多元混沌时间序列的相关状态机预测模型研究

韩 敏 [1] 许美玲 [1] 任伟杰 [1]

摘 要 针对多元混沌时间序列预测存在的过拟合问题及高维输入变量冗余问题, 提出一种新型的多变量稀疏化预测模型 — 多元相关状态机. 该模型采用主成分分析方法对相空间重构后的高维输入变量进行低维表示, 将动态储备池作为相关向量机的核函数, 充分映射多元混沌时间序列的动力学特性, 使得模型具有丰富的动态机制和良好的稀疏性能, 有效避免过拟合问题, 提高预测精度. 基于两组多元混沌时间序列的仿真实验验证了模型的有效性.

关键词 多元, 混沌时间序列, 储备池, 主成分分析, 相关向量机

引用格式 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

DOI 10.3724/SP.J.1004.2014.00822

Research on Multivariate Chaotic Time Series Prediction Using mRSM Model

HAN Min[1] XU Mei-Ling[1] REN Wei-Jie[1]

Abstract Considering that there may be overfitting problem, as well as the problem of high dimensional redundant input variables in multivariate chaotic time series prediction, we introduce a novel multivariate prediction model based on relevance vector machine and echo state network, named multivariate relevance state machine (mRSM). The proposed model reconstructs the multivariate chaotic time series into the phase space, then reduces the dimension of input variables with the principal component analysis method. Subsequently, the mRSM uses a reservoir, replacing kernel functions of relevance vector machine, to map the dynamic features of multivariate time series sufficiently. Therefore, the mRSM presents rich dynamics and good sparsity. Furthermore, it avoids overfitting, and improves the predictive accuracy. Simulation results, based on two multivariate time series, substantiate the effectiveness of the mRSM.

Key words Multivariate, chaotic time series, reservoir, principal component analysis, relevance vector machine

Citation Hao 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

多元混沌时间序列是一种短期内可以预测, 长期不可预测的貌似无规则、类似随机的非线性系统, 广泛存在于气象、水文、医学、经济和电子等众多领域的复杂系统中[1−2]. 如何在机理分析的基础上, 利用数学方法, 从多元混沌时间序列中挖掘出数据间的关联关系[3], 预测未来时刻的演化规律, 是当前复杂系统建模的研究热点. 为充分映射混沌时间序列的混沌特性, 通常采用相空间重构方法重构系统的状态空间[4]. 此方法在单变量混沌时间序列预测上取得了较好的效果. 但是, 实际观测的多元混沌时间序列可能存在共同的元组分及噪声成分, 重构后的高维输入变量存在冗余信息, 为建模带来了挑战. 主成分分析 (Principal component analysis, PCA)[5] 是一种有效的线性降维法. 采用 PCA 方法处理多元混沌时间序列, 可以保留原始混沌时间序列的本质特征, 去掉包含较少信息的冗余成分和噪声成分, 达到维数约简的目的[6], 为预测模型构造合适的输入变量.

在多元混沌时间序列预测模型建立方面, 由于神经网络模型能够以任意精度逼近非线性函数, 只需较少的统计学知识即可获得理想的预测效果, 因而普遍应用于多元混沌时间序列预测建模中[7−9]. 但是一般神经网络利用梯度下降法反复学习网络权值, 收敛速度慢, 易陷入局部最优. 针对这一问题, Jaeger 等[10−11] 提出一种新型的递归神经网络 — 回声状态网络 (Echo state network, ESN), 核心部分是一个大规模储备池, 只有输出权值需要通过线性回归方法求解, 学习算法简单快速, 弥补了一般神经网络收敛速度慢的缺点. 储备池的递归结构使得网络具有良好的动态性能和非线性映射能力[12], 但在预测大规模多元混沌时间序列时, 高维输入映射到高维空间, 有可能存在近似共线的成分, 状态矩阵不满秩[13]. 采用最小二乘方法求解, 经验风险最小化,

收稿日期 2013-01-17 录用日期 2013-08-1
Manuscript received January 17, 2013; accepted August 1, 2013
国家重点基础研究发展计划 (973 计划)(2013CB430403), 国家自然科学基金 (61374154) 资助
Supported by National Basic Research Program of China (973 Program) (2013CB430403), National Natural Science Foundation of China (61374154)
本文责任编委 王宏
Recommended by Associate Editor WANG Hong
1. 大连理工大学电子信息与电气工程学部 大连 116024
1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024

易出现过拟合现象, 影响泛化性能. 为解决这一问题, 文献 [14] 引入正则项, 提出正则化回声状态网络模型 (Regularized echo state network, RESN), 在目标函数上增加结构风险项, 以权衡结构风险和经验风险[15], 从而使模型具有更好的泛化能力. 但正则项系数往往需要采用交叉检验方法确定, 增加了计算复杂度.

2001 年, Tipping 等[16] 提出一种基于概率的稀疏模型 — 相关向量机 (Relevance vector machine, RVM), 在贝叶斯框架下, 给参数加上一个条件概率分布的限制[17], 使得大部分参数趋于零, 其相应的核函数被剔除, 进而控制模型的复杂度, 应用于回归估计, 能够有效避免过学习现象的出现. 此外, 由于工程领域的要求, 实际复杂系统产生的多元混沌时间序列需要实现多步预测. 迭代预测法将单步预测进行多次迭代, 对模型复杂性要求较低, 但随着预测步长的增加, 预测误差呈指数增加, 且模型的稳定性难以保证. 文献 [18] 提出一种直接预测思想, 根据预测原点和预测时域的映射关系, 直接估计多步之后的观测值. 此种方法较为简洁, 但要求模型具有很强的非线性映射能力. 因此, 模型多步预测的精度是衡量模型预测性能好坏的重要标准.

