

DOI:10.16356/j.1005-2615.2015.01.016

一类增量式支持向量机的分析

郑关胜^{1,2} 王建东¹ 顾彬² 於跃成³

(1. 南京航空航天大学计算机科学与技术学院, 南京, 210016;
2. 南京信息工程大学计算机与软件学院, 南京, 210044;
3. 江苏科技大学计算机科学与工程学院, 镇江, 212003)

摘要:针对训练数据发生增量改变时,标准一类支持向量机的批处理算法需要重新进行训练,不适合在线增量环境学习的问题,提出一种详细的增量式标准一类分类向量机算法,并通过理论分析对该算法的可行性和有限收敛性进行了证明,确保该算法的每步调整都是可靠的,并确保该算法通过有限步调整最终收敛到问题的最优解。在标准数据集上的实验结果验证了理论分析的正确性。

关键词:一类支持向量机;增量式学习;可行性分析;收敛性分析

中图分类号:TP301 **文献标志码:**A **文章编号:**1005-2615(2015)01-0113-06

Analysis of One Class Incremental Support Vector Machine

Zheng Guansheng^{1,2}, Wang Jiandong¹, Gu Bin², Yu Yuecheng³

(1. College of Computer Science & Technology, Nanjing University of
Aeronautics & Astronautics, Nanjing, 210016, China;

2. School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China;

3. School of Computer Science & Engineering, Jiangsu University of Science and Technology, Zhenjiang, 212003, China)

Abstract: Batch implementations of standard one class support vector machine (SVM) are inefficient on an on-line setting because they must be retrained from scratch every time when the training set is incremental modified. To solve this problem, a detailed incremental one class SVM algorithm is given, and the feasibility and the finite convergence of the algorithm are proven through theoretical analysis. It is ensured that each adjustment step in the C&P algorithm is reliable, and the algorithm will converge to the optimal solution within finite steps. The experimental results on benchmark datasets verify the correctness of theoretical analysis.

Key words: one class support vector machine; incremental learning; feasibility analysis; convergence analysis

由 Vapnik 等提出的支持向量机(Support vector machine, SVM)^[1], 目前已成为解决分类、回归和其他统计学习问题的一种流行的技术。在实际问题环境中, 训练样本数据大多为在线环境下增量提供, 所以有必要对支持一类分类支持向量机设计

相应的增量式算法。2001 年 Cauwenberghs 和 Poggio^[2]针对标准的支持向量机, 给出了一种精确的增量和减量式学习算法(C&P 算法)。为了对 C&P 算法有更深入的认识, 2006 年, Laskov 等^[3]对该算法增量式部分给出了部分理论分析, 并给出

基金项目:国家自然科学基金(61139002)重点资助项目;国家自然科学基金青年科学基金(61202137)资助项目;中国民航信息技术科研基地开放基金(CAAC-ITRB-201203)资助项目。

收稿日期:2013-04-16; **修订日期:**2013-09-26

通信作者:王建东, 男, 教授, 博士生导师, E-mail: aics@nuaa.edu.cn。

简单扩展到一类分类支持向量机算法;2010年, Karasuyama等^[4]扩展了C&P算法,使之能同时处理多个样本的增量式学习;2012年,Gu等^[5]扩展了C&P算法,使之能处理支持向量分类机。2013年,顾彬等^[6]详细讨论了增量式标准支持向量机的可行性以及收敛性分析。本文将在上述工作的基础上给出详细的增量式一类支持向量机算法,并从理论分析的角度,全面讨论增量式一类支持向量机算法的可行性及有限收敛性。

1 增量式一类支持向量机

若训练样本表示为: $\{x_1, x_2, x_3, \dots, x_n\}$, 其中 n 为训练样本的个数, $x_i \in \mathbf{R}^d$, d 为维数。若 φ 是将样本点映射到高维空间中的非线性映射, 则一类分类问题的原始目标函数可以定义为

$$\begin{aligned} \min R^2 + C \sum_i \xi_i \\ \text{s. t. } \|\varphi(x_i) - c\|^2 \leq R^2 + \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (1)$$

式中: c 为球心; R 为球的半径; ξ_i 为松弛变量; C 为惩罚系数。其相应的对偶问题可以表示为

$$\begin{aligned} \min L(\alpha) = \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) - \sum_i \alpha_i K(x_i, x_i) \\ \text{s. t. } \sum_i \alpha_i = 1 \\ 0 \leq \alpha_i \leq C \end{aligned} \quad (2)$$

