

12-12-2019

Chaotic Time Series Prediction Based on Gaussian Processes Mixture

Zhenjie Feng

1. School of Computer and Information Engineering, Anyang Normal University, Anyang 455000, China; ;

Fan Yu

2. Weinan Meteorological Bureau, Weinan 714000, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Chaotic Time Series Prediction Based on Gaussian Processes Mixture

Abstract

Abstract: Aiming at the problem that the existing learning algorithms of Gaussian processes mixture (GPM) model, such as Markov Chain Monte Carlo (MCMC), variation or leave one out, have high computational complexity, *a hidden variables posterior hard-cut iterative training algorithm is proposed, which simplifies the training process of the model. The GPM model based on the proposed algorithm is applied to chaotic time series prediction.* The effects of embedding dimension, time delay, learning sample number, and testing sample numbers on predictive ability are discussed. It is demonstrated by the experimental results that the prediction of the GPM model is more accurate than SVM, GP and RBF network, and the training speed of GPM model falls in between RBF network, GP model, and SVM.

Keywords

Gaussian processes mixtures model, chaotic time series, prediction, machine learning

Recommended Citation

Feng Zhenjie, Fan Yu. Chaotic Time Series Prediction Based on Gaussian Processes Mixture[J]. Journal of System Simulation, 2019, 31(7): 1387-1396.

混沌时间序列的高斯过程混合模型预测

冯振杰¹, 樊煜²

(1. 安阳师范学院计算机与信息工程学院, 河南 安阳 455000; 2. 陕西省渭南市气象局, 陕西 渭南 714000)

摘要: 高斯过程混合(Gaussian Processes Mixture, GPM)模型现有的学习算法如马尔科夫链蒙特卡洛法、变分法或留一法等, 计算复杂度偏高, 提出一种隐变量后验硬划分迭代学习算法, 简化模型的学习过程, 基于该算法将GPM模型用于混沌时间序列预测, 并讨论嵌入维、时间延迟、学习样本和测试样本数目等参数对预测性能的影响。实验结果表明, GPM模型预测精度高于支持向量机(Support Vector Machine, SVM)、高斯过程(Gaussian Process, GP)和径向基(Radical Basis Function, RBF)网络, 学习速度介于RBF网络、GP和SVM之间。

关键词: 高斯过程混合模型; 混沌时间序列; 预测; 机器学习

中图分类号: TM714 文献标识码: A 文章编号: 1004-731X(2019)07-1387-10

DOI: 10.16182/j.issn1004731x.joss.18-0543

Chaotic Time Series Prediction Based on Gaussian Processes Mixture

Feng Zhenjie¹, Fan Yu²

(1. School of Computer and Information Engineering, Anyang Normal University, Anyang 455000, China;

2. Weinan Meteorological Bureau, Weinan 714000, China)

Abstract: Aiming at the problem that the existing learning algorithms of Gaussian processes mixture (GPM) model, such as Markov Chain Monte Carlo (MCMC), variation or leave one out, have high computational complexity, a hidden variables posterior hard-cut iterative training algorithm is proposed, which simplifies the training process of the model. The GPM model based on the proposed algorithm is applied to chaotic time series prediction. The effects of embedding dimension, time delay, learning sample number, and testing sample numbers on predictive ability are discussed. It is demonstrated by the experimental results that the prediction of the GPM model is more accurate than SVM, GP and RBF network, and the training speed of GPM model falls in between RBF network, GP model, and SVM.

Keywords: Gaussian processes mixtures model; chaotic time series; prediction; machine learning

引言

混沌系统是一种可以从无序中产生出有序的高维复杂非线性动力系统。混沌固含的确定性因素使得貌似随机的混沌系统具有可预测性。但是描述

混沌系统随时间演化的状态变量往往未知, 在实际中能观测到的仅是一组混沌时间序列。根据 Takens 相空间延迟重构定理, 混沌系统的内在规律可以通过混沌时间序列进行重构预测^[1]。

如何选择或构造预测模型是混沌时间序列预测中的关键问题之一。近年来这些学习模型已被用于预测并取得良好效果。例如, 崔万照等^[2]将支持向量机(Support Vector Machine, SVM)用于混沌时间序列预测。最小二乘支持向量机(Least Square



收稿日期: 2018-08-08 修回日期: 2018-12-24;
基金项目: 河南省高等学校重点科研项目(17A510007);
作者简介: 冯振杰(1981-), 男, 河南安阳, 博士生, 讲师, 研究方向为无线通信技术、多维信号处理、机器学习等。

<http://www.china-simulation.com>

• 1387 •

Support Vector Machine, LS-SVM)是 SVM 的一种改进, 文献[3-4]曾提出一种多维递归 LS-SVM 模型并用于混沌时间序列预测。唐舟进等^[5]在预测中采用了一种基于迭代误差补偿的 LS-SVM 模型。田中大等^[6]将 LS-SVM 的核函数置换为多项式函数和径向基函数的组合, 因而预测精度更高。叶美盈等^[7]提出在线 LS-SVM, 以适应混沌系统参数随时间变化。Arash 等^[8]在预测中将 LS-SVM 和模糊神经网络结合, Shi 等^[9]则提出支持向量回声状态机的概念。尽管 SVM 和 LS-SVM 在实际预测中具有一定优势, 但不能输出预测置信区间。另外虽然 SVM 具有序贯最小优化等快速学习算法, 但当学习样本数目偏多时预测速度仍然偏慢。

人工神经网络(Artificial Neural Network, ANN)已被广泛用于混沌时间序列预测。例如 Li 等^[10]在预测中将反向传播(Back Propagation, BP)神经网络和遗传算法结合起来, 用遗传算法优化神经网络的权值和阈值。宋彤等^[11]将大规模递归神经网络和小波回声状态网络结合起来预测。Zhang^[12]提出一种带有时间延迟坐标的模糊神经网络模型并用于混沌时间序列预测。Yin 等^[13]使用径向基(Radical Basis Function, RBF)神经网络预测混沌时间序列。李松等^[14]利用改进粒子群算法优化 BP 神经网络后用于预测。但是 ANN 在实际预测中也存在一些不足, 例如网络拓扑结构需要人工指定, 初值选择对网络学习算法收敛性影响较大等。

