

基于极端学习机的多变量混沌时间序列预测*

王新迎 韩敏†

(大连理工大学电子信息与电气工程学部, 大连 116023)

(2011年7月17日收到; 2011年9月16日收到修改稿)

针对多变量混沌时间序列预测问题, 提出了一种基于输入变量选择和极端学习机的预测模型. 其基本思想是对多变量混沌时间序列进行相空间重构后, 采用互信息方法选择与预测输出统计相关最高的重构输入变量, 借助极端学习机的通用逼近能力建立多变量混沌时间序列的预测模型. 为进一步提高预测精度, 采用模型选择算法选择具有最小期望风险的极端学习机预测模型. 基于 Lorenz, Rössler 多变量混沌时间序列及 Rössler 超混沌时间序列的仿真结果证明所提方法的有效性.

关键词: 混沌时间序列预测, 输入变量选择, 极端学习机, 模型选择

PACS: 05.45.Tp

1 引言

非线性混沌时间序列预测是当今科学研究领域和工程应用领域中的一个研究热点, 并已在信号处理、自动控制等方面取得了广泛的应用^[1]. 虽然根据嵌入延时理论仅需单变量时间序列重构相空间即可描述系统动态, 但实际复杂系统往往由多个变量描述, 采用多变量时间序列建模可包含更多的系统动态信息, 并可得到更高的预测精度^[2]. 因此, 多变量混沌时间序列预测成为目前混沌时间序列预测的一个重要研究方向^[3]. 在混沌时间序列预测领域, 具有强大的非线性映射能力的神经网络已成为混沌时间序列预测的一种有效方法^[4-6]. 但是, 传统神经网络基于梯度下降的学习算法收敛速度慢, 易陷入局部最优, 这些缺陷制约了其预测精度, 限制了其更广泛的应用.

极端学习机^[7]是最近几年发展起来的一种有效的前馈神经网络学习方法, 网络输入权值和隐含层偏置值随机生成, 并在训练过程中保持不变, 仅需训练网络的输出权值, 简单的线性回归算法即可胜任训练任务. 极端学习机克服了传统神经网络的缺陷. 本文借助极端学习机的通用逼近能力^[7]拟合重构相空间中的多变量时间序列的输入、输出数据, 构建极端学习机多变量混沌时间序列预测模型.

同时考虑到多变量混沌序列相空间重构变量存在冗余或不相关变量问题, 本文首先对重构变量基于互信息方法进行输入变量选择, 保留和预测输出具有较大统计相关性的重构变量, 以减小预测模型输入, 提高预测模型精度.

为进一步提高极端学习机多变量混沌时间序列预测模型的预测精度并有效解决极端学习机的隐含层节点选择问题, 本文引入一种基于期望风险最小化的模型选择算法^[8], 构建具有最优预测精度的极端学习机多变量混沌时间序列预测模型.

2 相空间重构和输入变量选择

假设 M 维的多变量时间序列 $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)$, 其中

$$\mathbf{x}(i) = (\mathbf{x}_1(i), \mathbf{x}_2(i), \dots, \mathbf{x}_M(i)) \quad (i = 1, 2, \dots, N).$$

对该多变量时间序列进行相空间重构, 有

$$\begin{aligned} D(n) = & \{\mathbf{x}_1(n), \mathbf{x}_1(n - \tau_1), \dots, \\ & \mathbf{x}_1(n - (d_1 - 1)\tau_1), \\ & \mathbf{x}_2(n), \mathbf{x}_2(n - \tau_2), \dots, \\ & \mathbf{x}_2(n - (d_2 - 1)\tau_2), \dots, \\ & \mathbf{x}_M(n), \mathbf{x}_M(n - \tau_M), \dots, \end{aligned}$$

* 国家自然科学基金 (批准号: 61074096) 资助的课题.

† E-mail: minhan@dlut.edu.cn

$$\mathbf{x}_M(n - (d_M - 1)\tau_M)\}, \quad (1)$$

其中 τ_i 和 d_i ($i = 1, \dots, M$) 分别是对应时间序列的延迟时间和嵌入维数. 根据 Takens 的嵌入延时定理 [9], 如果 d 或 d_i 足够大, 存在一个映射 $F: \mathbb{R}^d \rightarrow \mathbb{R}^d$ ($d = \sum_{i=1}^M d_i$), 使得

$$\mathbf{D}(n+1) = F(\mathbf{D}(n)). \quad (2)$$

(2) 式也可写为

$$\begin{aligned} \mathbf{x}_1(n+1) &= F_1(\mathbf{D}(n)), \\ \mathbf{x}_2(n+1) &= F_2(\mathbf{D}(n)), \\ &\vdots \\ \mathbf{x}_M(n+1) &= F_M(\mathbf{D}(n)). \end{aligned} \quad (3)$$

由 (3) 式可知, 当选择好合适的嵌入维数 d_i 和延迟时间 τ_i ($i = 1, \dots, M$) 时, 可用重构相空间内的多变量数据进行建模预测. 有许多不同的方法 (如信息熵方法 [10] 等) 可以选择 d 和 τ .

虽然多变量时间序列相对单变量时间序列包含更多的系统动态信息, 但对多变量时间序列进行相空间重构预测时, 不相关、冗余或无用的变量往往也会进入到预测模型中, 并会掩盖重要变量的作用, 影响预测结果. 为了降低以上不利因素的影响, 提高模型的预测精度和效率 [11], 本文将互信息方法 [12] 应用于多变量时间序列预测模型的输入变量选择中, 利用互信息方法可有效测量变量间的统计依赖性的特点, 选择与预测输出之间统计相关度较大的输入变量.