为解决上述多元混沌时间序列预测过程中存在的问题, 本文提出一种适用于多元混沌时间序列的多步预测模型 — 多元相关状态机 (Multivariate relevance state machine, mRSM). 其基本思想是首先对相空间重构后的高维输入变量进行 PCA 降维, 将降维后的时间序列通过储备池映射到高维空间, 储备池代替相关向量机具体的核函数, 根据相关向量机的稀疏思想, 求解稀疏的储备池输出权值, 给出概率式的预测输出, 采用直接预测方式, 实现多元混沌时间序列的单步与多步预测.

1　储备池模型存在的过拟合问题

储备池模型是一种新型的递归神经网络, 典型代表是回声状态网络 ESN. 其结构如图 1 所示, 包含输入层、储备池和输出层三层, 核心部分储备池是一个大规模的稀疏连接的递归结构, 起到存储历史信息的 “回声” 作用, 成百上千个内部单元保证了网络强大的非线性映射能力, 递归连接保证了网络丰富的动态性能、良好的自适应性和自组织性[11].

对于多元混沌时间序列, 给定 n 时刻样本对 $[\boldsymbol{u}(n),\boldsymbol{y}(n)]$, 通常, $\boldsymbol{u}(n)$ 是多元输入向量, $\boldsymbol{y}(n)$ 是多元输出向量. 为不失一般性, 本文建立多输入单输出预测模型, $y(n)$ 为标量. 储备池模型状态方程为

$$\boldsymbol{x}(n) = \text{sigmoid}[W_{in}\boldsymbol{u}(n) + W_x\boldsymbol{x}(n-1)] \tag{1}$$

输出方程为

$$y(n) = \boldsymbol{x}(n)^{\mathrm{T}}\boldsymbol{w} \tag{2}$$

其中, $\boldsymbol{x}(n) \in \mathbf{R}^{B\times 1}$ 为 n 时刻储备池内部状态, 由当前输入 $\boldsymbol{u}(n)$ 和上一时刻状态 $\boldsymbol{x}(n-1)$ 激发产生. sigmoid 型函数一般选取 tanh 函数, W_{in} 和 W_x 分别为输入连接矩阵和储备池内部连接矩阵, 随机初始化后保持不变. 通常, 为保证网络丰富的动态性能, W_x 稀疏度设置在 [1 %, 5 %] 之间, 输入权值设置在 $[-\gamma_{in},\ \gamma_{in}]$ 之间, 其中, γ_{in} 为输入变换系数. 为保证网络稳定运行, 谱半径 $\rho(W_x)$ 设在 [0, 1] 之间. 输出权值向量 $\boldsymbol{w} \in \mathbf{R}^{B\times \mathbf{1}}$ 是唯一由训练求解的参数, 因此, 求解 ESN 模型转化为线性回归问题[19].

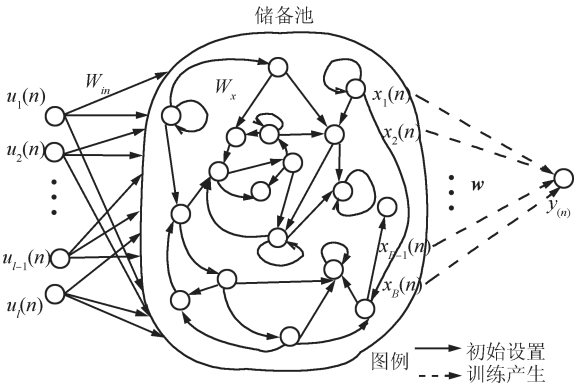

图 1　ESN 结构示意图

Fig. 1　The structure of ESN

将 $\boldsymbol{x}(n)$ 和 $y(n)$ 排列成矩阵, 得到 $X = [\boldsymbol{x}(Init+1),\boldsymbol{x}(Init+2),\cdots,\boldsymbol{x}(Init+N)]$, $\boldsymbol{Y} = [y(Init+1),y(Init+2),\cdots,y(Init+N)]^{\mathrm{T}}$, $X \in \mathbf{R}^{B\times N}$, $\boldsymbol{Y} \in \mathbf{R}^{N\times 1}$, 一般舍弃前 $Init$ 个样本产生的状态, 用于消除储备池初始暂态的影响[10], N 为实际的训练样本数. 以预测准确性为目标, 通过极小化经验风险来求解式 (2) 中的输出权值 $\boldsymbol{w}$:

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left\|\boldsymbol{Y} - X^{\mathrm{T}}\boldsymbol{w}\right\|_2^2 \tag{3}$$

伪逆解为

$$\hat{\boldsymbol{w}} = \left(X^{\mathrm{T}}\right)^{\dagger}\boldsymbol{Y} \tag{4}$$

设状态矩阵 X 的秩为 r, 约简奇异值分解表示如下:

$$X = U\Sigma V^{\mathrm{T}} = \sum_{i=1}^{r}\boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^{\mathrm{T}} \tag{5}$$

其中 $U = (\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r)$, $V = (\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r)$, $\Sigma = \text{diag}(\sigma_1,\cdots,\sigma_r)$ 是对角阵. 进而有:

$$\hat{\boldsymbol{w}} = (X^{\mathrm{T}})^{\dagger}\boldsymbol{Y} = \sum_{i=1}^{r}\frac{\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}}{\sigma_i}\boldsymbol{Y} \tag{6}$$

若实际观测的时间序列含有扰动 $\boldsymbol{\eta}$, 则输出为 $\boldsymbol{Y}+\boldsymbol{\eta}$, 输出权值计算公式为

$$\hat{\boldsymbol{w}} = (X^{\mathrm{T}})^{\dagger}(\boldsymbol{Y}+\boldsymbol{\eta}) = \sum_{i=1}^{r}\frac{\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}}{\sigma_i}\boldsymbol{Y} + \sum_{i=1}^{r}\frac{\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}}{\sigma_i}\boldsymbol{\eta} \quad (7)$$

当 X 的奇异值 σ_i 很小, 具有病态特性时, 式 (7) 求得的输出权值幅值较大, 扰动信号 $\boldsymbol{\eta}$ 严重影响 $\hat{\boldsymbol{w}}$ 的值, 网络的测试输入与训练输入即使具有微小的差别, 测试输出与训练输出也可能相差很大, 大大降低泛化性能.

