文章编号: 1001-0920(2009)06-0937-04

基于柯西分布加权的最小二乘支持向量机

邢永忠[1,2], 吴晓蓓[1], 徐志良[1]
(1. 南京理工大学 自动化学院, 南京 210094; 2. 中国人民解放军 63853 部队, 吉林 白城 137001)

摘 要: 针对 Suykens 等提出的加权最小二乘支持向量机(WLS-SVM)回归建模的不足和防止辨识模型的"过拟合",利用柯西分布函数的一些特性,提出了基于柯西分布加权的最小二乘支持向量机. 根据预测误差的统计特性,以确定加权规则的参数,从而赋予训练样本不同的权值. 由于考虑了生产过程中样本的实际特性,与已有的加权方法相比,新的加权最小二乘支持向量机更具有鲁棒性. 仿真结果验证了该方法的可行性和有效性.

关键词: 加权最小二乘支持向量机; 柯西分布函数; 过拟合; 预测误差
中图分类号: TP18 文献标识码: A

Weighted least square support vector machine based on Cauchy distribution

XING Yong-zhong[1,2], WU Xiao-bei[1], XU Zhi-liang[1]
(1. School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China; 2. Unit 63853 of the PLA, Baicheng 137001, China. Correspondent: XING Yong-zhong, E-mail: xyz21971@hotmail.com)

Abstract: Suykens et al. (2002) presented a weighted least-squares support vector machine (WLS-SVM) for regression problems and a weighted algorithm for robust approximation under the model bestirred enough. Therefore, by using the characters of Cauchy distribution function, a least-squares support vector machine based on Cauchy distritution weight is presented, which is prone to over-fitting under the model bestirred deficiently. Different values for weighted factor are selected based on statistical features of the prediction error. For the real features of the samples in the proceeding of production, the new WLS-SVM is more robust. The result of a numerical regression experiment shows the feasibility and effectiveness of this algorithm.

Key words: Weighted LS-SVM; Cauchy distribution function; Over-fitting; Prediction error

1 引 言

支持向量机(SVM)[1]融合了结构风险最小化原理、最优化理论和核函数映射等几项技术,有效地解决了在经典机器学习中出现的"维数灾难"和"局部极小"等传统困难,在很多领域得到了成功应用. 但是 SVM 在求解最优化问题时,需求解二次规划,计算复杂,难以满足海量数据和在线训练要求的实时性. 基于工程经验,Suykens 等[2]提出了一种最小二乘支持向量机(LS-SVM)算法. 该算法在优化指标中采用了平方项,且只有等式约束,将二次规划问题转化为求解线性方程组,因而简化了计算复杂性. 但同时失去了传统 SVM 的稀疏性和鲁棒性等优点. 为了克服 LS-SVM 丢失鲁棒性的问题,文献[3]提出了加权 LS-SVM, 较好地克服了输入信号能充分激励模态的对象噪声中野点带来的不良影响. 文献[4]利用非线性函数(Lyapunov exponents)作为加权因子,提出了改进的加权最小二乘支持向量机混沌系统建模方法,取得了一定效果.

然而在实际过程辨识中,建模数据对过程模态的描述往往是不充分的,用已有的加权方法容易导致辨识模型过度拟合. 为了提高最小二乘支持向量机动态建模的鲁棒性和防止辨识模型的"过拟合",本文利用柯西分布函数的一些特性,提出了基于柯西分布加权的最小二乘支持向量机. 仿真结果验证了该方法的可行性和有效性.