3 极端学习机预测模型

极端学习机是一种新的前馈神经网络学习方法, 并已成功地应用于函数逼近、时间序列预测以及模式分类等领域. 极端学习机特殊的学习机理可有效克服传统神经网络基于梯度学习算法的缺陷, 采用简单的线性回归即可得到令人满意的解.

3.1 极端学习机

极端学习机是一种新型的单隐层前馈神经网络, 其结构如图 1 所示, 数学表达式如下:

$$o_k = \mathbf{w}^T g(\mathbf{W}_{in} \cdot \mathbf{x}_k + \mathbf{b}) \quad (k = 1, 2, \dots, N), \quad (4)$$

其中 \mathbf{x}_k 为输入向量, o_k 为网络输出, \mathbf{W}_{in} 为连接输入层和隐含层的输入权值, \mathbf{b} 为隐含层偏置值, \mathbf{w} 为

连接隐含层和输出层的输出权值, g 为激活函数, 一般取 Sigmoid 函数, N 为样本数.

在训练开始时, \mathbf{W}_{in} 和 \mathbf{b} 随机生成并保持固定, 仅需训练确定 \mathbf{w} . 通常采用伪逆算法一步计算

$$\mathbf{w} = \mathbf{H}^\dagger \mathbf{T}, \quad (5)$$

其中

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{W}_{in} \cdot \mathbf{x}_1 + \mathbf{b}_1) & \cdots & g(\mathbf{W}_{in} \cdot \mathbf{x}_1 + \mathbf{b}_n) \\ \vdots & \ddots & \vdots \\ g(\mathbf{W}_{in} \cdot \mathbf{x}_N + \mathbf{b}_1) & \cdots & g(\mathbf{W}_{in} \cdot \mathbf{x}_N + \mathbf{b}_n) \end{bmatrix}_{N \times n}$$

为极端学习机的隐含层输出矩阵, n 为隐含层节点数; $\mathbf{T} = [t_1, t_2, \dots, t_N]^T$ 为期望输出向量.

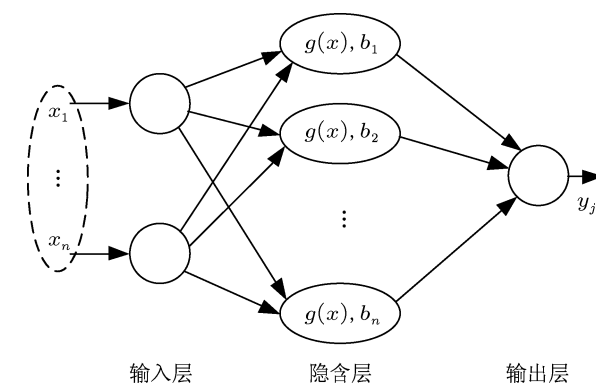


图 1 极端学习机结构示意图

极端学习机克服了传统神经网络训练速度慢、容易陷入局部最小等缺点. 与传统神经网络一样, 极端学习机的隐含层节点数也需在训练任务开始前人为设定. 隐含层的节点数决定着极端学习机的学习能力, 如果节点过少, 将导致网络不能有效地学习训练数据; 如果节点过多, 则很有可能造成过拟合, 影响网络的泛化能力.

3.2 模型选择

为获得更好的预测效果, 这里将对极端学习机预测模型进行模型选择. 所谓模型选择, 是指选择适当的模型规模, 从而使得模型的泛化误差 (期望风险) 最小, 以保证模型具有较好的泛化性能.

当极端学习机的损失函数选为二次损失时, 其期望风险可以分解为偏差项和方差项. 随着网络规模的增大, 偏差项逐渐减小而方差项逐渐增大. 当网络规模较小时, 由于具有较大的偏差项, 网络规模不是最优, 称为欠拟合. 当网络规模较大时, 由于具有较大的方差项, 网络规模也不是最优, 称为过拟合. 因此, 需在欠拟合和过拟合之间进行某种折

中,使得网络具有最优的规模.

当给定数据集 $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$, 其中 \mathbf{x}_i 为输入样本, $y_i = f(\mathbf{x}_i, \mathbf{w}_0) + \varepsilon_i$ 为输出样本, ε_i 为噪声, \mathbf{x}_i 和 ε_i 是相互独立的, 且分别由未知分布 $P(\mathbf{x}_i)$ 和 $P(\varepsilon_i)$ 产生. 建立预测模型 $f(\mathbf{x}, \mathbf{w}^*)$ 使得如下的期望风险最小:

$$R(f_n) = \int L(y, f(\mathbf{x}, \mathbf{w})) dP(\mathbf{x}) dP(\varepsilon), \quad (6)$$

其中 $R(f_n)$ 为预测模型隐节点数为 n 时的期望风险. $L(y, f(\mathbf{x}, \mathbf{w}))$ 是给定的损失函数, 通常取二次损失 $L(y, f_n(\mathbf{x})) = (y - f_n(\mathbf{x}, \mathbf{w}))^2$. 在极端学习机预测模型中, f_n 的表达式为

$$f_n(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^n w_i g(\mathbf{W}_{in} \cdot \mathbf{x} + b_i). \quad (7)$$

为使预测模型的期望风险最小, 需要辨识分布 $P(\mathbf{x})$ 和 $P(\varepsilon_i)$, 但其通常是未知的. 在实际中, 一般可通过对下式所表示预测模型的经验风险 $R_{\text{emp}}(f_n)$ 最小化得到 f_n :

$$R_{\text{emp}}(f_n) = \frac{1}{N} \sum_{i=1}^N L(y_i, f_n(\mathbf{x}_i, \mathbf{w})), \quad (8)$$

其中 $R_{\text{emp}}(f_n)$ 为预测模型隐节点数为 n 时的经验风险.

