

基于 L1-范数距离的最小二乘对支持向量机

周燕萍¹ 业巧林²

(无锡科技职业学院物联网与软件技术学院 江苏 无锡 214028)¹

(南京林业大学信息科学与技术学院 南京 210037)²

摘要 最小二乘对支持向量机(LSTSVM)是一种有效的分类技术。然而,该方法需计算点到平面的平方 L2-范数距离,从而易受野值或噪声的影响。为了缓解此问题,提出了一种有效的鲁棒 LSTSVM 方法,即基于 L1-范数距离的 LSTSVM(LSTSVM_{L1D})。该方法由于使用 L1 范数作为距离度量,因此不易受到野值或噪声数据的影响。此外,设计了一种有效的迭代算法,旨在求解目标问题,并从理论上证明了其收敛性。在人工数据集和 UCI 数据集上验证了 LSTSVM_{L1D} 的有效性。

关键词 最小二乘支持向量机,基于 L1-范数距离的 LSTSVM,L1 范数距离,L2 范数平方距离

中图分类号 TP181,TP39 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.04.015

L1-norm Distance Based Least Squares Twin Support Vector Machine

ZHOU Yan-ping¹ YE Qiao-lin²

(School of Internet of Things and Software Technology, Wuxi Professional College of Science and Technology, Wuxi, Jiangsu 214028, China)¹

(College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China)²

Abstract Recently, LSTSVM, as an efficient classification algorithm, was proposed. However, this algorithm computes squared L2-norm distances from planes to points, such that it is easily affected by outliers or noisy data. In order to avoid this problem, this paper presented an efficient L1-norm distance based robust LSTSVM method, termed as LSTSVM_{L1D}. LSTSVM_{L1D} computes L1-norm distances from planes to points and is not sensitive to outliers and noise. Besides, this paper designed an efficient iterative algorithm to solve the resulted objective, and proved its convergence. Experiments on artificial dataset and UCI dataset indicate the effectiveness of the proposed LSTSVM_{L1D}.

Keywords Least squares support vector machine, L1-norm distance based LSTSVM, L1-norm distance, Squared L2-norm distance

1 引言

对于分类和回归问题,支持向量机(Support Vector Machine, SVM)是一款公认的有效工具,其目标在于寻找一个最优平面以最大化不同样本间的间隔。然而,传统 SVM 存在两大问题:1)解不等式约束二次规划问题所需的高额的计算开销;2)无法保证有效求解异或(Exclusive OR, XOR)问题。尽管学者们提出了很多加速 SVM 学习的算法^[1-4],但它们都是单平面 SVM,仍然无法保证较好地求解 XOR 问题。

近年来,多平面支持向量机引起了机器学习领域学者的广泛关注。不同于传统 SVM,该类方法为每类样本寻找一个最优拟合平面,这些平面满足离本类样本尽可能近,而离其他类样本尽可能远的约束。此方面的工作起源于 Mangasarian 和 Wild 提出的基于广义特征值的近似支持向量机(Multi-surface Support Vector Machines via Generalized Eigenvalues, GEPSVM)^[5]。该方法因通过求解广义特征值问题获取

最优分类平面,而具较低的计算复杂度,且拥有处理 XOR 问题的优势。继 GEPSVM 之后,研究人员提出了很多多平面支持向量机算法,这些算法的目标问题都具有与 GEPSVM 相似的特征值形式。例如,文献[6]将 GEPSVM 进行模糊化,提出了模糊 GEPSVM(Fuzzy GEPSVM, FGEPSVM);Guarracino 提出了 ReGEC 算法^[7];杨绪兵等人提出了基于原型超平面的多类最接近支持向量机(Proximal Support Vector Machine Based on Prototypal Multi-classification Hyperplanes, MHPSVM)^[8]和局部化的广义特征值最接近支持向量机(Localized Proximal Support Vector Machine via Generalized Eigenvalues, LGEPSVM)^[9];Ye 等人^[10-11]提出了多权重向量投影支持向量机(Multi-Weight Vector Projection Support Vector Machines, MVSVM)。

另一方面,关于多平面支持向量机的研究是对支持向量机。此方面的研究起源于 Jayadeva 等提出的对支持向量机(Twin Support Vector Machine, TWSVM)^[12]。TWSVM 保

到稿日期:2017-05-01 返修日期:2017-08-16 本文受江苏省自然科学基金(BK20171453)资助。

周燕萍(1982-),女,讲师,主要研究领域为计算机应用技术,E-mail:zhouyping@163.com(通信作者);业巧林(1982-),男,副教授,主要研究领域为模式识别,E-mail:yeqiaolincom@163.com。

留了 GEPSVM 的几何解释以及求解 XOR 问题的优势,但在形式上两者却存在较大的差异。TWSVM 简单地将 GEPSVM 转换为相关支持向量机问题,但在性能上远好于 GEPSVM。学者们基于 TWSVM 提出了很多有效算法来提高 TWSVM 的性能,如文献[13-19]。从计算角度而言,TWSVM 比传统 SVM 快 4 倍,但在面对大规模问题时因其仍然需要解两个带约束的凸规划问题而需要高昂的计算代价。为了进一步缓解 TWSVM 的时间代价高的问题,Kumar 和 Gopal 提出了平滑孪生支持向量机(Smooth Twin Support Vector Machine, STWSVM)^[18],该算法用平滑近似函数代替目标函数的 Hinge 损失函数。随后,Kumar 和 Gopal 又提出了最小二乘孪生支持向量机(Least Squares Twin Support Vector Machine, LSTSVM)^[20]。LSTSVM 将 TWSVM 不等式约束转换为等式约束,该问题最终归结为求解线性方程组问题,其与 GEPSVM 的计算代价相当。所有这些方法的一个共同点在于应用 L2-范数平方距离度量模型中的距离,以至于具有对野值或噪声数据的敏感性^[21]。因 L1-范数距离对野值或噪声具有较好的抑制能力^[21],文献[22]提出了鲁棒 L1 范数 GEPSVM(Robust L1-norm GEPSVM, L1-GEPSVM),该算法考虑用 L1-范数距离替换 GEPSVM 中的 L2-范数平方距离。L1-GEPSVM 在形式上与特征提取算法线性判别分析 LDA(Linear Discriminant Analysis)相同,是一个非凸模型。为了获得解,文献[22]采用文献[21]的梯度上升(Gradient Ascending, GA)迭代算法。GA 容易实现,但却由于引入了非凸函数而难以保证结果的最优性^[23]。

