

一种基于训练数据的迭代改进核函数^{*}

周志祥¹, 韩逢庆²

(1. 重庆交通大学 土木建筑学院, 重庆 400074;

2. 重庆交通大学 理学院, 重庆 400074)

(吴承平推荐)

摘要: 为提高支持向量机性能, 提出一种支持向量机核函数的迭代改进新算法. 利用与数据有关的保角映射, 使核函数包含了全部学习样本的信息, 即核函数具有数据依赖性. 基本核函数的参数可取随机初值, 通过对核函数进行多次迭代改进, 直至得到满意的学习效果. 与传统方法相比, 新算法不需要筛选核函数的参数. 对一元连续函数和强地震事件的仿真计算结果表明, 改进 SVR (support vector regression) 的学习效果优于传统方法, 并且随着迭代次数的增加, 学习风险下降收敛, 收敛速度依赖于传统方法的基本参数和改进方法的参数.

关键词: 支持向量回归; 数据依赖; 核函数; 迭代

中图分类号: O241.82 **文献标识码:** A

引 言

近年来, 由 Vapnik 等^[1]提出的一种新的模式分类算法——支持向量机 (support vector machine (SVM)), 成为解决非线性分类、函数估计和密度估计等问题的有力工具. 与传统的神经网络不同, SVM 应用的是结构风险最小化原则, 而不是经验风险最小化原则, 即 SVM 寻求的是一般误差上界的最小化而不是单纯训练误差的最小化. 由于 SVM 具有全局最优解以及对于小样本良好的学习能力等特性, 引起了越来越多研究者的兴趣^[2-4], SVM 在回归 (support vector regression (SVR)), 尤其是非线性时间序列预测中的应用^[5-7]近年来颇受人们的关注.

SVM 的性能主要决定于核函数, Smola^[8]等指出了 SVM 的核函数方法与标准正规化理论之间的联系, 但目前还没有关于数据依赖的核函数选择理论. 文献[9]针对 SVM 用于模式分类, 根据核函数所导出的 Riemann 度量的性质对核函数进行了改进, 将初始训练得到的支持向量信息融进核函数, 用改进的核函数对样本重新训练, 扩大两类间的距离, 使得分类效果更好. 这种引入支持向量进行核函数的改进是保角的, 这是针对所给训练数据来改进核函数形式的开创性工作. 而文献[10]针对回归问题, 仍然利用 Riemann 度量的性质对核函数进行了局部的改进, 只是用一些关键点来替代初始支持向量进入核函数, 用改进的核函数对样本重新训练得到新的回归模型. 使用该方法对股票价格进行预测得到较常规 SVM 更好的结果.

* 收稿日期: 2008-07-18; 修订日期: 2008-12-03

基金项目: 国家自然科学基金资助项目 (50578168); 重庆市自然科学基金资助项目 (2007BB2396)

作者简介: 周志祥 (1958—), 男, 四川遂宁人, 教授, 博士, 博士生导师 (联系人, Tel: + 86-23-62652922; E-mail: zhixiangzhou@cqjtu.edu.cn);

韩逢庆 (1968—), 男, 重庆人, 教授, 博士 (E-mail: hanfengqing@cqit.edu.cn).

这些方法虽然比传统 SVM 算法效果要好,但仍然需要对核函数参数以及改进方法的参数进行仔细斟酌,以利于找到与训练样本相匹配的核函数,获得更好的结果.受文献[9-10]的启发,本文提出一种关于 SVM 用于回归的新改进算法,对核函数进行多次迭代改进,直至得到满意的学习效果.与传统 SVM 以及文献[10]的算法相比,其核函数的参数不需要筛选对比.数值模拟结果表明,迭代改进的 SVM 算法具有良好的学习效果.