随着隐节点数 n 的增大, 经验风险 $R_{\text{emp}}(f_n)$ 减小, 预测模型的偏差减小而方差增大. 然而建立预测模型的目的是对未知数据进行预测, 即应最小化期望风险. 因此, 必须在偏差和方差之间进行合理的折中以最小化期望风险. 选择合适的 n 即为模型选择问题. 为得到更小的期望风险, 需对下列的函数进行最小化:

$$R_{\text{emp}}^*(f_n) = R_{\text{emp}}(f_n) T(n, N), \quad (9)$$

其中 $T(n, N)$ 为进行模型选择的校正因子, $R_{\text{emp}}^*(f_n)$ 为预测期望风险.

假设

$$\hat{f}_n = \sum_{i=1}^n w_i \phi_i(\mathbf{x}) = \sum_{i=1}^n w_i g(\mathbf{W}_{in} \cdot \mathbf{x} + b_i),$$

其中

$$\phi_i(\mathbf{x}) = \mathbf{W}_{in} \cdot \mathbf{x} + b_i.$$

\hat{f}_n 为使得经验均方误差最小的预测模型, 即

$$\hat{f}_n = \arg \min_{f \in F_n} R_{\text{emp}}(f). \quad (10)$$

假设噪声 ε 服从均值为零、方差为 σ^2 的独立同分布, 则预测模型 \hat{f}_n 的期望风险可以写为

$$R(\hat{f}_n) = \int (f(\mathbf{x}) + \varepsilon - \hat{f}_n(\mathbf{x}))^2 d\mu(\mathbf{x}) dP(\varepsilon)$$

$$\begin{aligned} &= \sigma^2 + \int (f(\mathbf{x}) - \hat{f}_n(\mathbf{x}))^2 d\mu(\mathbf{x}) \\ &= \sigma^2 + \sum_{i=1}^n (w_i - \hat{w}_i)^2 + \sum_{i=n+1}^{\infty} w_i^2. \end{aligned} \quad (11)$$

当 $\phi_i(\mathbf{x})$ 相互正交时最后一个等号成立. (11) 式等号右端第一项相当于真实预测模型的期望风险 $R(f)$, 第二项是估计误差, 第三项是近似误差, 写为

$$r_d = \sum_{i=n+1}^{\infty} w_i^2. \quad (12)$$

对估计误差引入向量 $\beta_i = \hat{w}_i - w_i$, 并将经验风险 R_{emp} 表示为 β 的函数, 即

$$\begin{aligned} R_{\text{emp}}(\beta) &= \frac{1}{N} \sum_{i=1}^N \left(y_i - \sum_{j=1}^N (w_j + \beta_j) \phi_j(\mathbf{x}_i) \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^N \tilde{y}_i^2 - \frac{2}{N} \sum_{j=1}^N \beta_j \sum_{i=1}^N \tilde{y}_i \phi_j(\mathbf{x}_i) \\ &\quad + \sum_{j,k=1}^N \beta_j \beta_k \frac{1}{N} \sum_{j,k=1}^N \phi_j(\mathbf{x}_i) \phi_k(\mathbf{x}_i), \end{aligned} \quad (13)$$

其中

$$\tilde{y}_i = \varepsilon_i + \sum_{j=n+1}^{\infty} w_j \phi_j(\mathbf{x}_i).$$

引入 $N \times n$ 矩阵 Φ , 其中 $\Phi_{i,j} = \phi_j(\mathbf{x}_i)$ 为第 i 个样本在预测模型第 j 个隐节点的输出, 则预测模型经验风险最小时有

$$\beta = (\Phi^T \Phi)^{-1} \Phi^T \tilde{\mathbf{Y}},$$

其中 $\tilde{\mathbf{Y}} = [\tilde{y}_1, \dots, \tilde{y}_N]^T$. 经验风险的最小值为

$$R_{\text{emp}}(\hat{f}_n) = \frac{1}{N} \tilde{\mathbf{Y}}^T (\mathbf{I} - \Phi (\Phi^T \Phi)^{-1} \Phi^T) \tilde{\mathbf{Y}}. \quad (14)$$

矩阵 Φ 的奇异值分解 (SVD) 可写为 $\Phi = \mathbf{U} \mathbf{S} \mathbf{V}^T$, 其中 \mathbf{U} 和 \mathbf{V} 分别为 $N \times N$ 和 $n \times n$ 的正交矩阵, \mathbf{S} 为 $N \times n$ 的对角阵. 因而有

$$\Phi (\Phi^T \Phi)^{-1} \Phi^T = \mathbf{U} \mathbf{I}_n \mathbf{U}^T,$$

其中 $\mathbf{S} (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T = \mathbf{I}_n$ 是 $N \times N$ 的对角阵, 其前 n 个对角元素为 1, 其余为 0. 因此, (14) 式可写为

$$\begin{aligned} R_{\text{emp}}(\hat{f}_n) &= \frac{1}{N} \tilde{\mathbf{Y}}^T (\mathbf{I}_N - \mathbf{I}_n) \tilde{\mathbf{Y}} \\ &= \frac{1}{N} \sum_{j=n+1}^N \left(\sum_{i=1}^n \tilde{y}_i U_{ij} \right)^2. \end{aligned} \quad (15)$$