2 多元相关状态机预测模型

为解决上述问题, 本文提出一种新型的多元混沌时间序列预测模型 — 多元相关状态机 mRSM, 预测多元混沌时间序列. 首先采用相空间重构法和 PCA 方法为预测模型构建合适的输入变量, 在此基础上, 将 RVM 与储备池方法结合, 提出一种具有丰富动态性能和良好非线性映射能力的多元相关状态机模型, 解决储备池模型的不适定问题.

2.1 多元混沌时间序列输入变量低维表示

多元混沌时间序列的输入变量表示为 $\boldsymbol{u}(n) = [u_1(n),\cdots,u_l(n)],\ n = Init+1, Init+2,\cdots, Init+N$, 其中 l 为输入变量维数. 为充分获得多元混沌时间序列的动力学特性, 使用关联维数法计算各变量的相空间重构参数, 设第 i 元变量的嵌入维数为 m_i, 延迟时间为 τ_i, $i=1,2,\cdots,l$, 那么重构后的高维输入向量为 $M=\sum_{i=1}^{l}m_i$ 维, 具体表示如下:

$$\begin{aligned}\tilde{\boldsymbol{u}}(n) =& [u_1(n), u_1(n-\tau_1),\cdots,u_1(n-(m_1-1)\tau_1),\\ &\cdots,\\ &u_l(n), u_l(n-\tau_l),\cdots,u_l(n-(m_l-1)\tau_l)]\end{aligned}$$

一般地, 高维输入变量要求模型具有较多的隐层节点数, 大大增加了模型的复杂度. 相互关联的含噪声多元混沌时间序列重构后, 往往具有相似的元变量或者元组成成分, 若直接将重构后的 $\sum_{i=1}^{l}m_i$ 维变量作为模型的输入, 存在冗余变量和噪声干扰, 易导致过学习现象的出现, 降低模型的泛化性能. 基于此, 本文采用主成分分析方法[6] 对高维输入变量进行低维模式表示.

首先对样本进行标准化处理, 计算高维输入 $\tilde{\boldsymbol{u}}$ 的协方差矩阵:

$$\Sigma_{\tilde{\boldsymbol{u}}} = \mathrm{E}[(\tilde{\boldsymbol{u}}-\mathrm{E}[\tilde{\boldsymbol{u}}])(\tilde{\boldsymbol{u}}-\mathrm{E}[\tilde{\boldsymbol{u}}])^{\mathrm{T}}] \quad (8)$$

其中, $\mathrm{E}[\cdot]$ 表示期望. 由于协方差矩阵 $\Sigma_{\tilde{\boldsymbol{u}}}$ 为是实对称矩阵, 所以, 存在正交阵 Φ, 使得

$$\begin{aligned}\Sigma_{\tilde{\boldsymbol{u}}} &= \Phi\Lambda\Phi^{\mathrm{T}}\\ \Lambda &= \mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_M)\\ \Phi &= (\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\cdots,\boldsymbol{\phi}_M)\end{aligned} \quad (9)$$

其中, $\boldsymbol{\phi}_i$, $i=1,2,\cdots,M$ 为 $\Sigma_{\tilde{\boldsymbol{u}}}$ 对应于特征值 λ_i 的特征向量, $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_M$. 由于实际观测的混沌时间序列往往含有噪声, 噪声成分对应的某些较小特征值 λ_i 接近于零, 但不为零. 因此, 设定阈值为 δ_0, 要求特征值的累计误差贡献率满足:

$$\delta = \sum_{i}^{m}\lambda_i \Big/ \sum_{i}^{M}\lambda_i \geq \delta_0 \quad (10)$$

选出前 m 个主要特征值, 滤除噪声作用的影响. 相应的, $\hat{\Phi}=(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\cdots,\boldsymbol{\phi}_m)$, 则新的输入变量为

$$\widehat{\boldsymbol{u}} = \hat{\Phi}\tilde{\boldsymbol{u}} \quad (11)$$

可以看出, 主成分分析将分散在一组多元时间序列 $\tilde{\boldsymbol{u}}$ 上的信息集中到几个主要成分 $\widehat{\boldsymbol{u}}$ 上, 去除共线、冗余及噪声成分, 为预测模型构建合适的输入变量.

2.2 多元相关状态机

本文将储备池和 RVM 结合, 利用储备池映射代替 RVM 具体的核映射, 根据 RVM 的稀疏化思想求解输出权值, 提出多元相关状态机 (mRSM) 模型. mRSM 将 PCA 降维后的输入变量映射到储备池空间, 避免 RVM 中核函数的计算, 减少计算时间. 同时, 模型具有储备池丰富的动态性能和强大的非线性映射能力, 可实现多元混沌时间序列的单步或多步直接预测. 详细描述如下.

