

# 基于高斯过程的混沌时间序列单步与多步预测\*

李 军<sup>†</sup> 张友鹏

(兰州交通大学自动化与电气工程学院, 兰州 730070)

(2010年9月13日收到; 2010年11月4日收到修改稿)

针对混沌时间序列单步和多步预测, 提出基于复合协方差函数的高斯过程 (GP) 模型方法. GP 模型的确立由协方差函数决定, 通过对训练数据集的学习, 在证据最大化框架内, 利用矩阵运算和优化算法自适应地确定协方差函数和均值函数中的超参数. GP 模型与神经网络、模糊模型相比, 其可调整参数很少. 将不同复合协方差函数的 GP 模型应用在混沌时间序列单步及多步提前预测中, 并与单一协方差函数的 GP、支持向量机、最小二乘支持向量机、径向基函数神经网络等方法进行了比较. 仿真结果表明, 基于不同复合协方差函数的 GP 方法能精确地预测混沌时间序列, 具有稳健的特性. 因此, 它是研究复杂非线性动力系统辨识和控制的一种有效方法.

**关键词:** 高斯过程, 混沌时间序列, 预测, 模型比较

**PACS:** 05.45.Tp, 02.50.Ey, 05.45.-a, 05.45.Gg

## 1. 引言

混沌现象在自然界中广泛存在, 混沌动力学系统就是一类典型的复杂高维非线性动力系统, 具有“奇异”吸引子的混沌系统本身是确定性的, 但这种非线性动力系统即使维数很低, 其行为也有类似于随机性的复杂特性, 就实际观测而言, 描述混沌系统随时间演化的所有状态变量往往未知, 仅可得到一个状态变量的系统输出, 即一组含有噪声的混沌时间序列, 表示系统的某个状态变量随时间演化的过程<sup>[1]</sup>. 利用 Takens 的延迟-嵌入定理<sup>[2]</sup>, 原未知混沌动力学系统的吸引子和内在动态特性可从获得的一组含噪声的单变量时间序列输出中重构出, 因此混沌时间序列单步与多步预测、建模已经成为非常重要的研究领域<sup>[3,4]</sup>. 近年来, 国内外学者利用各种复杂非线性系统建模方法, 如径向基神经网络<sup>[5,6]</sup>、递归神经网络<sup>[4,7,8]</sup>、模糊逻辑<sup>[9]</sup>、支持向量机<sup>[10,11]</sup>、核学习机<sup>[12]</sup>等方法, 对混沌时间序列的预测与建模进行研究, 并且取得了成功的应用<sup>[13,14]</sup>.

目前, 高斯过程 (Gaussian processes, GP) 在机器学习领域逐步引起人们的密切关注<sup>[15-19]</sup>. GP 以概率分布来表示函数输出的先验知识, 并在泛函空

间建立模型, 基于数据间的相关性, 构造协方差函数, 通过 Bayesian 推理进行计算, 与典型的神经网络、支持向量机相比, GP 模型的优点是可调整参数少得多, 能够把先验知识和观测数据结合的完美无缺, 还能预先估计一些统计分布, 获得模型预测输出的方差估计, 从而很好地解释模型的置信水平, 这也是优于神经网络和模糊系统建模的特色之处<sup>[17,19]</sup>. 类似于支持向量机使用的“核函数”技巧<sup>[20,21]</sup>, GP 模型的协方差函数可称之为核函数, 它就是一种通用概率核学习机, 对模型提供了实用性、原理性的解释.

为验证 GP 模型方法的有效性, 本文将不同复合协方差函数的 GP 模型预测方法应用于混沌时间序列单步与多步直接预测、迭代预测中, 比较了 GP 模型的稳健性和泛化性能, 同时与常规 GP 模型、神经网络、支持向量机等预测模型方法在同等条件下也进行了比较, 从预测精度方面比较模型的性能.

## 2. 高斯过程

高斯过程建模是有限维空间中的高斯分布向无限维函数空间的推广, 建模的先验知识则通过协方差函数的选择和“超参数”的确定被合并于高斯

\* 甘肃省自然科学基金 (批准号: 0803RJZA023) 资助的课题.

<sup>†</sup> E-mail: lijun691201@mail.lzjtu.cn

过程中,当函数空间分布被认为符合高斯过程时,学习和推理所涉及的计算还会显得相对容易.在机器学习领域中的监督学习可以看作是一种直接在高斯过程框架下从样本数据中学习函数的方法<sup>[18]</sup>.高斯过程模型用贝叶斯方法选择模型并确定参数,为核学习机的学习提供一种实用的、基本的概率解释,而且对模型的选择和模型“超参数”的自适应建立了很好的贝叶斯框架,为模型的预测提供置信水平.

### 2.1. 高斯过程回归

高斯过程建模的思想就是无需给出  $f(\mathbf{x})$  的参数化或非参数化形式,直接在函数空间中将  $f(\mathbf{x})$  的取值看作是随机变量,将  $f(\mathbf{x})$  的先验概率分布  $p(f(\mathbf{x}))$  看作是高斯分布.若给定数据集  $D = \{\mathbf{x}^{(n)}, y_n\}_{n=1}^N, \mathbf{x} \in \mathbf{R}^d, y_i \in \mathbf{R}$ , 并定义输入数据矩阵  $\mathbf{X} \in \mathbf{R}^{d \times N}$ , 输出数据向量  $\mathbf{y} \in \mathbf{R}^{N \times 1}$ . 在给定数据集  $D$  的有限数据集合中,  $f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(N)})$  可构成随机变量的一个集合,且具有联合高斯分布,它们形成的随机过程就称之为高斯过程.即

$$f(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (1)$$

其中,  $m(\mathbf{x}) = E[f(\mathbf{x})], k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$ , 符号  $E$  表示数学期望.  $m(\mathbf{x})$  为均值函数,  $k(\mathbf{x}, \mathbf{x}')$  为协方差函数.