假设 \tilde{Y} 和 Φ 是统计无关的, 则 \tilde{Y} 和 U 是统计无关的, 因而 $\tilde{y}_i U_{ij}$ 的期望 $E\tilde{y}_i U_{ij} = E\tilde{y}_i E U_{ij} = 0$. 由 (15) 式可得

$$R_{\text{emp}}(\hat{f}_n) = \frac{1}{N} \sum_{j=n+1}^N \sum_{i=1}^n E\tilde{y}_i^2 E U_{ij}^2 = \left(1 - \frac{n}{N}\right)(r_d + \sigma^2). \quad (16)$$

(16) 式的第二个等号由 ε_i 和 \mathbf{x}_i 的独立性、基底的正交性 ($E\tilde{y}_i^2 = (r_d + \sigma^2)$) 以及矩阵 U 的正交性 ($\sum_{i=1}^N E U_{ij}^2 = 1$) 得出.

在 (11) 式中, 需估计

$$\sum_{i=1}^n (w_i - \hat{w}_i)^2 = \sum_{i=1}^n (\beta_i)^2.$$

为进行估计, 有

$$\|\beta\|^2 = \tilde{Y}^T \Phi (\Phi^T \Phi)^{-2} \Phi^T \tilde{Y}.$$

并用 $(\lambda_1, \dots, \lambda_n)$ 表示协方差矩阵

$$C = \frac{1}{N} \Phi^T \Phi$$

的特征值.

$$C_{jk} = \frac{1}{N} \sum_{i=1}^N \phi_j(\mathbf{x}_i) \phi_k(\mathbf{x}_i). \quad (17)$$

同理, 有

$$E \sum_{j=1}^n \beta_j^2 = \frac{\sum_{i=1}^n E(1/\lambda_i)}{N} (r_d + \sigma^2).$$

最后, 由 (11) 和 (17) 式可得

$$ER(\hat{f}_n) = ER_{\text{emp}}(\hat{f}_n) \left(1 - \frac{n}{N}\right)^{-1} \times \left(1 + \frac{\sum_{i=1}^n E(1/\lambda_i)}{N}\right). \quad (18)$$

由于对预测模型进行模型选择的目标是选择具有最小期望风险的预测模型, 由上面的分析, 误差修正项 $T(n, N)$ 可写为

$$T(n, N) = \left(1 - \frac{n}{N}\right)^{-1} \left(1 + \frac{E \sum_{i=1}^n (1/\lambda_i)}{N}\right), \quad (19)$$

其中 λ_i 是协方差矩阵 (17) 式的第 i 个特征值. 本文采用一种直接特征值估计 (DEE) 方法对特征值进

行计算. 当训练数据和测试数据均已得到时, 可计算训练数据协方差矩阵 C_{emp} 和测试数据协方差矩阵 C_{test} .

存在唯一的矩阵 P 满足

$$P^T C_{\text{test}} P = I, \\ P^T C_{\text{emp}} P = \Lambda,$$

其中 Λ 为对角矩阵, 且对角元素为 $\lambda_1, \dots, \lambda_n$.

将 (19) 式中的 $E \sum_{i=1}^n (1/\lambda_i)$ 用如下的经验值代替:

$$\sum_{i=1}^n 1/\lambda_i = \text{trace} \left(P^{-1} C_{\text{emp}}^{-1} (P^T)^{-1} P^T C_{\text{test}} P \right) = \text{trace} \left(C_{\text{emp}}^{-1} C_{\text{test}} \right). \quad (20)$$

(20) 式可处理非正交基底, 其中 $\text{trace}(C_{\text{emp}}^{-1} C_{\text{test}})$ 是一个标志经验协方差矩阵 C_{emp} 与其期望值 C_{test} 差别的指标.

将 (20) 与 (19) 式相结合, 可以有效地估计出极端学习网络的期望风险 (测试风险), 选择使得预测模型估计期望风险最小的模型规模就可得到具有最优预测性能的预测模型.

4 仿真结果及分析

为验证方法的有效性, 将本文提出的基于变量选择和极端学习机的多变量混沌时间序列预测方法分别应用于 Lorenz 和 Rössler 三变量混沌时间序列预测以及 Rössler 超混沌多变量时间序列预测中, 并引入均方根误差 E_{RMS} 来定量说明预测性能.

$$E_{\text{RMS}} = \left(\frac{1}{S-1} \sum_{t=1}^S [P_i - O_i]^2 \right)^{1/2},$$

其中 S 为测试样本数, O_i 是某个实际观测值, P_i 为对应预测输出.

4.1 Lorenz 混沌时间序列

Lorenz 混沌方程如下:

$$\dot{x} = a(y - x), \\ \dot{y} = (c - z)x - y, \\ \dot{z} = xy - bz. \quad (21)$$

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 法迭代产生混沌时间序列 (图 2), 使用 $x(t), y(t)$ 和 $z(t)$ 序列共同预测 $x(t + \eta)$.