将 PCA 降维后的 $\widehat{\boldsymbol{u}}(n)$ 映射到储备池中:

$$\boldsymbol{x}(n) = \tanh[W_{in}\widehat{\boldsymbol{u}}(n) + W_x\boldsymbol{x}(n-1)] \quad (12)$$

观测方程为

$$y(n) = \boldsymbol{x}(n)^{\mathrm{T}}\boldsymbol{w} + \eta(n) \quad (13)$$

其中, $\boldsymbol{w}\in\mathbf{R}^{B\times1}$, $\boldsymbol{x}(n)\in\mathbf{R}^{B\times1}$, $y(n)\in\mathbf{R}^{1\times1}$, 对于输入 $\widehat{\boldsymbol{u}}(n)$, 设定实际观测值为 $y_d(n+h)$, 用网络的预测输出 $y(n)$ 逼近 $y_d(n+h)$, 通过改变 h 的值实现不同时域的预测. $\eta(n)$ 是附加的独立同分布的高斯白噪声[20], 满足 $\eta(n)\sim\mathrm{N}(0,\sigma^2)$. 则有:

$$\begin{aligned}p[y(n)|\boldsymbol{x}(n),\boldsymbol{w}] = \mathrm{N}[y(n)|\boldsymbol{x}^{\mathrm{T}}(n)\boldsymbol{w},\sigma^2] =\\ \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{1}{2\sigma^2}[y(n)-\boldsymbol{x}^{\mathrm{T}}(n)\boldsymbol{w}]^2}\end{aligned} \quad (14)$$

对应 $X \in \mathbf{R}^{B\times N}$, $\boldsymbol{Y} \in \mathbf{R}^{N\times 1}$, 将 $\eta(n)$ 排列成矩阵 $\boldsymbol{\eta} = [\eta(Init+1), \cdots, \eta(Init+N)]^{\mathrm{T}}$, $\boldsymbol{\eta} \in \mathbf{R}^{N\times 1}$, 有:

$$\boldsymbol{Y} = X^{\mathrm{T}}\boldsymbol{w} + \boldsymbol{\eta}$$

似然函数为

$$p(\boldsymbol{Y}|X, \boldsymbol{w}) = \prod_{n=Init+1}^{Init+N} p(y(n)|\boldsymbol{w}, \boldsymbol{x}(n)) \tag{15}$$

假设 $\boldsymbol{w}$ 的每一个元素 w_i 服从高斯分布, 即 $w_i \sim \mathrm{N}(0, \alpha_i^{-1})$, $\boldsymbol{w} \sim \mathrm{N}(0, \alpha^{-1})$, 其中超参数 $\alpha = \mathrm{diag}(\alpha_1, \alpha_2, \cdots, \alpha_B)$, 则有先验分布:

$$p(\boldsymbol{w}|\alpha) = \prod_{i=1}^{B} \mathrm{N}(w_i|0, \alpha_i^{-1}) \tag{16}$$

由于 Gauss 正态分布方差倒数的共轭概率分布为 Gamma 分布, 因此设 $a_i \sim \mathrm{Gamma}(a_i|a, b)$, $\beta \sim \mathrm{Gamma}(\beta|c, d)$, 其中令 $\beta = \dfrac{1}{\sigma^2}$.

α 的先验分布表示为

$$\begin{aligned} &p(\alpha) = \prod_{i=1}^{B} \mathrm{Gamma}(\alpha_i|a, b) \\ &\mathrm{Gamma}(\alpha_i|a, b) = \Gamma(a)^{-1} b^a {\alpha_i}^{a-1} \mathrm{e}^{-b\alpha_i} \end{aligned} \tag{17}$$

由式 (16) 和式 (17) 得到权值 $\boldsymbol{w}$ 的先验分布为

$$\begin{aligned} &p(\boldsymbol{w}) = \prod_{i=1}^{B} p(w_i) = \\ &\prod_{i=1}^{B} \int_0^{\infty} \mathrm{N}(w_i|0, \alpha_i^{-1}) \mathrm{Gamma}(\alpha_i|a, b) \mathrm{d}\alpha_i \end{aligned} \tag{18}$$

其中, $\mathrm{N}(w_i|0, \alpha_i) = \sqrt{\dfrac{\alpha_i}{2\pi}} \mathrm{e}^{-\frac{\alpha_i}{2} w_i^2}$, 并且 $\int_0^{\infty} \mathrm{N}(w_i|0, \alpha_i^{-1}) \mathrm{Gamma}(\alpha_i|a, b) \mathrm{d}\alpha_i$ 服从 t 分布. 当选择 $a = b = 0$ 时, t 分布在 $w_i = 0$ 处附近获得峰值, 得到稀疏的权值 $\boldsymbol{w}$[16].

由后验概率 = 似然函数先验/边缘似然函数, 有:

$$p(\boldsymbol{w}|X, \boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|X, \boldsymbol{w}) p(\boldsymbol{w})}{p(\boldsymbol{Y}|X)} \tag{19}$$

式 (14)、式 (15) 和式 (18) 代入式 (19) 中得到后验概率:

$$\begin{aligned} &p(\boldsymbol{w}|X, \boldsymbol{Y}) \propto p(\boldsymbol{Y}|X, \boldsymbol{w}) p(\boldsymbol{w}) \propto \\ &\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y} - X^{\mathrm{T}}\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{Y} - X^{\mathrm{T}}\boldsymbol{w})\right] \times \\ &\exp\left(-\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\alpha\boldsymbol{w}\right) \propto \\ &\exp\left[-\frac{1}{2}(\boldsymbol{w} - \hat{\boldsymbol{w}})^{\mathrm{T}}\left(\frac{1}{\sigma^2}XX^{\mathrm{T}} + \alpha\right)(\boldsymbol{w} - \hat{\boldsymbol{w}})\right] \end{aligned} \tag{20}$$

其中, $\hat{\boldsymbol{w}} = \beta A^{-1}\boldsymbol{X}\boldsymbol{Y}$. 从式 (20) 可以导出 $\boldsymbol{w}$ 服从均值为 $\hat{\boldsymbol{w}}$, 方差为 A^{-1} 的高斯分布[21].

$$p(\boldsymbol{w}|X, \boldsymbol{Y}) \sim \mathrm{N}(\hat{\boldsymbol{w}}, A^{-1}) \tag{21}$$

其中, $A = \beta XX^{\mathrm{T}} + \alpha$, $\alpha = \mathrm{diag}(\alpha_1, \cdots, \alpha_B)$.

