第 24 卷 第 1 期 控 制 与 决 策 2009 年 1 月
Vol. 24 No. 1 Control and Decision Jan. 2009

文章编号: 1001-0920(2009)01-0145-04

一种改进的在线最小二乘支持向量机回归算法

郭振凯[1], 宋召青[2], 毛剑琴[1]
(1. 北京航空航天大学 第七研究室, 北京 100083; 2. 海军航空工程学院 控制工程系, 山东 烟台 264001)

摘 要: 针对一般最小二乘支持向量机处理大规模数据集会出现训练速度慢、计算量大、不易在线训练的缺点,将修正后的遗忘因子矩形窗方法与支持向量机相结合,提出一种基于改进的遗忘因子矩形窗算法的在线最小二乘支持向量机回归算法,既突出了当前窗口数据的作用,又考虑了历史数据的影响. 所提出的算法可减少计算量,提高在线辨识精度. 仿真算例表明了该方法的有效性.

关键词: 最小二乘支持向量机回归; 改进的遗忘因子矩形窗; 在线学习
中图分类号: TP274 **文献标识码**: A

An improved online least squares support vector machines regression algorithm

GUO Zhen-kai[1], *SONG Zhao-qing*[2], *MAO Jian-qin*[1]
(1. The Seventh Research Division, Beijing University of Aeronautic and Astronautics, Beijing 100083, China; 2. Department of Control Engineering, Naval Aeronautical and Astronautical University, Yantai 264001, China. Correspondent: GUO Zhen-kai, E-mail: zhkguo @126. com)

Abstract: Aiming at the problem that the large-scale samples training process is slow and large computation, and difficult to train online for the standard least squares support vector machines, a learning algorithm of online least squares support vector machines regression (OLS-SVMR) based on improved rectangular window with forgetting factor (IRWFF) method is proposed by combining the modified rectangular window with forgetting factor algorithm with support vector machines. The present and past window data are considered simultaneously. The proposed algorithm has less computation and high accuracy. The simulation results show effectiveness of the algorithm.

Key words: Least squares support vector machines regression; IRWFF; Online learning

1 引 言

支持向量机(SVM)一经提出[1],就受到理论研究和工程研究的双重重视,它是统计学习理论基础上发展起来的新一代机器学习方法. SVM 较好地解决了小样本、非线性、高维数、局部极小点等实际问题,具有很强的泛化能力. 近年来,Suykens[2]提出一种新的 SVM 方法——最小二乘支持向量机(LS-SVM)方法,LS-SVM 是标准 SVM 的一种扩展. 与传统的 SVM 不同,LS-SVM 算法将 SVM 的求解从二次规划问题转化为线性方程组,提高了 SVM 的求解效率,降低了 SVM 的学习难度. 对于回归问题,未知变量的数目仅相当于同等规模分类问题的未知变量数目,从而避免了传统 SVM 学习方法中回归问题未知变量数目膨胀的问题,而且 LS-SVM 的数值稳定性和容易控制的策略,使得核函数矩阵在非正定的情况下也能取得良好的效果[3-5]. LS-SVM 算法中 Ω 的维数等于训练目标集的长度,即 Ω 的元素个数等于训练目标集长度的平方,当训练样本较大时,计算量迅速增大,由此可能导致计算膨胀问题.

在 LS-SVM 算法的基础上,人们提出了许多改进算法,如动态加权 LS-SVM[6],迭代式 LS-SVM[7],在线 LS-SVM (OLS-SVM)[8],增量 LS-SVM[9]等算法. 其中叶美盈等[8]提出的在线 LS-SVM 回归算法本质上是一种矩形窗算法,样本是窗式移动的,即 k 时刻的参数估计只依据有限个过去

收稿日期: 2007-10-28; 修回日期: 2008-04-07.
基金项目: 国家自然科学基金重点项目(60534020); 国家 973 计划项目(2002CB312205); 高等学校博士学科点专项科研基金项目(20070006060); 北京市重点学科基金项目(XK100060526).
作者简介: 郭振凯(1976—),男,山东烟台人,博士生,从事支持向量机建模与控制、预测控制等研究; 毛剑琴(1940—),女,上海人,教授,博士生导师,从事鲁棒控制、智能控制理论等研究.

的数据,在这些数据之前的老数据完全被剔除.

OLS-SVMR 算法计算量小,计算速度快,可有效地解决 LS-SVM 算法的计算膨胀问题. 本文将 IRWFF 方法与 SVM 相结合,提出一种改进的在线最小二乘支持向量机回归(IOLS-SVMR)算法. 该算法既考虑了历史数据的影响,又突出了新数据的作用.

