

基于多重核学习支持向量回归的 混沌时间序列预测^{*}

张军峰[†] 胡寿松

(南京航空航天大学自动化学院, 南京 210016)

(2007 年 8 月 16 日收到, 2007 年 11 月 2 日收到修改稿)

鉴于标准支持向量回归应用于混沌时间序列预测时经常会遇到诸如核函数及其参数难以确定的问题, 提出了多重核支持向量回归的方法. 通过在混合核空间求解二次约束下的二次规划问题, 实现多重核支持向量回归算法. 该算法不仅可以减少支持向量的个数, 而且能够提高预测性能. 最后将该方法运用到 Lorenz, Henon 和 Mackey-Glass 混沌时间序列预测, 仿真结果表明该方法能够有效地提高预测精度, 增强预测模型的泛化性能.

关键词: 混沌时间序列, 支持向量机, 多重核学习, 优化

PACC: 0545

1. 引 言

混沌现象在自然界中广泛存在, 是一种由确定的非线性动力系统生成的复杂行为. 在混沌动力学系统中, 通过时间序列来研究整个系统的动力学行为, 并对混沌序列进行预测已成为一个重要的研究方向.

近年来, 在利用各种神经网络, 特别是前馈型神经网络预测混沌时间序列取得了成功的应用^[1, 2]. 与常规神经网络所基于的经验风险最小化原理不同, 由 Vapnik 等人^[3]创立的支持向量机 (support vector machines, SVM) 遵循了结构风险最小化原理, 使得 SVM 具有更好的泛化性能. 而且 SVM 集成了最大间隔超平面, Mercer 核, 凸二次规划, 稀疏解和松弛变量等多项技术, 使其可以将非线性问题转化为线性问题并能得到全局最优解. 得益于 Müller 等人^[4]的贡献, 从而使通过 SVM 预测混沌时间序列成为可能, 且已有多位学者对此进行了研究^[5, 6].

标准的 SVM 可以转化为线性约束下的一个凸二次规划问题, 其性能主要取决于核函数以及其参数的选择, 在学习时可以用交叉验证或考察训练集^[7]等方法来确定核函数以及其参数, 但其相当费

时且难以避免随意性. 近年来迅速发展的多重核学习 (multiple kernel learning, MKL)^[8-12]为解决该问题提供了可能. 毕竟, 多重核学习通过将标准的 SVM 学习的线性约束下二次规划 (quadratic programming, QP) 问题转化为半正定规划 (semi-definite programming, SDP) 问题或二次约束下二次规划 (quadratic constrained quadratic programming, QCQP) 问题, 可以同时解决 SVM 中模型选择和训练问题.

目前, 多重核学习 SVM 主要用于解决分类领域, 本文主要将其推广到回归领域并以此解决混沌时间序列预测问题. 由于多重核学习获得的混合核可以使数据信息在特征空间得到更为充分的表示, 一方面提高了回归预测模块的预测性能, 另一方面减少了支持向量的个数. 而支持向量数目的减少不仅体现在泛化性能的提高, 还有利于缩短预测时间. 最后将这种方法分别应用于 Mackey-Glass, Lorenz 和 Henon 混沌时间序列, 仿真结果表明运用本文方法可以更为有效的预测混沌时间序列.

2. 支持向量回归

对于回归问题, 假设给定的样本点

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n),$$

^{*} 国家自然科学基金重点项目 (批准号: 60234010), 航空科学基金 (批准号: 05E52031) 资助的课题.