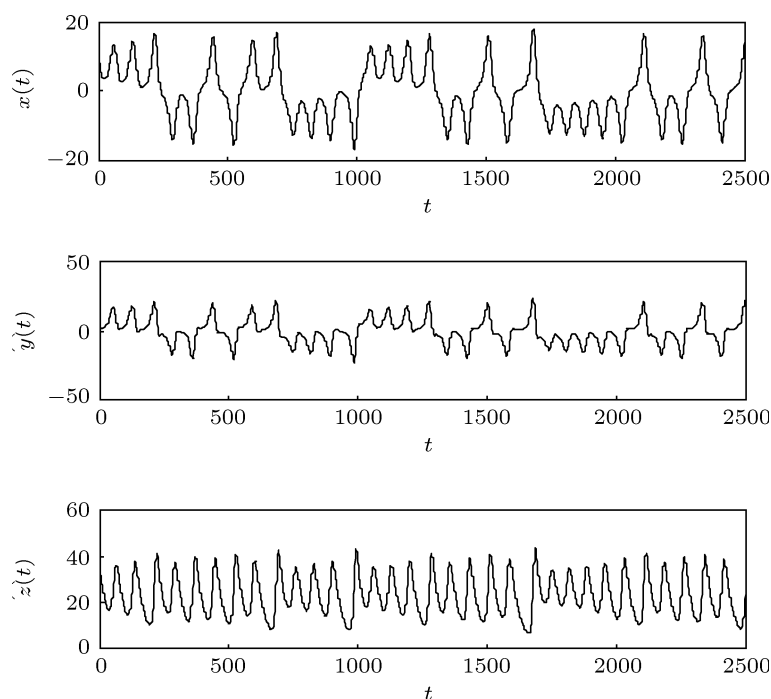


图2 Lorenz 三变量混沌时间序列

取延迟时间和嵌入维数分别为 $\tau_1 = 19$, $\tau_2 = 13$, $\tau_3 = 12$, $m_1 = 3$, $m_2 = 5$, $m_3 = 7$, 相应的 15 维延迟坐标向量为

$$\mathbf{d}(t) = \{x(t), \dots, x(t - 2\tau_1), y(t), \dots, y(t - 4\tau_2), z(t), \dots, z(t - 6\tau_3)\} \\ = \{\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T\}^T.$$

原序列长度为 2501, 当 $\eta = 1$ 时, 经相空间重构可以产生 2428 组数据. 令 $\mathbf{u}(k) = \mathbf{d}(k)$, 预测输出为 $x(t + \eta)$. 实际训练样本长度为 2000, 测试长度为 428.

首先通过互信息方法对输入变量 $\mathbf{d}(t)$ 进行输入变量选择. 从 15 维输入变量中选择了 6 维作为预测输入变量, 借助于极端学习机的通用逼近能力, 建立多变量时间序列预测模型, 通过 DEE 方法选择具有最小估计期望风险的预测模型以获得最好的预测精度. 实际预测风险和本文方法估计的期望风险如图 3 所示. 从图 3 可以看出, 本文所提算法可以很好地估计出预测模型的期望风险. 采用本文方法的预测曲线如图 4 所示. 从图 4 可以看出, 本文方法的预测曲线可以较好地拟合实际曲线. 预测误差如图 5 所示, 精确的预测误差为 3.0991×10^{-8} . 表 1 列出了本文方法与文献 [10] 中提到的其他常用多变量时间序列预测方法的预测结果比较. 从表 1 可知, 相对于其他方法, 本文所提方法在预测

精度上有了很大提高, 证明了本文将输入变量选择和预测模型优化相结合的多变量时间序列预测方法的有效性.

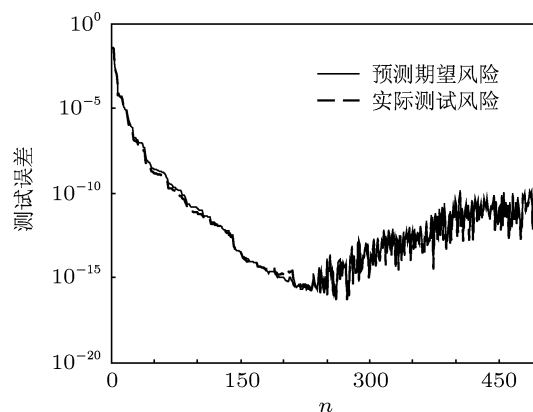


图3 Lorenz 混沌时间序列预测期望风险和实际测试风险

表1 本文 Lorenz 混沌时间序列预测结果与文献 [10] 中结果的比较

方 法	E_{RMS}
基于欧氏距离的输入 K 均值聚类	1.23×10^{-2}
基于高斯基距离的输入输出 K 均值聚类	8.5×10^{-3}
模糊树建模方法 I	7×10^{-4}
模糊树建模方法 II	3×10^{-4}
自适应模糊推理系统	5×10^{-4}
模糊边界模块化神经网络	2.046×10^{-5}
相空间重构参数的信息熵优化方法	1.0537×10^{-5}
本文方法	3.0991×10^{-8}