1 SVR 及其核函数

设 $\{(x_i, y_i)\} (i = 1, 2, \dots, l)$ 是容量为 l 的训练样本, $(x_i, y_i) \in R^n \times \mathbf{R}$, 则 SVR 模型为

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b, \quad (1)$$

其中, \mathbf{w} 为权值向量, b 为偏差, $\phi(\mathbf{x})$ 为输入空间 S 到特征空间 F 的非线性映射, $\langle \cdot, \cdot \rangle$ 表示 Hilbert 空间中的内积.

常用的损失函数是 Vapnik 引进的 ε -不敏感损失函数^[1], 即

$$L_\varepsilon(\mathbf{x}_i) = \begin{cases} |f(\mathbf{x}_i) - y_i| - \varepsilon, & |f(\mathbf{x}_i) - y_i| \geq \varepsilon, \\ 0, & |f(\mathbf{x}_i) - y_i| < \varepsilon. \end{cases} \quad (2)$$

为了训练 \mathbf{w} 和 b , 需要最小化下面的泛函:

$$\begin{aligned} \min Z &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i (\xi_i^+ + \xi_i^-) \\ \text{s. t. } &\begin{cases} y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^+, \\ -y_i + \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b \leq \varepsilon + \xi_i^-, \end{cases} \quad \xi_i^+, \xi_i^- \geq 0; i = 1, 2, \dots, l, \end{aligned} \quad (3)$$

其中, C 称为正规化常数, 为预先指定的. 它确定了经验误差项与正规化项之间的某种平衡关系.

在引入松弛变量以及 Lagrange 算子后, 模型(3)等价于一个标准二次规划问题. 将式(1)改写为

$$f(\mathbf{x}) = \sum_{x_i \in I_{SV}} (\alpha_i - \alpha'_i) k(x_i, \mathbf{x}) + b, \quad (4)$$

其中, $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$, I_{SV} 表示支持向量的集合, α_i 和 α'_i 是模型(3)的最优解.

并不是所有的函数都可作为 SVR 中的核函数, 理论上已经证明, 必须满足 Mercer 定理^[11]条件的函数才能作为 SVR 的核函数. 核函数在 SVR 中有重要的地位, 它的选择对学习精度的影响很大. 下面针对核函数的特点, 对其改造, 采用保角映射方法, 使核函数与给定的数据有关.

2 数据依赖的核函数

在传统的 SVM 和 SVR 中, 核函数的选择以及 Lagrange 算子的优化都未考虑数据的影响, 而对于某些时间序列, 如金融、气象数据等, 由于数据本身具有高噪声、非平稳等特点, 为了提高对其预测的精度, 有必要从数据的角度重新定义核函数. 本节介绍一种基于信息几何^[9,11]的方法, 对核函数进行数据依赖性改进.

从几何的观点看, 非线性映射是一个曲子流形, 它定义了从输入空间 S 到特征空间 F 的一个嵌入. 一般 F 为再生核 Hilbert 空间(RKHS), RKHS 是 Hilbert 空间的子空间. 因此, 可以在 S 空间中引入一个可以用核函数 $k(\mathbf{x}, \mathbf{x}')$ 表示的 Riemann 度量^[9]. 令

$$g_{ij}(\mathbf{x}) = \left\langle \frac{\partial}{\partial x_{(i)}} \phi(\mathbf{x}), \frac{\partial}{\partial x_{(j)}} \phi(\mathbf{x}) \right\rangle = \frac{\partial^2}{\partial x_{(i)} \partial x_{(j)}} k(\mathbf{x}, \mathbf{x}') \Big|_{\mathbf{x}' = \mathbf{x}}. \quad (5)$$

因此, 特征空间 F 中的微体元可表示为

$$dV = \sqrt{g(\mathbf{x})} dx_{(1)} dx_{(2)} \cdots dx_{(n)}, \quad (6)$$

其中, $g(\mathbf{x}) = \det(g_{ij}(\mathbf{x}))$, 而 $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ 是 \mathbf{x} 的全部分量, 因子 $\sqrt{g(\mathbf{x})}$ 表示在 $\phi(\mathbf{x})$ 映射作用下特征空间 F 中的局部缩放系数^[9].