高斯过程(Gaussian Process, GP)模型是核机器学习与贝叶斯推理相结合的典范^[15-16], 已被成功用于混沌时间序列预测^[17]。文献[18-19]基于小波分析理论, 提出一种多尺度 GP 模型并用于混沌时间序列预测。但是单个 GP 模型很难反映时间序列在不同时段间的细微差异。而高斯过程混合高斯过程混合(Gaussian Processes Mixture, GPM)模型是多个 GP 模型的混合, 可以分别用不同的混合成分描述序列的不同时段, 能细致刻画不同时段间的细微差异。目前 GPM 模型有 3 类常见学习算法, 即马尔科夫链蒙特卡洛法(Markov Chain Monte Carlo,

MCMC)、变分法和留一法。但是 MCMC 采用的抽样策略收敛慢且收敛条件难于判断。变分法的独立性假设比较粗略且对后验概率的逼近速率较慢。留一法是一种特殊的启发式 EM 算法, 但是受初值影响大且迭代收敛慢。

正是考虑到 GPM 模型兼具 GP、SVM 与 RBF 等模型的优势, 同时 GPM 现有的学习算法计算复杂度偏高, 本文提出一种 GPM 模型的隐变量后验硬划分迭代学习算法, 进而将之用于混沌时间序列预测。Mackey-Glass 和 Rossler 序列的预测实验结果表明, GPM 模型是一种预测混沌时间序列的有效模型。

1 高斯过程混合模型的原理

1.1 GP 模型原理简要回顾

已知的是 N 个学习样本 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 及它们对应的目标值 t_1, t_2, \dots, t_N ; $f(\mathbf{x})$ 是待求回归函数, 由 $f(\mathbf{x})$ 可预测给定测试样本 \mathbf{x}_{N+1} 处的函数值 $f(\mathbf{x}_{N+1})$ 。假设 $f(\mathbf{x})$ 是一个高斯过程, 将目标向量 $\mathbf{t}_N = (t_1, t_2, \dots, t_N)^T$ 看作另一个高斯过程 $\mathbf{t}(\mathbf{x})$ 在 $\mathbf{X}_N = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ 处的实现, 假设 $\mathbf{t}(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\varepsilon}(\mathbf{x})$, 其中 $\boldsymbol{\varepsilon}(\mathbf{x})$ 是一个方差为 σ_v^2 的高斯白噪过程。

GP 模型将 $f(\mathbf{x})$ 表示为:

$$f(\mathbf{x}) = \sum_{h=1}^H w_h \phi_h(\mathbf{x}) \quad (1)$$

并且假定随机向量 $\mathbf{w} = (w_1, w_2, \dots, w_H)^T$ 满足 $\mathbf{w} \sim N(0, \sigma_w^2 \mathbf{I}_H)$ 。设 $\mathbf{f}_N = (f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N))^T$ 是 $f(\mathbf{x})$ 在 $\{\mathbf{x}_n\}_{n=1}^N$ 处的 N 个取值, 可推知 $\mathbf{f}_N \sim N(0, \sigma_w^2 \mathbf{R}\mathbf{R}^T)$, \mathbf{R} 的元素 $R_{nh} = \phi_h(\mathbf{x}_n)$, 记 $\mathbf{Q}_N = \sigma_w^2 \mathbf{R}\mathbf{R}^T$ 。可推知

$$\mathbf{t}_N \sim N(0, \mathbf{C}_N) \quad (2)$$

式中: 矩阵 $\mathbf{C}_N = \mathbf{Q}_N + \sigma_v^2 \mathbf{I}_N$, 矩阵的每个元素 $C(\mathbf{x}_n, \mathbf{x}_{n'}) = Q(\mathbf{x}_n, \mathbf{x}_{n'}) + \delta_{nn'}$ (当 $n = n'$ 时 $\delta_{nn'} = 1$, 否则 $\delta_{nn'} = 0$)。其中矩阵 \mathbf{Q}_N 由下面的协方差函数决定

$$Q(\mathbf{x}_n, \mathbf{x}_{n'}) = \sigma_w^2 \sum_{h=1}^H \phi_h(\mathbf{x}_n) \phi_h(\mathbf{x}_{n'}) \quad (3)$$

当给定测试样本 \mathbf{x}_{N+1} 时有

$$P(t_{N+1} | \mathbf{t}_N) = P(t_{N+1}, \mathbf{t}_N) / P(\mathbf{t}_N) \quad (4)$$

由式(2)和 $t(\mathbf{x})$ 的性质推知式(4)中等号右边的分母与分子均服从高斯分布, 因此 $P(t_{N+1} | t_N)$ 也服从高斯分布, 并设式(4)中分子的表达式为

$$P(t_{N+1}, t_N) \propto \exp\left[-\frac{1}{2} [t_N \quad t_{N+1}] \mathbf{C}_{N+1}^{-1} \begin{bmatrix} t_N \\ t_{N+1} \end{bmatrix}\right] \quad (5)$$

现将矩阵 \mathbf{C}_{N+1} 分块, 并获得它的逆。将式(2)和(5)代入到式(4)后得到

$$P(t_{N+1} | t_N) \propto \exp\left(-\frac{(t_{N+1} - \hat{t}_{N+1})^2}{2\sigma_{\hat{t}_{N+1}}^2}\right) \quad (6)$$

式中: $\hat{t}_{N+1} = \mathbf{k}^T \mathbf{C}_N^{-1} t_N$, $\sigma_{\hat{t}_{N+1}}^2 = \kappa - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k}$, 其中 \mathbf{C}_N^{-1} 为逆矩阵。最后令 $f(\mathbf{x}_{N+1}) = \hat{t}_{N+1}$, 即取 $P(t_{N+1} | t_N)$ 分布的均值作为 GP 模型在 \mathbf{x}_{N+1} 处的预测值。