通过优化问题(2)的求解, 同样可以得到标准问题的解为

$$\begin{aligned} \min \sum_{0 \leq \alpha_i \leq C_i, j=1}^n \alpha_i \alpha_j K(x_i, x_j) - \\ \sum_{i=1}^n \alpha_i K(x_i, x_i) + \xi \left(\sum_{i=1}^n \alpha_i - 1 \right) \end{aligned} \quad (3)$$

对式(3)求一阶偏导, 可以导出下列的KKT(Karush-Kuhn-Tucker)条件为

$$\begin{aligned} g_i = 2 \sum_j \alpha_j K(x_i, x_j) - K(x_i, x_i) + \xi = \\ \begin{cases} \geq 0 & \alpha_i = C & S_R \\ 0 & 0 < \alpha_i < C & S_S \\ \leq 0 & \alpha_i = 0 & S_E \end{cases} \\ \sum_i \alpha_i = 1 \end{aligned} \quad (4)$$

式(4)将训练集 X 划分成3个子集 S_S, S_R 和 S_E , 其中 S_S 为边缘支持向量集, 即分布在超球边界的样本; S_E 为错分支持向量集, 指的是分布在超球外部的离异点; S_R 为剩余向量集, 表示样本点分布在超球内部。

当向训练集中 $X = \{x_1, x_2, \dots, x_n\}$ 增加一个新样本 x_c , 式(4)划分出的 S_S, S_R 和 S_E 可能不满足KKT条件, 所以需要调整 g_c 和 α_c 进行相应的调整, 在扩增样本集 $X' = \{x_1, x_2, \dots, x_n, x_c\}$ 中寻找最优解。在对 α_c 的调整过程中, 为了保证集合 X' 的所有样本都满足KKT等价条件, 那么仅有边缘支持向量集 S_S 中的 $|S_S|$ 个样本对应的 α 乘子可以调整。根据KKT条件(4), 则可以获得如下方程

$$\begin{cases} 2 \sum_{j \in S_S} \Delta \alpha_j K(x_i, x_j) + \Delta \xi = -2 \Delta \alpha_c K(x_i, x_c) \\ \sum_{j \in S_S} \Delta \alpha_j = -\Delta \alpha_c \end{cases} \quad (5)$$

将式(5)表示为矩阵形式为

$$\begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 2K(x_{s1}, x_{s1}) & 2K(x_{s1}, x_{s2}) & \dots & 2K(x_{s1}, x_{sc}) \\ 1 & 2K(x_{s2}, x_{s1}) & 2K(x_{s2}, x_{s2}) & \dots & 2K(x_{s2}, x_{sc}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2K(x_{sc}, x_{s1}) & 2K(x_{sc}, x_{s2}) & \dots & 2K(x_{sc}, x_{sc}) \end{bmatrix} \cdot \begin{bmatrix} \Delta \xi \\ \Delta \alpha_{s1} \\ \Delta \alpha_{s2} \\ \vdots \\ \Delta \alpha_{sc} \end{bmatrix} = - \begin{bmatrix} 1 \\ 2K(x_{s1}, x_c) \\ 2K(x_{s2}, x_c) \\ \vdots \\ 2K(x_{sc}, x_c) \end{bmatrix} \Delta \alpha_c \quad (6)$$

令 $Q_{ij} = K(x_i, x_j)$, 则有

$$\mathbf{Q}' \begin{bmatrix} \Delta \xi \\ \Delta \alpha_{SS} \end{bmatrix} = - \begin{bmatrix} 1 \\ 2Q_{Sc} \end{bmatrix} \Delta \alpha_c, \mathbf{Q}' = \begin{bmatrix} 0 & 1 \\ 1 & 2Q_{SS} \end{bmatrix} \quad (7)$$

这样, 可以得到 α_c, α_{SS} 与 ξ 的线性变化关系, 其中 α_{SS} 表示边缘支持向量集 S_S 中所有样本对应的 α 乘子向量, 即

$$\begin{cases} \Delta \alpha = \beta_S^c \Delta \alpha_c \\ \Delta \xi = \beta^c \Delta \alpha_c \end{cases}, \begin{bmatrix} \beta^c \\ \beta_S^c \end{bmatrix} = -\mathbf{R} \cdot \begin{bmatrix} 1 \\ 2Q_{Sc} \end{bmatrix}, \mathbf{R} = \mathbf{Q}'^{-1} \quad (8)$$