[†] E-mail: wufeng7919@163.com

$$\mathbf{x}_i \in R^l, y_i \in R. \quad (1)$$

首先用一个非线性映射 Φ 把数据映射到一个高维特征空间,然后在高维特征空间中进行线性回归

$$f(\mathbf{x}_{\text{new}}) = \mathbf{w}^T \Phi(\mathbf{x}_{\text{new}}) + b. \quad (2)$$

定义 ε 不敏感损失函数为

$$|y_i - f(\mathbf{x}_i)|_\varepsilon = \begin{cases} 0, & |y_i - f(\mathbf{x}_i)| \leq \varepsilon \\ |y_i - f(\mathbf{x}_i)| - \varepsilon, & |y_i - f(\mathbf{x}_i)| > \varepsilon. \end{cases} \quad (3)$$

根据结构风险最小化原则,并引入松弛变量 ξ_i 和 ξ_i^* ,回归问题在一阶范数软间隔约束下可归结为如下优化问题:

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{b}, \xi, \xi^*} & \mathbf{w}^T \Phi(\mathbf{x}_i) + b + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (4) \\ \text{s.t.} & y_i - \mathbf{w}^T \Phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i, \xi_i \geq 0 \\ & \mathbf{w}^T \Phi(\mathbf{x}_i) + b - y_i \leq \varepsilon + \xi_i^*, \xi_i^* \geq 0. \end{aligned}$$

应用拉格朗日乘子法求解这个具有线性不等式约束的二次规划问题,可得到该优化问题的对偶形式,即

$$\begin{aligned} \max_{\alpha, \alpha^*} & - \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) \\ & - 2\varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n 2y_i (\alpha_i - \alpha_i^*), \quad (5) \\ \text{s.t.} & \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, n, \end{aligned}$$

其中 $K(\mathbf{x}_i, \mathbf{x}_j)$ 为满足 Mercer 条件^[13]的核函数(5)式可以改写为以下形式:

$$\begin{aligned} \max_{\beta} & 2\mathbf{h}^T \beta - \beta^T Q(K) \beta, \quad (6) \\ \text{s.t.} & 0 \leq \beta \leq C, \beta^T \gamma = 0, \end{aligned}$$

其中 $Q(K) = \begin{bmatrix} K & -K \\ -K & K \end{bmatrix}$, $\beta = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}$, $\gamma = \begin{bmatrix} \mathbf{e} \\ -\mathbf{e} \end{bmatrix}$,

$$\mathbf{h} = \begin{bmatrix} -\varepsilon \mathbf{e} + \mathbf{y} \\ -\varepsilon \mathbf{e} - \mathbf{y} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in R^n.$$

3. 多重核支持向量回归

由于核的混合可以使数据的信息在特征空间得到更为充分的表示^[14],有助于提高学习性能,因此考虑具有如下形式的混合核,其中 μ_j 可视为权重系数:

$$K = \sum_{j=1}^m \mu_j K_j. \quad (7)$$

若假设(7)式中混合核的权重系数 $\mu_j \geq 0, j = 1, \dots, m$

,由核函数的性质^[13]可知 $K = \sum_{j=1}^m \mu_j K_j$ 是对称半正定矩阵,即 K 是有效的核函数.并且核矩阵 $K_j (j = 1, \dots, m)$ 可按下式进行标准化:

$$\bar{K}_j(\mathbf{x}_p, \mathbf{x}_q) = K_j(\mathbf{x}_p, \mathbf{x}_q) / \sqrt{K_j(\mathbf{x}_p, \mathbf{x}_p) \cdot K_j(\mathbf{x}_q, \mathbf{x}_q)}, \quad p, q = 1, \dots, n. \quad (8)$$

通过分析 Lanckriet^[10]提出的多重核支持向量分类,可推导出多重核学习在支持向量回归(support vector regression, SVR)中的应用,得到如下结论.