将高斯过程应用于一般的回归建模问题,可考虑含噪声的观测目标值  $y$ , 即

$$y = f(\mathbf{x}) + \varepsilon, \quad (2)$$

其中  $\varepsilon$  为附加的与  $f(\mathbf{x})$  不相关的独立的高斯白噪声,即服从均值为零、方差为  $\sigma_n^2$  的正态分布,可记作  $\varepsilon \sim N(0, \sigma_n^2)$ . 对于(2)式,由于噪声  $\varepsilon$  为独立于  $f(\mathbf{x})$  的高斯白噪声,若  $f(\mathbf{x})$  为高斯过程,则  $y$  同样服从高斯分布,其有限观测值联合分布的集合可形成一个高斯过程,即

$$y \sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}') + \sigma_n^2 \delta_{ij}), \quad (3)$$

其中,  $m(\mathbf{x})$  为均值函数,  $\delta_{ij}$  是狄拉克函数,  $i = j$  时,函数  $\delta_{ij} = 1$ . 协方差函数若以矩阵形式表达,则为

$$\mathbf{C}(\mathbf{X}, \mathbf{X}) = E[\mathbf{y}\mathbf{y}^T] = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}, \quad (4)$$

其中,  $\mathbf{I}$  表示  $N \times N$  的单位矩阵,  $\mathbf{C}(\mathbf{X}, \mathbf{X})$  表示  $N \times N$  的协方差矩阵,  $\mathbf{K}(\mathbf{X}, \mathbf{X})$  表示  $N \times N$  的核矩阵,称为 Gram 矩阵,其元素  $\mathbf{K}_{ij} = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ .

在高斯过程先验分布所定义的泛函空间中,在贝叶斯框架下可计算后验分布的函数预测输出值,

对于  $N_*$  个测试数据的集合  $D_* = \{\mathbf{x}^{(n)}, y_n\}_{n=N+1}^{N+N_*}$ , 由测试输入构成输入数据矩阵  $\mathbf{X}_* \in \mathbf{R}^{d \times N_*}$ , 相应函数输出值  $f(\mathbf{x}_*)$  可构成输出向量  $\mathbf{f}_* \in \mathbf{R}^{1 \times N_*}$ . 训练数据和测试数据集的均值向量分别为  $\mathbf{m}$  与  $\mathbf{m}_*$ . 由(3)式可知,训练数据集的观测输出值与测试数据集的观测输出值服从联合高斯分布,即

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{C}(\mathbf{X}, \mathbf{X}) & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right). \quad (5)$$

根据联合高斯分布的性质,若  $\mathbf{x}$  与  $\mathbf{t}$  是联合的具有高斯分布的随机向量,即

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{t} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{m}_x \\ \mathbf{m}_t \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{E} \\ \mathbf{E}^T & \mathbf{B} \end{bmatrix} \right). \quad (6)$$

则  $\mathbf{x}$  的边缘分布及给定  $\mathbf{t}$  的条件下  $\mathbf{x}$  的条件分布应为

$$\mathbf{x} \sim N(\mathbf{m}_x, \mathbf{A}), \quad (7a)$$

$$\begin{aligned} \mathbf{x} | \mathbf{t} &\sim N(\mathbf{m}_x + \mathbf{E}\mathbf{B}^{-1}(\mathbf{t} - \mathbf{m}_t), \\ &\mathbf{A} - \mathbf{E}\mathbf{B}^{-1}\mathbf{E}^T), \end{aligned} \quad (7b)$$

其中,  $\mathbf{A}, \mathbf{E}, \mathbf{B}$  是协方差矩阵,符号 T 表示转置.

应用(6)式,(7)式,由(5)式可相应得到相应测试数据集的输出向量  $\mathbf{f}_*$  所满足的后验条件分布,即 GP 模型回归方程为

$$\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim N(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*)), \quad (8a)$$

$$\bar{\mathbf{f}}_* = E[\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_*]$$

$$= \mathbf{m} + \mathbf{K}(\mathbf{X}_*, \mathbf{X})\mathbf{C}(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{y} - \mathbf{m}), \quad (8b)$$

$$\text{cov}(\mathbf{f}_*) = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*)$$

$$- \mathbf{K}(\mathbf{X}_*, \mathbf{X})\mathbf{C}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{X}, \mathbf{X}_*), \quad (8c)$$

其中,预测均值向量  $\bar{\mathbf{f}}_*$  就是 GP 模型回归方程的输出,也即输出向量  $\mathbf{f}_*$  的预测值.

对于单个测试数据  $x_*$ , 记  $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X}), \mathbf{K}_* = \mathbf{K}(\mathbf{X}_*, \mathbf{X})$ , 测试数据与所有训练数据集的数据形成的协方差向量  $\mathbf{k}(\mathbf{x}_*) = (k(\mathbf{x}_*, \mathbf{x}^{(1)}), \dots, k(\mathbf{x}_*, \mathbf{x}^{(N)}))^T$ , 记  $\mathbf{k}_* = \mathbf{k}(\mathbf{x}_*)$ , 由(8b),(8c)式可知相应的 GP 模型预测输出是具有均值  $\bar{f}(\mathbf{x}_*)$  和方差  $\sigma_f^2(\mathbf{x}_*)$  的高斯分布,即

$$\bar{f}(\mathbf{x}_*) = m(\mathbf{x}_*) + \mathbf{k}_*^T \mathbf{C}^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{x})), \quad (9)$$

$$\sigma_f^2(\mathbf{x}_*) = \mathbf{k}(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T \mathbf{C}^{-1} \mathbf{k}_*. \quad (10)$$

对于 GP 模型,(9)式即为含噪声的观测输出预测值,GP 模型预测输出的误差带由(10)式中的  $\sigma_f^2(\mathbf{x}_*)$  给出,这表明 GP 模型在给出输出预测值的同时还能给出模型预测的置信水平或不确定性.

## 2.2. 高斯过程的模型选择与超参数自适应

### 2.2.1. 高斯过程的模型选择

在 GP 模型定义的泛函空间中,协方差函数在 GP 模型中起到关键作用,它表达了一种相似性或接近性,对所学习的函数提供了假设信息. 协方差函数  $c(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$  唯一的约束条件就是,其构成的协方差矩阵  $C$  满足非负定条件. 在建模问题中,观测目标输出分布需考虑到噪声方差因素影响,即使训练数据集的输出观测值完全无误,由于数值稳定性原因,也可引入类似的噪声项,以确保协方差矩阵满足正定条件<sup>[18]</sup>. 基于(4)式,GP 模型中的全协方差函数可表示为

$$c(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \sigma_n^2 \delta_{ij}, \quad (11)$$

其中,噪声方差  $\sigma_n^2$  的取值需预先给出,  $k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$  也称为核函数,通常满足半正定和对称性即可. 由于具有可加结构的 GP 模型,其协方差函数也具有可加性,且是各自独立的 GP 模型协方差函数的和. GP 模型中,(11)式中的协方差函数可分别为如下形式:

$$k_{la}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = a_0 + a_1 \sum_{l=1}^d \mathbf{x}_l^{(i)} \mathbf{x}_l^{(j)}, \quad (12)$$

$$k_{se}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = v_0 \exp \left\{ -\frac{1}{2} \sum_{l=1}^d w_l (\mathbf{x}_l^{(i)} - \mathbf{x}_l^{(j)})^2 \right\}, \quad (13)$$

$$k_{rq}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = v_0 \left( 1 + \frac{1}{2\alpha_l} \sum_{l=1}^d w_l (\mathbf{x}_l^{(i)} - \mathbf{x}_l^{(j)})^2 \right)^{-\alpha}, \quad (14)$$

$$k_{pp3}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \frac{v_0}{3} (1-r)_+^{j+3} \left( (j^3 + 9j^2 + 23j + 15)r^3 + (6j^2 + 36j + 45)r^2 + (15j + 45)r + 15 \right), \quad (15)$$