当核函数取为径向基函数 (RBF):

$$k(\mathbf{x}, \mathbf{x}') = \exp[-\|\mathbf{x} - \mathbf{x}'\|^2 / (2\sigma^2)]. \quad (7)$$

此时, Riemann 度量为

$$g_{ij}(\mathbf{x}) = \delta_{ij} / \sigma^2, \quad (8)$$

$$\text{其中 } \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

定义

$$\phi(\mathbf{x}) = c(\mathbf{x}) \phi(\mathbf{x}), \quad (9)$$

其中, $c(\mathbf{x}): R^n \mapsto \mathbf{R}$ 为一正的保角映射, 则相应的改进核函数为

$$k(\mathbf{x}, \mathbf{x}') = c(\mathbf{x}) c(\mathbf{x}') k(\mathbf{x}, \mathbf{x}'). \quad (10)$$

容易证明, $k(\mathbf{x}, \mathbf{x}')$ 也能满足 Mercer 定理条件. 对于核函数 $k(\mathbf{x}, \mathbf{x}')$, 相应的 Riemann 度量为

$$g_{ij}(\mathbf{x}) = \frac{\partial c(\mathbf{x})}{\partial x_{(i)}} \frac{\partial c(\mathbf{x})}{\partial x_{(j)}} + c^2(\mathbf{x}) g_{ij}(\mathbf{x}). \quad (11)$$

在文献[9]中, 取

$$c(\mathbf{x}) = \sum_{x_i \in I_{SV}} h_i \exp[-\|\mathbf{x} - \mathbf{x}_i\|^2 / (2\tau^2)], \quad (12)$$

其中, I_{SV} 为指标集, 取为支持向量集^[9]. 在支持向量 x_i 附近, 有

$$\sqrt{g(\mathbf{x})} = (h_i^n / \sigma^4) \exp[-nr^2 / (2\tau^2)] \sqrt{1 + \sigma^2 r^2 / \tau^4}, \quad (13)$$

其中, $r = \|\mathbf{x} - \mathbf{x}_i\|$ 为两点距离.

3 迭代改进的核函数算法

为了提高 SVM 的性能, 核函数改进后要能够保证在支持向量 x_i 附近 $\sqrt{g(\mathbf{x})}$ 应该扩张. 在函数回归时, 在大部分区域 $\sqrt{g(\mathbf{x})}$ 应该压缩.

为保证缩放系数 $\sqrt{g(\mathbf{x})}$ 是压缩的, 令

$$c(\mathbf{x}) = \sum_{x_i \in I} h_i \exp[-\|\mathbf{x} - \mathbf{x}_i\|^2 / (2\tau^2)], \quad (14)$$

$$k_s(\mathbf{x}, \mathbf{x}') = c^s(\mathbf{x}) c^s(\mathbf{x}') k(\mathbf{x}, \mathbf{x}'), \quad (15)$$

其中, s 是一正整数, I 为指标集, 取为全部训练样本, 可以简化改进型 SVR 的学习过程. 此时有

$$g_{ij}^s(\mathbf{x}) = \frac{\partial^2}{\partial x_{(i)} \partial x_{(j)}} k_s(\mathbf{x}, \mathbf{x}') \big|_{\mathbf{x}' = \mathbf{x}}. \quad (16)$$

在本文中采用 RBF 核函数. 对于训练样本 x_i 附近的点 \mathbf{x} , 有

$$\sqrt{g_s(\mathbf{x})} = (h_i^{ns} / \sigma^4) \exp[-nsr^2 / (2\tau^2)] \sqrt{1 + s^2 \sigma^2 r^2 / \tau^4}. \quad (17)$$

容易验证如下结论:

- 1) 当 $h_i < \min\{\sigma, 1\}$ 时, $\forall s = 1, 2, \dots$, $\sqrt{g_s(\mathbf{x})}$ 在训练样本 x_i 附近是压缩的;
- 2) 当 $1 \leq h_i < e$, $s > M$ 时, 对 $\sqrt{g_s(\mathbf{x})}$ 在训练样本 x_i 附近是压缩的, M 是一个充分大的

正数;

3) 当 x 远离所有训练样本 x_i 时, $\sqrt{g_s(x)}$ 是压缩的.