2 加权 LS-SVM

设给定的输入样本 x 为 n 维向量, k 个样本,其输出值可表示为 $(x_1, y_1), \cdots, (x_k, y_k) \in R^n \times R$, 则最小二乘支持向量机的优化问题和约束条件为

收稿日期: 2008-06-26; 修回日期: 2008-10-07.
作者简介: 邢永忠(1971—),男,黑龙江集贤人,博士生,从事机器学习、模式识别等研究; 吴晓蓓(1958—),女,成都人,教授,博士生导师,从事控制理论与应用、建模与辨识等研究.

$$\min_{W,B,N} J(W,N) = \frac{1}{2}W^{T}W + C\frac{1}{2}\sum_{i=1}^{k}N_i^2, \quad (1)$$

$$\text{s.t. } y_i[W^{T}U(x_i)] + B = 1 - N_i,\ i = 1,\cdots,k. \quad (2)$$

其中: $U(\cdot): R^n \to R^h$ 是将样本空间映射到高维特征空间的映射函数; 权值向量 $W \in R^h$; 样本误差和偏置值满足 $N_i \in R$ 及 $B \in R$; $C > 0$ 为正则化因子, 对减小误差和提高泛化能力作出折衷. 由于没有标准支持向量机的 ε 不敏感区域, 最小二乘支持向量机的解虽然得到了简化, 却丢失了标准支持向量机的鲁棒性和稀疏性等优点. 为了获得鲁棒性, 对式(1)的误差 N_i 进行加权. 设 N_i 对应的权值为 C_i, 则式(1)的优化问题和约束条件变为

$$\min_{W^*,B^*,N^*} J(W^*,N^*) = \frac{1}{2}W^{*T}W^* + C\frac{1}{2}\sum_{i=1}^{k}N_i^{*2}, \quad (3)$$

$$\text{s.t. } y_i = W^{*T}U(x_i) + B^* + N_i^*,\ i = 1,\cdots,k. \quad (4)$$

对应的拉哥朗日函数为

$$L(W^*,B^*,N^*,A^*) = J(W^*,N^*) - \sum_{i=1}^{k}A_i[W^{*T}U(x_i) + B^* + N_i^* - y_i]. \quad (5)$$

根据最优化理论, 消去变量 W^* 和 N^* 后, 得到线性方程组, 可写成矩阵形式

$$\begin{bmatrix} 0 & \vec{1}^{T} \\ \vec{1} & \Omega + V_C \end{bmatrix}_{(k+1)\times(k+1)} \begin{bmatrix} B^* \\ A^* \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}. \quad (6)$$

其中: $Y = [y_1, y_2, \cdots, y_k]^{T}$, $A^* = [A_1^*, A_2^*, \cdots, A_k^*]^{T}$, $\vec{1} = [1,\cdots,1]^{T}_{1\times k}$; $\Omega_{i,j} = K(x_i,x_j) = U(x_i)^{T}U(x_j)$, $i,j = 1,\cdots,k$, $K(x_i,x_j)$ 为满足 Mercer 条件的核函数; $V_C = \mathrm{diag}\{\frac{1}{CC_1},\cdots,\frac{1}{CC_k}\}$, 加权系数 C_i 由误差变量 N_i 确定. 另称式(6)中的 $\Omega + V_C$ 为核相关矩阵, 记 $A = \Omega + V_C$. 由式(6)展开可得

$$B^* = \frac{\vec{1}^{T}A^{-1}Y}{\vec{1}^{T}A^{-1}\vec{1}}, \quad (7)$$

$$A^* = A^{-1}(Y - \vec{1}B^*). \quad (8)$$

由式(7)和(8)可知, 计算出 A^{-1} 便可确定 A^* 和 B^*, 进而得到函数模型

$$y = \sum_{i=1}^{k}A_i^*K(x,x_i) + B^*. \quad (9)$$

由上可知, 加权系数 C_i 对支持向量 A_i^* 的值具有调优的功能. Suykens 等采用下式来确定 C_i:

$$C_i = \begin{cases} 1, & |N_i/\hat{s}| < c_1; \\ \dfrac{c_2 - |N_i/\hat{s}|}{c_2 - c_1}, & c_1 < |N_i/\hat{s}| < c_2; \\ 10^{-4}, & \text{other}. \end{cases} \quad (10)$$