其中,  $r^2 = \sum_{l=1}^d w_l (\mathbf{x}_l^{(i)} - \mathbf{x}_l^{(j)})^2$ ,  $j = \lfloor d/2 \rfloor + q + 1$ ,  $q = 3$ ,符号  $\lfloor \cdot \rfloor$  表明向负无穷方向取整数,  $d$  是输入向量的维数.

(12)式为含偏置项的非平稳的线性协方差函数,(13)式为平方指数协方差函数,  $w_l$  决定输入的尺度特性,取值不同时,可以决定各维输入与目标值之间的相关性,表明对输入具有自动相关确定 (automatic relevance determination, ARD) 的特性<sup>[22]</sup>. 当参数  $w_l$  相同时,则称为高斯径向基(RBF)

核函数,具有各态历经特性. (14)式为有理二次协方差函数,要求参数  $\alpha > 0$ , 它控制协方差函数的衰减率,  $\alpha \rightarrow \infty$  时,它就为平方指数协方差函数;(15)式为一种具有紧凑支持特性的分段多项式协方差函数.

在常规 GP 模型中,取(13)式的高斯 RBF 核函数(参数  $w_l$  相同),此时就有  $c(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k_{se}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \sigma_n^2 \delta_{ij}$ . 单一高斯 RBF 核函数的 GP 模型很难描述复杂系统的输入各维的变化特性.

借助多核学习支持向量机的思想,本文 GP 模型的选择将考虑一类具有良好性能的复合协方差函数组合,复合协方差函数考虑随机过程中非平稳变化的趋势,为如下形式:

$$c_1(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k_{se}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + k_{la}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \sigma_n^2 \delta_{ij}, \quad (16)$$

$$c_2(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k_{rq}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + k_{la}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \sigma_n^2 \delta_{ij}, \quad (17)$$

$$c_3(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = k_{pp3}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + k_{la}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + \sigma_n^2 \delta_{ij}. \quad (18)$$

### 2.2.2. 高斯过程的模型超参数自适应

基于训练数据集对 GP 模型的训练就是对“超参数”的学习过程,由此决定了协方差函数的自适应性,即基于给定训练数据集自适应地确立 GP 模型. 基于证据最大化的贝叶斯框架,协方差函数中“超参数”的选择可通过对下式所示的对数似然函数极大化得到<sup>[17]</sup>,即

$$\begin{aligned} \theta_{opt} &= \arg \max_{\theta} \{ \log p(\mathbf{y} | \mathbf{X}, \theta) \} \\ &= \arg \max_{\theta} \left\{ -\frac{1}{2} \log(\det(\mathbf{K} + \sigma_n^2 \mathbf{I})) \right. \\ &\quad \left. - \frac{1}{2} (\mathbf{y} - \mathbf{m})^T [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} (\mathbf{y} - \mathbf{m}) \right. \\ &\quad \left. - \frac{N}{2} \log 2\pi \right\}, \quad (19) \end{aligned}$$

其中,  $\det$  是行列式符号. 对超参数初始化为随机值,然后使用共轭梯度优化算法迭代学习求解(19)式,搜索出超参数的最优值.

为求取(19)式中的超参数向量最优值  $\theta_{opt}$ , 需获取负对数似然函数关于  $\theta$  的梯度,即

$$\begin{aligned} &\frac{\partial}{\partial \theta_k} \log p(\mathbf{y} | \mathbf{X}, \theta) \\ &= \frac{1}{2} (\mathbf{y} - \mathbf{m})^T \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \mathbf{C}^{-1} (\mathbf{y} - \mathbf{m}) \end{aligned}$$

$$-\frac{1}{2}\text{tr}\left(\mathbf{C}^{-1}\frac{\partial\mathbf{C}}{\partial\theta_k}\right), \quad (20a)$$

$$\begin{aligned} & \frac{\partial}{\partial\theta_m}\log p(\mathbf{y}|\mathbf{X},\boldsymbol{\theta}) \\ &= -(\mathbf{y}-\mathbf{m})^T\mathbf{C}^{-1}\frac{\partial\mathbf{m}}{\partial\theta_m}, \end{aligned} \quad (20b)$$

其中,  $\theta_m$  和  $\theta_k$  分别是均值函数以及包含噪声在内的协方差函数各自的超参数, 符号  $\text{tr}$  表示矩阵的迹.

核函数中所有“超参数”必须为正数, 所以在对数空间中考虑“超参数”构成的向量. 如(17)式中, 超参数  $\boldsymbol{\theta}_\kappa = [\log v_0, \log w_1, \dots, \log w_d, \log a_0, \log a_1, \log \sigma_n, \log \alpha]^T$ . 若 GP 模型的均值函数为常数  $\mu$ , 则超参数  $\theta_m = \log \mu$ .

最后, GP 模型建模的训练学习算法概括如下: 第一步, 利用(16)式至(18)式, 确定高斯过程模型协方差函数先验形式, 并将对数超参数初始化为随机值, 均值函数参数初始值设为常数; 第二步, 由(19), (20)式, 在极大似然性的贝叶斯框架内, 使用基于迭代方法的共轭梯度优化算法搜索确定“超参数”的最优值; 第三步, 通过(8b)式或(9)式, 求出 GP 模型针对测试数据的预测输出, 同时还可得到模型预测输出的误差带, 即不确定性大小.

学习算法中会涉及到对核矩阵的求逆运算, 为加快求逆运算过程, 以及避免直接对核矩阵求逆影响数值解稳定性的问题, 可采用数值稳定性好的乔利斯基分解算法, 即分解  $[\mathbf{K} + \sigma_n^2\mathbf{I}] = \mathbf{L}\mathbf{L}^T$  的矩阵平方根分解算法.