1.2 GPM 模型的基本原理

GPM 模型由多个 GP 模型混合而成, 把学习样本分成若干组, 每组用一个 GP 模型来描述。为表述方便, 用隐变量 z_i 表示组别, $z_i = c$ 表示第 i 个样本属于第 c 组, $c = 1, 2, \dots, C$, 其中 C 是 GP 模型总个数。现将第 c 组样本的输入矩阵和目标向量记为 $\mathbf{X}_c = [\mathbf{x}_{k_c(1)}, \mathbf{x}_{k_c(2)}, \dots, \mathbf{x}_{k_c(N_c)}]^T$ 和 $\mathbf{T}_c = [t_{k_c(1)}, t_{k_c(2)}, \dots, t_{k_c(N_c)}]^T$, 其中 $k_c(\cdot)$ 和 N_c 表示第 c 个 GP 模型中学习样本的标号和个数, 则由 GP 模型的定义式(2)可得

$$\mathbf{T}_c | \mathbf{X}_c \sim N(\mathbf{0}, \mathbf{Q}_c + \sigma_{v,c}^2 \mathbf{I}_{N_c}) \quad (7)$$

式中: 下标 c 指代第 c 个 GP 模型, 核矩阵 \mathbf{Q}_c 的每个元素由如下协方差函数给出

$$Q_c(\mathbf{x}_n, \mathbf{x}_{n'}) = Q(\mathbf{x}_n, \mathbf{x}_{n'} | \theta_c) \quad (8)$$

式中: θ_c 为第 c 个 GP 模型的核参数。各 GP 模型的核参数并不相同, 以适应不同的学习样本模式。在工程应用中常使用的协方差函数是平方指数函数

$$Q(\mathbf{x}_n, \mathbf{x}_{n'} | \theta_c) = l_c^2 \exp(-\|\mathbf{x}_n - \mathbf{x}_{n'}\|^2 / 2\delta_c^2) \quad (9)$$

考虑到现有 GPM 模型结构复杂, 涉及较多的待估参数, 为了使模型更加实用, 文献[20]提出了一种精简的生成 GPM 模型。该模型首先采用如下多项分布门限函数生成隐变量 z_i 。

$$P(z_i = c) = \pi_c \quad (10)$$

然后每个成分按照相应的高斯分布独立生成

对应的输入 \mathbf{x}_i

$$(\mathbf{x}_i | z_i = c) \sim N(\boldsymbol{\mu}_c, \mathbf{S}_c) \quad (11)$$

当隐变量和输入都确定后, 由 GPM 的定义式(7)即得各成分的输出 T_c 。

2 基于高斯过程混合模型的混沌时间序列预测

2.1 GPM 模型的隐变量后验硬划分迭代学习算法

当 GPM 模型用于混沌时间序列时, 首先进行模型学习, 然后基于学习结果对给定样本进行预测。GPM 模型的常见学习算法有马尔科夫链蒙特卡洛法、变分贝叶斯法和期望最大化(Expectation Maximization, M)算法等。但前两种算法计算复杂度较高, 本文在 EM 算法的基础上, 提出 GPM 模型的一种隐变量后验硬划分迭代学习算法, 并将之用于混沌时间序列预测。

由于 GPM 模型中各样本的目标值是互相关的, 若用传统的 EM 学习算法, 算法中 Q 函数的计算复杂度为指数量级, 因此需要采用一定的近似策略。本文提出的硬划分迭代学习算法在 E 步根据最大后验概率准则分配样本, 在 M 步基于分配至各组的样本对各组参数分别进行最大似然估计。具体实现步骤如下:

第 1 步: 通过聚类算法将学习样本 $\{(\mathbf{x}_i, t_i)\}_{i=1}^N$ 分成 C 组。

第 2 步: 参与混合的各 GP 模型基于最大似然估计独立学习, 其中混合比例系数 π_c 以及输入分布的均值 $\boldsymbol{\mu}_c$ 与协方差阵 \mathbf{S}_c 分别用如下解析式求解:

$$\pi_c \leftarrow N_c / N \quad (12)$$

$$\boldsymbol{\mu}_c \leftarrow \frac{1}{N_c} \sum_{i: z_i=c} \mathbf{x}_i \quad (13)$$

$$\mathbf{S}_c \leftarrow \frac{1}{N_c} \sum_{i: z_i=c} (\mathbf{x}_i - \boldsymbol{\mu}_c)(\mathbf{x}_i - \boldsymbol{\mu}_c)^T \quad (14)$$

在式(7)基础上, 各 GP 模型的参数可以通过最大化如下所示的似然函数获得

$$\hat{\theta}_c, \hat{\sigma}_{v,c} = \arg \max_{\theta_c, \sigma_{v,c}} \ln N(T_c | \theta_c, Q_c(\theta_c) + \sigma_{v,c}^2 I_{N_c}) \quad (15)$$

第 3 步: 根据最大后验概率准则对学习样本重新指定组别

$$z_i \leftarrow \arg \max_{1 \leq c \leq C} P(z_i = c | \mathbf{x}_i, t_i) = \arg \max_{1 \leq c \leq C} \pi_c N(\mathbf{x}_i | \boldsymbol{\mu}_c, \mathbf{S}_c) N(t_i | 0, l_c^2 + \sigma_c^2) \quad (16)$$

如果重新指定的结果和上轮一致, 则迭代算法停止并输出最终结果; 否则返回第 2 步。

第 4 步: 上述参数的学习过程结束后, 对于给定的测试样本 \mathbf{x}_{N+1} , 若欲预测其对应的目标输出 t_{N+1} , 同样可以首先根据最大后验概率准则指定其组别

$$z = \arg \max_{1 \leq c \leq C} P(z^* = c | \mathbf{x}_{N+1}) = \arg \max_{1 \leq c \leq C} \pi_c N(\mathbf{x}_{N+1} | \boldsymbol{\mu}_c, \mathbf{S}_c) \quad (17)$$

然后将测试样本分配到第 z 组, 再由单个 GP 模型的预测式(6)可得 t_{N+1} 的预测分布。该预测式中所需学习样本为最后一次迭代时被分配到第 z 组的学习样本。

