

引用格式:沈洋,戴月明. 基于戴帽 L1 范数的双支持向量机[J]. 微电子学与计算机,2020,37(1):72-79. [SHEN Y,DAI Y M. Twin support vector machine based on capped L1-norm [J]. Microelectronics & Computer,2020,37(1):72-79.]

基于戴帽 L1 范数的双支持向量机

沈 洋,戴月明

(江南大学 物联网工程学院,江苏 无锡 214122)

摘 要:针对双支持向量机模型易受异常点影响导致泛化性能较低的问题,提出了一种基于戴帽 L1 范数的双支持向量机模型.采用带有上限值的戴帽 L1 范数代替 L2 范数来构造最优化问题,一定程度上削弱了离群点、噪音点对于两个超平面构造的影响,增强了模型的鲁棒性.另外,针对构造的新的双支持向量机模型最优化问题提出了一个简单有效的迭代算法并且在理论上证明了该算法的收敛性.在无噪以及有噪 UCI 数据集上的实验结果表明,与其它支持向量机模型相比,该模型有着更强的鲁棒性以及稳定性.

关键词:双支持向量机;L1 范数;L2 范数;戴帽 L1 范数;损失函数

中图分类号: TP391.1

文献标识码: A

文章编号: 1000-7180(2020)01-0072-08

Twin support vector machine based on capped L1-norm

SHEN Yang,DAI Yue-ming

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: In view of the low generalization performance of twin support vector machine model which is easily affected by outliers, a new twin support vector machine model based on capped L1-norm is proposed in this paper. The capped L1-norm with upper bound value is used to construct the optimization problem instead of L2-norm, which weakens the influence of outliers and noise points on the construction of two hyperplanes to a certain extent and enhances the robustness of the model. In addition, a simple and efficient iterative algorithm is proposed for the construction of a new twin support vector machine model optimization problem, and the convergence of the algorithm is proved theoretically. Experimental results on noiseless and noisy UCI data sets show that the proposed model is more robust and stable than other SVM models.

Key words: twin support vector machine; L1-norm; L2-norm; capped L1-norm; loss function

1 引言

支持向量机是由 Vapnik 于 1992 年提出的一种机器学习算法,它以结构风险最小化为原理,克服了之前算法采用经验风险容易过拟合的问题,而且通过引入核函数实现了对于非线性问题的求解^[1].但是由于支持向量机每次训练需要解决一个单独的二次规划问题,因此当它面对大规模的问题时会耗费大量的训练时间,另外当待区分的类别是不均衡分

布时,构造的超平面会产生一定的偏移,使得模型的性能受到一定的影响.

针对上面的两个问题,许多研究者对支持向量机进行改进并提出了一些新的支持向量机模型.其中,Jayadeva 和 Khemchandani 于 2007 年提出了一种双支持向量机模型^[2],与传统方法不同,它解决的不是一个单独的二次规划问题,而是两个规模更小的二次规划问题,这使得它的训练速度相比支持向量机提升了三倍.另外,由于构造了两个互不平行的超平面,使得它对于数据的不均衡分布不是特别敏

感,因此对于不均衡的分类也有比较好的适应性。

基于双支持向量机泛化性能好、训练效率高、不受不均衡分布影响等特性,许多改进的双支持向量机模型也纷纷出现。例如:Kumar 和 Gopal 提出了一个最小二乘法双支持向量机,他们利用等式约束替换原来的不等式约束通过求解线性方程组实现对原问题的求解^[3];Zaho 和 Chen 通过将正则化引入到双支持向量机的目标函数中,增强了模型的泛化性能^[4];Shao 等将加权线性损失函数引入到双支持向量机中,通过求解线性方程组来求解目标问题,极大的减少了训练中的计算量^[5];Xu 等提出了一个基于弹球损失函数的双支持向量机,利用弹球损失对于噪音数据不敏感的特性极大的增强了模型的鲁棒性^[6];除此之外,还有许多除此之外,还有许多对于双支持向量机的改进方法。

上述这些方法虽然对双支持向量机进行了部分改进,但是它们仍然保留了使用 L2 范数平方进行距离度量的方式,因此使得模型在遇到异常值时,会放大异常值对于超平面构造的影响,从而导致模型的泛化性能不强^[7],针对这一问题,本文提出了一个基于戴帽 L1 范数的双支持向量机(Twin Support Vector Machine Based on capped L1-norm, CLTS-VM)。利用戴帽 L1 范数代替原本双支持向量机的 L2 范数作为损失函数,一定程度上消除了异常值离群点对于超平面构造的影响,提高了模型的准确率;另外,针对该模型的最优化问题,提出了一个简易的最优化算法用于求解,并且从理论上证明了该优化算法的收敛性。在无噪以及有噪数据集上的实验结果证明了该模型的有效性以及鲁棒性。

2 双支持向量机

双支持向量机模型通过构造两个非平行的超平面来实现对于类别的划分,如图 1 所示。设 m_1 与 m_2 分别是正负类别的样本数量,那么 $A_{m_1 \times n}$ 与 $B_{m_2 \times n}$ 就是两个类别的向量矩阵模型,两个非平行超平面如下所示:

$$H_1: w_1 x + b_1 = 0 \quad (1)$$

$$H_2: w_2 x + b_2 = 0 \quad (2)$$

这两个超平面都是距离本类别的样本尽可能的近,与其它类别样本保持一定的距离。

为了找到这两个非平行的超平面,双支持向量机构造了如下最优化问题:

$$\min \frac{1}{2} (A w_1 + e_1 b_1)^T (A w_1 + e_1 b_1) + C_1 e_2^T \delta_1$$

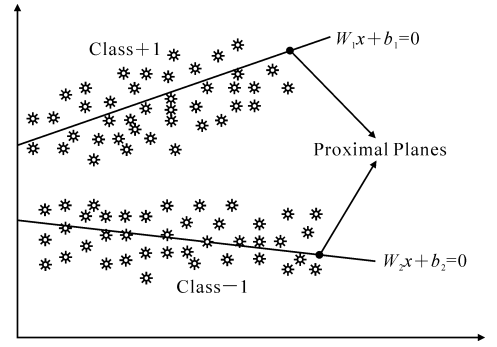


图 1 TWSVM 几何解释图

$$\text{s. t. } -(B w_1 + e_2 b_1) + \delta_1 \geq e_2 \quad (3)$$

$$\min \frac{1}{2} (B w_2 + e_2 b_2)^T (B w_2 + e_2 b_2) + C_2 e_1^T \delta_2$$

$$\text{s. t. } -(A w_2 + e_1 b_2) + \delta_2 \geq e_1 \quad (4)$$

式中, C_1 与 C_2 是两个惩罚系数; δ_1 与 δ_2 是松弛变量。这两个优化问题可以采用与支持向量机同样的解决方式,通过求解拉格朗日函数求解对偶问题进行解答。

$$\max e_2^T \alpha - \frac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha \quad (5)$$

$$\text{s. t. } 0 \leq \alpha \leq C_{1e2}$$

$$\max e_1^T \beta - \frac{1}{2} \beta^T H (G^T G)^{-1} H^T \beta \quad (6)$$

$$\text{s. t. } 0 \leq \beta \leq C_2 e_1$$

式中, $H = [A, e_1]$, $G = [B, e_2]$,最后可以得出两个二次规划问题的解为

$$\langle w_1 \rangle = -(H^T H)^{-1} G^T \alpha \quad (7)$$

$$\langle w_2 \rangle = -(G^T G)^{-1} H^T \beta \quad (8)$$

当两个超平面确定之后,只需要通过下列决策函数就可以对待测样本进行标记

$$f(x) = \operatorname{argmin} |w_i x + b_i|, i = 1, 2 \quad (9)$$

式中, $| \cdot |$ 用于计算样本与给定超平面的距离; w_i 与 b_i 是求得的法向量以及偏移量。

3 基于戴帽 L1 范数的双支持向量机

3.1 戴帽 L1 损失

双支持向量机由于采用了两个较小规模的二次规划问题代替了原来单个的问题,大大加快了支持向量机模型训练的速度,然而利用 L2 范数的平方来计算样本与平面之间的距离很容易扩大异常点对于超平面构造的影响,L1 范数^[8]损失相比 L2 来说,对于异常点的敏感程度较低一些,但是还不够,因此本文采用戴帽 L1 范数^[9-10]作为损失函数。戴帽 L1

范数损失函数如下所示

$$L_{\epsilon}(u) = \min(|u|, \epsilon) \quad (10)$$

式中, u 是输入的自变量; ϵ 是自选参数, 也就是该损失函数的上限。

图 2 显示的是 L2、L1 以及戴帽 L1 (ϵ 为 2.5) 三种函数的损失对比图, 我们发现, 随着自变量 u 的值不断增大或者不断减少, L1 以及 L2 两个函数的损失值都会越来越大, 但是变大的幅度不同, L1 变大的幅度比 L2 慢得多, 这就导致了 L1 损失函数相比 L2 对异常值的敏感度要低。但是, 随着自变量 u 的值不断增大或减少, 当达到一定程度时, 异常值还是会对 L1 损失函数造成影响。而戴帽 L1 损失就不会发生这种情况, 因为它设定了一个阈值 ϵ , 当自变量 u 的值达到这个阈值时, 损失值就不会再增加, 而是等于阈值所对应的 L1 损失的临界值^[11], 这就使得戴帽 L1 损失相比其它两种损失函数对于异常值的敏感度要低得多。

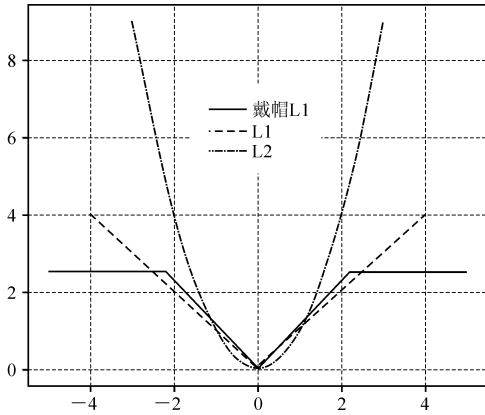


图 2 三种损失函数损失示例图

3.2 CLTSVM 模型

介绍了本文所采用的损失函数—戴帽 L1 损失函数, 采用该损失函数后, 双支持向量机的最优化问题转化为如下形式

$$\begin{aligned} \min \sum_{i=1}^{m_1} \min(|w_1 x_i + b_1|, \epsilon_1) + C_1 \sum_{i=1}^{m_2} \min(|\delta_{1,i}|, \epsilon_2) \\ \text{s. t. } -(Bw_1 + e_2 b_1) + \delta_1 \geq e_2 \end{aligned} \quad (11)$$

$$\begin{aligned} \min \sum_{i=1}^{m_2} \min(|w_2 x_i + b_2|, \epsilon_3) + C_2 \sum_{i=1}^{m_1} \min(|\delta_{2,i}|, \epsilon_4) \\ \text{s. t. } -(Aw_2 + e_1 b_2) + \delta_2 \geq e_1 \end{aligned} \quad (12)$$

式中, C_1 与 C_2 是惩罚系数; δ_1 与 δ_2 是松弛变量; m_1 与 m_2 分别是正负类别的样本数量; $A_{m_1 \times n}$ 与 $B_{m_2 \times n}$ 是两个类别的向量矩阵模型; $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ 是阈值, 用于