### 3. 混沌时间序列预测中的应用实例

混沌时间序列的预测可看作动力学系统研究的“反问题”, “正问题”是给定非线性动力学系统, 研究其相空间中轨道的各种性质. “反问题”是给定相空间中的一串迭代序列(轨道的演化过程)或一组观测序列, 要构造一个非线性映射来表达原系统, 这个映射就可作为预测用模型<sup>[1,3]</sup>. 若动力系统由下述离散时间非线性差分方程表示:

$$\mathbf{x}(n+1) = \mathbf{F}(\mathbf{x}(n)), \quad (21)$$

其中,  $\mathbf{x}(n)$  是系统在  $n$  时刻的状态向量,  $\mathbf{F}(\cdot)$  是非线性向量函数, 则观测到的时间序列  $y(n)$  为

$$y(n) = h(\mathbf{x}(n)) + w(n), \quad (22)$$

其中,  $w(n)$  为噪声,  $h(\cdot)$  是标量函数. 根据 Takens 定理<sup>[2]</sup>, 原动力学系统运动轨道的几何结构可由下述  $y(n)$  在  $D$  维空间“重构”:

$$\mathbf{y}_R(n) = [y(n), y(n-\tau), \dots, y(n-(D-1)\tau)]^T, \quad (23)$$

其中,  $\tau$  称为归一化嵌入时延宽度. 重构的条件是  $D \geq 2d_1 + 1$ ,  $d_1$  是动力系统的关联维数. 重构的结果可看为原系统相应的一条轨道在  $\mathbf{R}^D$  中的展开(或“嵌入”). 也可用下述  $q$  维向量作为重建向量, 即

$$\mathbf{y}_R(n) = [y(n), y(n-1), \dots, y(n-(q-1))]^T, \quad (24)$$

其中  $q$  为整数, 且  $q \geq D_E\tau$ ,  $D_E$  为最小嵌入维数.

为建立预测模型, 首先构造一个用于嵌入的短时记忆结构, 用于产生  $\mathbf{y}_R(n)$ , 其次是构造一个多输入、单输出的非线性系统模型  $f: \mathbf{R}^q \rightarrow \mathbf{R}$  可作为单步预测或多步预测. 即  $\hat{y}(n+h) = f(\mathbf{y}_R(n))$  或  $\hat{y}(n+h) = f(\mathbf{y}_R(n))$ , 其中  $h$  为预测步数. 在本文中则选择 GP 预测模型. 为了评价模型整体预测效果并与其他预测方法相比较, 可采用正则化均方根误差(normalized root mean square error, NRMSE) 两种指标, 表示为

$$E_{\text{NRMSE}} = \sqrt{\frac{\sum_{j=1}^N (y(j) - \hat{y}(j))^2}{N\sigma^2}} = \frac{E_{\text{RMSE}}}{\sigma}, \quad (25)$$

其中  $y(j)$  和  $\hat{y}(j)$  分别为混沌时间序列的实际值和预测值,  $\sigma$  为待预测混沌时间序列的标准差.  $E_{\text{RMSE}}$  是均方根误差,  $E_{\text{NRMSE}}$  是归一化的均方根误差, 其含义是均方根误差除以目标序列的标准差, 若  $E_{\text{NRMSE}} = 1$  则意味着预测输出是时间序列的均值.

在 GP 预测模型中, 对包含噪声方差在内的协方差函数“超参数”向量  $\boldsymbol{\theta}_\kappa$ , 对数初始值设置为从均值为 0, 方差为 1 的正态分布中随机选取. 常数  $C$  的均值函数初始值根据训练数据集的均值选取. GP 模型的学习就是基于训练数据集, 在极大化边际似然性的贝叶斯框架内, 对“超参数”向量  $\boldsymbol{\theta}$  优化选择的过程. 在本文的仿真实例中, 为搜索确定“超参数”向量  $\boldsymbol{\theta}$  的最优值, 使用共轭梯度优化算法最多迭代 100 次结束, 因为此时的最大似然性函数已经变化很慢. 为验证具有复合协方差函数 GP 模型的有效性, 在下面的仿真实例中均与取高斯 RBF 核函数的常规 GP 模型作了比较.

#### 3.1. Lorenz 混沌时间序列的单步预测

Lorenz 吸引子是最著名的 3 维自治非线性动力

学系统之一,能用于流体中湍流的建模. 描述它的动力学方程模型为<sup>[23]</sup>

$$\begin{aligned} \frac{dx(t)}{dt} &= -\sigma x(t) + \sigma y(t), \\ \frac{dy(t)}{dt} &= -x(t)z(t) + rx(t) - y(t), \\ \frac{dz(t)}{dt} &= x(t)y(t) - bz(t), \end{aligned} \quad (26)$$

其中参数设置  $(\sigma, r, b) = (10, 28, 8/3)$ , 此时系统为混沌状态. 对构成 Lorenz 吸引子的其中一个状态变量  $x(t)$  所构成的混沌时间序列进行单步预测, 系统状态变量  $x, y, z$  初始值设为  $[1; 2; 3]$ , 用四阶五级龙格-库塔 (Runge-Kutta) 法求解 (26) 式, 可以得到时间序列  $x(n)$  的数值解. 对 Lorenz 混沌时间序列计算最优的嵌入维<sup>[24]</sup>与归一化嵌入延迟<sup>[25]</sup>可分别得  $D_E = 3, \tau = 4$ , 由 (24) 式取  $q = 12$ . 因此, 用输入为 12 维的向量  $[x(n-11), \dots, x(n-1), x(n)]$  预测  $x(n+1)$ , 抽取 1400 对输入输出数据, 其中  $n = 5001$  到 6411. 前 400 个数据用作训练数据集, 其余的 1000 个数据用作测试数据集, 分别选取 (16), (17), (18) 式所示的复合协方差函数建立零均值 GP 预测模型. 在整个训练数据集、测试数据集上, 基于 (17) 式的复合协方差函数的 GP 模型对 Lorenz 混沌时间序列单步预测结果和实际输出值 (从  $n = 5013$  到 6412) 如图 1 所示, 其各点的预测误差分布如图 2 所示. 从图 1 中可以看出, 无论在训练数据还是测试数据集上, GP 模型对 Lorenz 混沌时间序列单步预测输出与实际输出吻合得很好, 因此, 预测模型具有很好的推广性. 表 1 给出了对 Lorenz 时间序列做单步预测时, 基于不同复合协方差函数的 GP 预测模型与与常规 GP 模型 (高斯 RBF 核函数)、最小二乘支持向量机<sup>[26]</sup> (LS-SVM),  $\epsilon$  不敏感支持向量机<sup>[20,21]</sup> ( $\epsilon$ -SVR),  $\nu$  支持向量机<sup>[21]</sup> ( $\nu$ -SVR), 采用正交最小二乘学习算法的 RBF 神经网络<sup>[27]</sup> (OLS-RBF) 等模型预测性能的比较.