相应地, 也可以得到 α_c 与 g_i 的线性变化关系, 即

$$\Delta g_i = \gamma_i \Delta \alpha_c, \gamma_i = 2 \sum_{j \in S_S} \beta_S^c Q_{ij} + 2Q_{ic} + \beta^c, \forall i \in S \quad (9)$$

α_c 的调整将改变边缘支持向量集 S_S , 所以 \mathbf{R} 矩阵相应地也会发生调整。对于集合 S_S 的变化有两种情形: 第一种是增加一个样本到集合 S_S ; 第二种是从集合 S_S 移出一个样本。相应地, 矩阵 \mathbf{R} 也需要随着更新。

对于第一种情形, 将新增样本 x_c 加入样本集

S_s 中时,矩阵 \mathbf{R} 规则扩展如下

$$\mathbf{R}^+ \leftarrow \begin{bmatrix} \mathbf{R} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} \beta^t \\ \beta^t_{S_s} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \beta^t \\ \beta^t_{S_s} \\ 1 \end{bmatrix}^T \quad (10)$$

对于第二种情形,矩阵 \mathbf{R} 根据如下规则进行收缩: $\forall i, j \in S_s$, 并且 $i \neq c, j \neq c$: $R_{ij}^- \leftarrow R_{ij} - R_{ij}^{-1} R_c R_{ij}$, 其中 c 代表集合 S_s 中一个移出的样本。

2 可行性分析

为了确保一类增量式支持向量机算法每步计算是可靠的,需要回答绝缘更新过程中一些关键的问题,具体包括:(1)在调整 α_c 过程中 S_s 始终不为空(定理 1);(2)矩阵 \mathbf{Q}' 始终存在逆矩阵(定理 2);(3)对矩阵 \mathbf{R} 的更新策略是正确的(定理 3);(4)向边缘支持向量集 S_s 新加入(移出)一个样本 t 后,在紧接着的调整中 t 不会反向移出(加入)边缘支持向量集(定理 4 和 5)。

定理 1 在 α_c 的调整过程中,边缘支持向量集 S_s 始终不为空。

证明:若 $S_s = \emptyset$,可改变 ξ 的值使得一个样本移入集合 S_s ,从而使 S_s 不为空。假设 $S_s = \{(x_t, y_t)\}$,根据式(8)可以证明 $\beta^t_i = -1$,再结合式(4),可得 ξ 值的改变方向必须为 -1 才能保证移入 S_s 的样本在接下来的调整中不会移出。据此最大可能的 $\Delta\xi$ 的绝对值可以表示为 $|\Delta\xi^{S_s}| = \min_{i \in S_s} |g_i|$;在候选样本 c 移入 S_s 前,最大可能的 $\Delta\xi$ 的绝对值可以表示为 $|\Delta\xi| = -g_c$,那么 ξ 的最后改变值为 $\Delta b^* = -\min\{|\Delta\xi^{S_s}|, |\Delta\xi|\}$ 。定理得证。

定理 2 矩阵 \mathbf{Q}' 始终存在逆矩阵。

证明 根据定理 1 知,在调整 α_c 过程中集合 S_s 始终不为空,则不失一般性假设 S_s 存在 l 个样本,则

$$\mathbf{Q}' = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 2Q_{11} & \cdots & 2Q_{1l} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2Q_{l1} & \cdots & 2Q_{ll} \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & \mathbf{I}^T \\ \mathbf{I} & 2\mathbf{Q} \end{bmatrix}$$

通过文献[6]假设 \mathbf{I} 和 \mathbf{Q} 正定,那么必存在 \mathbf{Q}^{-1} ,并且 \mathbf{Q}^{-1} 也是正定。因为

$$\begin{bmatrix} 1 & -0.5 \cdot \mathbf{I}^T \mathbf{Q}^{-1} \\ 0 & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} 0 & \mathbf{I}^T \\ 1 & 2\mathbf{Q} \end{bmatrix} = \begin{bmatrix} -0.5 \mathbf{I}^T \mathbf{Q}^{-1} \mathbf{I}^T & 0 \\ 1 & 2\mathbf{Q} \end{bmatrix}$$