假设问题(6)中的核矩阵 K 属于核空间 Ω ,且核空间 Ω 满足条件

$$\Omega = \{K = \sum_{j=1}^m \mu_j K_j \mid K_j \succcurlyeq 0, \mu_j \geq 0, \text{tr}(K) = c\}, \quad (9)$$

$K_j \succcurlyeq 0$ 表示矩阵 K_j 是半正定的,则问题(6)可以转化为以下的 QCQP 问题:

$$\begin{aligned} \max_{\beta, \mu} & 2\mathbf{h}^T \beta - c\mathbf{t}, \quad (10) \\ \text{s.t.} & 0 \leq \beta \leq C, \beta^T \gamma = 0, \\ & \mathbf{t} \geq \frac{1}{n} \beta^T Q(\bar{K}_j) \beta, i = 1, \dots, m. \end{aligned}$$

证明 由于对具有固定迹的核矩阵,SVR 的性能受限于(6)式,其值越小,所得 SVR 的性能越好.因此在核空间 Ω 中,为了优化问题(6)寻找最优的核矩阵 K ,可以等价于

$$\begin{aligned} \min_{\Omega} & \max_{\beta} 2\mathbf{h}^T \beta - \beta^T Q(K) \beta, \quad (11) \\ \text{s.t.} & 0 \leq \beta \leq C, \beta^T \gamma = 0, \mu_j \geq 0, \\ & K = \sum_{j=1}^m \mu_j K_j, \text{tr}(K) = c. \end{aligned}$$

显然,

$$\begin{aligned} \text{trac}(K) &= \text{trac}\left(\sum_{j=1}^m \mu_j K_j\right) \\ &= \sum_{j=1}^m [\mu_j \text{trac}(K_j)] = c. \end{aligned}$$

则优化问题(11)又可等价于

$$\begin{aligned} \min_{\mu} & \max_{\beta} \max_{\mathbf{t}} \min_{\beta \leq C, \beta^T \gamma = 0} 2\mathbf{h}^T \beta - \beta^T Q(K) \beta, \quad (12) \\ \text{s.t.} & \mu \geq 0, \mu^T \mathbf{t} = c, \mathbf{t}_j = \text{tr}(K_j). \end{aligned}$$

优化问题(12)可改写为

$$\begin{aligned} \min_{\mu} & \max_{\mathbf{t}: \mu \geq 0, \mu^T \mathbf{t} = c} \max_{\beta \leq C, \beta^T \gamma = 0} 2\mathbf{h}^T \beta \\ & - \beta^T Q\left(\sum_{j=1}^m \mu_j K_j\right) \beta, \\ &= \min_{\mu} \max_{\mathbf{t}: \mu \geq 0, \mu^T \mathbf{t} = c} \min_{\beta \leq C, \beta^T \gamma = 0} 2\mathbf{h}^T \beta \\ & - \sum_{j=1}^m \mu_j \beta^T Q(K_j) \beta, \quad (13) \\ &= \max_{\beta \leq C, \beta^T \gamma = 0} \min_{\mu: \mu \geq 0, \mu^T \mathbf{t} = c} 2\mathbf{h}^T \beta \\ & - \sum_{j=1}^m \mu_j \beta^T Q(K_j) \beta \end{aligned}$$

$$\begin{aligned}
&= \max_{\beta} \min_{\nu \leq \beta \leq C, \beta^T \gamma = 0} [2\mathbf{h}^T \boldsymbol{\beta} \\
&\quad - \max_{\mu_j \geq 0, \mu^T \tau = c} (\sum_{j=1}^m \mu_j \boldsymbol{\beta}^T Q K_j) \boldsymbol{\beta}] \\
&= \max_{\beta} \min_{\nu \leq \beta \leq C, \beta^T \gamma = 0} \left[2\mathbf{h}^T \boldsymbol{\beta} - \max_j \left(\frac{c}{\tau_j} \boldsymbol{\beta}^T Q K_j \boldsymbol{\beta} \right) \right].
\end{aligned} \tag{14}$$

以上推导中,由于(13)式对变量 $\boldsymbol{\mu}$ 是凸的,而对变量 $\boldsymbol{\beta}$ 是凹的,因此可以交换最小化和最大化的次序^[15].最终(14)式可以改写为

$$\begin{aligned}
&\max_{\beta, t} 2\mathbf{h}^T \boldsymbol{\beta} - ct, \\
&\text{s. t. } 0 \leq \boldsymbol{\beta} \leq C, \boldsymbol{\beta}^T \gamma = 0, \\
&t \geq \frac{1}{\tau_j} \boldsymbol{\beta}^T Q K_j \boldsymbol{\beta} \quad j = 1, \dots, m.
\end{aligned}$$

又因为在进行多重核学习时,首先对核矩阵进行标准化,因此 $\tau_j = n$,即可得到(10)式的形式.证毕.