$\nu$ -SVR 是  $\epsilon$ -SVR 的改进形式, 能自动调谐  $\epsilon$ ,  $\epsilon$ -SVR,  $\nu$ -SVR 及 LS-SVM 中的核函数均为高斯 RBF 核函数 ( $k(x, z) = \exp(-\|x - z\|^2/\sigma^2)$ ),  $\epsilon$ -SVR,  $\nu$ -SVR 及 LS-SVM 的参数选择可以基于训练数据集按照网格搜索、留一法或 5 折交叉验证的方法选取<sup>[21,26]</sup>.  $\epsilon$ -SVR 的参数选取按照文献<sup>[28]</sup>的经验选取, 正则化常数  $C = 3$ , 核参数  $\sigma^2 = 0.75 \times D_E = 2.2$ ,  $\epsilon = 0.01$ , 若  $\epsilon$  过小, 对输入噪声的抗干扰能力变差, 同时支持向量数据急剧增多.  $\nu$ -SVR 是  $\epsilon$ -

SVR 的改进形式, 其取值大小控制  $\epsilon$  的大小, 这里取 0.5, 其余参数与  $\epsilon$ -SVR 相同. LS-SVM 的取值按照 5 折交叉验证及单纯形搜索的方法选取参数  $\gamma = 8500, \sigma^2 = 3$ . OLS-RBF 选取 150 个中心. 从表 1 的比较结果看出, 复合协方差函数构成的 GP 模型在测试集和训练集上的预测性能均略好于其他方法, 显示了 GP 模型良好的预测效果和稳健性.

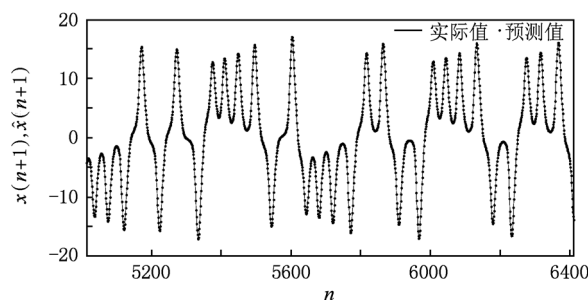


图 1  $n = 5013$  到 6412 的 Lorenz 混沌时间序列实际输出 (实线) 与单步预测输出 (点线)

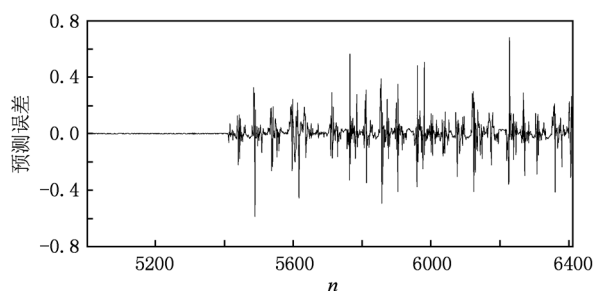


图 2  $n = 5013$  到 6412 的 Lorenz 混沌时间序列预测误差分布

表 1 GP 模型与其他方法对 Lorenz 时间序列单步预测的预测性能比较

预测模型方法	训练集 $E_{\text{NRMSE}}$	测试集 $E_{\text{NRMSE}}$
GP (使用 (16) 式核函数)	0.0053	0.0138
GP (使用 (17) 式核函数)	$8.0467 \times 10^{-5}$	0.0124
GP (使用 (18) 式核函数)	0.0086	0.0122
常规 GP 模型 (高斯 RBF 核函数)	0.0065	0.0142
$\epsilon$ -SVR ( $\epsilon = 0.01, C = 3, \sigma^2 = 2.2$ )	0.0091	0.0667
$\nu$ -SVR ( $\nu = 0.5, C = 3, \sigma^2 = 2.2$ )	0.0058	0.0545
LS-SVM ( $\gamma = 8500, \sigma^2 = 3$ )	0.0043	0.0244
OLS-RBF (150 个中心)	0.0041	0.0362

### 3.2. Mackey-Glass 混沌时间序列的提前 6 步和提前 84 步直接预测

时滞混沌系统 Mackey-Glass 方程<sup>[29]</sup> 可用来检验非线性动态系统模型的性能,用于描述白血病发病时,血液中白细胞的数量变化. 描述其混沌吸引子的动力学方程模型为

$$\frac{dy(t)}{dt} = \frac{0.2y(t-t_\Delta)}{1+y^{10}(t-t_\Delta)} - 0.1y(t), \quad (27)$$

其中  $t_\Delta$  为时滞参数. 由于白细胞新陈代谢生理机理方面的原因,该方程表现出某种周期性与混沌特征.  $t_\Delta > 17$  时呈现混沌性,  $t_\Delta$  值越大,混沌程度越高. 用四阶龙格-库塔法给出(27)式的数值解,得到时间序列  $y(t)$ , 其中  $t_\Delta = 17$ , 所用时间步长为 0.1, 初始条件  $y(0) = 1.2$ . 从 Mackey-Glass 时间序列中,抽取 1000 对输入输出数据,其中  $t = 118$  到 1117. 前 500 个数据用作训练数据集,其余的 500 个数据用作测试数据集,构造  $\hat{y}(n+h) = f(\mathbf{y}_R(n))$  的预测模型,分别令  $h=6$  及  $h=84$ ,作提前 6 步预测、提前 84 步直接预测. 在(23)式中,置  $D=4, \tau=6$ ,预测模型的输入与输出映射为

$$\hat{y}(t+h) = f(y(t-18), y(t-12), y(t-6), y(t)). \quad (28)$$

这也便于与以往文献比较. 利用复合协方差函数(16),(17),(18)式建立零均值 GP 预测模型. 在整个训练、测试数据集上,基于(17)式的复合协方差函数的 GP 模型对 Mackey-Glass 混沌时间序列提前 6 步预测结果和实际输出值(从  $t = 124$  到 1123)如图 3 所示,其各点的预测误差分布如图 4 所示. 从图 3 可以看出,无论在训练数据还是测试数据集上,GP 模型对 Mackey-Glass 混沌时间序列提前 6 步预测输出与实际输出几乎相同,二者的差别如图 4 所示,只能够在很小的数量级上才能区分.

基于(16)式的复合协方差函数的 GP 模型对  $t = 118$  到 1117 的 Mackey-Glass 混沌时间序列作提前 84 步直接预测结果和时间序列实际输出值(从  $t = 202$  到 1201)如图 5 所示,其各点预测误差分布如图 6 所示. 从图 5 中可以看出,无论在训练数据还是测试数据集上,GP 模型对 Mackey-Glass 混沌时间序列 84 步直接预测输出与实际输出符合得很好,因此,预测模型具有很好的推广性.