2.2 基于 GPM 模型的混沌时间序列预测过程

根据 Takens 相空间延迟重构定理, 将一维混沌时间序列嵌入到高维相空间中, 采用时延技术重构相空间, 相空间的演化能反映原混沌系统的演化规律^[1]。具体而言若给定一个混沌时间序列 $\{s(n)\}_{n=1}^{\infty}$, 确定时间序列重构相空间的嵌入维 d 和时延 τ 参数以后, 存在着一个光滑映射函数 $f: R^d \rightarrow R$, 且满足

$$s(i) = f[s(i-\tau), s(i-2\tau), \dots, s(i-d\tau)] \quad (18)$$

如果已知映射函数 f , 就可以用序列在 τ 时刻以前的 d 个值来预测当前目标值 $s(i)$ 。此时可用 GPM 模型来估计 f , 其中第 i 个样本是一个 d 维向量, 即 $\mathbf{x}_i = [s(i-\tau), s(i-2\tau), \dots, s(i-d\tau)]^T$, 该样本对应的目标值为 $y_i = s(i)$ 。

在预测中分别采用“均方根误差(RMSE)”和“预测误差(Err)”来度量预测准确度。同时利用 GPM 模型输出的预测置信区间来度量预测可信程度, 置信区间越窄预测结果越可信。RMSE 和 Err

这两个指标的定义式如下:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - \hat{t}_i)^2} \quad (19)$$

$$\text{Err} = t_i - \hat{t}_i \quad (20)$$

式中: t_i 和 \hat{t}_i 分别代表第 i 个测试样本的真实目标值和预测值。

在预测过程中嵌入维 d 和时延 τ 的取值较为关键。过大过小都影响预测精度和效率考虑到传统确定 d 和 τ 的时延互信息法和近邻法等比较耗时, 本文采用如下方式获得最优的 d 和 τ : 将 2 个参数视为二元组并在给定区间内逐一取值; 然后基于每组取值进行 GPM 模型学习并获得对应的 RMSE, 使 RMSE 最小的 d 和 τ 即为所求。

3 预测实例及结果分析比较

为了验证 GPM 模型预测混沌时间序列的性能, 本文分别使用该模型对 2 种典型的混沌时间序列即 Mackey-Glass 和 Rossler 序列进行预测, 并将之与 GP 模型、SVM 和 RBF 网络的预测结果进行比较。预测实验在一台华硕牌笔记本电脑上开展, CPU 为 Intel Core i3 2.53 GHz, 内存为 4 GB, 软件平台为 Matlab 2010 a。

3.1 Mackey-Glass 混沌时间序列预测

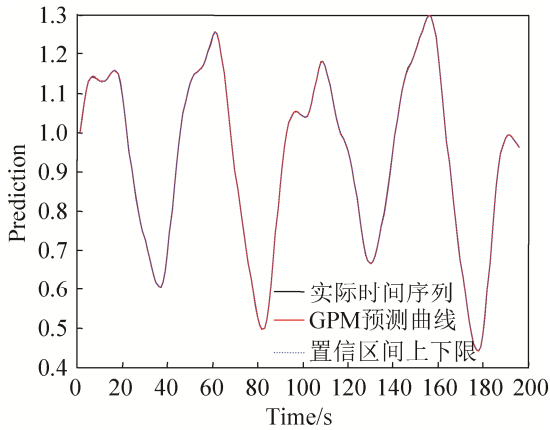
Mackey-Glass 是一种典型的混沌时间序列, 产生该序列的时滞微分方程如下:

$$\frac{ds(n)}{dn} = \frac{0.2s(n-\Delta)}{1+s^{10}(n-\Delta)} - 0.1s(n) \quad (21)$$

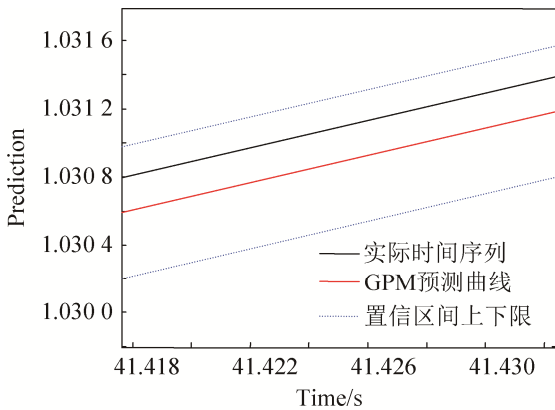
式中: Δ ($\Delta > 17$) 为时滞系数, 系统的混沌程度随着 Δ 的增大而变高。图 1 展示的是 Mackey-Glass 序列的 GPM 模型预测结果与预测置信区间。各预测参数取值分别为: 嵌入维 $d=5$, 时延 $\tau=1$, GPM 模型混合成分数 $C=2$ 。在预测中从序列的第 200 个采样点开始, 选取连续的 300 个点组合成学习样本, 并用接下来的 200 个点组合成测试样本。

从图 1(a)中可见预测序列和实际序列很吻合, 并且 GPM 模型能输出预测置信区间, 根据该置信区间能判断预测结果的可信程度。为了更清楚展示

预测结果, 将图中横坐标[41.418, 41.432]区域放大后显示于图 1(b)中, 这样就能清楚看到实际序列、预测曲线及预测置信区间上下限之间的关系。



(a) 预测结果与置信区间图



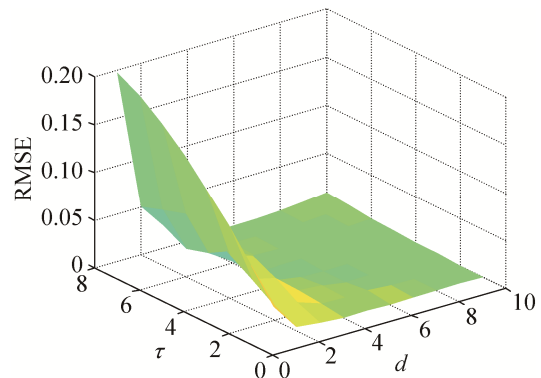
(b) 预测结果与置信区间局部区域放大

图 1 Mackey-Glass 序列的 GPM 模型预测结果与预测置信区间

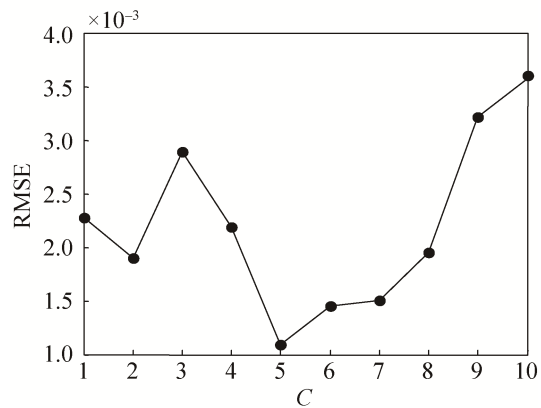
Fig. 1 GPM prediction results and the confidence interval of Mackey-Glass series