如上所述,LSTSVM 不仅继承了 GEPSVM 和 TWSVM 的性能优势,而且保证了速度优势。正因如此,近年来,该算法得到了广泛的关注。然而,LSTSVM 仍然应用不鲁棒的 L2-范数平方度量点到平面的距离,因此对噪声或者野值具有敏感性。为了缓解此问题,本文提出了一种基于 L1-范数距离的鲁棒 LSTSVM,称为 LSTSVM_{L1D},其有以下 4 个优势:1) LSTSVM_{L1D} 通过 L1 范数计算点到平面的距离。为了解决所生成的最优化问题,本文提出了一种简单而有效的迭代算法,该算法易实现且其收敛性在理论上得到了保证。2) 使用 L1 范数距离使得 LSTSVM_{L1D} 对野值或噪声与 L2 范数平方距离相比更鲁棒,从而可以有效抑制野值或噪声的负面影响。3) 人工数据和公共数据集上的实验结果表明,与 GEPSVM, TWSVM, LSTSVM 和 L1-GEPSVM 相比, LSTSVM_{L1D} 在分类准确性上更优,进一步表明 L1-范数距离促进了 LSTSVM_{L1D} 的泛化能力。4) 本文提出的方法易扩展到其他多平面相关方法中,这也是我们未来要研究的工作。

2 相关工作

在 n 维实空间 \mathbf{R}^n 定义一个二分类问题,训练样本表示为 $T = \{(\mathbf{x}_j^{(i)}, y_i) \mid i = 1, 2, j = 1, 2, \dots, m_i\}$, 其中 $\mathbf{x}_j^{(i)} \in \mathbf{R}^n$, 且 $y_i \in \{-1, 1\}$, $\mathbf{x}_j^{(i)}$ 表示 i 类中的第 j 个样本。假设尺寸为 $m_1 \times n$ 的矩阵 $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{m_1}]^T$ 代表正类的数据,尺寸为 $m_2 \times n$ 的矩阵 $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_{m_2}]^T$ 代表负类的数据, \mathbf{A} 和 \mathbf{B} 组合矩阵代表了所有训练数据,即 $m_1 + m_2 = m$, 其中 m_1 代表属于正类样本的数目, m_2 代表负类样本的数目。下面回顾两

个众所周知的多平面支持向量机算法:TWSVM 和 LSTSVM。

2.1 TWSVM

TWSVM^[12]继承了 GEPSVM^[5]的几何解释,即:本类平面离本类样本尽可能近,而离其他类样本尽可能远。具体地,其目标问题归结为如下两个凸规划问题:

$$\min_{\mathbf{w}_1, b_1, q_1} \frac{1}{2} \|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_2^2 + c_1 \mathbf{e}_2^T \mathbf{q}_1 \quad (1)$$

$$\text{s. t. } -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 \geq \mathbf{e}_2, \mathbf{q}_1 \geq 0$$

$$\min_{\mathbf{w}_2, b_2, q_2} \frac{1}{2} \|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|_2^2 + c_2 \mathbf{e}_1^T \mathbf{q}_2 \quad (2)$$

$$\text{s. t. } (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \mathbf{q}_2 \geq \mathbf{e}_1, \mathbf{q}_2 \geq 0$$

其中, $\|\cdot\|_2$ 代表 L2 范数, q_1 和 q_2 是松弛变量, c_1 和 c_2 是两个非负惩罚系数。值得注意的是,在式(1)和式(2)的目标函数中,TWSVM 通过 L2 范数度量点到平面的距离。通过解决式(1)和式(2),可以得到如下两个平面:

$$\mathbf{w}_1^T \mathbf{x} + b_1 = 0, \mathbf{w}_2^T \mathbf{x} + b_2 = 0 \quad (3)$$

平面 $\mathbf{w}_1^T \mathbf{x} + b_1 = 0$ 到正类的距离小于其到负类的距离,同时平面 $\mathbf{w}_2^T \mathbf{x} + b_2 = 0$ 到负类的距离小于其到正类的距离。一个新的数据点 \mathbf{x} 要分类于正类还是负类,取决于它临近于两个非平行平面中的哪一个。式(1)和式(2)的解可通过解下列对应的 Wolfe 对偶问题获取。

$$\max_{\alpha} \mathbf{e}_2^T \alpha - 0.5 \alpha^T \mathbf{G} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{G}^T \alpha \quad (4)$$

$$\text{s. t. } 0 \leq \alpha \leq c_1 \mathbf{e}_2$$

$$\max_{\beta} \mathbf{e}_1^T \beta - 0.5 \beta^T \mathbf{H} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{H}^T \beta \quad (5)$$

$$\text{s. t. } 0 \leq \beta \leq c_2 \mathbf{e}_1$$

其中, $\alpha \in \mathbf{R}^{m_2}$, $\beta \in \mathbf{R}^{m_1}$ 是拉格朗日乘子向量, $\mathbf{G} = [\mathbf{B} \ \mathbf{e}_2]$, $\mathbf{H} = [\mathbf{A} \ \mathbf{e}_1]$ 。最后,得出:

$$\mathbf{z}_1 = [\mathbf{w}_1^T \ b_1]^T = -(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{G}^T \alpha \quad (6)$$

$$\mathbf{z}_2 = [\mathbf{w}_2^T \ b_2]^T = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{H}^T \beta$$

这样,可以获得式(3)中的两个平面。在处理相反矩阵 $(\mathbf{H}^T \mathbf{H})^{-1}$ 和 $(\mathbf{G}^T \mathbf{G})^{-1}$ 时,容易遇到奇异性问题。可通过向其引入正则项 $\epsilon \mathbf{I}$ 来缓解此问题^[12],其中 ϵ 是一个足够小的正数。