表 2 给出对  $t = 124$  到 1123 的 Mackey-Glass 时间序列分别做提前 6 步预测与提前 84 步直接预测时,所建立的 GP 预测模型与常规 GP 模型、自适应

神经模糊系统<sup>[10]</sup> (ANFIS),采用 OLS 学习算法的 RBF 神经网络(150 个中心),基于 RBF 核函数的 LS-SVM,  $\varepsilon$ -SVR,  $\nu$ -SVR 等方法的预测性能比较. ANFIS 的输入隶属度函数个数选取与文献[9]相同.  $\varepsilon$ -SVR 的参数选取按照文献[28]的经验选取,  $\varepsilon = 0.01$  正则化常数  $C = 20$ ,核参数  $\sigma^2 = 3$ . 取  $\nu$ -SVR 的  $\nu = 0.5$ ,正则化常数  $C = 3$ ,核参数  $\sigma^2 = 6$ . LS-SVM 的取值按照基于单纯形搜索的 5 折交叉验证方法选取,参数  $\gamma = 59595$ ,  $\sigma^2 = 0.9$ . OLS-RBF 选取 150 个中心.

从表 2 的比较结果看出,基于复合协方差函数的 GP 模型在测试集和训练集上的预测性能略好于其他方法,LS-SVM,  $\varepsilon$ -SVR,  $\nu$ -SVR 也达到了较好的效果,但这是在其参数仔细调谐后取得的,自适应神经模糊系统<sup>[9]</sup> (ANFIS)虽然性能也不错,但随着输入维数的增加,隶属度函数的数目呈几何级数增长,为训练带来困难. GP 模型是概率核学习机,其超参数的选取是基于训练数据集学习,优化其对数“边际似然性函数”获得的,对 GP 模型的训练就是对均值函数“超参数”以及些方差函数的“超参数”选取的过程,同时 GP 模型还能给出预测的误差带,这些特性也显示出 GP 模型良好的预测性能和稳健性.

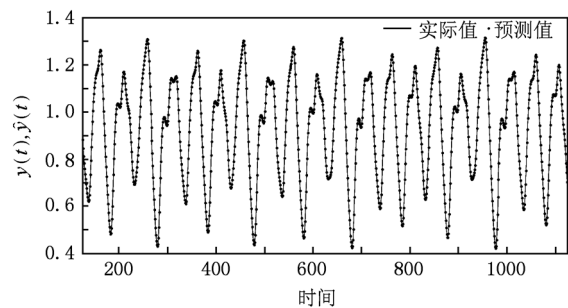


图 3  $t = 124$  到 1123 的 Mackey-Glass 混沌时间序列实际输出(实线)与提前 6 步预测输出(点线)

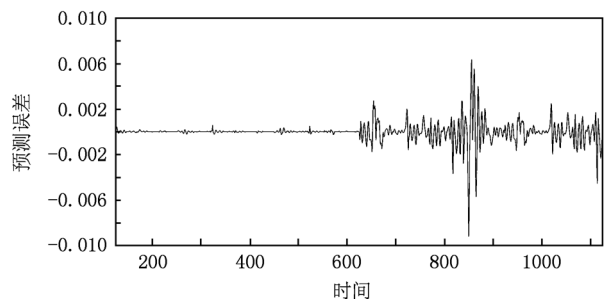


图 4  $t = 124$  到 1123 的 Mackey-Glass 混沌时间序列各点预测误差分布

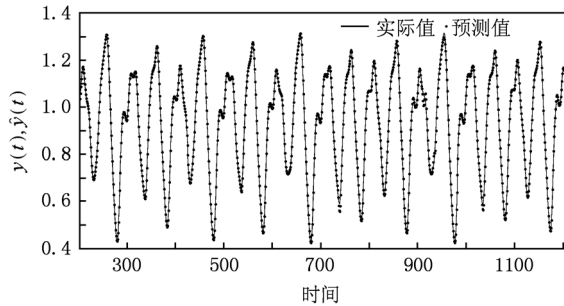


图5  $t = 202$  到  $1201$  的 Mackey-Glass 混沌时间序列实际输出 (实线) 与提前 84 步直接预测输出 (点线)

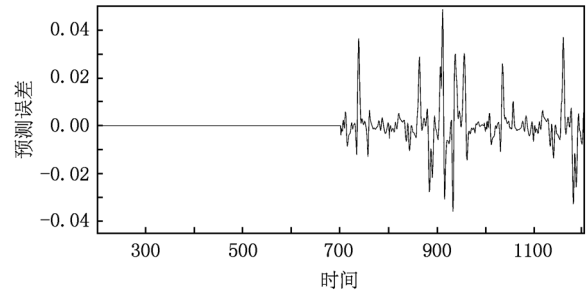


图6  $t = 202$  到  $1201$  的 Mackey-Glass 混沌时间序列各点预测误差分布

表2 GP 模型与其他预测模型对 Mackey-Glass 时间序列提前多步预测的性能比较

预测模型方法	提前 6 步预测的训	提前 6 步预测的测	提前 84 步预测的训	提前 84 步预测的测
	练集 $E_{\text{NRMSE}}$	试集 $E_{\text{NRMSE}}$	练集 $E_{\text{NRMSE}}$	试集 $E_{\text{NRMSE}}$
GP (式(16)核函数)	$4.0860 \times 10^{-4}$	0.0063	$5.2684 \times 10^{-5}$	0.0803
GP (式(17)核函数)	$3.8916 \times 10^{-4}$	0.0058	$1.0607 \times 10^{-5}$	0.0812
GP (式(18)核函数)	$4.0083 \times 10^{-4}$	0.0058	0.0027	0.0528
常规 GP 模型 (高斯 RBF 核函数)	$5.3814 \times 10^{-4}$	0.0062	0.0035	0.0607
$\epsilon$ -SVR ( $\epsilon = 0.01, C = 20, \sigma^2 = 1.5$ )	0.0080	0.0076	0.0397	0.1173
$\nu$ -SVR ( $\nu = 0.5, C = 3, \sigma^2 = 6$ )	0.0076	0.0071	0.0939	0.1846
LS-SVM ( $\gamma = 59595, \sigma^2 = 0.9$ )	$9.9877 \times 10^{-4}$	0.0016	0.0053	0.0767
OLS-RBF (150 个中心)	0.0359	0.0343	0.0158	0.1090
ANFIS <sup>[9]</sup>	0.0071	0.0067	0.0822	0.2294