对于新的观测数据 $\boldsymbol{x}^*$, 预测输出为

$$y^* = (\boldsymbol{x}^*)^{\mathrm{T}}\hat{\boldsymbol{w}} = \beta(\boldsymbol{x}^*)^{\mathrm{T}}A^{-1}\boldsymbol{X}\boldsymbol{Y} \tag{22}$$

预测方差为

$$\begin{aligned} \mathrm{var}(y^*) = \frac{1}{\beta} + (\boldsymbol{x}^*)^{\mathrm{T}}A^{-1}\boldsymbol{x}^* = \\ \frac{1}{\beta} + (\boldsymbol{x}^*)^{\mathrm{T}}(\beta XX^{\mathrm{T}} + \alpha)^{-1}\boldsymbol{x}^* \end{aligned} \tag{23}$$

2.3　模型参数估计

本文将权值 $\boldsymbol{w}$ 当作隐变量, 借助 EM 算法[16], 通过积分将似然函数转化为 $\boldsymbol{w}$ 的隐式表达式, 进而估计式 (22) 和式 (23) 中的超参数 α 和 β.

E 步: 根据条件分布求期望 E.

ML-Ⅱ型边缘似然函数为

$$\begin{aligned} E = \log[p(\boldsymbol{Y}|X)] = \\ \int p(\boldsymbol{Y}|X, \boldsymbol{w}) p(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \end{aligned} \tag{24}$$

将式 (15) 和式 (18) 代入式 (24), 有:

$$\begin{aligned} &p(\boldsymbol{Y}|X, \boldsymbol{w}) p(\boldsymbol{w}) = \\ &p(\boldsymbol{Y}|X, \boldsymbol{w}, \beta) p(\boldsymbol{w}|\alpha) p(\alpha) p(\beta) = \\ &\prod_{n=Init+1}^{Init+N} p(y(n)|\boldsymbol{w}, \boldsymbol{x}(n)) \times \\ &\prod_{i=1}^{B} \int_0^{\infty} \mathrm{N}(w_i|0, \alpha_i^{-1}) \mathrm{Gamma}(\alpha_i|a, b) \mathrm{d}\alpha_i \end{aligned} \tag{25}$$

整理得:

$$E = \log[p(\boldsymbol{Y}|X)] = -\frac{1}{2}\left[\log\left|\beta^{-1}I + X^{\mathrm{T}}\alpha^{-1}X\right| + \boldsymbol{Y}^{\mathrm{T}}(\beta^{-1}I + X^{\mathrm{T}}\alpha^{-1}X)^{-1}\boldsymbol{Y}\right] + \sum_{i=0}^{N}(a\log\alpha_i - b\alpha_i) + c\log\beta - d\beta \tag{26}$$

M 步: 将 $E = \log[p(\boldsymbol{Y}|X)]$ 极大化, 估计 α 和 β.

$$\frac{\partial E}{\partial \log\alpha_i} = \frac{1}{2}[1 - \alpha_i(\mu_i^2 + [A^{-1}]_{ii})] + a - b\alpha_i \tag{27}$$

$$\frac{\partial E}{\partial \log\beta} = \frac{1}{2}\left[\frac{N}{\beta} - \left\|\boldsymbol{Y} - X^{\mathrm{T}}\boldsymbol{w}\right\|^2 - \beta^{-1}\sum_{i=1}^{B}\gamma_i\right] + c - d\beta \tag{28}$$

解得:

$$\frac{1}{\alpha_i^{\mathrm{new}}} = \frac{w_i^2 + 2b}{\gamma_i + 2a} \tag{29}$$

$$\frac{1}{\beta^{\mathrm{new}}} = \frac{\|\boldsymbol{Y} - X^{\mathrm{T}}\boldsymbol{w}\| + 2d}{N - \sum_{i=1}^{B}\gamma_i + 2c} \tag{30}$$

式中, $\gamma_i \triangleq 1 - \alpha_i[A^{-1}]_{ii}$. 令 $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_B]$, 对 $\boldsymbol{\gamma}$, α 和 β 交替更新, 直至算法收敛. 由式 (29) 和式 (30), 可以看出, 当取 $a = b = c = d = 0$ 时, 可促进权值 $\boldsymbol{w}$ 的稀疏[16].

由上可知, 本文所提 mRSM 模型能够实现参数自适应估计, 避免交叉检验方法确定模型参数带来的计算复杂度高的问题. 并且, 稀疏的权值进一步保证了模型具有较少的计算量. 因此, mRSM 模型可适用于数据量比较多的多元混沌时间序列预测问题.

3 仿真实例

为验证本文所提模型的有效性, 将其应用于含噪声 Lorenz 时间序列和大连气温-降雨量序列的预测仿真中. 同时, 在相同的数据集上, 仿真结果与 ESN[10], RESN[14], RVM[16], SVESM[22] 和 PCAESN[23] 进行比较. ESN 及其改进模型储备池参数设置与本文所提模型储备池参数设置一致. 采用均方根误差 (Root mean square error, RMSE) 和对称平均绝对百分率误差 (Symmetric mean absolute percentage error, SMAPE) 作为预测性能的评价指标.

均方根误差反应模型预测值对观测值的平均偏离程度. 计算公式如下:

$$\mathrm{RMSE} = \sqrt{\sum_{n=1}^{S}[\hat{y}(n) - y(n)]^2/(S-1)} \tag{31}$$

其中, $\hat{y}(n)$ 为模型预测值, $y(n)$ 为观测值, S 为样本个数.