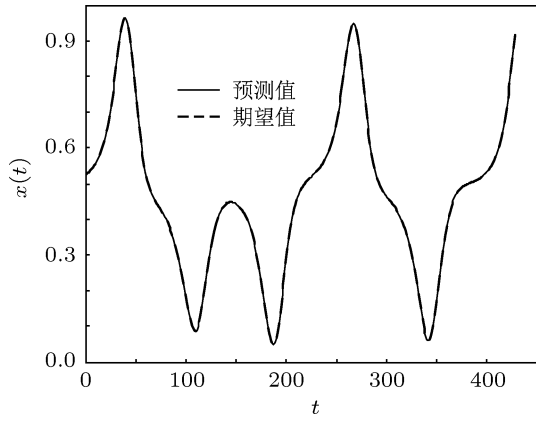


图4 Lorenz 混沌时间序列预测曲线

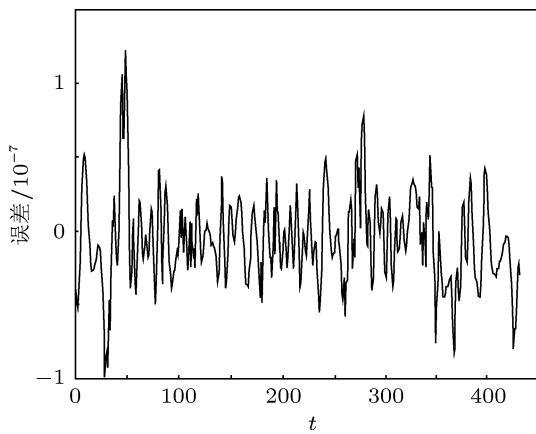


图5 Lorenz 混沌时间序列预测误差

4.2 Rössler 混沌时间序列

Rössler 混沌方程如下:

$$\begin{aligned} \dot{x} &= -y - z, \\ \dot{y} &= x + ay, \\ \dot{z} &= b + z(x - c). \end{aligned} \quad (22)$$

当取 $a = 0.2, b = 0.2, c = 5.7, x(0) = 1.0, y(0) = z(0) = 0$ 时, 系统产生混沌. 利用四阶 Runge-Kutta 法迭代产生混沌时间序列 (图 6), 使用 $x(t), y(t)$ 和 $z(t)$ 序列共同预测 $x(t + \eta)$.

取延迟时间和嵌入维数分别为 $\tau_1 = 30, \tau_2 = 15, \tau_3 = 12, m_1 = 3, m_2 = 5, m_3 = 7$, 相应的 15 维延迟坐标向量为

$$\begin{aligned} \mathbf{d}(t) &= \{x(t), \dots, x(t - 2\tau_1), y(t), \dots, y(t - 4\tau_2), \\ &\quad z(t), \dots, z(t - 6\tau_3)\} \\ &= \{\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T\}^T. \end{aligned}$$

原序列长度为 2000, 当 $\eta = 1$ 时, 经相空间重构可以产生 1927 组数据. 令 $\mathbf{u}(k) = \mathbf{d}(k)$, 预测输出为 $x(t + \eta)$. 实际训练样本长度为 1400, 测试长度为 527.

采用互信息方法对输入变量 $\mathbf{d}(t)$ 进行输入变量选择, 从 15 维输入变量中选择了 6 维作为预测输入变量, 并采用通过 DEE 方法进行模型选择的

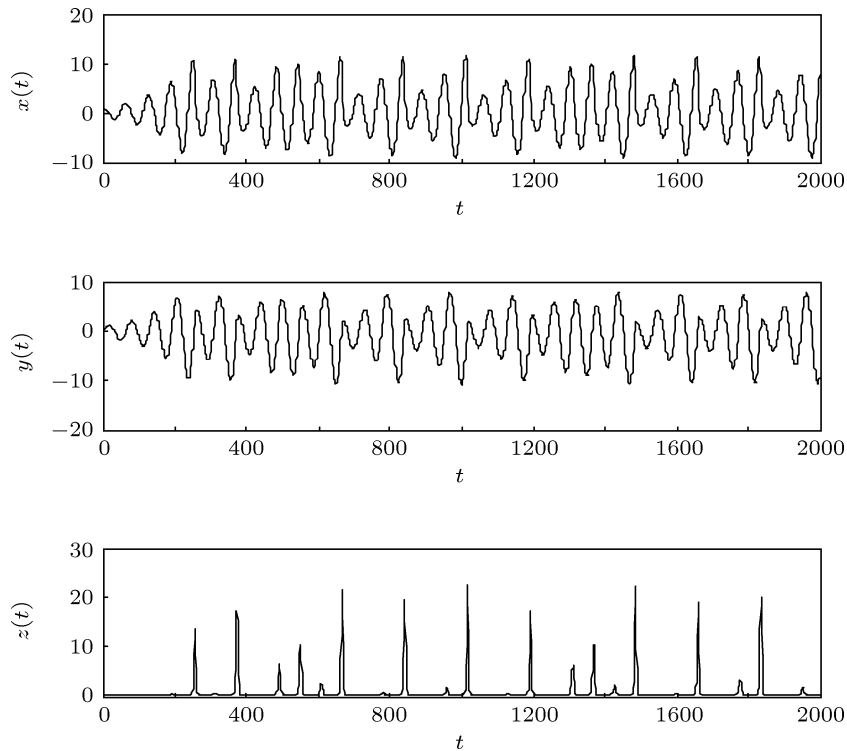


图6 Rössler 三变量混沌时间序列

极端学习机建立预测模型, 预测曲线如图 7 所示. 从图 7 可以看出, 本文方法的预测曲线可以较好地

拟合实际曲线. 预测误差如图 8 所示, 精确的预测误差为 2.9431×10^{-5} .