2.2 LSTSVM

TWSVM 因仍要求解两个带约束的凸规划问题而需高额的计算代价。于是,LSTSVM^[20]被提出以解决此问题,其用等式约束代替 TWSVM 中的不等式约束,即:

$$\min_{\mathbf{w}_1, b_1, q_1} \frac{1}{2} \|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_2^2 + \frac{1}{2} c_1 \|\mathbf{q}_1\|_2^2 \quad (7)$$

$$\text{s. t. } -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 = \mathbf{e}_2$$

$$\min_{\mathbf{w}_2, b_2, q_2} \frac{1}{2} \|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|_2^2 + \frac{1}{2} c_2 \|\mathbf{q}_2\|_2^2 \quad (8)$$

$$\text{s. t. } (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \mathbf{q}_2 = \mathbf{e}_1$$

将等式约束代入目标函数,式(7)和式(8)转换为:

$$\min_{\mathbf{w}_1, b_1} \frac{1}{2} \|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_2^2 + \frac{1}{2} c_1 \|\mathbf{e}_2 + \mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1\|_2^2 \quad (9)$$

$$\min_{\mathbf{w}_2, b_2} \frac{1}{2} \|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|_2^2 + \frac{1}{2} c_2 \|\mathbf{e}_1 - \mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2\|_2^2 \quad (10)$$

计算式(9)关于 (\mathbf{w}_1, b_1) 的梯度,且设置其为 0,可以得到:

$$\mathbf{A}^T (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{B}^T (\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1 + \mathbf{e}_2) = 0 \quad (11)$$

$$\mathbf{e}_1^T (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{e}_2^T (\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1 + \mathbf{e}_2) = 0 \quad (12)$$

组合式(11)和式(12),有:

$$\mathbf{z}_1 = [\mathbf{w}_1^\top \ b_1]^\top = -(\mathbf{G}^\top \mathbf{G} + 1/c_1 \mathbf{H}^\top \mathbf{H})^{-1} \mathbf{G}^\top \mathbf{e}_2$$

通过相同的方式,可以得到:

$$\mathbf{z}_2 = [\mathbf{w}_2^\top \ b_2]^\top = -(\mathbf{H}^\top \mathbf{H} + 1/c_2 \mathbf{G}^\top \mathbf{G})^{-1} \mathbf{H}^\top \mathbf{e}_1$$

从解的形式可以看出,LSTSVM主要解两个 $(n+1) \times (n+1)$ 维矩阵的逆。显然,在大规模($m \gg n$)问题中,LSTSVM保证了计算优势。

3 基于L1范数距离的LSTSVM(LSTSVM_{L1D})

在LSTSVM模型中,计算点到平面的距离由L2范数平方度量,其强调远离于本类的样本(野值或噪声)到本类平面距离的最小化,从而容易增加它们在模型中的影响,降低类的准确性。为了缓解此问题,我们提出基于L1范数距离的LSTSVM,称为LSTSVM_{L1D}。L1范数距离的应用被认为是一种减少野值影响的简单有效的方式^[21-22]。另外,LSTSVM_{L1D}继承了LSTSVM解决XOR问题和计算时间远远小于TWSVM的优点。最终,LSTSVM_{L1D}归结为求解如下两个问题:

$$\min_{\mathbf{w}_1, b_1, \mathbf{q}_1} \|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_1 + 0.5c_1 \|\mathbf{q}_1\|_2^2 \quad (13)$$

$$\text{s. t. } -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 = \mathbf{e}_2$$

$$\min_{\mathbf{w}_2, b_2, \mathbf{q}_2} \|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|_1 + 0.5c_2 \|\mathbf{q}_2\|_2^2 \quad (14)$$

$$\text{s. t. } (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \mathbf{q}_2 = \mathbf{e}_1$$

其中, $\|\cdot\|_1$ 代表L1范数。显然,LSTSVM_{L1D}并未改变LSTSVM的几何解释。问题(13)和问题(14)可以转换为形如TWSVM模型的两个带不等式约束的凸规划问题,因此这种转换将继承TWSVM具高额计算代价的缺陷。为了保证与LSTSVM相当的计算开销,下面提出一种有效的迭代算法对问题(13)和问题(14)进行求解。以求解问题(13)为例,我们阐述了该算法的设计思路。由于问题(14)与问题(13)具有相似的形式,因此所设计的迭代算法同样可以应用于求解问题(14)。

将式(13)中的等式约束代入目标后,有:

$$\min_{\mathbf{w}_1, b_1} \|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_1 + 0.5c_1 \|\mathbf{e}_2 + \mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1\|_2^2 \quad (15)$$

将式(15)中目标函数关于 \mathbf{w}_1 和 b_1 的梯度设置为0:

$$\sum_{i=1}^{m_1} \frac{2\mathbf{A}_i (\mathbf{A}_i^\top \mathbf{w}_1 + b_1)}{\mathbf{D}_i} + c_1 \mathbf{B}^\top (\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1 + \mathbf{e}_2) = 0 \quad (16)$$

$$\sum_{i=1}^{m_1} \frac{2(\mathbf{A}_i^\top \mathbf{w}_1 + b_1)}{\mathbf{D}_i} + c_1 \mathbf{e}_2^\top (\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1 + \mathbf{e}_2) = 0 \quad (17)$$

其中, $\mathbf{D}_i = 2|\mathbf{A}_i^\top \mathbf{w}_1 + \mathbf{e}_1 b_1|$ 。令 $\mathbf{K} = \text{diag}(1/\mathbf{D}_1, 1/\mathbf{D}_2, \dots, 1/\mathbf{D}_{m_1})$,有:

$$2\mathbf{A}^\top \mathbf{K} (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{B}^\top (\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1 + \mathbf{e}_2) = 0 \quad (18)$$

$$2\mathbf{e}_1^\top \mathbf{K} (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{e}_2^\top (\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1 + \mathbf{e}_2) = 0 \quad (19)$$