其中 $\hat{s} = IQR/1.349$ 是误差变量 N_i 的标准方差的鲁棒估计值. IQR(Inter quartile range) 是 N_i 的 75% 分点与 25% 分点的间距, IQR 受两端的极值影响较小, 因此更为稳定. IQR 反映了处于中间位置的一半数据的范围, 该范围的大小可以反映整个数列的离散程度. 误差变量 $\hat{s}$ 的另一种估计为 $\hat{s} = 1.483\mathrm{MAD}(N_i)$, 其中 MAD 为平均绝对误差. 与 IQR 相同, 平均绝对误差也是用来反映数据的紧密程度. 常数 c_1 和 c_2 通常设为 $c_1 = 2.5$, $c_2 = 3$.

为获得辨识模型的鲁棒性, 这种加权方法的本质是根据个体样本的预测误差来进行加权. 如果误差小, 则表明该样本对模型的贡献大, 相应的权值就大, 对于输入信号能充分激励模态的研究对象. 该方法对提升模型的鲁棒性是很有效的. 但在过程辨识中, 建模数据对过程模态的描述往往是不充分的, 信息重复出现在几个工作点附近, 用于建模的样本具有冗余的特性. 与冗余的样本相比, 有用的/野点0或/拐点0的量太少, 模型对这些特殊点的表达不明显, 容易导致辨识模型的过度拟合. 如果按照这种方法进行加权后, 模型对这些特殊点的表达反而更加恶化了. 为兼顾特殊点对辨识模型的贡献, 以及避免过拟合, 下面提出一种基于柯西分布加权的最小二乘支持向量机.

3 基于柯西分布加权的 LS2SVM

3.1 柯西分布简介

通过研究随机变量的统计和分布特性, 有学者提出了柯西分布数学方法. 与正态分布相似, 柯西分布也是统计学中应用较为广泛的分布. 设连续随机变量 X 的概率密度为

$$f(x) = \frac{a}{P}\cdot\frac{1}{a^2 + (x-L)^2}. \quad (11)$$

其中: $-\infty < x < +\infty$, $-\infty < L < +\infty$, $a > 0$. 则称 X 服从参数为 L, a 的柯西分布, 记为 $X \sim C(L,a)$. 如果参数 L, a 已知, 则可画出柯西分布的密度曲线, 如图 1 所示.

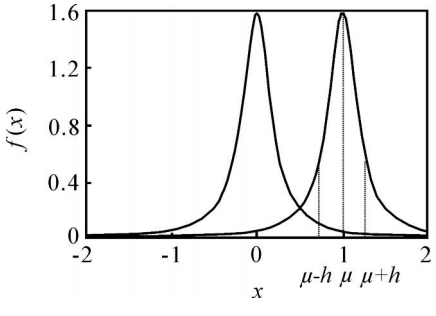


图 1 柯西分布的密度曲线

由图 1 可见, 柯西分布的密度曲线具有与正态分布的密度曲线相似的性质. x 离 L 越远, f(x) 的值越小. 这表明, 对于同样长度的区间, 当区间离 L 越

远时, X 落在这个区间的概率越小. 曲线以 x 轴为渐近线. 另外, 如果固定 a, 改变 L 的值, 则曲线沿着 x 轴平移, 而不改变其形状(如图 1). 可见, 柯西分布的概率密度曲线 y = f(x) 的位置完全由参数 L 确定. L 称为位置参数.

如果固定 L, 改变 a, 则由最大值 $f(L) = \frac{1}{Pa}$ 可知, 当 a 越小时图形变得越尖(如图 2), 因而 X 落在 L 附近的概率也越大.