该 QCQP 问题可以通过优化软件包 MOSEK 基于内点法求解^[16],由于 QCQP 问题与 QP 问题求解算法的复杂度相当^[15],而(10)式在训练 SVR 的同时解决了模型选择的困扰,显然较之标准 SVR,多重核支持向量回归(MKL-SVR)具有优越性.在求解(10)式时,不仅可以得到 $\boldsymbol{\beta}, t$,还可以通过其对偶解得到反映备选核函数重要性的权重系数 μ_j .此时(2)式所示的回归函数具有如下形式:

$$f(\mathbf{x}_{new}) = \sum_{j=1}^m \sum_{i=1}^{|sv|} \mu_j (\alpha_i - \alpha_i^*) K_j(\mathbf{x}_i, \mathbf{x}_{new}) + b, \tag{15}$$

其中 $|sv|$ 表示支持向量的个数.

4. 混沌时间序列预测

对于混沌时间序列的预测问题,其首要任务应根据嵌入定理重构输入输出空间.混沌时间序列的时间延迟嵌入状态空间为

$$\mathbf{x}(t) = [x(t) \quad x(t - td) \quad \dots \quad x(t - (Em - 1)td)], \tag{16}$$

其中 $\mathbf{x}(t)$ 是嵌入矢量, $x(t)$ 是第 t 时间点的序列值, td 是时延量, Em 是嵌入维数.此时可以通过嵌入矢量 $\mathbf{x}(t)$ 预测第 $t + v$ 时间点的序列值 $x(t + v)$, v 表示预测步长.

下面将本文提出的方法应用于 Mackey-Glass, Lorenz 和 Henon 混沌时间序列预测,并以均方根误差(root mean square error, RMSE)作为性能指标来衡量不同方法的预测效果,

$$RMSE = \sqrt{\sum_{i=1}^N (y(i) - \hat{y}(i))^2 / N}, \tag{17}$$

其中 $y(i)$ 和 $\hat{y}(i)$ 分别为混沌时间序列的实际值和预测值.注意,在使用训练样本进行学习时,首先要对样本实行预处理,那么对数据的预测结果也要进行相应的反处理.该仿真实验中所涉及的对偶优化问题是利用 MOSEK 优化软件包基于内点法^[16]求解,可在 <http://www.mosek.com> 下载 MOSEK 及其使用手册.

4.1. Mackey-Glass 序列

Mackey-Glass 时间序列如下:

$$x(t + 1) = (1 - b)x(t) + \frac{ax(t - \tau)}{1 + x(t - \tau)^{10}}. \tag{18}$$

该时间序列当 $\tau > 17$ 时呈现混沌性, τ 值越大,混沌程度越高.本文分别选取 $\tau = 30$ 作单步预测 ($Em = 4; td = 1; v = 1$), $\tau = 20$ 作多步预测 ($Em = 6; td = 3; v = 10$),参数 $a = 0.2, b = 0.1$,产生 3000 个数据,选取其中 150 个数据作为训练样本,1000 个数据作为测试样本.标准 SVR 中核函数为高斯核,核参数 σ 均由基于交叉验证的网格搜索方法获得. MKL-SVR 中备选核函数的个数为 3,包括一个反映样本全局信息的线性核与两个反映样本局部信息的高斯核,且核参数分别为 $\sigma_1 = 1, \sigma_2 = 2.5$. Mackey-Glass 时间序列的预测结果如表 1 所示,其中 $\#sv$ 表示支持向量的个数.