组合式(18)和式(19),得到:

$$\frac{1}{c_1} \begin{bmatrix} 2\mathbf{A}^\top \mathbf{K} \mathbf{A} & 2\mathbf{A}^\top \mathbf{K} \mathbf{e}_1 \\ 2\mathbf{e}_1^\top \mathbf{K} \mathbf{A} & 2m_1 \mathbf{K} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} \mathbf{B}^\top \mathbf{B} & \mathbf{B}^\top \mathbf{e}_2 \\ \mathbf{e}_2^\top \mathbf{B} & m_2 \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} \mathbf{B}^\top \mathbf{e}_2 \\ m_2 \end{bmatrix} = 0 \quad (20)$$

所以,有:

$$\begin{bmatrix} \mathbf{w}_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} 2/c_1 \mathbf{A}^\top \mathbf{D}_1 \mathbf{A} + \mathbf{B}^\top \mathbf{B} & 2/c_1 \mathbf{A}^\top \mathbf{D}_1 \mathbf{e}_1 + \mathbf{B}^\top \mathbf{e}_2 \\ 2/c_1 \mathbf{e}_1^\top \mathbf{D}_1 \mathbf{A} + \mathbf{e}_2^\top \mathbf{B} & 2/c_1 m_1 \mathbf{D}_1 + m_2 \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{B}^\top \mathbf{e}_2 \\ -m_2 \end{bmatrix} \quad (21)$$

即:

$$\begin{bmatrix} \mathbf{w}_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} \frac{2}{c_1} \mathbf{A}^\top \\ \mathbf{e}_1^\top \end{bmatrix} \mathbf{K} [\mathbf{A} \ \mathbf{e}_1] + \begin{bmatrix} \mathbf{B}^\top \\ \mathbf{e}_2^\top \end{bmatrix} [\mathbf{B} \ \mathbf{e}_2]^{-1} \begin{bmatrix} -\mathbf{B}^\top \mathbf{e}_2 \\ -m_2 \end{bmatrix} \quad (22)$$

所以,可以得到 \mathbf{w}_1 和 b_1 的更新的解,如式(23)所示:

$$\mathbf{z}_1 = [\mathbf{w}_1^\top \ b_1]^\top = -\left(\frac{2}{c_1} \mathbf{H}^\top \mathbf{K} \mathbf{H} + \mathbf{G}^\top \mathbf{G}\right)^{-1} \mathbf{G}^\top \mathbf{e}_2 \quad (23)$$

其中, $\mathbf{G} = [\mathbf{B} \ \mathbf{e}_2]$ 和 $\mathbf{H} = [\mathbf{A} \ \mathbf{e}_1]$ 由式(4)定义。比较LSTSVM问题,即式(7)和式(8),可以看到,LSTSVM_{L1D}可被视为LSTSVM的加权版。

算法1 求解问题(13)的有效迭代算法

输入: $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbf{R}^{m \times n}$, $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{m_1}]^\top \in \mathbf{R}^{m_1 \times n}$ 和 $\mathbf{B} =$

$$[\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_{m_2}]^\top \in \mathbf{R}^{m_2 \times n}$$

输出: \mathbf{z}_1

初始化 $\mathbf{z}_1 = [\mathbf{w}_1^\top \ b_1]^\top$ 且迭代次数 $p=1$ 。

while 不收敛 do

1. 计算 $\mathbf{K} = \text{diag}(1/\mathbf{D}_1, 1/\mathbf{D}_2, \dots, 1/\mathbf{D}_{m_1})$,其中 $\mathbf{D}_i = 2|\mathbf{A}_i^\top \mathbf{w}_1 + b_1|$ 。

2. 通过解式(23)得到解 \mathbf{z}_1 。

3. 设置 $p=p+1$ 。

end while

显然,在 \mathbf{K} 为已知矩阵的情况下,问题(13)有着与LSTSVM相似的解,即都是线性方程。然而, \mathbf{K} 包含了变量 (\mathbf{w}_1, b_1) ,其可被视为目标问题(13)的潜在变量,可以利用本文的迭代算法替换优化策略对其进行优化。具体地,通过用上次迭代计算的 (\mathbf{w}_1, b_1) 计算当前的 \mathbf{K} 。最后,算法1概括了整个计算程序。从算法1中可以看到,我们的迭代算法每次迭代更新 \mathbf{K} ,以保证每次迭代求解线性方程系统。从算法1中还可以看到,式(13)中的目标被迭代地优化。注意,如果存在 \mathbf{A}_i 使得 $\mathbf{A}_i^\top \mathbf{w}_1 + b_1 = 0$,那么算法1中步骤2的 \mathbf{K} 无法被很好地定义。自然地,同文献[21],解决此问题的一个简单的方法是通过计算 $\mathbf{z}_1 = (\mathbf{z}_1 + \Delta) / \|\mathbf{z}_1 + \Delta\|_2$ 达到移动 \mathbf{z}_1 的目的,其中 Δ 是一个小的随机向量。考虑到此方法需要尝试不固定的次数,我们正则化 \mathbf{K} 中的每个对角元素为 $\mathbf{K}_{ii} = 1/\sqrt{(\mathbf{A}_i^\top \mathbf{w}_1 + b_1)^2 + \nu}$,其中 ν 是一个小的正常数。

接下来,我们证明LSTSVM_{L1D}目标问题的单调递减性,首先引入定理1。

定理1 对于任意非零向量 $\mathbf{z}, \mathbf{z}^p \in \mathbf{R}^n$,下面不等式成立:

$$\|\mathbf{z}\|_1 - \frac{1}{2} \frac{\|\mathbf{z}\|_1^2}{\|\mathbf{z}^p\|_1} \leq \|\mathbf{z}^p\|_1 - \frac{1}{2} \frac{\|\mathbf{z}^p\|_1^2}{\|\mathbf{z}^p\|_1} \quad (24)$$