为了分析嵌入维 d 、时延 τ 、混合成分数 C 等参数对 GPM 模型预测精度的影响, 首先将待定参数组成三元组 (d, τ, C) , 并在给定区间 $[1, 10] \times [1, 8] \times [1, 10]$ 内逐一取值。然后基于每组取值进行 GPM 模型学习并获得对应的 RMSE。当固定 C 以后, RMSE 随 d 和 τ 的变化规律如图 2(a)所示, 由于 d 过大会导致计算量增加, d 过小会导致重构相空间不能准确刻画原混沌系统。时延 τ 过大会造成预测中丢失混沌序列信息, τ 过小会使相空间中各分量的冗余度增大。可看出当 $d > 4$ 时 RMSE 的值可达到最小,

并且 RMSE 随时延 τ 的变化亦不明显, 此时 GPM 模型能获得很好的预测性能。类似地当固定 $d=6$ 和 $\tau=7$ 以后, RMSE 随混合成分数 C 的变化如图 2(b)所示, 由于 GP 模型的个数影响预测精度, 从图中看出 $C=5$ 时 RMSE 达到最小, 过大或过小的混合成分数都会导致模型预测精度降低。



(a) 预测均方根误差随嵌入维和时延变化



(b) 预测均方根误差随混合成分数 C 变化

图 2 GPM 模型的预测均方根误差随嵌入维、时延、模型混合成分数的变化

Fig. 2 Variation of RMSE of GPM with embedding dimension, time delay and the number of mixed components

GP 模型^[17-19]、SVM^[2-9]和 RBF 网络^[13]均已被用于混沌时间序列预测。现比较上述 3 种模型和 GPM 间的性能。当固定嵌入维 $d=6$ 和时延 $\tau=4$ 以后, 将 Mackey-Glass 序列的 4 种模型预测结果显示于图 3(a)中。为了更清楚展示 4 者间的差异, 将图中横坐标 [72.10, 72.60] 区域放大后显示于图 3(b)中。

为了进一步分析嵌入维 d 和时延 τ 参数对 4 种模型预测性能的影响, 首先固定学习样本和测试样

本数目, 即选取 Mackey-Glass 序列第 501 个采样点至第 800 个采样点组合成学习样本, 紧接着的 200 个采样点组合成测试样本。然后在不断改变 d 和 τ 取值的前提下分别学习 4 种模型。在学习模型的内含参数(如 GPM 的混合成分数、SVM 的核宽度和正则化因子、RBF 的径向基个数等)均取最优值。现将 RMSE 和学习时间随 d 和 τ 变化的 10 组典型记录列于表 1 中, 表中还列出了 GPM 模型的混合成分数 C 的最优取值。

从表 1 中的 RMSE 值能看出:

(1) 对于大部分的 d 和 τ 取值组合, GPM 模型的预测均方根误差要小于其他模型;

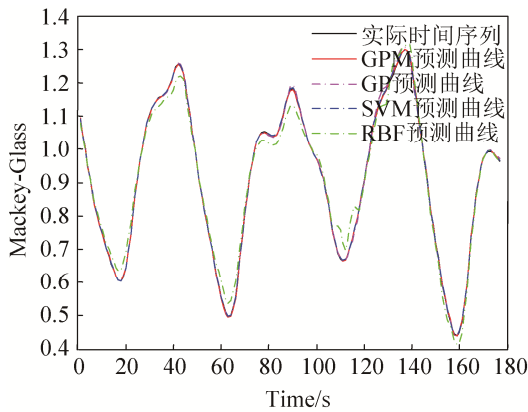
(2) 存在着某些 d 和 τ 的取值组合, 例如 $d=6, \tau=4$ 或 $d=7, \tau=3$ 时, 4 种模型的预测均方

误差均较低。

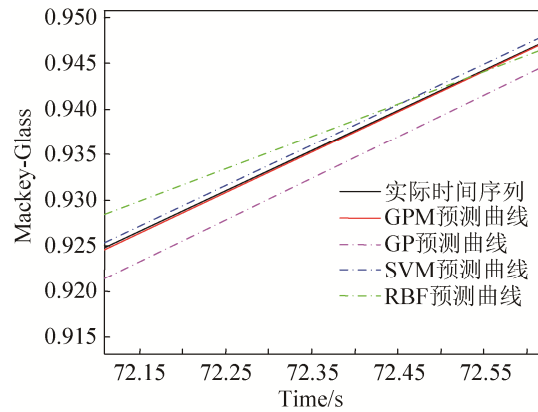
从表 1 中的模型学习时间能看出:

(1) 由于 RBF 网络拓扑结构简单, 故 RBF 网络的学习时间最短; 其次是 GP 模型, 再次是 GPM 模型, SVM 的学习时间最长。因 GPM 是由多个 GP 模型混合而成, 故耗时长于 GP 模型, SVM 训练时需要求解 2 次规划, 故耗时最长。

(2) 随着 d 取值逐步增大, 样本维数在逐步增高, SVM 的学习时间也随之增长。GPM 模型的学习时间虽然随 d 和 τ 取值变化的幅度较大, 但没有呈现出象 SVM 那样的规律, 可能是因为两种模型的学习方式不一样。



(a) 预测结果比较图



(b) 预测结果比较图局部区域放大

图 3 Mackey-Glass 序列的 4 种模型预测结果比较

Fig. 3 Comparison of the prediction results of Mackey-Glass series obtained by four models

表 1 预测 Mackey-Glass 序列时嵌入维和时延参数对 4 种模型的预测性能影响

Tab. 1 Influence of embedding dimension and delay parameters in predicting Mackey-Glass sequence on the predictive performance of 4 models