所以有 $\det(\mathbf{Q}') = \det(2\mathbf{Q}) \cdot (-0.5 \mathbf{I}^T \mathbf{Q}^{-1} \mathbf{I}^T) \neq 0$,那么易知矩阵 \mathbf{Q}' 始终存在逆矩阵。

引理 1 当向集合 S_s 增加一个样本 t ,相应的 $r'_t > 0$ 。

证明 首先证明结论

$$y_t \varphi(x_t) + \sum_{i \in S_s} \beta'_i y_i \varphi(x_i) \neq 0$$

假设结论不成立,那么有 $y_t \varphi(x_t) = -\sum_{i \in S_s} \beta'_i y_i \varphi(x_i)$,代入式(9),可以得到 α_c ,这和前提条件向集合 S_s 增加一个样本 t 相矛盾。基于此,进一步有

$$\begin{aligned} \gamma'_t &= 2Q_u + \sum_{j \in S_s} 2Q_{uj} \beta'_j + \beta^t = \\ & \sqrt{2} y_t \varphi(x_t) \cdot (\sqrt{2} y_t \varphi(x_t) + \\ & \sqrt{2} \sum_{j \in S_s} \beta'_j y_j \varphi(x_j)) + \beta^t = \\ & \sqrt{2} y_t \varphi(x_t) \cdot (\sqrt{2} y_t \varphi(x_t) + \\ & \sqrt{2} \sum_{j \in S_s} \beta'_j y_j \varphi(x_j)) - \beta^t \sum_{j \in S_s} \beta'_j = \\ & \sqrt{2} y_t \varphi(x_t) \cdot (\sqrt{2} y_t \varphi(x_t) + \sqrt{2} \sum_{j \in S_s} \beta'_j y_j \varphi(x_j)) - \\ & \sum_{j \in S_s} \beta'_j (-\sqrt{2} y_j \varphi(x_j) \cdot (\sqrt{2} y_t \varphi(x_t) + \\ & \sqrt{2} \sum_{j \in S_s} \beta'_j y_j \varphi(x_j))) = \\ & 2(y_t \varphi(x_t) + \sum_{j \in S_s} \beta'_j y_j \varphi(x_j))^2 > 0 \end{aligned}$$

引理 2 当向集合 S_s 移出一个样本 t ,相应的 $R_u > 0$ 。

根据逆矩阵的定义和假设 1,可以很容易地得出该结论。

定理 3 对于集合 S_s 增加一个样本或移出一个样本 t 的两种情形,其相应的矩阵 \mathbf{R} 的两种更新策略均是正确的。

证明 对于集合 S_s 增加一个样本 t ,可以证明

$$\mathbf{Q}^+ \cdot \mathbf{R}^+ = \begin{bmatrix} 0 & 1 & \cdots & 1 & 1 \\ 1 & 2Q_{11} & \cdots & 2Q_{1n} & 2Q_{1t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2Q_{n1} & \cdots & 2Q_{nn} & 2Q_{nt} \\ 1 & 2Q_{t1} & \cdots & 2Q_{tn} & 2Q_{tt} \end{bmatrix} \cdot$$

$$\begin{bmatrix} 0 \\ 0 & \mathbf{R} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} \beta^t \\ \beta^t_1 \\ \vdots \\ \beta^t_n \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \beta^t \\ \beta^t_1 \\ \vdots \\ \beta^t_n \\ 1 \end{bmatrix}^T = \mathbf{I}$$

$$\begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \gamma'_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 1 \end{bmatrix}$$

其中:=应用了式(6), \mathbf{I} 为相应的单位矩阵。对于集合 S_s 移出一个样本 t ,可以证明

$$\mathbf{Q}^- \cdot \mathbf{R}^- = \mathbf{Q}^- \cdot \mathbf{R}_{\setminus u} - \frac{\mathbf{Q}^-}{R_u} (\mathbf{R}_{*t} \cdot \mathbf{R}_{t*})_{\setminus u} = \mathbf{Q}^- \cdot \mathbf{R}_{\setminus u} + (\mathbf{Q}_{*t} \cdot \mathbf{R}_{t*})_{\setminus u} = \mathbf{I}$$