证明:由不等式 $(\sqrt{u} - \sqrt{u^p})^2 \geq 0$,得到:

$$\begin{aligned} (\sqrt{u} - \sqrt{u^p})^2 \geq 0 &\Rightarrow u - 2\sqrt{uu^p} + u^p \geq 0 \\ &\Rightarrow \sqrt{u} - \frac{u}{2\sqrt{u^p}} \leq \sqrt{u^p} - \frac{u^p}{2\sqrt{u^p}} \end{aligned} \quad (25)$$

用 $\|\mathbf{z}\|_1^2$ 和 $\|\mathbf{z}^p\|_1^2$ 替换式(25)中的 u 和 u^p ,得到不等式(24)。

定理2 在每次迭代中,目标问题式(13)具有单调递减性。

证明:对于每次迭代,我们定义 $(\bar{\mathbf{w}}_1, \bar{b}_1)$ 是 (\mathbf{w}_1, b_1) 的更新。根据算法1中的步骤2,可以得到不等式:

$$\frac{1}{2} (\mathbf{A} \bar{\mathbf{w}}_1 + \mathbf{e}_1 \bar{b}_1)^\top \mathbf{K} (\mathbf{A} \bar{\mathbf{w}}_1 + \mathbf{e}_1 \bar{b}_1) + \frac{1}{2} c_1 \|\bar{\mathbf{q}}_1\|_2^2$$

$$\begin{aligned} &\leq \frac{1}{2} (\mathbf{A}\bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1)^\top \mathbf{K} (\mathbf{A}\bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1) + \frac{1}{2} c_1 \|\mathbf{q}_1\|_2^2 \\ &= \frac{1}{2} \|\mathbf{A}\bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1\|_1 + \frac{1}{2} c_1 \|\mathbf{q}_1\|_2^2 \end{aligned} \quad (26)$$

$$\begin{aligned} &\frac{1}{2} \sum_{i=1}^{m_1} \frac{(\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1)^2}{|\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1|} + 0.5 c_1 \|\bar{\mathbf{q}}_1\|_2^2 \leq \\ &\frac{1}{2} \sum_{i=1}^{m_1} \frac{(\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1)^2}{|\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1|} + 0.5 c_1 \|\mathbf{q}_1\|_2^2 \end{aligned} \quad (27)$$

根据定理 1, 对于每个 i , 有:

$$\begin{aligned} &|\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1| - 0.5 \frac{(\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1)^2}{|\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1|} \leq |\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1| - \\ &0.5 \frac{(\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1)^2}{|\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1|} \end{aligned} \quad (28)$$

因此,

$$\begin{aligned} &\sum_{i=1}^{m_1} (|\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1| - 0.5 \frac{(\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1)^2}{|\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1|}) \leq \\ &\sum_{i=1}^{m_1} (|\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1| - 0.5 \frac{(\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1)^2}{|\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1|}) \end{aligned} \quad (29)$$

组合式(27)和式(29), 得到:

$$\sum_{i=1}^{m_1} |\mathbf{A}_i \bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1| + 0.5 c_1 \|\bar{\mathbf{q}}_1\|_2^2 \leq \sum_{i=1}^{m_1} |\mathbf{A}_i \mathbf{w}_1 + \mathbf{e}_i b_1| + 0.5 c_1 \|\mathbf{q}_1\|_2^2 \quad (30)$$

矩阵化式(30)得到:

$$\|\mathbf{A}\bar{\mathbf{w}}_1 + \mathbf{e}_i \bar{b}_1\|_1 + 0.5 c_1 \|\bar{\mathbf{q}}_1\|_2^2 \leq \|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_i b_1\|_1 + 0.5 c_1 \|\mathbf{q}_1\|_2^2 \quad (31)$$

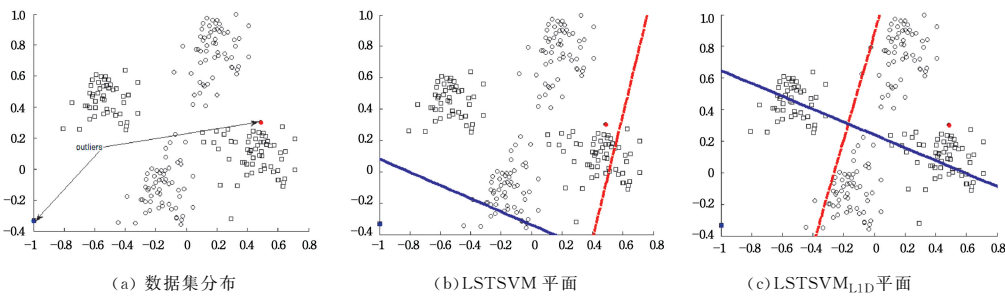
定义式(13)的目标函数为 $J(\mathbf{w}_1, b_1)$ 。显然, 根据式

(31), 有 $J(\bar{\mathbf{w}}_1, \bar{b}_1) \leq J(\mathbf{w}_1, b_1)$ 。

用类似于算法 1 的迭代算法, 可以求解目标问题(14)。具体地, 在每次迭代时更新解为:

$$\mathbf{z}_2 = [\mathbf{w}_2^\top \quad b_2]^\top = (\mathbf{H}^\top \mathbf{H} + 2/c_2 \mathbf{G}^\top \mathbf{S} \mathbf{G})^{-1} \mathbf{H}^\top \mathbf{e}_1$$

其中, $\mathbf{S} = \text{diag}(1/\mathbf{F}_1, 1/\mathbf{F}_2, 1/\mathbf{F}_{m_2})$, $\mathbf{F}_i = 2|\mathbf{B}_i^\top \mathbf{w}_2 + b_2|$ 。与 LSTSVM 和 TWSVM 相同, 每次迭代中的解 \mathbf{z}_1 和 \mathbf{z}_2 包含了



注: 红色平面是类“○”样本的最优平面, 蓝色的平面是类“□”样本的最优平面

图 1 人工数据集

Fig. 1 Artificial dataset

图 1 显示, 与 LSTSVM 相比, LSTSVM_{L1D} 在引入野值后得到了更好的分类平面, 从而有力地解释了基于 L2 范数平方距离的分类器对野值的敏感性, 而 L1 范数距离能有效抑制野值对分类的影响, 从而也验证了我们的想法的可行性。