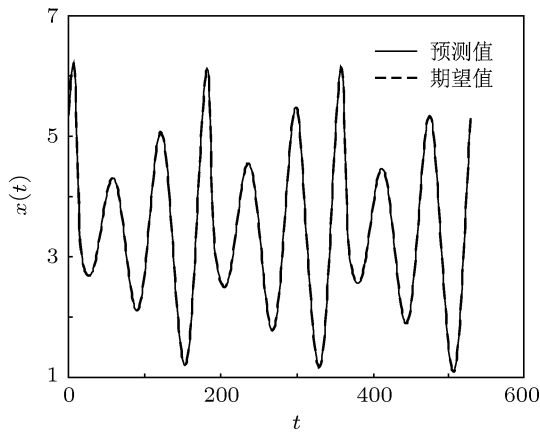


图 7 Rössler 混沌时间序列预测曲线

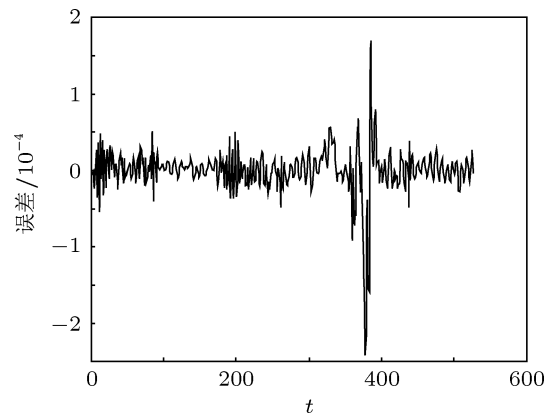


图 8 Rössler 混沌时间序列预测误差

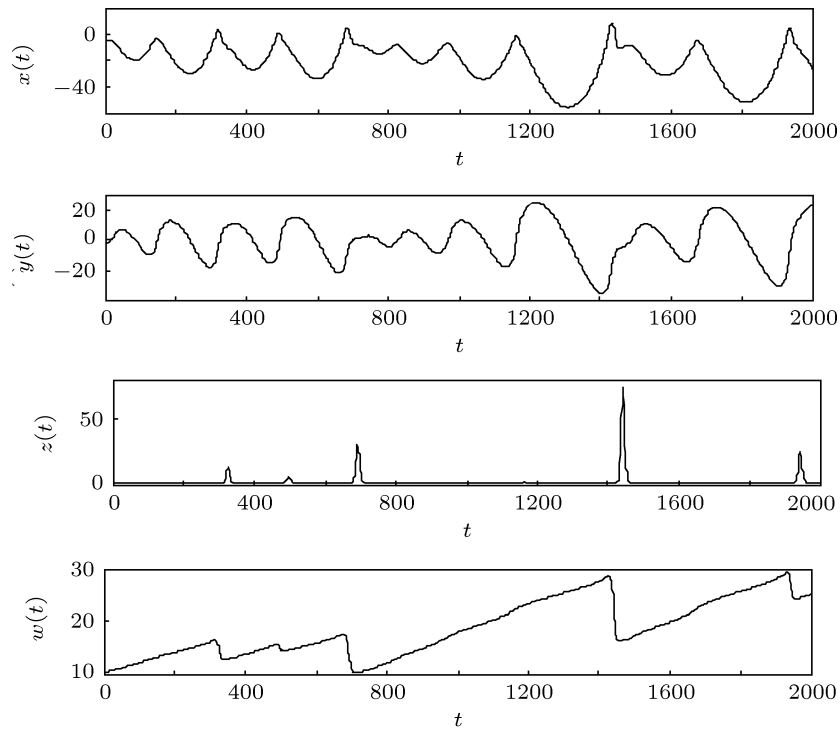


图 9 Rössler 超混沌四变量时间序列

4.3 Rössler 超混沌时间序列

超混沌系统具有更强的非线性特性, 其运动行为更加难以预测. Rössler 超混沌系统如下:

$$\begin{aligned}
 \dot{x} &= -y - z, \\
 \dot{y} &= x + ay + w, \\
 \dot{z} &= b + xz, \\
 \dot{w} &= -cz + dw.
 \end{aligned} \tag{23}$$

当取 $a = 0.25, b = 3, c = 0.5, d = 0.05$ 时, 系统产

生超混沌. 利用四阶 Runge-Kutta 法迭代产生混沌时间序列 (图 9). 使用 $x(t), y(t), z(t)$ 和 $w(t)$ 序列共同预测 $x(t + \eta)$.

取延迟时间和嵌入维数分别为 $\tau_1 = 30, \tau_2 = 15, \tau_3 = 12, \tau_4 = 10, m_1 = 3, m_2 = 5, m_3 = 7, m_4 = 5$, 相应的 20 维延迟坐标向量为

$$\begin{aligned}
 \mathbf{d}(t) &= \{x(t), \dots, x(t - 2\tau_1), y(t), \dots, y(t - 4\tau_2), \\
 &\quad z(t), \dots, z(t - 6\tau_3), w(t), \dots, w(t - 4\tau_4)\} \\
 &= \{\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T, \mathbf{w}^T\}^T.
 \end{aligned}$$