参数取值	GPM 模型	GP 模型	SVM	RBF 网络
$d=2, \tau=3$	0.031 0, 46.340 4 ($C=3$)	0.032 7, 8.140 1	0.032 2, 37.899 1	0.032 5, 0.900 6
$d=2, \tau=4$	0.045 9, 13.258 5 ($C=8$)	0.048 7, 8.713 6	0.047 7, 33.768 6	0.047 6, 1.111 0
$d=3, \tau=2$	0.011 0, 21.459 8 ($C=8$)	0.011 7, 9.845 6	0.010 9, 32.563 6	0.018 1, 1.382 6
$d=3, \tau=5$	0.031 6, 13.728 9 ($C=6$)	0.042 3, 8.540 2	0.034 6, 27.855 1	0.042 9, 0.976 2
$d=4, \tau=2$	0.003 1, 41.569 9 ($C=7$)	0.003 0, 11.050 3	0.003 7, 193.130 9	0.021 4, 3.177 9
$d=4, \tau=5$	0.001 1, 18.244 0 ($C=9$)	0.001 2, 9.995 0	0.001 2, 179.004 4	0.030 5, 4.902 4
$d=4, \tau=6$	0.002 1, 8.869 4 ($C=7$)	0.004 8, 6.760 4	0.002 0, 138.759 5	0.028 1, 4.750 2
$d=6, \tau=4$	0.000 4, 3.542 3 ($C=5$)	0.000 6, 2.366 5	0.000 4, 242.931 0	0.040 3, 4.904 9
$d=6, \tau=5$	0.000 7, 10.732 1 ($C=8$)	0.001 3, 6.493 5	0.001 2, 211.624 5	0.044 9, 4.756 4
$d=7, \tau=3$	0.000 4, 8.203 1 ($C=7$)	0.000 4, 4.292 9	0.000 3, 208.133 3	0.049 4, 3.152 1

注: 每个纵栏中第 1 列为 RMSE 值, 第 2 列为模型学习时间

3.2 Rossler 混沌时间序列预测

Rossler 混沌方程组是 O. E. Rossler 于 1976 年将 Lorenz 方程组简化后, 得到的简单且具有混沌特性的非线性方程组:

$$\begin{cases} \frac{ds_1}{dz} = -s_2 - s_3 \\ \frac{ds_2}{dz} = s_1 + as_2 \\ \frac{ds_3}{dz} = b + s_3(s_1 - c) \end{cases} \quad (22)$$

式中: 3 个常量 $a=0.2$, $b=0.2$, $c=5.7$; 3 个变量 s_1 、 s_2 和 s_3 的初值均取为 0.1, 离散采样间隔取为 0.5, 利用四阶龙格库塔法产生长度为 1 000 个点的 Rossler 序列。

图 4 是当嵌入维 $d=5$ 、时延 $\tau=1$ 、混合成分数 $C=3$ 时, Rossler 序列的预测结果与预测置信区间, 可看出 GPM 模型的预测结果和实际序列很吻合。图 5 是与图 4 对应的模型预测误差曲线, 可看出最大预测误差约为 0.08, 和 Rossler 序列的实际最大值相比很小, 说明 GPM 模型能准确预测该序列。

图 6 为 Rossler 序列的 4 种模型预测误差曲线比较, 其中图 6(b)是将图 6(a)横坐标[62.0, 87.0]区域放大后的显示结果。由于 GPM 模型可通过不同 GP 参数的模型描述 Rossler 序列不同时间段的细微变化, 因此能进一步提升预测精度。

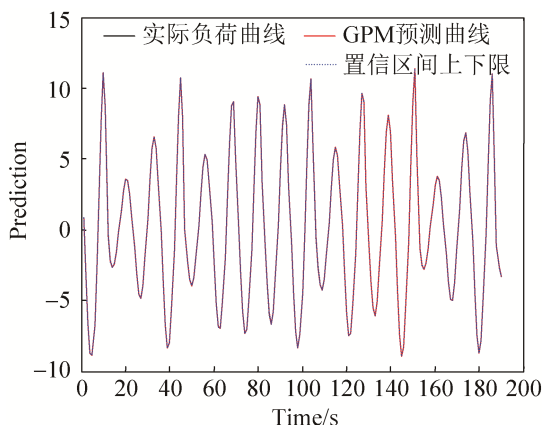


图 4 Rossler 序列的预测结果与预测置信区间
Fig. 4 Prediction results and the confidence interval of Rossler series

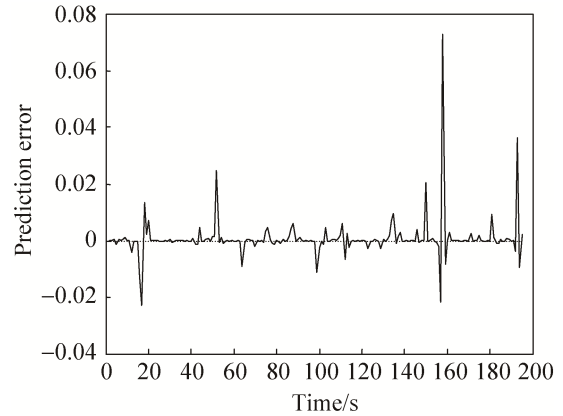
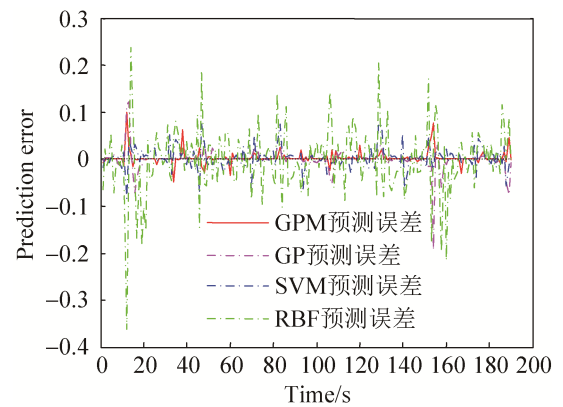
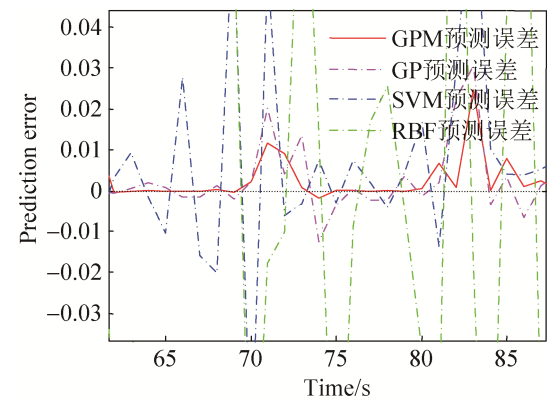