2 最小二乘支持向量机

LS-SVM[3] 本质上是一种变形算法. 变形算法主要是通过增加函数项、变量或系数等方法使公式变形,从而产生出各种具有某一方面优势或者一定应用范围的算法.

假设学习样本集为

$$S = \{s_i \mid s_i = (x_i, y_i),\ x_i \in R^n,\ y_i \in R\}_{i=1}^{l},$$

回归函数的形式为

$$y(x) = w\phi(x) + b. \tag{1}$$

其中:$\phi(x)$ 是特征映射,w 和 b 是待求的回归参数.

Suykens 等提出的 LS-SVM 方法相当于求解下面的最小值问题:

$$\begin{cases} \min Q(w,e) = \dfrac{1}{2}\|w\|^2 + \dfrac{\gamma}{2}\sum_{i=1}^{l} e_i^2, \\ \text{s.t. } y_i = w\phi(x_i) + b + e_i,\ i = 1,2,\cdots,l, \end{cases} \tag{2}$$

其中 γ 为正则化参数. 最小值问题(2) 的 Lagrange 函数为

$$L(w,b,e,a) = \frac{1}{2}\|w\|^2 + \frac{\gamma}{2}\sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l} a_i(w\phi(x_i) + b + e_i - y_i), \tag{3}$$

其中 $a = [a_1\ \ a_2\ \ \cdots\ \ a_l]^T$. 由式(2) 的平衡条件可知

$$\begin{cases} \dfrac{\partial L}{\partial w} = w - \sum_{i=1}^{l} a_i\phi(x_i) = 0, \\ \dfrac{\partial L}{\partial b} = -\sum_{i=1}^{l} a_i = 0, \\ \dfrac{\partial L}{\partial e_i} = \gamma e_i - a_i = 0, \\ \dfrac{\partial L}{\partial a_i} = w\phi(x_i) + b + e_i - y_i = 0, \end{cases} \tag{4}$$

即

$$\begin{bmatrix} I & \mathbf{0} & \mathbf{0} & -Z \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\bar{\mathbf{1}} \\ \mathbf{0} & \mathbf{0} & \gamma I & -I \\ Z & \bar{\mathbf{1}} & I & \mathbf{0} \end{bmatrix} \begin{bmatrix} w \\ b \\ e \\ a \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ y \end{bmatrix}. \tag{5}$$

其中

$Z = [\phi(x_1)\ \ \cdots\ \ \phi(x_l)]^T$, $y = [y_1\ \ \cdots\ \ y_l]^T$, $\bar{\mathbf{1}} = [1\ \ \cdots\ \ 1]^T \in R^l$, $e = [e_1\ \ \cdots\ \ e_l]^T$, $a = [a_1\ \ \cdots\ \ a_l]^T$.

由式(4) 可知 $w = \sum_{i=1}^{l} a_i\phi(x_i)$, $e_i = \dfrac{1}{\gamma}a_i$, 消去 w 和 e_i 之后,可得线性方程组

$$\begin{bmatrix} 0 & \bar{\mathbf{1}}^T \\ \bar{\mathbf{1}} & ZZ^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}. \tag{6}$$

记 $\Omega = ZZ^T$, $\Omega_{ij} = K(\cdot,\cdot)$, 称式(6) 中的 $\Omega + \gamma^{-1} I$ 为核相关矩阵. 若记 $A \equiv \Omega + \gamma^{-1} I$, 则式(6) 等价于

$$\begin{bmatrix} \mathbf{0} & \bar{\mathbf{1}}^T \\ \bar{\mathbf{1}} & A \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ y \end{bmatrix}. \tag{7}$$

由式(7) 可以求得

$$b = \frac{\bar{\mathbf{1}}^T A^{-1} y}{\bar{\mathbf{1}}^T A^{-1} \bar{\mathbf{1}}}, \tag{8}$$

$$a = A^{-1}(y - b\bar{\mathbf{1}}). \tag{9}$$

由式(8), (9) 和 $w = \sum_{i=1}^{l} a_i\phi(x_i)$ 可得式(1) 的回归函数

$$y(x) = w\phi(x) + b = \sum_{i=1}^{l} a_i K(x_i, x) + b. \tag{10}$$

将确定回归函数的参数 a, b 统称为回归参数. 由式(10) 可以看出,确定回归参数的关键在于计算核相关矩阵的逆 A^{-1}.