下面将依赖于数据的迭代改进 SVR 方法的执行步骤简述如下:

步骤 1 $\sigma, \varepsilon, C, h_i, \tau$ 任取 5 个随机值, 给定学习误差 ε_r ;

步骤 2 $k_0(x, x') = k(x, x'), j = 0$;

步骤 3 函数 $k_j(x, x')$ 训练 SVR, 对全体训练样本计算学习风险

$$\delta = \sqrt{\sum_i [y_i - f(x_i)]^2 \sum_i (y_i - y)^2};$$

步骤 4 $\sigma > \varepsilon_r$ 时, $k_{j+1}(x, x') = c(x)c(x')k_j(x, x'), j = j + 1$, 返回步骤 3; 否则迭代结束.

上述算法结束后, 改进的核函数为

$$k(x, x') = c^s(x)c^s(x')k(x, x'), \quad (18)$$

其中, s 为迭代次数.

4 仿 真

这一节将进行一些仿真实验以观察改进方法的效果, 在同样的数据上对比改进方法与传统方法, 两种方法的基本参数取相同值. 为了评价模型的逼近效果和推广能力, 学习风险采用如下定义:

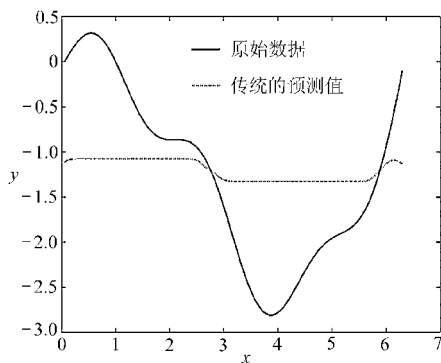
$$\delta = \sqrt{\sum_i [y_i - f(x_i)]^2 \sum_i (y_i - y)^2}, \quad (19)$$

其中, $\{(x_i, y_i) \mid i = 1, 2, \dots, m\}$ 是测试集合, $y = \sum_i y_i / m$.

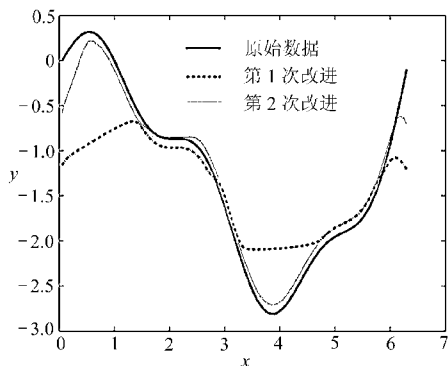
4.1 一元函数学习

图 1 显示了改进的 SVR 和传统方法对于函数

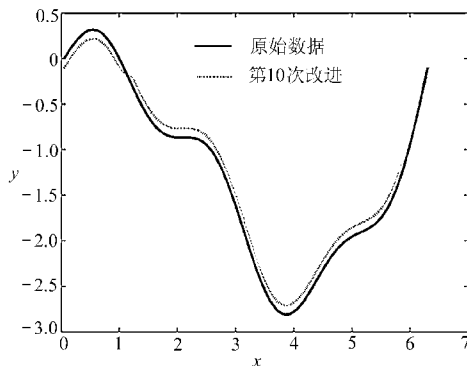
$$f(x) = \sin x + (\sin 3x)/3 - 2\sin(x/2), \quad x \in [0, 2\pi]$$



