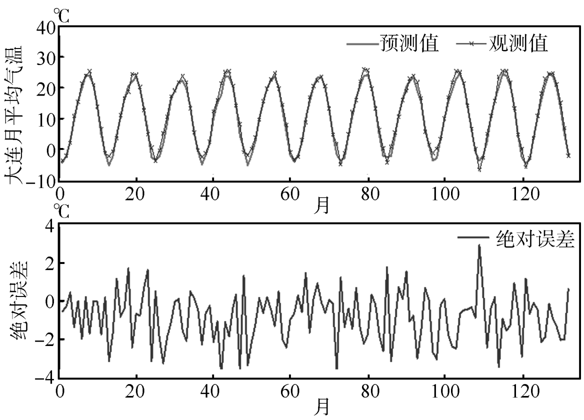

图 5　FESN 预测曲线及误差曲线 (大连月平均气温)
Fig. 5　Prediction and error curves for monthly average temperature series in Dalian obtained by FESN

从表 2 中可以看出, 本文所提 FESN 模型在测试集上的均方根误差较小, 具有明显的优势. 储备池规模为 250 时,

ESN 的条件数为 2.365×10^4, 病态特性严重, 预测误差很大. FESN 模型克服了这一问题, 条件数减小为 3.002×10^3. 从表中还可以看出, FESN 模型仅采用大连平均气温历史数据预测下一时刻的大连平均气温的效果远不如综合利用大连气温 – 降雨量数据预测的效果好, 验证了降雨量可以影响平均气温的事实.

4 结论

本文结合因子分析算法和回声状态网络模型, 提出了一种新型的多元混沌时间序列预测模型 — 因子回声状态网络. 采用因子分析方法降维回声状态网络的高维储备池, 得到高维储备池的本征成分, 解决病态解问题. 储备池本征维数通过 HQ 准则估计, 实现拟合精度和模型复杂度的平衡, 提高泛化能力. 基于 Lorenz 序列和大连月平均气温 – 降雨量序列的仿真实验, 验证了本文所提方法的有效性. 相对于其他降维方法和改进模型, 本文模型具有更高的预测精度.

References

1 Rong T Z, Xiao Z. Nonparametric interval prediction of chaotic time series and its application to climatic system. *International Journal of Systems Science*, 2013, **44**(9): 1726−1732

2 Han Min, Xu Mei-Ling, Ren Wei-Jie. Research on multivariate chaotic time series prediction using mRSM model. *Acta Automatica Sinica*, 2014, **40**(5): 822−829
(韩敏, 许美玲, 任伟杰. 多元混沌时间序列的相关状态机预测模型研究. 自动化学报, 2014, **40**(5): 822−829)

3 Inoussa G, Peng H, Wu J. Nonlinear time series modeling and prediction using functional weights wavelet neural network-based state-dependent AR model. *Neurocomputing*, 2012, **86**(1): 59−74

4 Li P H, Li Y G, Xiong Q Y, Chai Y, Zhang Y. Application of a hybrid quantized Elman neural network in short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, 2014, **55**: 749−759

5 Yeh W C. New parameter-free simplified swarm optimization for artificial neural network training and its application in the prediction of time series. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(4): 661−665

6 Xuan Zhao-Yan, Yang Gong-Xun. Application of EMD in the atmoshere time series prediction. *Acta Automatica Sinica*, 2008, **34**(1): 97−101
(玄兆燕, 杨公训. 经验模态分解法在大气时间序列预测中的应用. 自动化学报, 2008, **34**(1): 97−101)

7 Zeng Z G, Wang J. Improved conditions for global exponential stability of recurrent neural networks with time-varying delays. *IEEE Transactions on Neural Networks*, 2006, **17**(3): 623−635

8 Zhang H G, Liu J H, Ma D Z, Wang Z S. Data-core-based fuzzy min-max neural network for pattern classification. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2339−2352

9 Zhang H G, Liu D R, Luo Y H, Wang D. *Adaptive Dynamic Programming for Control: Algorithms and Stability*. London: Springer, 2013. 1−19

10 Jaeger H. The "Echo State" Approach to Analysing and Training Recurrent Neural Networks — with An Erratum Note, GMD Report 148, German National Research Center for Information Technology, Germany, 2001.