### 3.3. Santa Fe 混沌时间序列的单步与 100 步迭代预测

Santa Fe 混沌时间序列<sup>[14]</sup>是一个远红外 NH<sub>3</sub> 激光器产生的混沌振荡的强度波形,由于它的混沌动态行为只表现出三个重要的剧烈变化,使得它成为很难预测的时间序列,如图 7 所示. 它的混沌特性属于强 Lorenz 混沌系统和弱 Rossler 混沌系统之间. 对 Santa Fe 混沌时间序列计算得最优的嵌入维<sup>[24]</sup>  $D_E = 9$ , 归一化嵌入延迟<sup>[25]</sup>  $\tau = 2$ , 由(24)式可知  $q = 18$ , 令  $h = 1$ , 在训练数据集上可构造  $\hat{y}(n+h) = f(\mathbf{y}_R(n))$  的单步预测模型, 用输入为 18 维的向量  $\mathbf{y}_R(n)$  预测  $y(n+1)$ , 训练数据为该时间序列的前 1000 个数据. 测试数据集为后续的 100 个数据. 测试时分别使用单步预测和迭代预测方法对该混沌时间序列进行单步预测以及进行 100 步的提前预测.

迭代预测的方法就是在训练数据集上首先用单步预测方法进行训练, 训练结束后, 先用混沌吸

引子中的一个点送给已训练好的预测用模型, 然后把模型输出反馈给模型的输入端构成一个闭环自激系统, 使用递归的方式迭代预测输出. 在本例中, 即利用训练好的 GP 预测模型, 使用迭代预测方式预测后续的 100 个数据点.

为说明 GP 模型预测的稳健性, 分别选取(16), (17), (18)式的复合协方差函数建立零均值 GP 预测模型. 图 8 给出了基于(18)式的复合协方差函数的 GP 模型, 对  $t = 1001$  到  $t = 1100$  的 Santa Fe 时间序列进行单步预测的输出结果与时间序列实际输出结果的比较曲线. 图 9 则给出了应用(18)式的复合协方差函数 GP 模型对该时间序列进行 100 步迭代预测的结果与时间序列实际输出结果的比较曲线. 从图 8、图 9 中可以看出, GP 模型对 Santa Fe 时间序列进行单步预测与多步迭代预测的效果均较好, 预测模型具有良好的推广性. 表 3 给出了对  $t = 1001$  到  $t = 1100$  的 Santa Fe 时间序列分别做单步预测和 100 步迭代预测时, 所建立的基于不同复合协方差函数的 GP 预测模型与常规 GP 模型、基于 RBF

核函数的 LS-SVM,  $\epsilon$ -SVR,  $\nu$ -SVR, OLS-RBF 神经网络(150 个中心)等预测模型方法的性能比较.  $\epsilon$ -SVR 的参数选取交叉验证方法选取,  $\epsilon = 0.01$ , 正则化常数  $C = 32$ , 核参数  $\sigma^2 = 16$ .  $\nu$ -SVR 取  $\nu = 0.5$ , 其余参数与  $\epsilon$ -SVR 相同. LS-SVM 的取值按照 5 折交叉验证方法选取, 参数  $\gamma = 200$ ,  $\sigma^2 = 13$ . OLS-RBF 选取 150 个中心. 从表 3 的比较结果看出, 复合协方差函数(18)式构成的 GP 模型, 无论在测试集和训练集上, 其预测性能均略好于常规 GP 方法和其他模型. 而其他基于复合协方差函数的 GP

模型, 也均能给出较好的性能, 其超参数是通过训练数据集的学习获取的, LS-SVM,  $\epsilon$ -SVR,  $\nu$ -SVR 在其参数仔细调谐后也可达到较好的效果, 这也显示出 GP 模型良好的预测性能和稳健性.

表 3 GP 模型与其他预测模型对  $n = 1000$  到 1100 的 Santa Fe 时间序列预测的性能比较

预测模型方法	单步预测的测试集 $E_{\text{NRMSE}}$	100 步迭代预测的测试集 $E_{\text{NRMSE}}$
GP ((16)式核函数)	0.0648	0.7586
GP ((17)式核函数)	0.0645	0.7611
GP ((18)式核函数)	0.0875	0.2177
常规 GP 模型(高斯 RBF 核函数)	0.0991	0.2657
LS-SVM( $\gamma = 200, \sigma^2 = 13$ )	0.0994	0.2736
$\epsilon$ -SVR ( $\epsilon = 0.01, C = 32, \sigma^2 = 16$ )	0.1432	0.7234
$\nu$ -SVR ( $\nu = 0.5, C = 32, \sigma^2 = 16$ )	0.1901	0.7073
OLS-RBF (150 个中心)	0.1030	0.3202

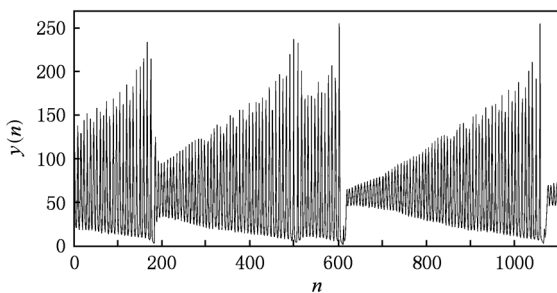


图 7 远红外 NH<sub>3</sub> 激光器产生的 Santa Fe 混沌时间序列

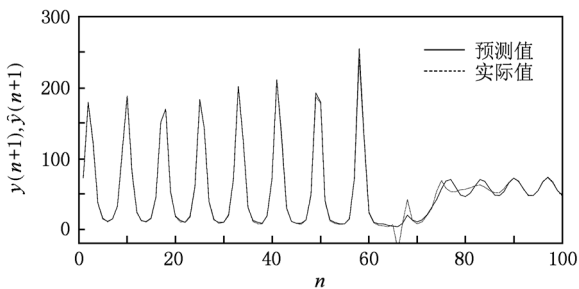


图 8  $n = 1000$  到 1100 的 Santa Fe 混沌时间序列的实际输出(实线)与单步预测输出(虚线)

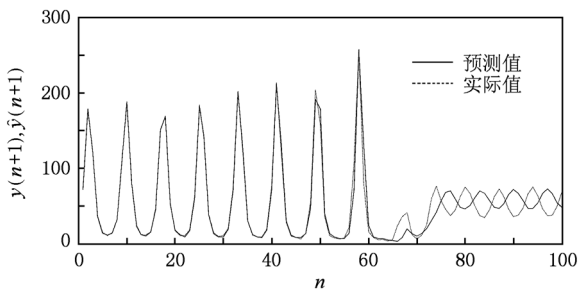


图 9 Santa Fe 混沌时间序列的实际输出(实线)与 100 步提前预测输出(虚线)