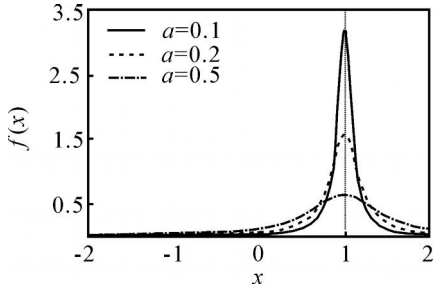
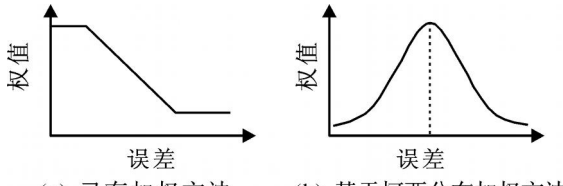


图 2　参数 a 对柯西分布密度曲线的影响

3.2　基于柯西分布加权的 LS-SVM

根据随机柯西分布函数曲线所具有的特性, 考虑了/ 野点0 对模型的贡献, 以及削弱冗余数据造成模型的/ 过度拟合0, 在取加权值方面, 图 3 给出了已有加权方法与基于柯西分布加权方法的不同之处. 新方法的权值最大点取预测误差处于中间的样本, 而误差很小或很大的样本其取值反而很小. 因此, 基于柯西分布加权方法更侧重于训练样本的实际特性, 从而对干扰(野点) 很敏感. 但根据柯西分布曲线加权, 会削弱干扰(野点) 带来的不良影响.



(a) 已有加权方法　　(b) 基于柯西分布加权方法

图 3　两种方法的对比示意

柯西分布函数的参数 L 和 a 决定了曲线的特性, 因此如何确定这两个参数是该方法的主要内容.

1) 参数 L 值的确定

加权的目的是调和/ 野点0 和/ 过拟合0 的作用, 因此参数 L 应取误差区域的中间值, 即

$$L = \frac{1}{k}\sum_{i=1}^{k} |N|,$$

或

$$L = \frac{\max(|N|) + \min(|N|)}{2}. \qquad (12)$$

2) 参数 a 的确定

参数 a 决定了柯西分布函数曲线的形状, 即样本点越靠近 L, 该点的权值也越大. 因此, 考虑预测误差的统计特性应遵循以下取值规则: 如果预测误差的分布较紧密, 则 a 的取值应偏大; 如果预测误差的分布较分散, 则 a 的取值应偏小.

综上所述, 可以得到基于柯西分布加权的最小二乘支持向量机的算法, 其步骤如下:

Step1: 选择合适的核函数 $K(x, xc) = \exp(-\|x - xc\|^2/2R^2)$ 和模型参数核宽度 R 以及正则化因子 C, 用最小二乘支持向量机建立初始模型;

Step2: 计算训练样本的预测误差 N, 根据式(12) 和误差的统计特性确定加权函数的参数 L 和 a;

Step3: 根据式(11) 计算各训练样本的加权值 G;

Step4: 将各加权值 G 代入式(7) 和(8) 计算模型的系数 A^* 和 B^*, 并获得最终的回归模型(9).

4　仿真结果与分析

为验证新算法的可行性和有效性, 采用以下的仿真例子(数据由一个非线性分段函数产生):

$$y(t) = \begin{cases} 0.06e^{-0.5u(t)}, & -10 \le u(t) < -7.5; \\ y(t-1), & -7.5 \le u(t) < -6; \\ -2.186u(t) - 12.864, & -6 \le u(t) < -2; \\ 4.246u(t), & -2 \le u(t) < 0; \\ 10e^{-0.05u(t)-0.5}\sin[(0.03u(t) + 0.7)u(t)], & \\ 0 \le u(t) < 10. & \end{cases} \qquad (13)$$

输入信号为

$$u(t) = -10 + 0.04(t-1). \qquad (14)$$

本文算法是在学习样本不能充分激励过程的前提下提出的, 因此对建模的训练样本进行了处理, 人为造成模拟数据存在/ 野点0 和/ 拐点0. 同时假定训练样本在 $u(t) \in [-10, -7.5]$ 区间较稀疏, 而在 $u(t) \in [-7.5, 10]$ 区间较密集. 训练样本的特点如图 4, 图 5 所示. 以下实验中, 取 C = 50, R = 0.2.