(a) 传统 SVR 的学习效果



(b) 第 1 次改进和第 2 次改进 SVR 的学习效果



(c) 第 10 次改进后的 SVR 学习效果

(参数取 $\sigma = 0.1, \varepsilon = 0.1, C = 0.05, h_i = 0.1, \tau = 1$)

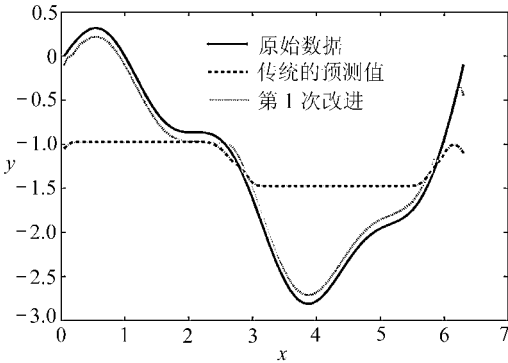
图 1 改进 SVR 和传统方法用于一元函数的学习效果

的学习效果,改进方法分别取了3组不同的 h_i 和 τ , 学习样本是区间 $[0, 2\pi]$ 上均匀采样的 63 个点, 测试数据是区间 $[0, 2\pi]$ 上均匀采样的 126 个点. 两种方法的参数和学习风险见表 1.

图 2 显示了对上述函数的另一个学习效果, 相应的参数和学习风险 $\sigma = 0.1, \varepsilon = 0.1, C = 0.1, \delta = 0.7902$ (见表 1).

表 1 对一元函数的学习效果及相关参数

传统方法的相关参数及学习风险 $\sigma, \varepsilon, C, \delta$	改进方法的相关参数 h_i, τ	第 1 次改进方法的学习风险 δ_1	第 2 次改进方法的学习风险 δ_2	第 10 次改进后的学习风险 δ_{10}
0.1, 0.1, 0.05, 0.895	0.1, 1	0.571 9	0.159 8	0.101 4
	0.1, 0.8	0.682 1	0.349 1	0.101 4
	0.08, 0.8	0.784 1	0.609 9	0.143 9
0.1, 0.1, 0.1, 0.790 2	0.8, 1	0.104 4	0.103 8	0.103 8



(参数取 $\sigma = 0.1, \varepsilon = 0.1, C = 0.1, h_i = 0.8, \tau = 1$)

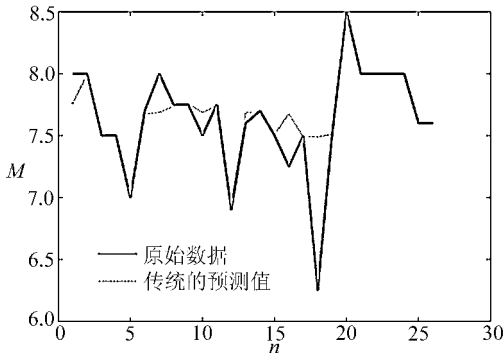
图 2 改进 SVR 和传统方法用于一元函数的学习效果

4.2 多元函数学习

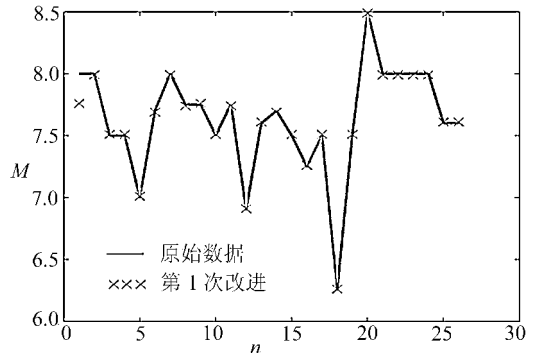
将上述改进的 SVR 和传统方法用于青藏高原强震震级的学习, 从 1411 年至 2001 年共 26 次 6 级以上的强震事件. 由断裂力学理论可知, 地震震级与地表破裂长度及垂直位错量有关^[12], 因此, 输入空间有二维, 分别是地表破裂长度及垂直位错量, 输出为地震震级. 图 3 显示了改进的 SVR 和传统方法的学习效果, 学习样本取 20 次强震事件, 测试数据是 26 次强震事件, 两种方法的参数和学习风险见表 2, 强震事件的相关参数和学习结果见表 3.