定理4 向集合 S_s 新加入一个样本 t 后,在紧接着的调整中 t 不会反向移出集合 S_s 。

证明 假设前面一次调整记为 k ,则下次调整记为 $k+1$,有

$$\beta^{[k+1]} = \begin{bmatrix} \beta_{\xi}^{[k+1]} \\ \beta_1^{[k+1]} \\ \vdots \\ \beta_n^{[k+1]} \\ \beta_t^{[k+1]} \end{bmatrix} = -\mathbf{R}^+ \cdot \begin{bmatrix} 1 \\ 2Q_{1c} \\ \vdots \\ 2Q_{nc} \\ 2Q_{tc} \end{bmatrix} = - \left(\begin{bmatrix} \mathbf{R} & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} \beta_{\xi}^k \\ \beta_1^k \\ \vdots \\ \beta_n^k \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_{\xi}^k \\ \beta_1^k \\ \vdots \\ \beta_n^k \\ 1 \end{bmatrix}^T \right) \cdot \begin{bmatrix} 1 \\ 2Q_{1c} \\ \vdots \\ 2Q_{nc} \\ 2Q_{tc} \end{bmatrix} = \begin{bmatrix} \beta_{\xi}^{[k]} \\ \beta_1^{[k]} \\ \vdots \\ \beta_n^{[k]} \\ 0 \end{bmatrix} - \frac{1}{\gamma'_t} \begin{bmatrix} \beta_{\xi}^k \\ \beta_1^k \\ \vdots \\ \beta_n^k \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_{\xi}^k \\ \beta_1^k \\ \vdots \\ \beta_n^k \\ 1 \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ 2Q_{1c} \\ \vdots \\ 2Q_{nc} \\ 2Q_{tc} \end{bmatrix} = \begin{bmatrix} \beta_b^{[k]} \\ \beta_1^{[k]} \\ \vdots \\ \beta_n^{[k]} \\ 0 \end{bmatrix} - \frac{\gamma'_c}{\gamma'_t} \begin{bmatrix} \beta_b^k \\ \beta_1^k \\ \vdots \\ \beta_n^k \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_b^k \\ \beta_1^k \\ \vdots \\ \beta_n^k \\ 1 \end{bmatrix}^T$$

那么有 $\beta_t^{[k+1]} = -\frac{\gamma'_c}{\gamma'_t}$ 。又因为

$$\gamma_t^{[k+1]} = 2Q_c + \sum_{j \in S_s^{[k+1]}} 2Q_{jt} \beta_j^{[k+1]} + \beta_{\xi}^{[k+1]} = 2Q_c + \sum_{j \in S_s} 2\beta_j^{[k]} Q_{jt} + \beta_b^{[k]} - \left(\frac{\gamma'_c}{\gamma'_t}\right) \cdot (2Q_u + \sum_{j \in S_s} 2\beta_j^k Q_{jt} + \beta_{\xi}^k) =$$

$$\gamma_t^c - \frac{(\gamma'_c)}{\gamma'_t} \cdot \gamma_t^c = \gamma_t^c - \gamma_t^c = 0$$

所以, $\beta_t^{[k+1]} = \frac{\gamma'_c}{\gamma'_t}$ 。这意味着向集合 S_s 新加入一个样本 t 后,在紧接着的调整中 t 不会反向移出 S_s 。因为对于 t 加入 S_s 的两种方式:一是从 S_R 移向 S_s ,二是从 S_E 移向 S_s 。对于第一种方式有 $\alpha_t^k = 0$, $g_t^k > 0$ 且 $\gamma_t^k < 0$,那么通过引理2可得 $\beta_t^{[k+1]} > 0$;对于第二种方式有 S_s ,那么通过引理2可得 t 。所以在紧接着的调整中 t 不会反向移出 S_s 。