对称平均绝对百分率误差衡量时间序列趋势拟合的准确度. 计算公式如下:

$$\mathrm{SMAPE} = \frac{1}{S}\sum_{n=1}^{S}\left|\frac{y(n) - \hat{y}(n)}{y(n) + \hat{y}(n)}\right| \tag{32}$$

3.1 含噪声 Lorenz 时间序列

Lorenz 系统是美国气象学家 Lorenz 模拟大气环流模型建立的三元一阶常微分方程组, 表达式如下:

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = a(-x + y) \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = bx - y - xz \\ \dfrac{\mathrm{d}z}{\mathrm{d}t} = xy - cz \end{cases} \tag{33}$$

当 $a = 10$, $b = 28$, $c = 8/3$, $x(0) = 12$, $y(0) = 2$, $z(0) = 9$ 时, 系统具有混沌特性. 取步长 0.02, 选用四阶 Runge-Kutta 方法产生 2 500 个 Lorenz 序列样本点. 本文使用前 75 % 样本训练, 后 25 % 样本测试, 比较不同模型对 $x(t)$ 的预测效果. 考虑到网络初始暂态的影响[10], 舍弃前 100 个训练样本产生的状态. 为了测试本文所提模型的抗噪声干扰能力, 训练数据中添加噪声水平为 5% 的高斯白噪声. $x(t)$, $y(t)$, $z(t)$ 的相空间重构参数为 $m_x = 3$, $\tau_x = 4$, $m_y = 5$, $\tau_y = 4$, $m_z = 7$, $\tau_z = 1$. 在 ESN 及其改进模型中, 储备池参数设置如表 1 所示. 相关向量机的核函数通过交叉检验方法确定为高斯型函数, 核宽为 25. 相空间重构后, PCA 方法将重构后的 15 维变量降至 10 维, 其中 PCA 方法累积误差贡献率选为 99.5 %.

表 1 储备池参数设置 (Lorenz 序列)

Table 1 Configuration of the employed reservoirs in our experiment (Lorenz series)

储备池参数	取值
储备池规模	100
稀疏度	0.05
谱半径	0.98
输入变换系数	0.01

表 2 给出了不同模型对 Lorenz-$x(t)$ 单步预测的测试误差 RMSE 和 SMAPE. 可以看出, mRSM 的两种预测误差均最小. 同时使用 $[x(t), y(t), z(t)]$ 预测 $x(t+1)$ 的测试均方根误差为 0.0055, 远远好于仅使用单变量 $x(t)$ 时的测试均方根误差 0.0257, 这与 Lorenz 系统方程式 (33) 是相符的, Lorenz 序列的三个成分 $[x(t), y(t), z(t)]$ 含有较强的耦合关系, 多元成分相对于一元成分, 包含了系统更多的有用信息, 本实验验证了多元混沌时间序列预测相对于一元混沌时间序列预测的优越性. 此外, 由于储备池状态矩阵含有冗余状态, 使得 ESN 求解过程中存在病态解问题, 测试误差较大, 出现过拟合现象. 而 ESN 改进模型及 RVM 几乎没有出现过拟合现象. 另一方面, 从表 2 还可以看出, mRSM 的相关向量个数少于 RVM 的相关向量个数, 说明 mRSM 具有更为稀疏的结构.

表 2　Lorenz-$x(t)$ 单步预测结果

Table 2　Prediction results of Lorenz-$x(t)$ obtained by the evaluated models for one-step ahead

模型	多元/单元	测试 RMSE	测试 SMAPE	相关向量
RSM	单元	0.0257	0.0065	15
mRSM	多元	0.0055	0.0011	23
RVM	多元	0.0075	0.0029	32
SVESM	多元	0.0129	0.0031	–
PCAESN	多元	0.0165	0.0064	–
RESN	多元	0.0084	0.0049	–
ESN	多元	0.1495	0.0603	–

图 2 给出了本文所提模型对 Lorenz-$x(t)$ 测试数据的单步预测曲线和误差曲线. 从图中可以看出, 本文模型能够很好的拟合 Lorenz-$x(t)$ 曲线.

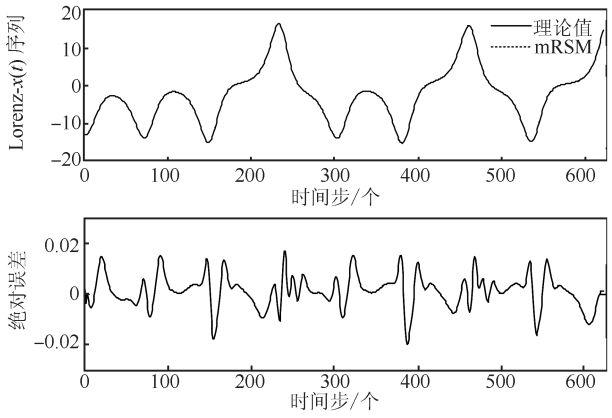

图 2　mRSM 单步预测 Lorenz-$x(t)$ 效果图

Fig. 2　Prediction curves of Lorenz-$x(t)$ obtained by mRSM model for one-step ahead

为进一步验证本文所提模型的有效性, 将其应用于 Lorenz 序列 10 步预测, 并与其它模型进行比较. 模型参数设置同 Loren-$x(t)$ 单步预测. 仿真结果如表 3 所示. 可以看出, 本文所提模型 mRSM 具有较小的 RMSE 和 SMAPE, 虽然 PCAESN 也具有较小的 RMSE, 但是其 SMAPE 较大, 说明 mRSM 能够更好的预测多元混沌时间序列未来时刻的变化趋势, 具有更强的非线性映射能力.