从上述几个结果对比可以发现, 在学习模型的基本参数都固定的情况下, 传统 SVR 的学习效果远不如迭代改进的 SVR, 对于不同的参数取值,

有的一次改进就能取得较好效果, 有的需要多次改进. 由于文献[10]的算法就是一次改进的方法, 因此, 迭代改进的 SVR 效果也是优于文献[10]的算法. 随着迭代次数的增加, 学习风险下降收敛, 收敛速度依赖于传统方法的基本参数和改进方法的参数.



(a) 传统 SVR 的学习效果



(b) 第 1 次改进 SVR 的学习效果

(参数取 $\sigma = 0.04, \varepsilon = 0.01, C = 2, h_i = 4, \tau = 0.06$)

图 3 改进 SVR 和传统方法用于多元函数的学习效果

表 2 对多元函数的学习效果及相关参数

传统方法的相关参数及 学习风险 $\sigma, \epsilon, C, \delta$	改进方法的相关 参数 h, τ	第 1 次改进方法的 学习风险 δ_1	第 2 次改进方法的 学习风险 δ_2
0.04, 0.01, 4, 0.629 1	4, 0.1	0.111 3	0.111 3

表 3 青藏高原地区强地震事件的地表破裂参数和震级预测值

破裂长度 L / km	垂直位移 H / m	实际震级 M_r	传统方法预测 的震级 M_1	第 1 次改进 SVR 预测 的震级 M_1	第 2 次改进 SVR 预测 的震级 M_2
81		8.000 0	7.760 0	7.760 0	7.760 0
136	8~ 9	8.000 0	7.990 0	7.990 0	7.990 0
57.7	5~ 5.5	7.500 0	7.510 0	7.510 0	7.510 0
50	0.3	7.500 0	7.510 0	7.510 0	7.510 0
14	0.9	7.000 0	7.010 0	7.010 0	7.010 0
60	0.5	7.700 0	7.676 2	7.690 0	7.690 0
90	2	8.000 0	7.686 2	7.990 0	7.990 0
150	2	7.750 0	7.740 0	7.740 0	7.740 0
81		7.750 0	7.760 0	7.760 0	7.760 0
90	5.6	7.500 0	7.686 2	7.510 0	7.510 0
35	2~ 5	7.750 0	7.740 0	7.740 0	7.740 0
44		6.900 0	6.910 0	6.910 0	6.910 0
90	0.5	7.600 0	7.686 2	7.610 0	7.609 6
90	5~ 6	7.700 0	7.686 2	7.690 0	7.690 0
180	5~ 6	7.500 0	7.510 0	7.510 0	7.510 0
60	1~ 2	7.250 0	7.676 2	7.260 0	7.260 0
38	4	7.500 0	7.490 0	7.510 0	7.510 0
38	0.6	6.250 0	7.490 0	6.260 0	6.260 0
100	1.5	7.500 0	7.510 0	7.510 0	7.510 0
237	4.7	8.500 0	8.490 0	8.490 0	8.490 0
110	7.4	8.000 0	7.990 0	7.990 0	7.990 0
190	4	8.000 0	7.990 0	7.990 0	7.990 0
130	2.7	8.000 0	7.990 0	7.990 0	7.990 0
120	1.7	8.000 0	7.990 0	7.990 0	7.990 0
73	5.4	7.600 0	7.610 0	7.610 0	7.609 6
25	4	7.600 0	7.610 0	7.610 0	7.609 6