定理5 从集合 S_s 移出一个样本 t 后,在紧接着的调整中 t 不会反向加入集合 S_s 。

证明 假设前面一次调整记为 k ,则下次调整记为 $k+1$,则有

$$\beta^{[k+1]} = \begin{bmatrix} \beta_{\xi}^{[k+1]} \\ \beta_1^{[k+1]} \\ \vdots \\ \beta_{n-1}^{[k+1]} \end{bmatrix} = -\mathbf{R}^- \cdot \begin{bmatrix} 1 \\ 2Q_{1c} \\ \vdots \\ 2Q_{(n-1)c} \end{bmatrix} = - \left(\mathbf{R}_{n \times n} - \frac{1}{R_u} \begin{bmatrix} R_{1t}R_{t1} & \cdots & R_{nt}R_{t1} \\ \vdots & \ddots & \vdots \\ R_{1t}R_{tn} & \cdots & R_{nt}R_{tn} \end{bmatrix} \right) \cdot \begin{bmatrix} 1 \\ 2Q_{1c} \\ \vdots \\ 2Q_{(n-1)c} \end{bmatrix} = \begin{bmatrix} \beta_{\xi}^{[k]} - 2R_{1t}Q_{tc} \\ \beta_1^{[k]} - 2R_{2t}Q_{tc} \\ \vdots \\ \beta_{n-1}^{[k]} - 2R_{nt}Q_{tc} \end{bmatrix} + \frac{1}{R_u} \begin{bmatrix} -R_{1t}\beta_t^{[k]} + 2R_{1t}R_uQ_{tc} \\ -R_{2t}\beta_t^{[k]} + 2R_{2t}R_uQ_{tc} \\ \vdots \\ -R_{nt}\beta_t^{[k]} + 2R_{nt}R_uQ_{tc} \end{bmatrix} \cdot \gamma_t^{[k+1]} = 2Q_c + \sum_{j \in S_s^{[k+1]}} 2Q_{tj} \beta_j^{k+1} + \beta_{\xi}^{k+1} = 2Q_c + \sum_{j \in S_s^{[k+1]}} 2Q_{tj} (\beta_j^k - R_{j+1t}Q_{tc}) + (\beta^k - R_{1t}2Q_c) + \frac{1}{R_u} (-R_{1t}\beta_t^k + 2R_{1t}R_uQ_{tc}) + \frac{1}{R_u} \left(\sum_{j \in S_s^{[k+1]}} 2Q_{tj} (-R_{j+1}\beta_t^k + 2R_{j+1}R_uQ_{tc}) \right) = \gamma_t^c - \frac{\beta_t^k}{R_u} \cdot (2Q_u R_u + \sum_{j \in S_s^{[k+1]}} 2Q_{tj} R_{jt} + R_{1t}) = -\frac{\beta_t^k}{R_u}$$

其中:=应用了 γ_t^c 的定义;:=应用了等式 $2Q_u R_u + \sum_{j \in S_s^{[k+1]}} 2Q_{tj} R_{jt} + R_{1t} = 1$ 。类似于定理4后面的分析,可以得出结论:从集合 S_s 移出一个样本 t 后,在紧接着的调整中 t 不会反向加入集合 S_s 。

3 有限收敛性分析

有限收敛性将确保一类增量式支持向量机算法在有限步后收敛到问题的最优解。本文首先证明了在调整过程中目标函数 W 是严格单调递减的(定理 6);在此基础上,证明了一类增量式标准支持向量机算法经过有限步迭代后,将收敛于相应问题的最优解(定理 7)。

引理 3 在一类增量式标准支持向量机算法中, $\forall i \in S_s, \beta_i^c \neq \pm\infty$, 并且 $\forall i \in S - S_s, \gamma_i \neq \pm\infty$ 。

证明 根据定理 5 的更新策略, $\forall i \in S_s$, 有

$$\begin{aligned} \beta_i^c = & - \sum_{k \in S_s} 2R_{ik} Q_{kc} - R_{ii} = \\ & - \frac{1}{\det(Q')} \left(\sum_{k \in S_s} (-1)^{i+k} 2\det(Q'_{ki}) Q_{kc} + \right. \\ & \left. (-1)^{i+1} \det(Q'_{i1}) \right) \end{aligned}$$

假设存在一个样本 $i \in S_s$, 使得 $\beta_i^c \neq \pm\infty$, 那么必然有 $\det(Q') = 0$, 这和定理 2 相矛盾。所以 $\forall i \in S_s, \beta_i^c \neq \pm\infty$, 并依据定理 5, 也有: $\forall i \in S - S_s, \gamma_i \neq \pm\infty$ 。