控制各个戴帽 L1 损失函数的上限, 设置合适的阈值, 可以使得损失函数有很强的鲁棒性。

上述显示的最优化问题相比之前的二次规划问题更难解答, 我们可以采用文献[9, 10, 12] 中的重新加权技巧, 将公式(11)(12) 中的问题转化为

$$\begin{aligned} \min \frac{1}{2} (Aw_1 + e_1 b_1)^T O (Aw_1 + e_1 b_1) + C_1 \delta_1^T P \delta_1 \\ \text{s. t. } -(Bw_1 + e_2 b_1) + \delta_1 \geq e_2 \end{aligned} \quad (13)$$

$$\begin{aligned} \min \frac{1}{2} (Bw_2 + e_2 b_2)^T Q (Bw_2 + e_2 b_2) + C_2 \delta_2^T R \delta_2 \\ \text{s. t. } -(Aw_2 + e_1 b_2) + \delta_2 \geq e_1 \end{aligned} \quad (14)$$

其中, O, P, Q, R 是四个对角矩阵, 它们的对角元素由下面公式获得

$$O_i = \begin{cases} \frac{1}{|w_1 x_i + b_1|}, & |w_1 x_i + b_1| \leq \epsilon_1 \\ sv, & |w_1 x_i + b_1| > \epsilon_1 \end{cases} \quad (15)$$

$$P_i = \begin{cases} \frac{1}{|\delta_{1,i}|}, & |\delta_{1,i}| \leq \epsilon_2 \\ sv, & |\delta_{1,i}| > \epsilon_2 \end{cases} \quad (16)$$

$$Q_i = \begin{cases} \frac{1}{|w_2 x_i + b_2|}, & |w_2 x_i + b_2| \leq \epsilon_3 \\ sv, & |w_2 x_i + b_2| > \epsilon_3 \end{cases} \quad (17)$$

$$R_i = \begin{cases} \frac{1}{|\delta_{2,i}|}, & |\delta_{2,i}| \leq \epsilon_4 \\ sv, & |\delta_{2,i}| > \epsilon_4 \end{cases} \quad (18)$$

这四个对角矩阵主要是用来判断样本是否是异常点, 如果一个样本点到超平面的距离大于阈值, 则将该样本认为是异常点, 不采用该样本点的数据, 而是给它赋予一个很小的值 sv , 以消除它对于超平面构造的影响。

上述最优化问题的拉格朗日函数如下所示(以第一个问题为例)

$$\begin{aligned} L(w_1, b_1, \alpha, \delta_1) = \\ \frac{1}{2} (Aw_1 + e_1 b_1)^T O (Aw_1 + e_1 b_1) + \frac{C_1}{2} \delta_1^T P \delta_1 - \alpha^T (-(Bw_1 + e_2 b_1) + \delta_1 - e_2) \end{aligned} \quad (19)$$

式中, α 是拉格朗日系数。本问题的 KKT 条件如下所示

$$\frac{\partial L}{\partial w_1} = A^T O (Aw_1 + e_1 b_1) + B^T \alpha = 0 \quad (20)$$

$$\frac{\partial L}{\partial b_1} = e_1^T O (Aw_1 + e_1 b_1) + e_2^T \alpha = 0 \quad (21)$$

$$\frac{\partial L}{\partial \delta_1} = C_1 P \delta_1 - \alpha = 0 \quad (22)$$

$$\frac{\partial L}{\partial \alpha} = -(Bw_1 + e_2 b_1) + \delta_1 - e_2 = 0 \quad (23)$$

$$\alpha^T (Bw_1 + e_2 b_1 - \delta_1 + e_2) = 0 \quad (24)$$

$$\text{定义 } H = [A, e_1], G = [B, e_2], W = [w_1, b_1]^T,$$

结合公式(20) 与(21),可以得出

$$[A^T, e_1^T] O [A, e_1] W + [B^T, e_2^T] \alpha = 0$$

从而得到

$$W = -(H^T O H)^{-1} G^T \alpha \quad (25)$$

从上面得知, O 是一个非零的对角矩阵,所以 $H^T O H$ 是一个半正定矩阵,为了避免在一些特殊情况下表现不好,我们引入一个正则化项,公式(25) 变为如下所示

$$W = -(H^T O H + \theta I)^{-1} G^T \alpha \quad (26)$$

式中, θ 是一个小的常数, I 是单位矩阵.

通过上述的 KKT 条件以及拉格朗日函数,我们可以得到公式(13) 的对偶问题为

$$\max e_2^T \alpha - \frac{1}{2} \alpha^T (G (H^T O H)^{-1} G^T + \frac{1}{C_1} P^{-1}) \alpha$$

$$\text{s. t. } \alpha \geq 0 \quad (27)$$

$W = [w_1, b_1]^T$ 是我们所要求的法向量以及偏移值, O, P 依据当前计算出的 W 值进行更新,重复这个优化算法直到收敛,详细过程如算法 1 所示.

算法 1: 迭代算法寻找最优值(以问题(13) 为例)

输入: 训练数据 $A \in R^{m_1 \times n}, B \in R^{m_2 \times n}, C_1, C_2, \epsilon_1, \epsilon_2$.

$$H \in R^{m_1 \times (n+1)}, G \in R^{m_2 \times (n+1)}.$$

$$\text{初始化 } O \in R^{m_1 \times m_1}, P \in R^{m_2 \times m_2}$$

Do Until 目标收敛

1. Calculate α via formula(27)

2. Calculate $W = -(H^T O H + \theta I)^{-1} G^T \alpha$

3. Updated O, P , the element of O, P are calculated by

$$O_i = \begin{cases} \frac{1}{|w_1 x_i + b_1|}, & |w_1 x_i + b_1| \leq \epsilon_1 \\ s\nu, & |w_1 x_i + b_1| > \epsilon_1 \end{cases}$$

$$P_i = \begin{cases} \frac{1}{|\delta_{1,i}|}, & |\delta_{1,i}| \leq \epsilon_2 \\ s\nu, & |\delta_{1,i}| > \epsilon_2 \end{cases}$$

结束循环

输出 $W \in R^{(n+1) \times 1}$

3.3 算法 1 收敛性分析

3.2 节中提出了一个新的基于上限 $L1$ 损失函数的双支持向量机模型,并且提出了一个求解该模型的优化算法,本节就该算法的收敛性进行分析.