#### 4. 结 论

针对混沌时间序列单步与多步预测, 本文提出基于复合协方差函数的 GP 模型预测方法. 从泛函空间分析 GP 模型, 它更好地展现输入与输出数据之间的非线性映射关系, 进一步提高了参数回归与非参数回归方法的泛化能力, 模型的“解释性”好. GP 模型方法只需选择不同的协方差函数, 本文选择具有平稳性和非平稳性的复合协方差核函数, 保持了对先验函数光滑性的概率解释优点. 其学习过程就是基于非线性优化算法自适应地确定其模型“超参数”的过程. 具有复合协方差函数的 GP 模型应用于混沌时间序列单步与多步预测中, 并与常规 GP 模型, SVM, LS-SVM, RBF 神经网络等模型方法进行比较, 多个混沌时间序列实例的仿真结果表明了 GP 模型方法的有效性和稳健性. GP 模型方法保持利用了贝叶斯建模的优点, 是实用的、强有力的针对数据建模的贝叶斯工具. 基于贝叶斯方法的 GP 建模作为神经网络与 SVM 方法的延伸与替代, 在复杂非线性动力系统的建模、辨识与控制方面, 具有一定的优越性和很好的应用潜力.



- [1] Abarbanel H D I 1996 *Analysis of Observed Chaotic Data* (New York: Springer-Verlag)
- [2] Takens F 1981 *Dynamical Systems and Turbulence* **898** 366
- [3] Haykin S, Principe J 1998 *IEEE Signal Processing Magazine* **15** 66
- [4] Jaeger H 2004 *Science* **308** 78
- [5] Schilling R J, Carroll J J 2001 *IEEE Trans. Neural Networks* **12** 1
- [6] Li J, Liu J H 2005 *Acta Phys. Sin.* **54** 4569 (in Chinese) [李军、刘君华 2005 物理学报 **54** 4569]
- [7] Ma Q L, Zheng Q L, Peng H Z, Tan W, Qin J W 2008 *Chin. Phys. B* **17** 536
- [8] Han M, Shi Z W, Guo W 2007 *Acta Phys. Sin.* **56** 43 (in Chinese) [韩敏、史志伟、郭伟 2007 物理学报 **56** 0043]
- [9] Jang J S R 1993 *IEEE Trans. Syst., Man, Cybern.* **23** 665
- [10] Cui W Z, Zhu C C, Bao W X, Liu J H 2005 *Chin. Phys.* **14** 922
- [11] Ye M Y, Wang X D, Zhang H R 2005 *Acta Phys. Sin.* **54** 2568 (in Chinese) [叶美盈、汪晓东、张浩然 2005 物理学报 **54** 2568]
- [12] Li J, Dong H Y 2008 *Acta Phys. Sin.* **57** 4756 (in Chinese) [李军、董海鹰 2008 物理学报 **57** 4756]
- [13] Ding G, Zhong S S, Li Y 2008 *Chin. Phys. B* **17** 1998
- [14] Weigend A S, Gershenfeld N A 1994 *Time Series Prediction: forecasting the future and understanding the past* (Harlow, UK: Addison Wesley)
- [15] Williams C K I, Barber D 1998 *IEEE Trans. PAMI* **20** 1342
- [16] Seeger M 2004 *International Journal of Neural System* **14** 69
- [17] Rasmussen C E, Williams C K I 2006 *Gaussian Processes for Machine Learning* (Cambridge, MA: The MIT Press)
- [18] MacKay D J C 1999 *Neural Computation* **11** 1035
- [19] Gregoric G, Lightbody G 2009 *Engineering Applications of Artificial Intelligence* **22** 522
- [20] Cristianini N, Shawe-Taylor J 2000 *An introduction to support vector machines and other kernel-based learning* (Cambridge, UK: Cambridge University Press)
- [21] Scholkopf B, Smola A J 2002 *Learning with Kernels* (Cambridge MA: MIT Press)
- [22] Neal R M 1996 *Bayesian Learning for Neural Networks* (New York: Springer-Verlag)
- [23] Lorenz E N 1963 *J. Atmos. Sciences* **20** 130
- [24] Kennel M B, Brown R, Abarbanel H D I 1992 *Phys. Rev. A* **45** 3403
- [25] Fraser A M 1989 *IEEE Trans. on Information Theory* **35** 245
- [26] Suykens J A K, Gestel T V, Brabanter J De, Moor B De, Vandewalle J 2002 *Least Squares Support Vector Machines* (Singapore: World Scientific Pub. Co.)
- [27] Chen S, Cowan C F N, Grant P M 1991 *IEEE Trans. Neural Networks* **2** 302
- [28] Scholkopf B, Burges C J C, Smola A J 1999 *Advances in Kernel Methods — Support Vector Learning* (Cambridge MA: MIT Press) 211
- [29] Mackey M C, Glass L 1977 *Science* **197** 287

# Single-step and multiple-step prediction of chaotic time series using Gaussian process model<sup>\*</sup>

Li Jun<sup>†</sup> Zhang You-Peng

(School of Automation & Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

(Received 13 September 2010; revised manuscript received 4 November 2010)

## Abstract

For the chaotic time series single-step and multi-step prediction, Gaussian processes (GPs) method based on composite covariance function is proposed. GP priors over functions are determined mainly by covariance functions, and through learning from training data sets, all hyperparameters that define the covariance function and mean function can be estimated by using matrix operations and optimal algorithms within evidence maximum bayesian framework. As a probabilistic kernel machine, the number of tunable parameters for a GP model is greatly reduced compared with those for neural networks and fuzzy models. GP models with different composite covariance functions are applied to chaotic time series single-step and multi-step ahead prediction and compared with other models such as standard GP model with single covariance function, standard support vector machines, least square support vector machine, radial basis functional (RBF) neural networks, *etc.* Simulation results reveal that GP method with using different composite covariance functions can be used to accurately predict the chaotic time series and show stable performance with robustness. Hence, it provides an effective approach to studying the properties of complex nonlinear system modeling and control.

**Keywords:** Gaussian process, chaotic time series, prediction, comparison of model

**PACS:** 05.45.Tp, 02.50.Ey, 05.45.-a, 05.45.Gg

---

<sup>\*</sup> Project supported by the Natural Science Foundation of Gansu Province, China (Grant No. 0803RJZA023).

<sup>†</sup> E-mail: lijun691201@mail.lzjtu.cn