4.2 UCI 数据集

为了进一步证明 LSTSVM_{L1D} 的有效性和实用性, 我们在 17 个 UCI 数据集上对 LSTSVM_{L1D}, GEPSVM, TWSVM, LSTSVM 和 L1-GEPSVM 进行比较。表 1 记录了所有算法

矩阵 $2/c_1 \mathbf{H}^\top \mathbf{K} \mathbf{H} + \mathbf{G}^\top \mathbf{G}$ 和 $\mathbf{H}^\top \mathbf{H} + 2/c_2 \mathbf{G}^\top \mathbf{S} \mathbf{G}$ 的逆计算。然而, 在某些情况下, 它们可能是病态矩阵, 所以可能会得到不正确或不稳定的解。同于文献[12]和文献[20], 我们引入正则项 $\epsilon \mathbf{I}$ 解决此问题。实验显示, 本文的迭代算法一般在 5 步内收敛。因为在每次迭代中本文的模型在形式上同于 LSTSVM 模型, 都是求解线性方程, 因此在计算效率上我们的算法与 LSTSVM 是相当的, 从而保留了 LSTSVM 模型的计算优势。同于文献[12]和文献[20], LSTSVM_{L1D} 将被推广到非线性情况, 该项工作将在未来的研究中进行深入探讨。

4 实验结果

为了估价 LSTSVM_{L1D} 的分类性能, 我们将其与 GEPSVM, TWSVM, LSTSVM 和 L1-GEPSVM 在 UCI 数据集和人工数据集上进行比较, 所有数据均被归一化。L1-GEPSVM 和 LSTSVM_{L1D} 是两个迭代算法, 自然需要设置其初始解。L1-GEPSVM 和 LSTSVM_{L1D} 的思想来源于基于 L1 范数距离的线性判别分析 LDA, 因此我们依据文献[21, 23], 设置这两种算法的初始解为其对应的 L2 范数距离模型。设置最大迭代次数为 50, 且当两次迭代目标值的差小于 0.001 时迭代终止。对于 TWSVM, 其凸问题由内部点方法进行求解。所有算法中的参数在 $\{2^i | i = -7, -6, -5, \dots, 7\}$ 范围内, 通过在训练集上执行 10 折交叉验证进行选择, 所记录的分类精度为 10 折测试精度的平均值。

4.1 人工数据集

本节通过一个人工数据集来说明 LSTSVM_{L1D} 处理野值的有效性。考虑文献[11]中的 2 维 XOR 数据集“Cross-planes”, 并在该数据集的每类样本中分别引入 1 个野值, 如图 1(a)所示, 其中两类数据分别由“○”和“□”表示。该数据集共有 232 个样本。图 1(b)和图 1(c)分别给出了 LSTSVM 和 LSTSVM_{L1D} 的分类平面。

在原数据集上的分类结果。为了评估所有算法对野值的鲁棒性, 依据文献[22, 24], 分别向数据集加入 10% 和 20% 的高斯噪声来模仿野值, 表 2 和表 3 分别记录了在这两种情况下的分类结果。实验中, 统计置信水平设置为 0.05。P 值小于 0.05 指两种算法在分类精度上存在明显的差异。在所有表中, 黑体指最好的分类精度, * 表示与 LSTSVM_{L1D} 相比统计上存在明显的差异。

表1 GEPSVM, TWSVM, LSTSVM, L1-GEPSVM 和 LSTSVM_{L1D}在原数据集上的分类性能Table 1 Classification performance of GEPSVM, TWSVM, LSTSVM, L1-GEPSVM and LSTSVM_{L1D} on original dataset

(单位: %)

Dataset($m \times n$)	GEPSVM	TWSVM	LSTSVM	L1-GEPSVM	LSTSVM _{L1D}
	精度±标准差	精度±标准差	精度±标准差	精度±标准差	精度±标准差
Heart(270×13)	80.74±6.1532	75.93±3.7037*	82.96±3.9545*	68.15±7.0662	84.07±3.4347
Monks2(601×6)	68.72±2.1208*	62.73±4.0855*	65.23±2.2525	68.39±2.9216*	66.23±2.2851
Monks3(554×6)	80.14±3.7623*	80.86±3.0458*	83.57±5.1715	84.28±5.4189*	82.13±2.8645
Cancer(683×9)	95.76±2.6617*	92.54±0.9527*	96.05±1.9855	86.23±10.2011*	96.34±1.5933
Ktest(1130×5)	49.29±5.4988*	54.60±5.7665	54.51±5.6623	53.19±4.3615*	54.96±2.7760*
Pidd(768×8)	73.83±5.7464*	75.39±2.3068*	77.21±3.6558	74.87±5.0051*	77.47±2.7060*
Clave(963×19)	73.41±3.0727	73.52±2.1024	74.04±2.1596	73.31±3.3311	73.93±1.9505
Sonar(208×60)	74.12±8.1330*	68.70±5.5519*	72.08±4.9362*	74.14±9.9677*	75.04±5.2053*
Pimadata(768×8)	73.83±5.7464	74.39±2.3068	76.56±2.9052	73.70±4.0178	77.34±2.6012*
Ionodat(351×34)	78.07±5.5333*	85.75±5.6590*	90.60±1.9303*	80.06±6.3632*	82.62±8.4077
Clevedata(297×13)	85.89±4.1119	83.85±3.7094*	81.49±3.3255*	84.53±5.1730*	85.88±5.6529
Housing(506×13)	74.31±3.9873*	78.66±5.4781*	86.18±3.3184	72.32±5.2794*	86.37±3.2940
Germ(1000×24)	70.80±4.2615*	75.40±1.0198*	76.40±1.5620*	70.70±4.6054*	72.60±2.4779
Mush(8124×22)	78.01±10.3617*	80.75±0.9712*	81.60±0.8751*	78.90±9.2556*	83.11±1.1703
Spect(267×44)	78.69±5.7435*	79.43±5.6063	78.69±6.2299*	76.45±5.9462*	79.43±5.6063
Bright(2462×14)	95.65±1.6714*	98.58±0.5748*	98.74±0.3504*	95.66±2.6206*	99.27±0.2757
Dim(4192×14)	90.89±1.3756*	95.13±0.6521	94.68±0.4698	89.74±1.2963*	94.30±0.4752