$$S = \min \sum_{i=1}^{m_1} \min(|w x_i + b|_1, \epsilon_1) +$$

$$C_1 \sum_{i=1}^{m_2} \min(|\delta_{1,i}|_1, \epsilon_2)$$

$$= \min \sum_{i=1}^{m_1} \min(|t_i u|_1, \epsilon_1) +$$

$$C_1 \sum_{i=1}^{m_2} \min(|\delta_{1,i}|_1, \epsilon_2) \quad (28)$$

式中, $t_i = (x_i, 1); u = (w, b)^T$. 当 $|t_i u|_1$ 与 $|\delta_{1,i}|_1$ 分别小于 ϵ_1 与 ϵ_2 时,公式(28) 又等于

$$S(u) = \min \left(\sum_{i=1}^{m_1} |t_i u|_1 + C_1 \sum_{i=1}^{m_2} |\delta_{1,i}|_1 \right) \quad (29)$$

将公式(29) 分别对 u 以及 $\delta_{1,i}$ 进行求导,得出

$$\sum_{i=1}^{m_1} \left(\frac{t_i^T t_i u}{|t_i u|} \right) = 0 \quad (30)$$

$$C_1 \sum_{i=1}^{m_2} \frac{(\delta_{1,i})}{|\delta_{1,i}|} = 0 \quad (31)$$

由(30)、(31) 可得

$$\sum_{i=1}^{m_1} \left(\frac{t_i^T t_i u}{|t_i u|} \right) + C_1 \sum_{i=1}^{m_2} \frac{(\delta_{1,i})}{|\delta_{1,i}|} = 0 \quad (32)$$

公式(32) 又可以写为

$$H^T O H u + C_1 P \delta_1 = 0 \quad (33)$$

很明显,公式(33) 是下列问题的最优解

$$\min \frac{1}{2} u^T H^T O H u + \frac{C_1}{2} \delta_1^T P \delta_1 \quad (34)$$

令 $\hat{u} = (\hat{w}, \hat{b})^T$ 和 $\hat{\delta}$ 分别表示 u 与 δ 经过算法 1 迭代之后更新的值,可以得出

$$\frac{1}{2} \hat{u}^T H^T O H \hat{u} + \frac{C_1}{2} \hat{\delta}_1^T P \hat{\delta}_1 \leq \frac{1}{2} u^T H^T O H u + \frac{C_1}{2} \delta_1^T P \delta_1 \quad (35)$$

公式(35) 可以重写为

$$\sum_{i=1}^{m_1} \frac{(\hat{t}_i \hat{u}^T)(\hat{t}_i \hat{u})}{2(\hat{t}_i \hat{u})} + \frac{1}{2} C_1 \sum_{i=1}^{m_2} \frac{\hat{\delta}_{1,i}^T(\hat{\delta}_{1,i})}{|\hat{\delta}_{1,i}|} \leq \sum_{i=1}^{m_1} \frac{(t_i u^T)(t_i u)}{2(t_i u)} + \frac{1}{2} C_1 \sum_{i=1}^{m_2} \frac{\delta_{1,i}^T(\delta_{1,i})}{|\delta_{1,i}|} \quad (36)$$

定理. 对于任意 $x \neq a \in R^n, f(x) = x - \frac{x^2}{2a}$, 则

$f(x) \leq f(a)$ 成立^[13].

$$\text{证明: } f(x) = x - \frac{x^2}{2a}, f(a) = \frac{a}{2}.$$

$$f(a) - f(x) = \frac{x^2}{2a} - x + \frac{a}{2} > 0$$

由上式知,该二次函数是一个开口向上的图像,

而该函数的顶点是 $(\frac{1}{2a}, \frac{4ac - b^2}{4a})$, 代入上式知,顶

点为 $(a,0)$,说明该函数图象与 x 轴相切, y 最小值为零,因此, $f(a)-f(x)\geqslant 0$.

$f(a)\geqslant f(x)$,得证.

由上述定理可知,对于任意的 $(x,y)\neq (a,b)$,
 $f(x,y)=x-\frac{x^2}{2a}+c(y-\frac{y^2}{2b})$, 有 $f(x,y)\leqslant f(a,b)$.

设 $x=|t_i u|, a=|t_i u|, c=C_1, y=|\delta_1|, b=|\delta_1|$. 由定理可得

$$\begin{aligned} &|t_i u|-\frac{(|t_i u|)^T(|t_i u|)}{2|t_i u|}+C_1(|\delta_1|-\frac{\delta_1^T(\delta_1)}{2|\delta_1|}) \\ &\leqslant |t_i u|-\frac{(|t_i u|)^T(|t_i u|)}{2|t_i u|}+C_1(|\delta_1|-\frac{\delta_1^T(\delta_1)}{2|\delta_1|}) \end{aligned} \tag{37}$$

由公式(36)、(37) 易知

$$\begin{aligned} &\sum_{i=0}^{m_1}|t_i u|+C_1\sum_{i=1}^{m_2}|\delta_{1,i}|\leqslant \sum_{i=0}^{m_1}|t_i u|+C_1\sum_{i=1}^{m_2}|\delta_{1,i}| \\ &\text{上式又等于} \\ &\sum_{i=0}^{m_1}||t_i u||_1+C_1\sum_{i=1}^{m_2}||\delta_{1,i}||_1\leqslant \sum_{i=0}^{m_1}||t_i u||_1+C_1\sum_{i=1}^{m_2}||\delta_{1,i}||_1 \\ &\rightarrow \sum_{i=1}^{m_1}\min(||t_i u||_1,\epsilon_1)+C_1\sum_{i=1}^{m_2}\min(||\delta_{1,i}||_1,\epsilon_2)\leqslant \\ &\sum_{i=1}^{m_1}\min(||t_i u||_1,\epsilon_1)+C_1\sum_{i=1}^{m_2}\min(||\delta_{1,i}||_1,\epsilon_2) \end{aligned} \tag{38}$$

由公式(38) 以及(29) 可知, $S(u)\leqslant S(u)$. 另外,当 $||t_i u||_1$ 与 $||\delta_{1,i}||_1$ 大于 ϵ_1 与 ϵ_2 时, $S(u)=S(u)$. 所以,在全集上, $S(u)\leqslant S(u)$,这说明算法 1 在每次更新之后都会取得更加小的值,直到最后完全收敛.