5 总 结

综上,本文提出了一种关于SVR的改进算法,通过对核函数的改造,利用与数据有关的保角映射,使核函数具有了数据依赖性.在处理回归问题和数据具有高噪声的情形,用支持向量是不适宜的.本文工作与当前已有的研究不同,在构造保角映射时不是利用支持向量,而是利用全部学习样本.并且基本核函数的参数 σ 可取任意随机值,通过对核函数进行多次迭代改进,直至得到满意的效果.应用于仿真计算的结果表明,无论是一维连续函数还是多维强震数据,改进的SVR都比传统方法具有更好的学习效果,随着迭代次数的增加,学习风险下降收敛.

[参 考 文 献]

- [1] Vapnik V. 统计学习理论的本质[M]. 张学工译. 北京:清华大学出版社, 2000.

- [2] Scholkopf B, Sung K, Burges C. Comparing support vector machines with Gaussian kernels to radial basis function classifiers[J]. IEEE Trans Signal Processing, 1997, **45**(11): 2758-2765.
- [3] Perez-Cruz F, Navia-Vazquez A, Figueiras-Vidal AR, et al. Empirical risk minimization for support vector classifiers[J]. IEEE Trans on Neural Networks, 2003, **14**(2): 296-303.
- [4] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997, **19**(7): 711-720.
- [5] CAO Li-juan, Tay Francis E H. Financial forecasting using support vector machines[J]. Neural Computing & Applications, 2001, **10**(2): 184-192.
- [6] Tay Francis E H, CAO Li-juan. ϵ -Descending support vector machines for financial time series forecasting[J]. Neural Processing Letters, 2002, **15**(2): 179-195.
- [7] YANG Hai-qin, CHAN Lai-wan, King Irwin. Support vector machine regression for volatile stock market prediction[A]. In: Yin H, Allinson N, Freeman R, et al, Eds. Proceedings of Intelligent Data Engineering and Automated Learning [C]. Berlin: Springer-Verlag, 2002. 319-396.
- [8] Smola A J, Schölkopf B, Müller K R. The connection between regularization operators and support vector kernels[J]. Neural Network, 1998, **11**(4): 637-649.
- [9] Amari S, Wu Si. Improving support vector machine classifiers by modifying kernel functions[J]. Neural Networks, 1999, **12**(6): 783-789.
- [10] LIANG Yan-chun, SUN Yan-feng. An improved method of support vector machine and its applications to financial time series forecasting[J]. Progress in Natural Science, 2003, **13**(9): 696-700.
- [11] Colin C. Kernel methods: a survey of current techniques[J]. Neurocomputing, 2002, **48**: 63-84.
- [12] 马润勇, 彭建兵. 震级与破裂尺度及位错量关系的讨论[J]. 西北大学学报(自然科学版), 2006, **36**(5): 799-802.

An Iterative Modified Kernel Based on Training Data

ZHOU Zhi-xiang¹, HAN Feng-qing²

(1. College of Civil Engineering and Architecture, Chongqing Jiaotong University,
Chongqing 400074, P. R. China;

2. School of Science, Chongqing Jiaotong University, Chongqing 400074, P. R. China)

Abstract: In order to improve the performance of a support vector regression, a new method for modified kernel function is proposed. In this method the information of whole samples is included in kernel function by conformal mapping. So the kernel function is data-dependent. With random initial parameter of kernel function, the kernel function is modified repeatedly until a satisfactory effect is achieved. Compared with the conventional model, the improved approach needs no selecting parameters of kernel function. Simulation was finished for one-dimension continuous function and strong earthquake even. The results show that the improved approach has better learning ability and forecasting precision than the traditional model. With the iteration number increasing, the figure of merit will decrease and converge. The speed of convergence depends on the parameters in the algorithm.

Key words: support vector regression; data-dependent; kernel function; iteration