表 1 Mackey-Glass 时间序列预测

	C	ϵ	σ	RMSE	$\#sv$
$\tau = 30$ (单步预测)					
SVR	1000	0.01	1.5	0.0070	10
MKL-SVR	1000	0.01	- / 1.0 / 2.5	0.0062	6
$\tau = 20$ (多步预测)					
SVR	1000	0.005	2.0	0.0034	15
MKL-SVR	1000	0.005	- / 1.0 / 2.5	0.0028	8
SVR	1000	0.01	2.5	0.0129	26
MKL-SVR	1000	0.01	- / 1.0 / 2.5	0.0096	19

考察表 1 可以得出以下结论:首先, MKL-SVR 同时解决了参数确定和模型学习的问题.其次,基于标准 SVR 的预测结果不仅预测精度较低(RMSE 均值较 MKL-SVR 高 20.18%),而且所得支持向量的个数也较多($\#sv$ 均值较 MKL-SVR 多 6 个),说明该预测模型的泛化性能较差.

4.2. Lorenz 序列

Lorenz 时间序列如下:

$$\begin{aligned} \frac{dx}{dt} &= s(y - x), \\ \frac{dy}{dt} &= rx - y - xz, \\ \frac{dz}{dt} &= xy - bz. \end{aligned} \quad (19)$$

本文选取 $s = 16$, $r = 45.92$, $b = 4$, 运用四阶 Runge-Kutta 法得到微分方程的离散序列, 积分步长为 0.01, 初始值为 $[1 \ 0 \ -1]$. 选择其中 150 个数据作为训练样本, 1000 个数据作为测试样本. 仿真中使用以下参数: 不敏感参数 $\epsilon = 0.01$, 正则化因子 $C = 1000$, 嵌入维数 $Em = 6$ 时延参数 $td = 1$, 预测步长 $v = 1$. 标准 SVR 中核函数为高斯核, 核参数 σ 均由基于交叉验证的网格搜索方法获得. MKL-SVR 中备选核函数的个数为 3, 包括一个线性核与两个高斯核, 且核参数分别为 $\sigma_1 = 1$, $\sigma_2 = 2.5$. Lorenz 时间序列的预测结果如表 2 和图 1 所示, 其中实线表示目标序列, 虚线表示基于标准 SVR 获得的预测序列, 折线表示基于 MKL-SVR 获得的预测序列.

表 2 Lorenz 时间序列预测

	x		y		z	
	SVR	MKL-SVR	SVR	MKL-SVR	SVR	MKL-SVR
RMSE	0.3131	0.2595	0.4312	0.3648	0.3402	0.2305
# sv	13	7	15	8	14	8

由表 2 和图 1 中各误差分布图可以看出: MKL-SVR 的预测效果优于标准 SVR 的预测效果, 且所需的支持向量的个数要少.

4.3. Henon 序列

Henon 时间序列如下:

$$\begin{aligned} x(t+1) &= 1 - ax^2(t) + y(t), \\ y(t+1) &= bx(t). \end{aligned} \quad (20)$$

当 $a = 1.4$, $b = 0.3$ 时系统成为混沌状态, 初始值为 $[0.1 \ 0.1]$ 产生 3000 个数据, 选取其中 150 个数据作为训练样本, 另 2000 个数据作为测试样本. 仿真中使用以下参数: 不敏感参数 $\epsilon = 0.01$, 正则化因子 $C = 1000$, 嵌入维数 $Em = 3$ 时延参数 $td = 1$, 预测步长 $v = 1$, 备选核函数个数 $m = 3$. 针对 Henon 时间序列, 基于 MKL-SVR 的预测序列的误差分布及混