11 Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*, 2004, **304**(5667): 78−80

12 Massar M, Massar S. Mean-field theory of echo state networks. *Physical Review E*, 2013, **87**(4): 042809

13 Lukoševičius M, Jaeger H, Schrauwen B. Reservoir computing trends. *Kl-Künstliche Intelligenz*, 2012, **26**(4): 365−371

14 Qiao Jun-Fei, Bo Ying-Chun, Han Guang. Application of ESN-based multi indices dual heuristic dynamic programming on wastewater treatment process. *Acta Automatica Sinica*, 2013, **39**(7): 1146−1151
(乔俊飞, 薄迎春, 韩广. 基于 ESN 的多指标 DHP 控制策略在污水处理过程中的应用. 自动化学报, 2013, **39**(7): 1146−1151)

15 Shi Z W, Han M. Support vector echo-state machine for chaotic time-series prediction. *IEEE Transactions on Neural Networks*, 2007, **18**(2): 359−372

16 Chatzis S P, Demiris Y. Echo state Gaussian process. *IEEE Transactions on Neural Networks*, 2011, **22**(9): 1435−1445

17 Han Min, Wang Ya-Nan. Prediction of multivariate time series based on reservoir principal component analysis. *Control and Decision*, 2009, **24**(10): 1526−1530
(韩敏, 王亚楠. 基于储备池主成分分析的多元时间序列预测研究. 控制与决策, 2009, **24**(10): 1526−1530)

18 Ghahramani Z, Hinton G E. The EM Algorithm for Mixtures of Factor Analyzers, Technical Report CRG-TR-96-1, University of Toronto, Canada, 1996.

19 He Liang, Shi Yong-Zhe, Liu Jia. Eigenchannel space combination method of joint factor analysis. *Acta Automatica Sinica*, 2011, **37**(7): 849−856
(何亮, 史永哲, 刘加. 联合因子分析中的本征信道空间拼接方法. 自动化学报, 2011, **37**(7): 849−856)

20 Meng Qing-Fang, Peng Yu-Hua, Qu Huai-Jing, Han Min. The neighbor point selection method for local prediction based on information criterion. *Acta Physica Sinica*, 2008, **57**(3): 1423−1430
(孟庆芳, 彭玉华, 曲怀敬, 韩民. 基于信息准则的局域预测法邻近点的选取方法. 物理学报, 2008, **57**(3): 1423−1430)

21 Yang Xu-Kui, Qu Dan, Zhang Wen-Lin. An orthogonal Laplacian language recognition approach. *Acta Automatica Sinica*, 2014, **40**(8): 1812−1818
(杨绪魁, 屈丹, 张文林. 正交拉普拉斯语种识别方法. 自动化学报, 2014, **40**(8): 1812−1818)

22 Varadhan R, Roland C. Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scandinavian Journal of Statistics*, 2008, **35**(2): 335−353

23 Mirmomeni M, Lucas C, Araabi B N, Moshiri B, Bidar M R. Recursive spectral analysis of natural time series based on eigenvector matrix perturbation for online applications. *IET Signal Processing*, 2011, **5**(6): 515−526

24 van der Maaten L J P, Postm E O, van den Herik H J. Dimensionality Reduction: a Comparative Review. Technical Report TiCC-TR 2009-005, Tilburg University, Tilburg, The Netherlands, 2009.

许美玲 大连理工大学电子信息与电气工程学部博士研究生. 主要研究方向为神经网络和多元时间序列预测.
E-mail: xuml@mail.dlut.edu.cn
(**XU Mei-Ling** Ph. D. candidate at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her research interest covers neural networks and multivariate time series prediction.)

韩 敏 大连理工大学电子信息与电气工程学部教授. 主要研究方向为神经网络, 模式识别和混沌时间序列预测. 本文通信作者.
E-mail: minhan@dlut.edu.cn
(**HAN Min** Professor at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology now. Her research interest covers neural networks, pattern recognition, and chaotic time series prediction. Corresponding author of this paper.)