4 实验结果与分析

4.1 实验设置

为了全面的评估 CLTSVM 模型的有效性以及鲁棒性,本文准备了两组实验. 第一组实验是通过对比 CLTSVM 与另外五种双支持向量机模型的准确率以及训练速度来评估本文提出的 CLTSVM 模型的有效性;第二组实验是将六种模型在含有不同程度噪音的数据集中进行实验,评估 CLTSVM 模型的鲁棒性.

实验的开发环境是 Inter Core i5 2. 5GHz,内存

4GB,硬盘 1TB 的 PC,开发平台选用的是 Matlab R2013a. 实验采用的数据集是来自于 UCI 中的 8 个数据集,详细情况如表 1 所示,实验采用十折交叉验证方法进行实验. 实验采用的是高斯核函数,C 设置为 1, γ 选择范围是[2-6, 26]. 公式(15)~(18)中的阈值 $\epsilon_i(i=1,2,3,4)$ 由[0. 001, 0. 2]中选择. 本文所采用的五种对比算法如下所示:

表 1 UCI 数据集

name	# dimension	# number
Sonar	61	208
Monk3	7	554
Haberman	4	306
Vote	17	435
Spect	45	267
Ionosphere	35	351
German	25	1000
BUPA	7	345

(1)TWSVM:由 Jayadeva 和 Khemchandani 于 2007 年提出的经典双支持向量机模型.

(2)WLTSVM:由 Shao 和 Chen 等提出的加权线性损失双支持向量机模型^[5],通过引入加权线性损失,大大降低了计算复杂度,加快了训练速度.

(3)PTSVM:由 Ma 和 Ye 等提出的一种新的基于 L2 范数损失的双支持向量机模型^[14].

(4)LSTSVM:由 Kumar 和 Gopal 提出了一个最小二乘法双支持向量机,他们利用等式约束替换原来的不等式约束通过求解线性方程组实现对原问题的求解^[3].

(5)NLPTSVM:是一种使用 L1 范数作为损失函数的双支持向量机^[15],既加快了训练的速度,由一定程度上增强了模型的稳定性.

4.2 CLTSVM 模型有效性评估

表 2 显示的六种双支持向量机模型在无噪 UCI 数据集上的准确率对比(黑体显示的是最优值). P 值是在显著性水平 5%以下由 T 检验所得,原假设为本文提出的模型与另外五种模型的准确率没有显著不同. 我们发现本文的模型 CLTSVM 在大多数的数据集上表现都是最好的(大多数 P 值小于 0. 05),这证明了本文采用戴帽 L1 范数代替 L2 范数作为损失函数的有效性. PTSVM 是除了本文算法之外表现最好的一种模型,这是因为它采用了改进的 L2 范数测量距离一定程度上屏蔽了噪音点的干扰,而在 PTSVM 的两个表现最好的数据集 So-

nar 与 Spect 上它的 P 值分别为 0.845 7 以及 0.975 1,远大于 0.05,这说明在这两个数据集上模型的准确率没有显著不同. WLTSVM 模型主要是引入加

权线性损失针对双支持向量机的训练速度进行改进,准确率上并没有太大变化,与传统 TWSVM 模型差不多.

表 2 无噪 UCI 数据集准确率比较

%

Datasets	Index	CLTSVM	TWSVM	WLTSVM	PTSVM	NLPTSVM	LSTSVM
Sonar	Accuracy	79.43±9.34	74.35±7.48	75.73±7.29	80.27±7.08	71.24±6.39	72.29±6.52
	p-value	—	0.8164	0.4631	0.8457	6.43e-4	0.0153
Monk3	Accuracy	85.52±3.17	77.31±5.63	76.25±4.81	79.38±6.27	77.82±3.39	78.82±3.67
	p-value	—	0.0039	0.0858	5.31e-6	0.0323	0.4281
Haberman t	Accuracy	79.35±4.71	74.27±8.27	72.31±7.83	78.59±8.39	73.18±3.44	72.80±5.13
	p-value	—	0.0013	0.0017	0.8729	0.0209	0.0091
Vote	Accuracy	93.67±2.67	88.47±4.45	91.29±4.69	92.38±2.51	88.39±3.91	87.53±3.27
	p-value	—	1.8453e-7	0.0057	0.9536	0.4671	0.1462
Spect	Accuracy	83.35±8.42	81.69±8.57	84.37±8.37	84.69±10.48	79.57±5.45	78.37±5.38
	p-value	—	0.9327	0.0054	0.9751	0.8329	0.0062
Ionosphere	Accuracy	91.37±3.87	85.37±6.21	83.45±7.74	81.32±8.37	88.58±7.21	86.72±8.21
	p-value	—	0.0653	0.0013	0.0071	0.2571	0.3176
German	Accuracy	82.31±4.13	78.83±3.06	76.32±4.37	79.41±3.79	79.81±3.94	76.57±4.12
	p-value	—	0.5791	4.03e-7	0.0018	0.0193	0.0471
BUPA	Accuracy	73.45±7.71	64.35±8.43	66.58±8.39	71.19±8.18	69.28±6.79	66.29±7.37
	p-value	—	0.0291	0.0173	0.0459	0.1006	0.0341

图 3 显示的是六种算法在 8 个数据集上训练时间对比图,纵轴显示的是 CPU 运行时间的对数值,六种颜色代表六种支持向量机模型. WLTSVM 模型是六种算法中训练速度最快的,因为它主要是对双支持向量机的计算复杂度进行了改进,每次只要求解简单的线性方程组即可,不需要耗费大量的时间. 本文模型 CLTSVM 表现中规中矩,相比传统 TWSVM,CLTSVM 需要进行算法迭代来求得最优解,因此耗费的时间比较多一点. PTSVM 与 NLPTSVM 是训练时间较长的,因为它们都采用了不同的迭代算法来求解最优解,计算量相对较大.