图 5 Rossler 序列的 GPM 模型预测误差
Fig. 5 Prediction error of Rossler series obtained by GPM



(a) 预测误差曲线比较图



(b) 预测误差曲线比较图局部区域放大

图 6 Rossler 序列的 4 种模型预测误差曲线比较
Fig. 6 Comparison of the prediction errors of Rossler series obtained by four models

从图 6 中发现 GPM 模型的预测误差要小于其他 3 种模型, 意味着该模型有更高的预测精度。下面分析嵌入维 d 和时延 τ 对 4 种模型预测性能的影响。将 Rossler 序列的第 1~300 个点组合成学习样

本, 第 301~500 个点组合成测试样本, 然后在不断改变 d 和 τ 取值的前提下进行模型学习。表 2 列出了 RMSE 和模型学习时间随 d 和 τ 变化的 10 组典型记录, 从表中发现对于大部分 d 和 τ 取值组合, GPM 模型的 RMSE 要小于其他 3 种模型; GPM 模型的学习时间比 RBF 网络和 GP 模型长但比 SVM 短, 这和表 1 展现出的规律一致。

在固定嵌入维 $d=4$ 和时延 $\tau=2$ 的前提下, 表 3

展示了 4 种模型预测 Rossler 序列时的 RMSE 值随学习样本数目和测试样本数目变化情形。从表中看出, 与预测 Mackey-Glass 序列类似, 学习样本和测试样本数目直接影响训练参数的选定与样本的分配, 从而对 4 种模型的预测精度均有一定影响, 而且 GPM 模型因其稳定性能好的优势, 在大部分情形下的预测精度高于其他 3 种模型。

表 2 预测 Rossler 序列时嵌入维和时延参数对 4 种模型的预测性能影响

Tab. 2 Influence of embedding dimension and delay parameters in predicting Rossler sequence on the predictive performance of four models

参数取值	GPM 模型	GP 模型	SVM	RBF 网络
$d=1, \tau=2$	4.559 0, 9.144 9 (C=2)	4.560 1, 8.621 6	4.560 3, 58.042 6	2.134 4, 0.214 5
$d=2, \tau=7$	1.451 5, 51.507 3 (C=2)	1.496 4, 15.923 4	1.537 2, 29.907 8	1.229 7, 0.290 9
$d=3, \tau=1$	0.012 4, 31.329 3 (C=3)	0.032 5, 35.398 5	0.018 6, 29.369 0	0.144 5, 0.414 2
$d=3, \tau=3$	0.198 1, 39.744 1 (C=4)	0.312 4, 42.100 8	0.342 5, 28.277 2	0.543 8, 2.002 9
$d=4, \tau=2$	0.020 0, 74.880 2 (C=3)	0.034 3, 30.817 2	0.027 6, 25.495 5	0.239 1, 0.866 4
$d=4, \tau=4$	0.036 0, 22.720 6 (C=7)	0.049 1, 29.914 9	0.076 5, 22.757 9	0.326 6, 2.727 0
$d=4, \tau=7$	0.402 5, 19.681 0 (C=6)	0.367 4, 29.492 7	0.212 9, 26.922 3	0.526 9, 0.994 4
$d=5, \tau=2$	0.014 2, 18.149 4 (C=6)	0.023 2, 42.773 6	0.023 2, 25.581 3	0.265 6, 0.639 8
$d=5, \tau=4$	0.046 0, 94.5903 (C=2)	0.044 6, 36.597 5	0.072 7, 29.423 8	0.335 6, 1.649 1
$d=7, \tau=6$	0.928 0, 14.644 3 (C=4)	1.400 4, 41.959 6	0.954 2, 149.891 9	1.031 5, 0.308 1

注: 每个纵栏中第 1 列为 RMSE 值, 第 2 列为模型学习时间

表 3 预测 Rossler 序列时 4 种模型的 RMSE 值随学习样本数和测试样本数变化

Tab. 3 RMSE values of four models varying with the number of learning samples and test samples when predicting Rossler sequence

学习样本与测试样本数	时序起点	GPM 模型	GP 模型	SVM	RBF 网络
NL=100, NT=100	601	0.034 7 (C=2)	0.194 7	0.101 2	0.181 1
NL=100, NT=100	701	0.116 0 (C=5)	0.084 6	0.257 0	0.337 7
NL=100, NT=100	801	0.110 6 (C=2)	0.056 7	0.067 3	0.175 7
NL=200, NT=100	301	0.006 7 (C=4)	0.009 2	0.018 1	0.045 1
NL=200, NT=100	601	0.016 9 (C=3)	0.030 1	0.047 7	0.092 6
NL=200, NT=200	201	0.035 5 (C=4)	0.048 6	0.042 1	0.104 6
NL=200, NT=200	401	0.047 9 (C=3)	0.024 1	0.044 1	0.124 1
NL=200, NT=300	201	0.035 3 (C=4)	0.053 2	0.046 5	0.112 5
NL=200, NT=300	601	0.025 9 (C=4)	0.016 1	0.046 7	0.115 1
NL=300, NT=200	1	0.013 0 (C=7)	0.030 1	0.026 9	0.057 1