3 基于 IRWFF 算法的 IOLS-SVMR

3.1 基于 RWFF 算法的 OLS-SVMR

基于矩形窗算法的 OLS-SVMR 的样本是随着时刻 k 的递进而翻滚的,即每进来一个新样本,便同时丢弃一个旧样本,样本个数保持不变[8]. 但该算法对历史数据仅是简单的抛弃,并没有考虑历史数据的影响. 遗忘因子法又称衰减记忆法,其基本思想是对历史数据加上遗忘因子,既考虑了历史数据的影响,又突出了新数据的作用[10].

假设学习样本集为

$$S = \{s_i \mid s_i = (x_i, y_i), x_i \in R^n, y_i \in R\}_{i=1}^{l},$$

矩形窗宽度为 m,则 k 时刻的学习样本集可表示为 $\{x(k), y(k)\}$. 其中:$x(k) = [x_{k-m+1}\ \ x_{k-m+2}\ \ \cdots\ \ x_k]$, $y(k) = [y_{k-m+1}\ \ y_{k-m+2}\ \ \cdots\ \ y_k]^T$, $x_k \in R^n$, $y_k \in R$. 核函数矩阵 Ω,待求的 Lagrange 乘子 α 和常值偏差 b 可表示为

$$\Omega_K(i,j) = K(x_{k-m+i}, x_{k-m+j}),\quad i,j = 1,2,\cdots,m; \tag{11}$$

$$\alpha(k) = [\alpha_{k-m+1}\ \ \alpha_{k-m+2}\ \ \cdots\ \ \alpha_k]^T; \tag{12}$$

$$b(k) = b_k. \tag{13}$$

则 k 时刻 ORLS-SVM 的输出为

$$y_k = \sum_{i=k-m+1}^{k} \alpha_i K(x, x_i) + b(k). \tag{14}$$

令 $Q_k = \Omega_k + I/\gamma$,则有

$$\begin{bmatrix} 0 & \bar{1}^{T} \\ \bar{1} & Q_k \end{bmatrix} \begin{bmatrix} b(k) \\ a(k) \end{bmatrix} = \begin{bmatrix} 0 \\ y_k \end{bmatrix}. \tag{15}$$

令

$$\theta_k = \begin{bmatrix} b(k) \\ \alpha(k) \end{bmatrix}, \tag{16}$$

$$\Phi_k = \begin{bmatrix} 0 & \bar{1}^{T} \\ \bar{1} & Q_k \end{bmatrix}, \tag{17}$$

$$z_k = \begin{bmatrix} 0 \\ y_k \end{bmatrix}. \tag{18}$$

假设 $J_k(\theta)$ 可表示成二次型函数

$$J_k(\theta) = (\theta - \theta_k)^{T} M_k^{-1} (\theta - \theta_k) + \beta_k. \tag{19}$$

选取参数估计的指标函数

$$J_{k+1}(\theta) = \alpha J_k(\theta) + (z_{k+1} - \Phi_{k+1}\theta)^{T}(z_{k+1} - \Phi_{k+1}\theta) - \alpha(z_{k+1} - \Phi_{k+1}\theta_k)^{T}(\alpha I + \Phi_{k+1} M_k \Phi_{k+1})^{-1}(z_{k+1} - \Phi_{k+1}\theta_k). \tag{20}$$

其中:$0 < \alpha < 1$,称为遗忘因子,也称衰减因子或加权因子;M_k^{-1} 和 Φ_{k+1} 均为正定对称矩阵. 则基于 RWFF 算法的 OLS-SVMR 回归参数的递推公式为

$$\theta_{k+1} = \theta_k + N_{k+1}(z_{k+1} - \Phi_{k+1}\theta_k), \tag{21}$$

$$N_{k+1} = M_k \Phi_{k+1}(\alpha I + \Phi_{k+1} M_k \Phi_{k+1})^{-1}, \tag{22}$$

$$M_{k+1} = \frac{M_k}{\alpha} - \frac{M_k}{\alpha}\Phi_{k+1}(\alpha I + \Phi_{k+1} M_k \Phi_{k+1})^{-1}\Phi_{k+1} M_k. \tag{23}$$

初始参数 $M_0 = c^2 I$,其中 c 是一个充分大的实数. 可以看出,OLS-SVMR 的 RWFF 算法是一种在线递推算法,它不仅考虑了历史数据的影响,还突出了当前数据的作用.