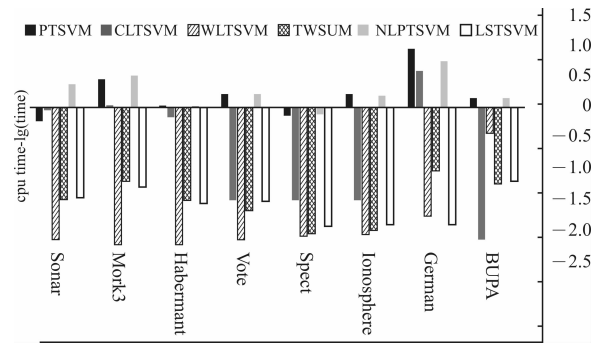


图 3 六种模型训练时间对比图

之前在 3.3 节,我们已经对本文模型优化算法的收敛性做了理论上的证明,现在,我们观察一下算法 1 在 UCI 数据集上的收敛效果. 图 4 显示的是本文提出的算法 1 在 Sonar、Vote、Spect、Haberman 四个数据集上的收敛性效果图. 我们发现算法 1 每次只需要很少次数的迭代就可以取得稳定的目标函

数值,这一方面说明了本文优化算法的收敛性很好,另一方面证明了本文模型 CLTSVM 的实用性.

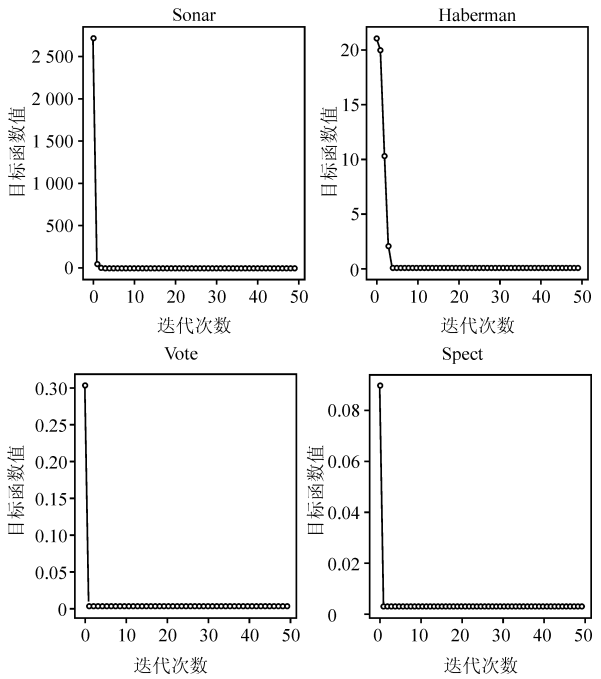


图 4 算法 1 迭代收敛效果图

4.3 CLTSVM 模型鲁棒性评估

针对该模型对于噪音点以及异常点的鲁棒性进行评估,对比算法采用的还是另外五种模型.

为了模仿异常点样本,本文采取在原始数据集中加入高斯噪音的方式^[16],详细操作如下所示. 设给定数据为 $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$,噪音矩阵

$\tilde{X} \in R^{d \times n}$, 其中的数据都是独立且同分布的标准高斯变量. 最终用于实验的数据集为 $X + \sigma \tilde{X}$, 其中, $\sigma = nf ||X||_F / ||\tilde{X}||_F$, nf 是自设的噪音系数.

表 3 与表 4 是分别加入 20% 以及 25% 高斯噪音后, 六种算法在 UCI 数据集上的表现情况, P 值是在显著性水平 5% 以下由 T 检验所得, 原假设为本文提出的模型与另外五种模型的准确率没有显著不同. 从表 3 可以得出, 本文模型在八个数据集上的六个数据集上得到了最高的准确率, 这说明了本文算法对于带有噪音的数据集有很好的效果. 另外, 本

文模型的标准偏差相比其它算法要更小一些, 这说明本文算法的稳定性也不错. 在 Spect 以及 BUPA 两个数据集的结果中, 本文模型未取得最高的准确率, 最高的准确率由 TWSVM 取得, 但是我们发现在这两个数据集上的 P 值都接近 70%, 远大于 0.05, 这说明在这两个数据集上, 本文模型与另外五种模型的准确率没有显著地不同. 表 4 显示的效果与表 3 差不多, 唯一一个本文模型准确率不是最高的数据集是 Spect, Spect 中表现最好的是 PTSVM 模型, 它与本文算法一致都是对损失函数进行了改进.