引理 4 对于一类增量式支持向量机算法, 相应的最大调整量 $\Delta\alpha_c^{\max} > 0$ 。

根据 1.3 节和 1.4 节的内容, 以及引理 4, 较容易推导出引理 4 的结论。

定理 6 对于一类增量式标准支持向量机算法, 调整过程中目标函数 W 是严格单调递减。

证明 假设前面一次调整记为 k , 则下次调整记为 $k+1$, 并令 $\beta_{S_R} = 0, \beta_{S_E} = 0, \beta_c = 1$, 那么有

$$\begin{aligned} W^{[k+1]} = & \sum_{i_1, i_2 \in S} (\alpha_{i_1}^{[k]} + \beta_{i_1}^{[k]} \Delta\alpha_c^{[k]}) (\alpha_{i_2}^{[k]} + \beta_{i_2}^{[k]} \Delta\alpha_c^{[k]}) Q_{i_1 i_2} + \\ & (\zeta^{[k]} + \beta_\zeta^{[k]} \Delta\alpha_c^{[k]}) \left(\sum_{i \in S} \alpha_i^{[k]} + \beta_i^{[k]} \Delta\alpha_c^{[k]} \right) - \\ & \sum_{i \in S} (\alpha_i^{[k]} + \beta_i^{[k]} \Delta\alpha_c^{[k]}) Q_{ii} = \\ & W^{[k]} + \sum_{i \in S} g_i^{[k]} \beta_i^{[k]} \Delta\alpha_c^{[k]} + \\ & \frac{1}{2} \sum_{i \in S} \gamma_i^{[k]} \beta_i^{[k]} (\Delta\alpha_c^{[k]})^2 + \frac{1}{2} \beta_\zeta^{[k]} (\Delta\alpha_c^{[k]})^2 \sum_{i \in S} \beta_i^{[k]} = \\ & W^{[k]} + g_c^{[k]} \Delta\alpha_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} (\Delta\alpha_c^{[k]})^2 + 0 = \\ & W^{[k]} + \left(g_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} \Delta\alpha_c^{[k]} \right) \Delta\alpha_c^{[k]} \end{aligned}$$

容易验证 $g_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} \Delta\alpha_c^{[k]} < 0$, 因此根据引理 5 可知 $W^{[k+1]} - W^{[k]} < 0$ 。

定理 7 对于一类增量式支持向量机算法, 经

过有限步迭代后, 将收敛于相应问题的最优解。

证明 类似于文献[6]中定理 8 的证明, 可知该定理成立。

4 实验验证

本文基于 MATLAB 实现了增量式一类支持向量机算法。实验对可行性与有限收敛性的主要结论分别做了验证, 包括: (1) 边缘支持向量集 S_s 不为空; (2) 矩阵 Q' 始终存在逆矩阵; (3) 向边缘支持向量集 S_s 新加入(移出)一个样本 t 后, 在紧接着的调整中 t 不会反向移出(加入)边缘支持向量集; (4) 本文算法在有限步内能收敛到最优解。

所有实验在 3.2 GHz, Pentium-4, 具有 4 GB 的内存和 MATLAB 7.5 平台的机器上进行的。表 1 给出了在实验中使用的 4 个 UCI 标准数据集的特性 1, 这些数据有些为两类数据集, 有些为多类数据集, 一类支持向量机统一不区分样本的标签。实验中使用了 3 种核函数, 分别为线性核 $K(x_i, x_j) = [x_i, x_j]$; 多项式核 $K(x_i, x_j) = ([x_i, x_j] + 1)^d$, 其中 $d = 2$; 高斯核 $K(x_i, x_j) = \exp\left(\frac{-x_i - x_j^2}{2\sigma^2}\right)$, 其中 $\sigma = 0.707$ 。

表 1 实验中使用的标准数据集

数据集名称	数据集的大小	属性集的大小
Iris	150	4
Wine	178	13
Sonar	208	60
Ionosphere	351	34

为了说明在本文算法学习中边缘支持向量集 S_s 不为空, 矩阵 Q' 存在逆矩阵, 以及向边缘支持向量集 S_s 新加入(移出)一个样本 t 后, 在紧接着的调整中 t 不会反向移出(加入)边缘支持向量集, 实验在样本数量分别为 50, 100, 150 的 Iris 数据集, 以及数据样本数量为 25, 75, 125, 175 的 Wine 数据集, 样本数量为 50, 100, 150, 200 的 Sonar 数据集和数据量为 80, 160, 240, 320 的 Ionosphere 数据集上模拟增量式变化的数据集, 分别尝试 200 次学习, 测试结果均验证了上述的结论。图 1 说明了本文算法在不同训练数据大小情况下的平均迭代次数, 数据结果验证了算法的有限收敛性, 尤其是算法在高斯核下具有更快速的收敛速度。