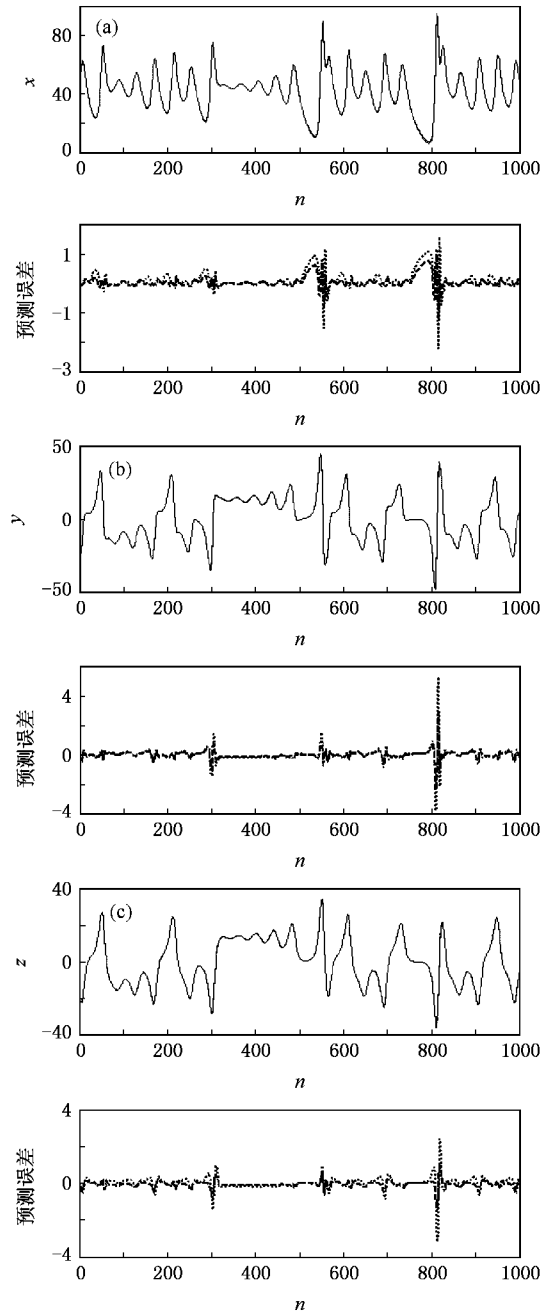


图 1 Lorenz 时间序列预测效果(各误差分布图中“虚线”表示基于标准 SVR 的预测误差;“折线”表示基于 MKL-SVR 的预测误差)

沌吸引子如图 2 所示.

对于 Henon 时间序列预测而言, 标准 SVR 运用 21 个支持向量获得的预测序列 RMSE 为 0.0074, 而 MKL-SVR 仅需 14 个支持向量便可使预测序列 RMSE 达到 0.0041. 因此, 无论从预测精度还是从支持向量的个数来考虑, 都说明该方法的确是有效的.



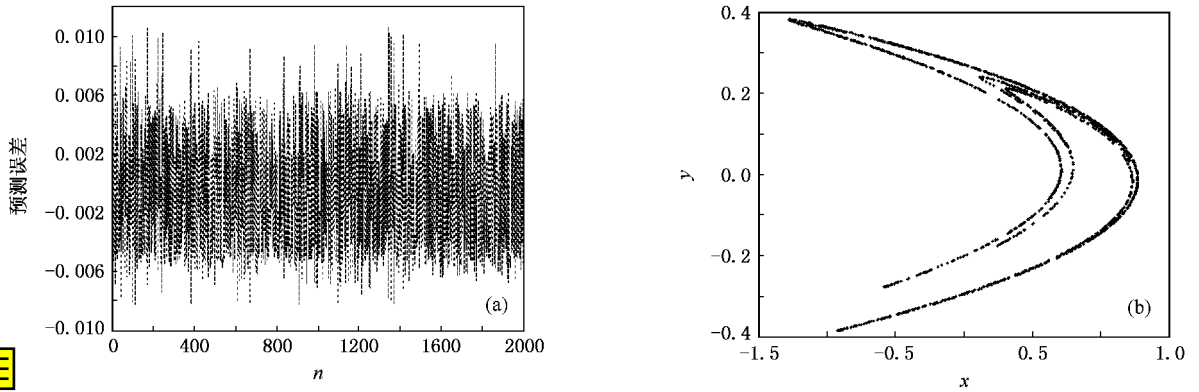


图2 Henon 时间序列预测效果 (a)误差分布 (b)预测序列吸引子

5. 结 论

基于统计学习理论的支持向量机拥有通过把输入向量映射到一个高维特征空间中实现数据线性可分,来提取信息的能力,再结合混沌动力系统的相空间延迟坐标重构理论,可以建立混沌时间序列的支

持向量机预测模型.