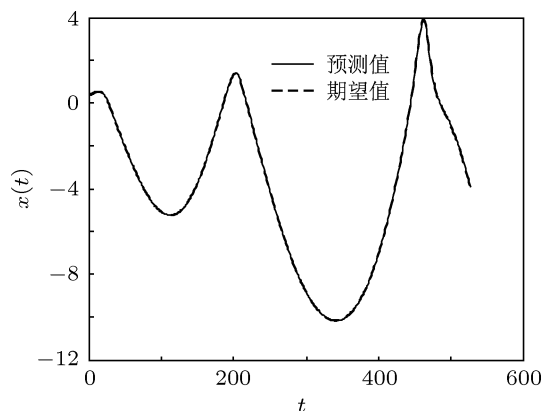


图 10 Rössler 超混沌时间序列预测曲线

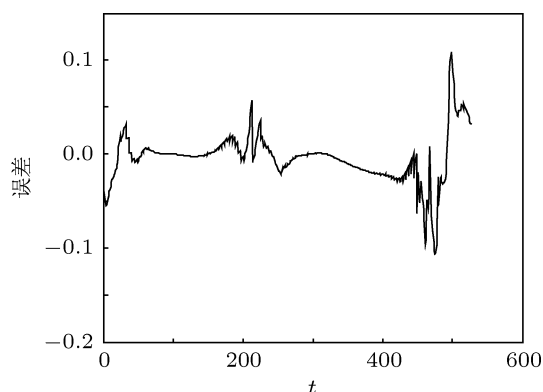


图 11 Rössler 超混沌时间序列预测误差曲线

原序列长度为 2000, 当 $\eta = 1$ 时, 经相空间重构可以产生 1927 组数据. 令 $\mathbf{u}(k) = \mathbf{d}(k)$, 预测输出

为 $x(t + \eta)$. 实际训练样本长度为 1400, 测试长度为 527.

采用互信息方法对输入变量 $\mathbf{d}(t)$ 进行输入变量选择, 从 20 维输入变量中选择了 6 维作为预测输入变量, 并采用通过 DEE 方法进行模型选择的极端学习机建立预测模型, 预测曲线如图 10 所示. 从图 10 可以看出, 本文方法的预测曲线可以较好地拟合实际曲线. 预测误差如图 11 所示, 精确的预测误差为 0.0541.

5 结论

针对多变量混沌时间序列预测问题, 本文提出一种基于输入变量选择和极端学习机的预测方法. 对多变量混沌时间序列进行相空间重构后, 选择与预测输出包含互信息最多的重构变量作为预测模型输入. 然后采用一种有效的神经网络——极端学习机, 建立多变量混沌时间序列预测模型. 为进一步提高预测精度, 引入直接特征值模型选择算法, 选择具有最小期望风险的极端学习机预测模型. 基于 Lorenz 和 Rössler 多变量混沌时间序列及 Rössler 超混沌时间序列的仿真结果证明所提方法的有效性. 本文将输入变量选择和预测模型的模型选择算法有效结合, 为多变量混沌时间序列预测提供了一种新方法.

- [1] Zhang S Q, Jia J, Gao M, Han X 2010 *Acta Phys. Sin.* **59** 1576 (in Chinese) [张淑清, 贾健, 高敏, 韩叙 2010 物理学报 **59** 1576]
- [2] Cao L, Mees A, Judd K 1998 *Physica D* **121** 75
- [3] Chakraborty K, Mehrotra K, Mohan C K, Ranka S 1992 *Neural Networks* **5** 961
- [4] Zhang J F, Hu S S 2008 *Acta Phys. Sin.* **57** 2708 (in Chinese) [张军峰, 胡寿松 2008 物理学报 **57** 2708]
- [5] 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]
- [6] Han M, Shi Z W, Guo W 2007 *Acta Phys. Sin.* **56** 43 (in Chinese)

- [7] 韩敏, 史志伟, 郭伟 2007 物理学报 **56** 43]
- [7] Huang G B, Zhu Q Y, Siew C K 2006 *Neurocomputing* **70** 489
- [8] Chapelle O, Vapnik V, Bengio Y 2002 *Mach. Learn.* **48** 9
- [9] Takens F 1981 *Dynam. Syst. Turbul.* **898** 366
- [10] Zhang C T, Ma Q L, Peng H 2010 *Acta Phys. Sin.* **59** 7623 (in Chinese) [张春涛, 马千里, 彭宏 2010 物理学报 **59** 7623]
- [11] Kohavi R, John G H 1997 *Artif. Intell.* **97** 273
- [12] Peng H, Long F, Ding C 2005 *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226

Multivariate chaotic time series prediction based on extreme learning machine*

Wang Xin-Ying Han Min[†]

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

(Received 17 July 2011; revised manuscript received 16 September 2011)

Abstract

For multivariate chaotic time series prediction problem, a prediction based on input variable selection and extreme learning machine is proposed in this paper. The multivariate chaotic time series is reconstructed in phase space, and a mutual information based method is used to select the input variables, which have high statistics information with the output variables. The extreme learning machine is conducted to model the multivariate chaotic time series in the phase space by utilizing its approximation capability. In order to improve the prediction accuracy, a model selection algorithm is conducted for extreme learning machine to choose an expected minimum risk prediction model. Simulation results based on Lorenz, Rössler multivariate chaotic time series and Rössler hyperchaotic time series show the effectiveness of the proposed method.

Keywords: chaotic time series prediction, input variables selection, extreme learning machine, model selection

PACS: 05.45.Tp

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

[†] E-mail: minhan@dlut.edu.cn