3.2　基于 IRWFF 算法的 IOLS-SVMR

RWFF 算法虽然考虑了历史数据的影响,但该算法在每次迭代过程中,需要进行多个矩阵的运算,增加了每次迭代的运算量,而计算机有限字长等因素会使计算的舍入误差累加,从而降低辨识的精度. 出于对上述原因的考虑,对 OLS-SVMR 的 RWFF 算法作如下简化:

在式(22)中,将 N_{k+1} 修改为

$$N_{k+1} = M_k \Phi_{k+1}(\alpha \Phi_{k+1} M_k \Phi_{k+1} + \Phi_{k+1} M_k \Phi_{k+1})^{-1} = \Phi_{k+1}^{-1}/(1+\alpha) = \beta \Phi_{k+1}^{-1}, \tag{24}$$

则式(21)可简化为

$$\theta_{k+1} = \theta_k + \beta \Phi_{k+1}^{-1}(z_{k+1} - \Phi_{k+1}\theta_k). \tag{25}$$

由以上推导可得如下定理:

定理 1　基于 IRWFF 算法的 IOLS-SVMR 参数的递推公式为

$$\theta_{k+1} = \theta_k + \beta \Phi_{k+1}^{-1}(z_{k+1} - \Phi_{k+1}\theta_k). \tag{26}$$

其中:$0 < \beta \le 1$,称为遗忘因子,也称衰减因子或加权因子;Φ_{k+1} 为正定对称矩阵.

定理 2　如果基于 IRWFF 算法的 IOLS-SVMR 参数通过式(26)进行调整,当 $0 < \beta \le 1$ 时,误差是单调递减的;当 $\beta = 1$ 时,则基于 IRWFF 算法的 IOLS-SVMR 的误差将收敛到零.

证明　记 $e_k = z_{k+1} - \Phi_{k+1}\theta_k$ 为第 $k+1$ 次参数修正前的学习误差,即将前一时刻学习得到的参数 θ_k 作为当前时刻的参数所带来的误差;记 $e_{k+1} = z_{k+1} - \Phi_{k+1}\theta_{k+1}$ 为第 $k+1$ 次参数修正后的学习误差. 则有

$$\begin{aligned} e_{k+1} &= z_{k+1} - \Phi_{k+1}\theta_{k+1} = \\ & z_{k+1} - \Phi_{k+1}[\theta_k + \beta\Phi_{k+1}^{-1}(z_{k+1} - \Phi_{k+1}\theta_k)] = \\ & (1-\beta) z_{k+1} - (1-\beta)\Phi_{k+1}\theta_k = \\ & (1-\beta) e_k. \end{aligned}$$

因此,当 $0 < \beta \le 1$ 时,误差是单调递减的;当 $\beta = 1$ 时,参数修正后的学习误差收敛到零. □

显然,基于 IRWFF 算法的 IOLS-SVMR 不仅保留了原算法的优点,而且简化了算法,减小了计算量.

4　仿真实例

考虑如下非线性系统:

$$\begin{cases} \dot{x}_1 = x_1 + 0.8 x_2, \\ \dot{x}_2 = -x_1^3 - 0.1 x_2 + u, \\ y = x_1. \end{cases}$$

输入信号为 $u = 2\sin(0.5t) + \sin t + \sin(2t)$,时间为 $t = [0, 5]$ s,仿真步长为 0.01 s.

采用以下 3 种 LS-SVMR 算法建立该系统的模型. $\gamma = 10000$, 核函数取为 $K(x, x_i) = \exp\left(\frac{\| x - x_i \|^2}{2\sigma^2}\right)$,$\sigma = 0.25$.

1) 采用矩形窗 LS-SVMR 算法[8]. 矩形窗长度 $m = 2$,其他参数不变. 辨识误差如图 1 所示.

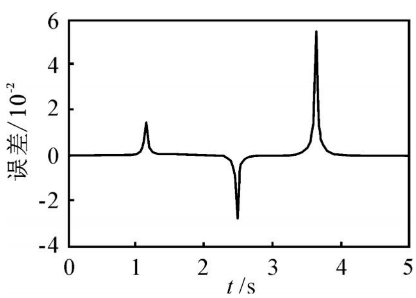

图 1　辨识误差

2) 采用基于 RWFF 的 OLS-SVMR 算法(21)～(23). 矩形窗长度 $m = 2$,遗忘因子 $\alpha = 0.05$,初

始参数 $c = 100$,其他参数不变. 辨识误差如图 2 所示.