4 结论

本文提出了 GPM 模型的一种隐变量后验硬划分迭代学习算法, 并以 Mackey-Glass 和 Rossler 序

列为例将之用于混沌时间序列预测。由于 GPM 模型既能细致刻画混沌时间序列不同时段间的细微差异, 又能灵活调节模型中的组合成分数, 同时还

能输出预测置信区间, 因此取得了良好的预测效果。实验结果表明 GPM 模型的预测精度总体上要高于 SVM、GP 模型和 RBF 网络; 模型的学习速度介于 RBF 网络、GP 模型和 SVM 之间, 因而是一种预测混沌时间序列的有效模型, 可用于不需要对混沌时间序列进行实时预测的场合。下一步研究工作为: 混沌时间序列在含有随机噪音情形下的预测问题, 以及定量比较 4 种预测模型对随机噪音的敏感度。另外还要研究混沌时间序列的先验知识在 MGP 模型预测中的导入问题, 包括似周期性的概率化描述以及基于贝叶斯推理的导入等。

参考文献:

- [1] Wang W X, Lai Y C, Grebogi C. Data based identification and prediction of nonlinear and complex dynamical systems[J]. *Physics Reports* (S0370-1573), 2016, 644(7): 1-76.
- [2] 崔万照, 朱长纯, 保文星, 等. 混沌时间序列的支持向量机预测[J]. *物理学报*, 2004, 53(10): 3303-3310.
Cui Wanzhao, Zhu Changchun, Bao Wenxing, et al. Prediction of the chaotic time series using support vector machines[J]. *Acta Physica Sinica*, 2004, 53(10): 3303-3310.
- [3] Sun J C, Zhou Y T, Luo J G. Prediction of chaotic systems with multidimensional recurrent least squares support vector machines[J]. *Chinese Physics* (S1674-1056), 2006, 15(6): 1208-1215.
- [4] Sun J, Zheng C, Zhou Y, et al. Nonlinear noise reduction of chaotic time series based on multidimensional recurrent LS-SVM[J]. *Neurocomputing* (S0925-2312), 2008, 71(16): 3675-3679.
- [5] 唐舟进, 任峰, 彭涛, 等. 基于迭代误差补偿的混沌时间序列最小二乘支持向量机预测算法[J]. *物理学报*, 2014, 63(5): 050505.
Tang Zhoujin, Ren Feng, Peng Tao, et al. A least square support vector machine prediction algorithm for chaotic time series based on the iterative error correction[J]. *Acta Physica Sinica*, 2014, 63(5): 050505.
- [6] 田中大, 高宪文, 石彤. 用于混沌时间序列预测的组合核函数最小二乘支持向量机[J]. *物理学报*, 2014, 63(16): 160508.
Tian Zhongda, Gao Xianwen, Shi Tong. Combination kernel function least squares support vector machine for chaotic time series prediction[J]. *Acta Physica Sinica*, 2014, 63(16): 160508.
- [7] 叶美盈, 汪晓东, 张浩然. 基于在线最小二乘支持向量机的混沌时间序列预测[J]. *物理学报*, 2005, 54(6): 2568-2573.
Ye Meiyong, Wang Xiaodong, Zhang Haoran. Chaotic time series forecasting using online least squares support vector machine regression[J]. *Acta Physica Sinica*, 2005, 54(6): 2568-2573.
- [8] Arash M, Majid A. Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series prediction[J]. *IEEE Transactions on Neural Networks and Learning Systems* (S2162-237X), 2013, 24(2): 207-218.
- [9] Shi Z W, Han M. Support vector echo-state machine for chaotic time-series prediction[J]. *IEEE Transactions on Neural Networks* (S1045-9227), 2007, 18(2): 359-372.
- [10] Li S, Luo Y, Zhang M R. Prediction method for chaotic time series of optimized BP neural network based on genetic algorithm[J]. *Computer Engineering and Applications* (S1002-8331), 2011, 47(29): 52-55.
- [11] 宋彤, 李菡. 基于小波回声状态网络的混沌时间序列预测[J]. *物理学报*, 2012, 61(8): 080506.
Song Tong, Li Han. Chaotic time series prediction based on wavelet echo state network[J]. *Acta Physica Sinica*, 2012, 61(8): 080506.
- [12] Zhang J, Chung S H, Lo W L. Chaotic Time Series Prediction Using a Neuro-Fuzzy System with Time-Delay Coordinates[J]. *IEEE Transactions on Knowledge and Data Engineering* (S1041-4347), 2008, 20(7): 956-964.
- [13] Yin L S, He Y G, Dong X P, et al. Adaptive chaotic prediction algorithm of RBF neural network filtering model based on phase space reconstruction[J]. *Journal of Computers* (S1796-203X), 2013, 8(6): 1449-1455.
- [14] 李松, 刘力军, 刘颖鹏. 改进 PSO 优化 BP 神经网络的混沌时间序列预测[J]. *计算机工程与应用*, 2013, 49(6): 245-248.
Li Song, Liu Lijun, Liu Yingpeng. Prediction for chaotic time series of optimized BP neural network based on modified PSO[J]. *Computer Engineering and Applications*, 2013, 49(6): 245-248.
- [15] Li S, Zhao H, Ru Z, et al. Probabilistic back analysis based on Bayesian and multi-output support vector machine for a high cut rock slope[J]. *Engineering Geology* (S0013-7952), 2016, 203(3): 178-190.
- [16] Rasmussen C E, Nickisch H. Gaussian Processes for Machine Learning (GPML) Toolbox[J]. *Journal of*

- Machine Learning Research (S1532-4435), 2010, 11(6): 3011-3015.
- [17] 李军, 张友鹏. 基于高斯过程的混沌时间序列单步与多步预测[J]. 物理学报, 2011, 60(7): 070513.
Li Jun, Zhang Youpeng. Single-step and multiple-step prediction of chaotic time series using Gaussian process model[J]. Acta Physica Sinica, 2011, 60(7): 070513.
- [18] Zhou Y T, Zhang T Y, Sun J C. Multi-scale Gaussian processes: a novel model for chaotic time series prediction[J]. Chinese Physics Letters (S0256-307X), 2007, 24(1): 42-45.
- [19] Zhou Y T, Zhang T Y, Li X H. Prediction of chaotic time series based on multi-scale Gaussian processes[J]. Lecture Notes in Computer Science (S0302-9743), 2006, 4224(9): 183-190.
- [20] Chen Z, Ma J, Zhou Y. A Precise Hard-Cut EM Algorithm for Mixtures of Gaussian Processes[C]. International Conference on Intelligent Computing. Cham: Springer, 2014: 68-75.