与 Suykens 加权方法进行对比实验, 表 1 为基于柯西分布加权方法与 Suykens 加权方法建模的预

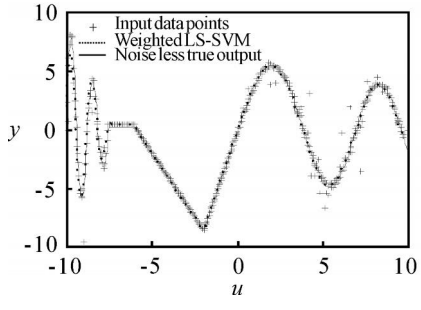


图 4　Suykens 加权方法的回归结果

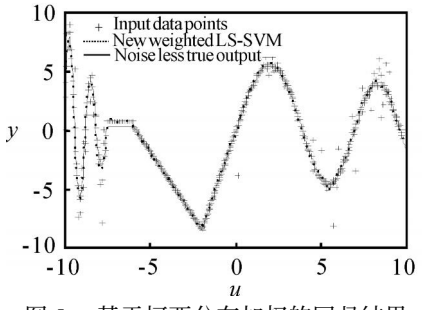


图 5 基于柯西分布加权的回归结果

表 1 预报误差的统计分析结果

方 法	最大绝对误差	平均相对误差	均方误差
Suykens 加权方法	0.432	0.0318	0.0661
本文方法(a = 0.5)	0.377	0.0207	0.0523

报误差分析结果. 图 4 与图 5 分别给出了两种加权方法的回归预报值. 图中, 实线为函数曲线, 虚线为加权后的回归结果. 通过比较可以发现, 在数据密集点处, 两种方法都取得了较好的拟合效果; 而在数据较稀疏的区域, 本文方法则具有更好的拟合效果.

实验结果表明, 虽然冗余区的训练样本受到噪声污染, 不能很好地反映函数的真实特性, 但相对于样本稀疏的区段, 密集的数据对函数的表征还是很有效的. 当加权规则侧重于这些点而忽略其他样本时, 会产生/ 过度拟合0, 因此基于随机分布(柯西分布、正态分布) 加权的方法可有效地解决这类问题. 仿真结果也验证了新方法的可行性和有效性.

5 结 论

考虑实际用于建模的样本不能充分激励过程, 本文利用随机柯西分布函数的一些特性, 提出了基于随机柯西分布加权的最小二乘支持向量机. 根据预测误差的统计特性来确定加权规则的参数, 从而赋予训练样本不同的权值. 由于考虑了/ 过度拟合0和/ 非高斯噪声0对辨识模型的影响, 相对已有的加权方法, 基于随机(柯西分布)加权的最小二乘支持向量机更具有鲁棒性和快速性, 因而具有很大的应用潜力.

参考文献(References)

[1] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32242.
(Zhang X G. Introduction to statistical learning theory and support vector machines [J]. Acta Automatica Sinica, 2000, 26(1): 32242.)

[2] Suykens J A K, Van Gestel T, De Brabanter J, et al. Least squares support vector machines[M]. Singapore: World Scientific Publishing Co Pte, 2002.

[3] Suykens J A K, Branbanter J K, Lukas L, et al, Weighted least squares support vector machine: Robustness and sparse approximation [J]. Neuro Computing, 2002, 48(1): 852105.

[4] Sun Jiancheng, Zhang Taiyi, Liu Feng. Modeling of chaotic systems based on modified weighted recurrent least squares support vector machines [J]. Chinese Physics, 2004, 13(12): 2045222053.

[5] Chen A J, Song Z H, Li P. Soft sensor modeling base on DICA2SVR[C]. Advances in Intelligent Computing, PT 1. Hefei, 2005, 3644: 8682877.

[6] 王连详, 方德植, 张鸣镛, 等. 数学手册[M]. 北京: 高等教育出版社, 2004: 7982801.
(Wang L X, Fang D Z, Zhang M Y, et al. Mathematics manual[M]. Beijing: Higher Education Press, 2004: 7982801.)

下 期 要 目