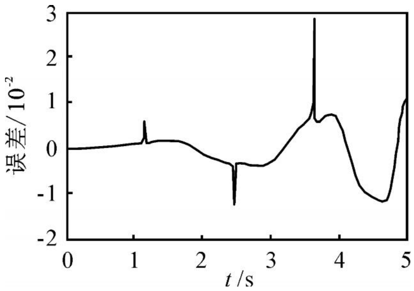

图 2 辨识误差

3) 采用基于 IRWFF 的 IOLS-SVMR 算法. 遗忘因子 $\beta = 1$,其他参数不变. 辨识误差如图 3 所示.

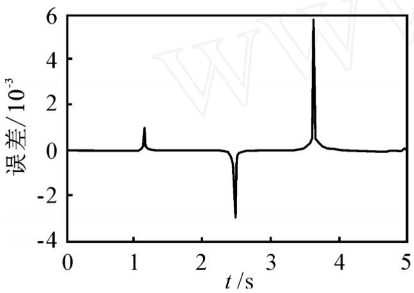

图 3 辨识误差

上述 3 种算法的仿真结果比较如表 1 所示.

表 1 3 种算法的仿真结果比较

	矩形窗算法的 LS-SVMR	基于 RWFF 算法的 OLS-SVMR	基于 IRWFF 算法的 IOLS-SVMR
最大误差	0.0547	0.0287	0.0058
MSE	3.2124×10^{-5}	2.3548×10^{-5}	1.3729×10^{-7}
运行时间/s	9.763	5.162	4.562

由仿真结果可以看出,基于 IRWFF 算法的 IOLS-SVMR 算法优于矩形窗 LS-SVMR 算法与基于 RWFF 算法的 OLS-SVMR 算法. 基于 RWFF 算法的 OLS-SVMR 算法,在每次迭代过程中,需要进行多个矩阵的运算,增加了每次迭代的运算量,因计算的舍入误差累加,从而降低了辨识的精度. 而基于 IRWFF 算法的 IOLS-SVMR 算法简化了迭代过程,降低了累加舍入误差,从而可减小运算时间与最大误差,提高辨识的精度.

5 结 论

矩形窗 LS-SVMR 算法采用在线递推的方法,有效解决了 LS-SVMR 算法的计算膨胀问题. 本文将修正的最小二乘遗忘因子方法与支持向量机相结合,提出了基于IRWFF的IOLS-SVMR算法. 该算法简化了计算步骤,提高了辨识速度和建模精度.

参考文献(References)

[1] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.

[2] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers [J]. Neural Processing Letters, 1999, 9(3): 293-300.

[3] Gestel T V, Suykens J A K, Baesens B, et al. Benchmarking least squares support vector machine classifiers[J]. Machine Learning, 2004, 54(1): 5-32.

[4] Anguital D, Boni A. Digital least squares support vector machines[J]. Neural Processing Letters, 2003, 18(1): 65-72.

[5] Tsujinishi D, Abe S. Fuzzy least squares support vector machines for multiclass problems[J]. Neural Networks, 2003, 16(5/6): 785-792.

[6] 范玉刚,李平,宋执环. 动态加权最小二乘支持向量机[J]. 控制与决策, 2006, 21(10): 1129-1134.
(Fan Y G, Li P, Song Z H. Dynamic weighted least squares support vector machines [J]. Control and Decision, 2006, 21(10): 1129-1134.)

[7] 吴春国. 广义染色体遗传算法与迭代式最小二乘支持向量机回归算法研究[D]. 长春: 吉林大学, 2006.
(Wu C G. Study on generalized chromosome genetic algorithm and iterative least squares support vector machine regression [D]. Changchun: Jilin University, 2006.)

[8] 叶美盈,汪晓东,张浩然. 基于在线最小二乘支持向量机回归的混沌时间序列预测[J]. 物理学报, 2005, 54(6): 2568-2573.
(Ye M Y, Wang X D, Zhang H R. Chaotic time series forecasting using online least squares support vector machine regression[J]. Acta Physica Sinica, 2005, 54(6): 2568-2573.)

[9] 张浩然,汪晓东. 回归最小二乘支持向量机的增量和在线式学习算法[J]. 计算机学报, 2006, 29(3): 400-406.
(Zhang H R, Wang X D. Incremental and online learning algorithm for regression least squares support vector machine[J]. Chinese J of Computers, 2006, 29(3): 400-406.)

[10] 李言俊,张科. 系统辨识理论及应用[M]. 北京: 国防工业出版社, 2003.
(Li Y J, Zhang K. System identification theory and application [M]. Beijing: National Defence Industry Press, 2003.)