表 3 20%噪音 UCI 数据集准确率比较

%

Datasets	Index	CLTSVM	TWSVM	WLTSVM	PT SVM	NLPT SVM	LST SVM
Sonar	Accuracy	80.76±5.65	70.63±8.09	73.81±6.21	78.83±8.47	71.24±5.43	69.91±8.37
	p-value	—	0.0281	0.00631	0.9372	6.43e-4	0.0153
Monk3	Accuracy	89.36±4.71	78.52±6.67	76.81±6.53	88.29±6.27	77.39±4.07	83.71±926
	p-value	—	0.0072	0.0013	0.9972	0.0323	0.8681
Haberman t	Accuracy	77.87±4.38	76.39±8.07	73.49±5.91	73.27±7.63	74.38±5.26	72.13±4.94
	p-value	—	2.78e-6	0.6819	0.1936	0.3271	0.4791
Vote	Accuracy	92.47±3.01	89.09±4.81	85.71±8.17	89.31±5.37	85.73±4.38	86.53±5.07
	p-value	—	1.8453e-7	0.0082	0.8946	0.0171	0.0237
Spect	Accuracy	80.47±8.27	86.93±8.73	78.71±7.01	81.32±9.53	79.39±6.03	77.93±7.84
	p-value	—	0.7927	0.1391	0.9427	0.2693	0.0062
Ionosphere	Accuracy	90.46±5.09	83.91±5.63	82.37±8.41	85.61±8.62	86.73±7.36	85.39±7.63
	p-value	—	0.0197	0.4327	0.0239	7.12e-5	0.0431
German	Accuracy	80.79±6.12	74.35±5.63	72.29±5.07	76.81±9.36	75.86±4.72	77.69±6.69
	p-value	—	0.0291	6.27e-6	0.0016	0.5182	0.0373
BUPA	Accuracy	67.93±5.92	68.47±6.90	63.92±7.74	65.64±9.05	67.29±7.63	66.68±8.02
	p-value	—	0.7291	0.7591	0.6975	0.7052	5.23e-5

表 4 25%噪音 UCI 数据集准确率比较

%

Datasets	Index	CLTSVM	TWSVM	WLTSVM	PT SVM	NLPT SVM	LST SVM
Sonar	Accuracy	79.72±6.21	70.31±6.84	78.47±6.27	74.26±5.39	74.93±6.07	71.35±8.64
	p-value	—	0.1261	0.5673	0.0039	0.9842	0.0271
Monk3	Accuracy	84.82±4.80	78.62±5.22	80.25±6.31	79.67±5.39	78.93±4.15	76.82±5.73
	p-value	—	0.0179	0.0439	0.1327	3.72e-7	0.6892
Haberman t	Accuracy	76.71±5.09	71.41±4.93	73.82±6.71	74.38±9.26	72.91±4.65	71.63±6.09
	p-value	—	0.0273	0.0017	0.3761	0.0136	0.0187
Vote	Accuracy	91.53±2.97	89.50±5.37	90.15±3.62	87.64±3.06	86.72±4.92	87.69±4.07
	p-value	—	0.3943	0.0174	0.0236	0.1738	0.0482
Spect	Accuracy	79.47±9.68	78.92±11.24	84.37±8.37	85.37±10.52	77.63±6.31	76.34±4.86
	p-value	—	0.8353	0.0342	0.9823	2.93e-6	0.0236
Ionosphere	Accuracy	90.47±4.06	86.72±5.09	83.67±9.03	89.58±7.63	84.61±6.34	85.25±9.07
	p-value	—	0.0065	0.0029	0.8730	0.4861	0.0385
German	Accuracy	71.62±4.06	70.10±4.65	67.39±3.82	68.26±4.05	64.53±5.97	65.03±3.93
	p-value	—	0.9738	5.13e-4	0.0364	0.1458	0.0128
BUPA	Accuracy	71.47±8.69	68.29±9.38	64.07±10.29	67.36±9.04	65.35±6.37	68.42±8.61
	p-value	—	0.0318	0.0157	8.27e-5	0.0206	0.0417

图 5 显示的是 Vote、Ionosphere 两个数据集的噪音系数 η 为 $[0, 0.5]$ 时,六种模型的准确率折线图。从表中可以看出,本文模型 CLTSVM 在不同比例的噪音情况下,都有着最优的准确率,这是因为本文所采用的戴帽 L1 损失函数相比其它损失函数有着更优的噪音屏蔽能力。另外本文的模型相比其它算法稳定性也要更高一点,噪音系数从 $0 \sim 0.5$,准确率相差了 5% 不到,其他模型基本都大于 5%。另外,PTSVM 模型也针对损失函数进行了改进,因此对于噪音数据有一定的屏蔽作用。

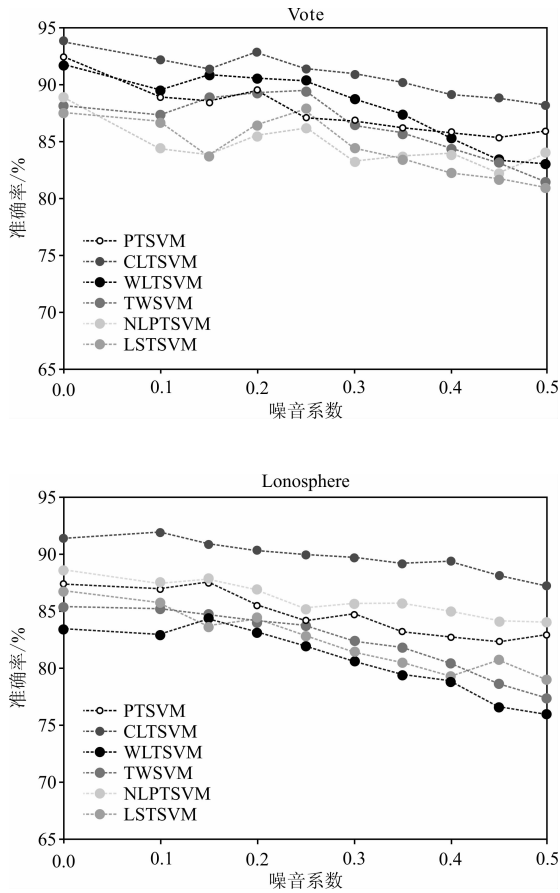


图 5 不同噪音系数对应准确率图

5 结束语

本文提出了一个基于戴帽 L1 损失函数的双支持向量机模型,利用戴帽 L1 范数代替原本双支持向量机的 L2 范数作为损失函数,一定程度上消除了异常值、离群点对于超平面构造的影响,提高了模型的准确率;另外,针对该模型的最优化问题,提出了一个简易的优化算法用于求解,并且从理论上证明了该优化算法的收敛性。在无噪以及有噪数据集上的实验结果证明了该模型的有效性以及鲁棒性。但是在计算成本方面,本文模型相比传统模型需要

进行迭代来淘汰数据集中的异常样本,因此耗费了更多的时间,后续将针对这方面进行改进,以完善该模型。

参考文献:

[1] CORINNA CORTES, VLADIMIR VAPNIK. Support-vector networks[J]. Machine learning, 1995, 20(3):273-297.