表2 GEPSVM, TWSVM, LSTSVM, L1-GEPSVM 和 LSTSVM_{L1D}在加入10%高斯噪声的数据集上的分类性能Table 2 Classification performance of GEPSVM, TWSVM, LSTSVM, L1-GEPSVM and LSTSVM_{L1D} on dataset with 10% Gaussian noise

(单位: %)

Dataset($m \times n$)	GEPSVM	TWSVM	LSTSVM	L1-GEPSVM	LSTSVM _{L1D}
	精度±标准差	精度±标准差	精度±标准差	精度±标准差	精度±标准差
Heart(270×13)	68.52±9.2593*	70.37±5.4935*	74.44±3.1860*	70.74±6.5105*	76.67±3.2288
Monks2(601×6)	67.73±3.9638	62.39±3.4785*	61.23±2.8834*	68.24±6.9758*	66.22±2.1834
Monks3(554×6)	77.61±4.5604*	80.14±2.3286*	80.68±2.1872*	84.31±4.2190*	82.31±2.8777
Cancer(683×9)	95.46±2.9649*	92.10±0.6979*	93.42±1.6514*	85.94±8.8914*	96.49±1.7468
Ktest(1130×5)	53.36±3.8584*	55.04±5.4308	55.13±2.6929	53.36±4.4079*	55.31±2.9351*
Pidd(768×8)	74.48±4.5257*	73.57±1.5978*	75.12±3.8776	75.39±3.7020	75.78±2.5273
Clave(963×19)	73.51±2.9302	73.73±1.8125	73.83±2.1509	73.73±2.7286	73.83±1.7802
Sonar(208×60)	74.12±7.8492	68.75±4.8806*	70.13±5.7602*	74.64±10.5609	74.08±3.5758
Pimadata(768×8)	75.13±4.4184	74.86±4.6422*	74.86±3.8403*	75.30±3.7020	75.38±3.9405
Ionodat(351×34)	79.21±6.3443*	86.32±4.1192*	89.74±2.4617*	81.17±8.2317*	90.88±3.4566
Clevedata(297×13)	85.54±4.2138	75.76±7.7899*	77.45±6.4367*	85.53±5.1674	85.54±5.2382
Housing(506×13)	71.16±4.9628*	71.15±6.2795*	81.03±5.2794*	72.32±5.2794*	82.21±5.2769
Germ(1000×24)	70.00±4.9800*	70.60±3.0232	72.90±2.6533*	70.60±4.5651*	71.10±1.8815
Mush(8124×22)	75.89±12.0174*	78.52±1.3667*	81.52±0.9431*	70.47±7.6785*	82.37±1.5020
Spect(267×44)	75.64±6.0309*	79.06±5.8798	79.43±6.5118	75.70±8.2287*	79.77±5.5370
Bright(2462×14)	97.20±1.4676	97.08±0.3299	96.10±0.2359*	95.16±3.7022*	97.48±0.7119
Dim(4192×14)	91.84±1.2244*	94.27±0.7364	92.27±0.8520*	89.67±1.6551*	93.32±0.6760

表3 GEPSVM, TWSVM, LSTSVM, L1-GEPSVM 和 LSTSVM_{L1D}在加入20%高斯噪声的数据集上的分类性能Table 3 Classification performance of GEPSVM, TWSVM, LSTSVM, L1-GEPSVM and LSTSVM_{L1D} on dataset with 20% Gaussian noise

(单位: %)

Dataset($m \times n$)	GEPSVM	TWSVM	LSTSVM	L1-GEPSVM	LSTSVM _{L1D}
	精度±标准差	精度±标准差	精度±标准差	精度±标准差	精度±标准差
Heart(270×13)	64.81±7.9866*	75.93±4.2229*	72.59±5.4181	68.52±8.8036*	72.96±5.1852
Monks2(601×6)	67.23±2.7161	63.73±2.7558*	65.23±2.2525*	67.74±6.3451	66.06±2.4170
Monks3(554×6)	80.86±3.6583*	83.75±2.7549*	83.39±4.9195*	84.28±5.0300*	81.77±2.8810
Cancer(683×9)	95.61±2.7792*	96.19±0.8460	96.05±1.9855*	86.97±6.8319*	96.35±1.8415
Ktest(1130×5)	53.63±6.3223*	52.57±1.5681*	55.22±5.0014	53.72±4.1328*	55.66±1.6650
Pidd(768×8)	73.83±3.6682*	76.18±3.4303	75.91±2.3471*	74.74±5.6852*	76.43±2.6215
Clave(963×19)	72.89±3.2446*	72.38±2.6255*	72.69±2.3777*	73.21±3.1700	73.93±1.7384
Sonar(208×60)	76.02±7.8910*	73.01±7.6527*	73.53±4.2148	74.60±9.9146	74.07±6.1799
Pimadata(768×8)	74.74±3.8145*	75.13±4.1262*	75.77±3.4134*	74.62±3.6471*	76.56±2.5607
Ionodat(351×34)	77.21±6.2307*	87.75±4.6662*	90.03±1.5659*	82.03±7.0347*	92.02±3.5736
Clevedata(297×13)	84.87±4.9698	84.19±3.6981*	81.49±4.2372*	85.18±5.0238	85.21±5.3146
Housing(506×13)	74.31±4.5802*	80.44±3.7275*	79.84±4.2134*	72.32±5.2794*	80.83±4.2140
Germ(1000×24)	69.60±4.3863*	73.90±3.0232*	74.20±2.4207*	70.60±4.6733*	71.50±3.3317
Mush(8124×22)	82.02±0.9832	80.64±0.9119*	81.25±0.9873*	73.20±2.6432*	82.59±1.2071
Spect(267×44)	78.69±7.5663*	79.01±5.5570	78.30±5.6828*	76.05±9.6514*	79.43±5.6063
Bright(2462×14)	97.16±1.2820*	97.68±0.2748*	93.46±1.2915*	96.79±1.7290*	98.09±0.6631
Dim(4192×14)	91.27±0.9633*	93.77±0.3651*	91.56±1.0550*	89.79±1.3299*	92.87±0.8865