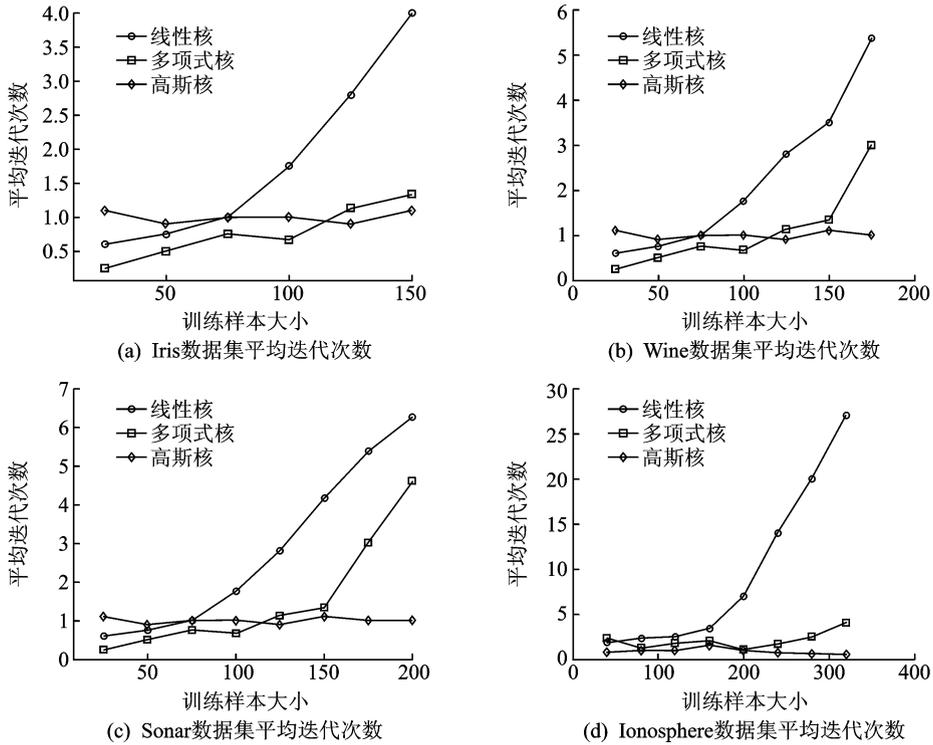


图1 数据集上的平均迭代次数

Fig. 1 Average iterations on datasets

5 结束语

一类增量式支持向量机算法是一种精确的增量式算法,在很多任务中有着广泛的应用前景。为了更好地认识该算法,本文从理论和实验上较为全面地分析了一类增量式支持向量机算法的可行性与有限收敛性,实验结果验证了理论分析的正确性。

参考文献:

- [1] Vapnik V. The nature of statistical learning theory [M]. New York: Springer, 1999.
- [2] Cauwenberghs G, Poggio T. Incremental and decremental support vector machine learning [J]. Advances in Neural information Processing Systems, 2001(12): 409-415.
- [3] Laskov P, Gehl C, Krüger S, et al. Incremental sup-

port vector learning: Analysis, implementation and applications[J]. The Journal of Machine Learning Research, 2006, 7: 1909-1936.

- [4] Karasuyama M, Takeuchi I. Multiple incremental decremental learning of support vector machines[J]. Neural Networks, IEEE Transactions on, 2010, 21 (7): 1048-1059.
- [5] Gu B, Wang J D, Yu Y C, et al. Accurate on-line ν -support vector learning [J]. Neural Networks, 2012, 27: 51-59.
- [6] 顾彬,郑关胜,王建东. 增量和减量式标准支持向量机的分析[J]. 软件学报, 2013, 24(7): 1601-1613.
Gu Bin, Zheng Guansheng, Wang Jiandong. Analysis for incremental and decremental standard support vector machine[J]. Journal of Software, 2013, 24 (7): 1601-1613.