[2] JAYADEVA, KHEMCHANDANI R, CHANDRA S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(5):905.

[3] KUMAR M A, GOPAL M. Least squares twin support vector machines for pattern classification[J]. Expert Systems with Applications, 2009, 36(4): 7535-7543.

[4] YE Q, ZHAO C, YE N, et al. A feature selection method for nonparallel plane support vector machine classification[J]. Optimization Methods and Software, 2012, 27(3):431-443.

[5] SHAO Y H, CHEN W J, WANG Z, et al. Weighted linear loss twin support vector machine for large-scale classification[J]. Knowledge-Based Systems, 2015 (73):276-288.

[6] XU Y, YANG Z, PAN X. A novel twin support-vector machine with pinball loss[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017,28(2):359-370.

[7] KWAK N. Principal component analysis based on L1-Norm maximization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(9):1672-1680.

[8] GAO S, YE Q, YE N. 1-Norm least squares twin support vector machines[J]. Neurocomputing, 2011, 74(17):3590-3597.

[9] JIANG W, NIE F, HUANG H. Robust dictionary learning with capped L1 -norm[C]// International Conference on Artificial Intelligence. San Diego,2015.

[10] LU Q, LI X, DONG Y, et al. Subspace clustering by capped L(1) norm[C]// Chinese Conference on Pattern Recognition. Springer, Singapore, 2016.

[11] NIE F, HUO Z, HUANG H. Joint capped norms minimization for robust matrix recovery[C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia, AAAI Press, 2017.

- phrase representations using RNN Encoder Decoder for statistical machine translation [C] //Computer Science 2014,[S. L]. 2014:1 -12.
- [12] EBRAHIMI J, DOU D. Chain based RNN for relation classification [C] //Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, USA, 2015: 1244 -1249.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997 (9): 1735-1780.
- [14] 曹宇慧. 基于深度学习的文本情感分析研究[D]. 哈尔滨:哈尔滨工业大学,2016.
- [15] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language. Doha: CoRR, 2014: 1724-1734.
- [16] 曹宇,李天瑞,贾真,等. BGRU: 中文文本情感分析的新方法[J/OL]. 计算机科学与探索: 1-11 [2019-04-07]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20180723.1614.008.html>.
- [17] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. Computer Science, 2014,10(4):<http://arxiv.org/abs/1409.0473>.
- [18] MNH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Advances in Neural Information Processing Systems. [S. L]. 2014: 2204-2212.
- [19] 胡荣磊,芮璐,齐筱,等. 基于循环神经网络和注意力模型的文本情感分析[J/OL]. 计算机应用研究,2019 (11): 1-7. [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20180811.1330.064.html>.
- [20] 王伟,孙玉霞,齐庆杰,等. 基于 BiGRU-Attention 神经网络的文本情感分类模型[J/OL]. 计算机应用研究:1-10 [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1196.TP.20181011.1246.010.html>.
- [21] ZHOU C, SUN C, LIU Z, et al. A C-LSTM Neural Network for Text Classification [J]. Computer Science,2015,1(4): 39-44.
- [22] 李洋,董红斌. 基于卷积神经网络和 BLSTM 网络特征融合的文本情感分析[J/OL]. 计算机应用:1-7. [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1307.tp.20180719.1604.080.html>.
- [23] 陈洁,邵志清,张欢欢,等. 基于并行混合神经网络模型的短文本情感分析[J/OL]. 计算机应用:1-6. [2019-04-07]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20190329.1643.008.html>.
- [24] 任勉,甘刚. 基于双向 LSTM 模型的文本情感分类[J]. 计算机工程与设计,2018,39(7):2064-2068.
- [25] 王煜涵,张春云,赵宝林,等. 卷积神经网络下的 Twitter 文本情感分析[J]. 数据采集与处理,2018,33(5): 921-927.

作者简介:

王丽亚 女,(1994-),硕士研究生. 研究方向为文本挖掘. E-mail:1578938844@qq.com.

刘昌辉 男,(1965-),博士,教授,硕士生导师. 研究方向为机器人控制技术、软件系统与智能计算、信息处理.

蔡敦波 男,(1981-),博士,教授,硕士生导师. 研究方向为人工智能的智能规划、自动推理、约束优化、文本挖掘.

赵彭洲 女,(1970-),硕士,副教授,硕士生导师. 研究方向为机器视觉、数据分析、文本挖掘.

王梦 女,(1994-),硕士研究生. 研究方向为数据挖掘.

(上接第 79 页)

- [12] CAO Y, BO J, JIN T, et al. Image set representation with ℓ_1/ℓ_2 -Norm optimal mean robust principal component analysis[C]// International Conference on Image & Graphics. Shanghai, China,2017.
- [13] LAI Z, XU Y, YANG J, et al. Rotational invariant dimensionality reduction algorithms[J]. IEEE Transactions on Cybernetics, 2016 (99): 1-14, DOI: 10.1109/TCYB.2016.2578642.
- [14] MA X, YE Q, YAN H. L2P-norm distance twin support vector machine [J]. IEEE Access, 2017 (5): 23473-23483.
- [15] TANVEER M. Robust and sparse linear programming

twin support vector machines[J]. Cognitive Computation, 2015, 7(1):137-149.

- [16] HUA W, NIE F, HUANG H. Robust distance metric learning via simultaneous ℓ_1 -norm minimization and maximization[C]// International Conference on International Conference on Machine Learning. Banff, Canada,2014.

作者简介:

沈洋 男,(1994-),硕士研究生. 研究方向为人工智能、机器学习. E-mail:1825472824@qq.com.

戴月明 男,(1964-),硕士,副教授. 研究方向为人工智能、软件工程.