然而,诸如训练费时和核函数及其参数难以确定的问题会阻碍标准的支持向量机运用于混沌时间序列预测.通过引入多重核学习实现支持向量回归以及在 Mackey-Glass, Lorenz 和 Henon 混沌时间序列预测上运用表明,本文的方法是有效的且在混沌时间序列的研究中具有很强的实用价值.

- [1] Yu Z H , Cai Y L 2006 *Acta Phys. Sin.* **55** 1659 (in Chinese) [于振华、蔡远利 2006 物理学报 **55** 1659]
- [2] Zhang J F , Hu S S 2007 *Acta Phys. Sin.* **56** 713 (in Chinese) [张军峰、胡寿松 2007 物理学报 **56** 713]
- [3] Vapnik V 1995 *The Nature of Statistical Learning Theory* (New York : Springer)
- [4] Müller K R , Smola A J , Rätsch G *et al* 1997 *Proceedings of ICANN* **1327** 999
- [5] Cui W Z , Zhu C C , Bao W X , Liu J H 2004 *Acta Phys. Sin.* **53** 3303 (in Chinese) [崔万照、朱长纯、保文星、刘君华 2004 物理学报 **53** 3303]
- [6] 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]
- [7] Chapelle O , Vapnik V , Bousquet O *et al* 2002 *Machine Learning* **46** 131
- [8] Ong C , Smola A , Williamson R 2003 *Journal of Machine Learning Research* **3** 1001
- [9] Bach F R , Lanckriet G , Jordan M I 2004 *Proceedings of the 21st ICML* **227** 775
- [10] Lanckriet G , Bie T D , Cristianini N *et al* 2004 *Bioinformatics* **20** 2626
- [11] Lanckriet G , Cristianini N , Bartlett P *et al* 2004 *Journal of Machine Learning Research* **5** 27
- [12] Sonnenburg S , Rätsch G , Schäfer C *et al* 2006 *Journal of Machine Learning Research* **1** 1
- [13] Schölkopf B , Smola A 2002 *Learning with Kernels* (Cambridge : MIT Press)
- [14] Genton M G 2001 *Journal of Machine Learning Research* **2** 299
- [15] Boyd S , Vandenberghe L 2004 *Convex Optimization* (Cambridge : Cambridge University Press)
- [16] Anderson E D , Anderson A D 2000 *High Performance Optimization*

Chaotic time series prediction based on multi-kernel learning support vector regression *

Zhang Jun-Feng[†] Hu Shou-Song

(*College of Automation Engineering Nanjing University of Aeronautics and Astronautics , Nanjing 210016 ,China*)

(Received 16 August 2007 ; revised manuscript received 2 November 2007)

Abstract

Multi-kernel learning support vector regression (MKL-SVR) are proposed for chaotic time series prediction to solve the problems of kernel selection and hyper-parameter determination when using the standard SVR. The algorithm is realized through quadratic constrained quadratic programming (QCQP) in the hybrid kernel space , which not only reduces the number of support vectors , but also improves the prediction performance. Finally , it is applied to Mackey-Glass , Lorenz and Henon chaotic time series prediction. The results indicate that the proposed method can effectively increase the prediction precision , accelerate the convergency of cascade learning and enhance the generalization of prediction model.

Keywords : chaotic time series , support vector machines , multi-kernel learning , optimization

PACC : 0545

* Project supported by the National Natural Science Key Foundation (Grant No.60234010) and the Aeronautic Science Foundation (Grant No.05E52031).

[†] E-mail : wufeng7919@163.com