表 3　Lorenz-$x(t)$10 步预测结果

Table 3　Prediction results of Lorenz-$x(t)$ obtained by the evaluated models for ten-step ahead

模型	多元/单元	测试 RMSE	测试 SMAPE	相关向量
RSM	单元	0.0627	0.0108	18
mRSM	多元	0.0245	0.0053	26
RVM	多元	0.0380	0.0079	36
SVESM	多元	0.0437	0.0138	–
PCAESN	多元	0.0313	0.0161	–
RESN	多元	0.0393	0.0127	–
ESN	多元	0.5104	0.0812	–

3.2　大连月平均气温-降雨量时间序列

大连月平均气温-降雨量数据是一组二元混沌时间序列, 本文选用 1951 年到 2001 年大连市月平均气温和月降雨量共 612 个样本点进行仿真实验, 412 组数据用于训练, 200 组数据用于测试, 其中训练样本中舍弃前 12 个样本产生的状态以消除初始暂态的影响. 以大连市月平均气温-降雨量作为输入, 单步或多步预测大连市月平均气温. 大连气温和降雨量相空间重构参数为 $m_1 = 12$, $\tau_1 = 1$, $m_2 = 12$, $\tau_2 = 1$. 在 ESN 改进模型中, 储备池参数设置如表 4 所示. 相关向量机的核函数通过交叉检验方法确定为 r 型函数, 核宽 $width$ 为 650, 表达式为 $K(\boldsymbol{x}, \boldsymbol{y}) = {}^{1}\!/\!_{(width)} \times \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$. PCA 方法将相空间重构后的 24 维变量降低至 19 维, 其中 PCA 方法累积误差贡献率设为 99.0 %.

表 4　储备池参数设置 (大连月平均气温-降雨量)

Table 4　Configuration of the employed reservoirs in our experiment (The average monthly temperature-rainfall in Dalian)

储备池参数	取值
储备池规模	100
稀疏度	0.05
谱半径	0.98
输入变换系数	1

不同模型单步预测误差如表 5 所示. 从表中可以看出, 对于实际观测的含噪声的大连月平均气温-降雨量序列, RVM 的测试误差很大. 分析原因为 RVM 是一种静态模型, 无法模拟大连市月平均气温-降雨量的较强的动力学特性. ESN 具有较大的测试误差, 发生过拟合. 而 mRSM 弥补了 RVM 和 ESN 的缺点, 预测精度高于其他改进模型. 值得注意的是, mRSM 多元预测精度略高于 RSM 预测精度, 多元的优势不如 Lorenz 序列明显. 究其原因在于虽然降雨量对气温有隐式的影响, 但是影响程度不像 Lorenz 系统 $x(t)$, $y(t)$, $z(t)$ 之间的影响程度高. 大连市月平均气温测试数据预测曲线及误差曲线如图 3 所示, mRSM 模型能够有效的单步预测实际观测的大连月平均气温序列.

表 5 大连市月平均气温单步预测结果

Table 5 Prediction results of the average monthly temperature in Dalian obtained by the evaluated models for one-step ahead

模型	多元/单元	测试 RMSE	测试 SMAPE	相关向量
RSM	单元	1.6934	0.3319	21
mRSM	多元	1.5194	0.2352	33
RVM	多元	2.5669	0.6982	37
SVESM	多元	1.5915	0.3296	–
PCAESN	多元	2.1021	0.4155	–
RESN	多元	1.5938	0.4237	–
ESN	多元	3.4389	0.7598	–

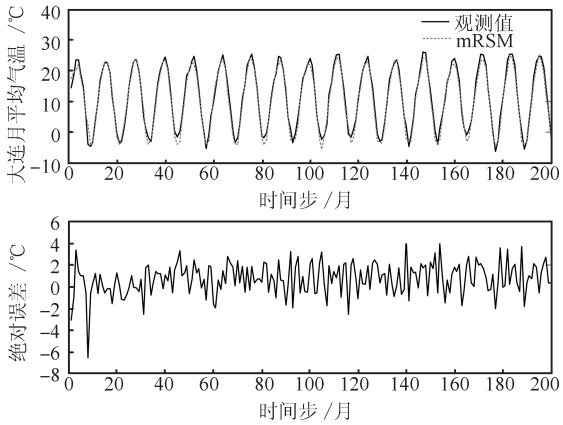

图 3 mRSM 单步预测大连月平均气温效果图

Fig. 3 Prediction curves of the average monthly temperature in Dalian obtained by mRSM model for one-step ahead

对于大连气温-降雨量时间序列, 实际生产中往往需要实现中长期预测, 气候出现异常时, 便于相关部门提前做好应对措施. 本文以大连月平均气温-降雨量作为输入, 预测 6 个月后的大连月平均气温. 相关向量机的核函数通过交叉检验方法确定为 tps 型, 核宽 $width$ 为 2, 表达式为 $K(x,y) = 0.5\times k\times\log(k+(k==0))$. 其中, $k = 1/(width)^2\times \|\boldsymbol{x}-\boldsymbol{y}\|_2^2$. 其他模型参数设置与大连市月平均气温单步预测相同. 不同模型的预测结果比较如表 6 所示. 可以看出, mRSM 精度仍然高于其他模型. SVESM 与 mRSM 在 RMSE 上, 相差不多, 但在 SMAPE 上, mRSM 明显优于 SVESM. 仿真结果表明, 本文所提模型适用于多元混沌时间序列的中长期预测.

表 6 大连月平均气温 6 步预测结果

Table 6 Prediction results of the average monthly temperature in Dalian obtained by the evaluated models for six-step ahead

模型	多元/单元	测试 RMSE	测试 SMAPE	相关向量
RSM	单元	1.8963	0.5174	20
mRSM	多元	1.7212	0.2752	28
RVM	多元	2.6073	0.7568	31
SVESM	多元	1.9112	0.5035	–
PCAESN	多元	2.4135	0.6015	–
RESN	多元	1.9181	0.7707	–
ESN	多元	3.8320	1.2232	–

4 结论

本文提出了一种多元相关状态机预测模型, 进行多元混沌时间序列单步与多步预测. 多元相关状态机为预测模型构建了合适的低维输入变量, 充分刻画了多元混沌时间序列的动力学特性, 基于贝叶斯的概率式优化算法有效地平衡了经验风险和结构风险, 实现了模型的稀疏表达, 解决了病态解问题, 避免了过拟合现象的出现, 提高了预测精度. 通过对 Lorenz 序列和大连月平均气温-降雨量序列的仿真实验, 证明了本文所提模型的优越性.

References

1 Gao Chuan-Hou, Jian Ling, Chen Ji-Ming, Sun You-Xian. Data-driven modeling and predictive algorithm for complex blast furnace ironmaking process. *Acta Automatica Sinica*, 2009, **35**(6): 725−730
(郜传厚, 渐令, 陈积明, 孙优贤. 复杂高炉炼铁过程的数据驱动建模及预测算法. 自动化学报, 2009, **35**(6): 725−730)

2 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

3 Han Min, Liu Xiao-Xin. Stepwise input variable selection based on mutual information for multivariate forecasting. *Acta Automatica Sinica*, 2012, **38**(6): 1000−1005
(韩敏, 刘晓欣. 基于互信息的分步式输入变量选择多元序列预测研究. 自动化学报, 2012, **38**(6): 1000−1005)

4 Shen M, Chen W N, Zhang J, Chung S H, Kaynak O. Optimal selection of parameters for nonuniform embedding of chaotic time series using ant colony optimization. *IEEE Transactions on Cybernetics*, 2013, **43**(2): 790−802

5 Ma Z M. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 2013, **41**(2): 772−801

6 Abdi H, Williams L J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, **2**(4): 433−459

7 Deng Wan-Yu, Zheng Qing-Hua, Chen Lin, Xu Xue-Bin. Research on extreme learning of neural networks. *Chinese Journal of Computers*, 2010, **33**(2): 279−287
(邓万宇, 郑庆华, 陈琳, 许学斌. 神经网络极速学习方法研究. 计算机学报, 2010, **33**(2): 279−287)

8 Buonomano D V. Harnessing chaos in recurrent neural networks. *Neuron*, 2009, **63**(4): 423−425

9 Li Peng-Hua, Chai Yi, Xiong Qing-Yu. Quantum gate Elman neural network and its quantized extended gradient backpropagation training algorithm. *Acta Automatica Sinica*, 2013, **39**(9): 1511−1522
(李鹏华, 柴毅, 熊庆宇. 量子门 Elman 神经网络及其梯度扩展的量子反向传播学习算法. 自动化学报, 2013, **39**(9): 1511−1522)

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

11 Yildiz I B, Jaeger H, Kiebel S J. Re-visiting the echo state property. *Neural Networks*, 2012, **35**: 1−9

12 Boccato L, Lopes A, Attux R, Von Zuben F J. An extended echo state network using Volterra filtering and principal component analysis. *Neural Networks*, 2012, **32**: 292−302

13 Wang Y G, Cao F L, Yuan Y B. A study on effectiveness of extreme learning machine. *Neurocomputing*, 2011, **74**(16): 2483−2490

14 Dutoit X, Schrauwen B, Van Campenhout J, Stroobandt D, Van Brussel H, Nuttin M. Pruning and regularization in reservoir computing. *Neurocomputing*, 2009, **72**(7−9): 1534−1546

15 Xu Z B, Guo H L, Wang Y, Zhang H. Representative of $L_{1/2}$ regularization among $L_q(0 < q \leq 1)$ regularizations: an experimental study based on phase diagram. *Acta Automatica Sinica*, 2012, **38**(7): 1225−1228

16 Tipping M E. Sparse Bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 2001, **1**: 211−244

17 Xu Dan-Lei, Du Lan, Liu Hong-Wei, Hong Ling, Li Yan-Bing. Joint feature selection and classification design based on variational relevance vector machine. *Acta Automatica Sinica*, 2011, **37**(8): 932−943
(徐丹蕾, 杜兰, 刘宏伟, 洪灵, 李彦兵. 一种基于变分相关向量机的特征选择和分类结合方法. 自动化学报, 2011, **37**(8): 932−943)

18 Han Min, Shi Zhi-Wei, Guo Wei. Reservoir neural state reconstruction and chaotic time series prediction. *Acta Physica Sinica*, 2007, **56**(1): 43−50
(韩敏, 史志伟, 郭伟. 储备池状态空间重构与混沌时间序列预测. 物理学报, 2007, **56**(1): 43−50)

19 Peng Yu, Wang Jian-Min, Peng Xi-Yuan. Survey on reservoir computing. *Acta Electronica Sinica*, 2011, **39**(10): 2387−2396
(彭宇, 王建民, 彭喜元. 储备池计算概述. 电子学报, 2011, **39**(10): 2387−2396)

20 Girard A, Rasmussen C E, Candela J Q, Murray-Smith R. Gaussian Process priors with uncertain inputs-Application to multiple-step ahead time series forecasting. In: Proceedings of Advances in Neural Information Processing Systems. Cambridge, Massachusetts, USA: MIT Press 2003. 529−536

21 Nakajima S, Sugiyama M. Theoretical analysis of Bayesian matrix factorization. *The Journal of Machine Learning Research*, 2011, **12**: 2583−2648

22 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

23 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)

韩 敏 大连理工大学电子信息与电气工程学部教授. 1999 年获得日本国立九州大学电气电子系工学博士学位. 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 Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. She received her Ph. D. degree in Kyushu University, Japan in 1999. Her research interest covers neural networks, pattern recognition and chaotic time series prediction. Corresponding author of this paper.)

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. 2011 年获得大连理工大学工学学士学位. 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 in the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. She received her B. S. degree from Dalian University Technology, China in 2011. Her research interest covers neural networks and multivariate time series prediction.)

任伟杰 大连理工大学电子信息与电气工程学部硕士研究生. 2012 年获得大连理工大学工学学士学位. 主要研究方向为神经网络和多元时间序列预测.
E-mail: renweijie@mail.dlut.edu.cn
(**REN Wei-Jie** Master student at Dalian University of Technology. He received his B. S. degree from Dalian University Technology, China in 2012. His current research interest covers neural networks and multivariate time series prediction.)