从表中可以看到, $LSTSVM_{L1D}$ 的分类性能优于其他算法,特别是在引入高斯噪声后。比较表中的结果容易发现,在引入高斯噪声后, $GEPSVM$, $TWSVM$ 和 $LSTSVM$ 在大多数情况下的分类性能有所降低。然而, $L1-GEPSVM$ 和 $LSTSVM_{L1D}$ 的分类准确性变化很小,这表明 $L1$ 范数距离在镇压野值方面较 $L2$ 范数距离有更好的鲁棒性。与 $L1-GEPSVM$ 相比, $LSTSVM_{L1D}$ 在性能上更优,这可能有以下两个原因: 1) $L1-GEPSVM$ 继承了 $GEPSVM$ 性能的不足; 2) $L1-GEPSVM$ 采用性能无法保证的 GA 迭代算法。实验结果有力地证明了我们之前的观点,即 $LSTSVM_{L1D}$ 有着极佳的鲁棒性和抗噪能力。在引入高斯噪声的例子中,其鲁棒优势尤其明显。

上述现象已表明了 $LSTSVM_{L1D}$ 的性能优势。正如前所述, $LSTSVM_{L1D}$ 的另一个优势是继承了 $LSTSVM$ 的计算优势。表 4 列出了 Mush 和 Dim 数据集上 $GEPSVM$, $TWSVM$, $LSTSVM$, $L1-GEPSVM$ 和 $LSTSVM_{L1D}$ 的 10 折平均计算时间。 $LSTSVM_{L1D}$ 具有与 $LSTSVM$ 相当的时间。

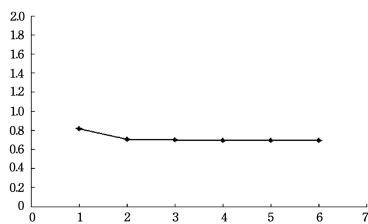
表 4 所有算法的计算时间

Table 4 Computing time of algorithms

(单位: s)

Dataset ($m \times n$)	GEPSVM	TWSVM	LSTSVM	L1-GEPSVM	$LSTSVM_{L1D}$
Mush (8124 × 22)	0.0294	78.64	0.0187	1.2546	0.09432
Dim (4192 × 14)	0.0213	69.25	0.0156	0.9263	0.05732

$LSTSVM_{L1D}$ 的目标问题由一个迭代算法求解。从理论上证明了,在每次迭代中该算法都单调减少 $LSTSVM_{L1D}$ 的目标值。图 2 展示了在 Dim 上的收敛过程,水平轴代表迭代次数,垂直轴代表目标值。从图 2 中可知, $LSTSVM_{L1D}$ 能较快地收敛,其也是 $LSTSVM_{L1D}$ 能进行快速计算的一个重要因素。

图 2 $LSTSVM_{L1D}$ 的收敛性分析Fig. 2 Convergence analysis of $LSTSVM_{L1D}$

结束语 本文提出了一种基于 $L1$ 范数距离的 $LSTSVM$ 分类算法,称为 $LSTSVM_{L1D}$ 。相比于基于 $L2$ 范数平方距离的 $LSTSVM$,应用 $L1$ 范数距离使得 $LSTSVM_{L1D}$ 对于野值更具鲁棒性,且增加了模型的灵活性。而且,还设计了一个简单有效的迭代算法来求解所构建的最优化问题(其容易实现),并且从理论上保证了其收敛性。实验也证实了 $LSTSVM_{L1D}$ 的有效性。目前, $LSTSVM_{L1D}$ 被应用于求解线性分类问题。未来的工作是将其拓展到非线性情况,并将其应用于求解更多的实际问题。

参 考 文 献

- [1] SMITH R S, KITTLER J, HAMOUZ M, et al. Face Recognition Using Angular LDA and SVM Ensembles[C]// 18th International Conference on Pattern Recognition (ICPR 2006). 2006: 1008-1012.
- [2] C J, LIN C W, HSU, et al. A practical guide to support vector classification[OL]. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [3] IVOR W T, JAMES T K, CHEUNG P K. Fast SVM Training on Very Large Data Sets[J]. Journal of Machine Learning Research, 2005(6): 363-392.
- [4] FRANC V, SONNENBURG S. Optimized cutting plane algorithm for large-scale risk minimization[J]. Journal of Machine Learning Research, 2009(10): 2157-2192.
- [5] MANGASARIAN O, WILD E. Multisurface proximal support vector machine classification via generalized eigenvalues [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 69-74.
- [6] JAYADEVA, KHEMCHANDANI R, CHANDRA S. Fuzzy multi-category proximal support vector classification via generalized eigenvalues[J]. Soft Comput, 2007, 11: 679-685.
- [7] GUARRACINO M R, CIFARELLI C, SEREF O, et al. A Parallel Classification Method for Genomic and Proteomic Problems [C]// 20th International Conference on Advanced Information Networking and Applications(AINA'06). 2006: 588-592.
- [8] YANG X B, CHEN S C. Proximal Support Vector Machine Based on Prototypal Multiclassification Hyperplanes[J]. Journal of Computer Research and Development, 2006, 43(10): 1700-1705. (in Chinese)
杨绪兵, 陈松灿. 基于原型超平面的多类最接近支持向量机[J]. 计算机研究与发展, 2006, 43(10): 1700-1705.
- [9] YANG X B, CHEN S C, YANG Y M. Localized proximal support vector machine via generalized eigenvalues [J]. Chinese Journal of Computers, 2007, 30(8): 1227-1234. (in Chinese)
杨绪兵, 陈松灿, 杨益民. 局部化的广义特征值最接近支持向量机[J]. 计算机学报, 2007, 30(8): 1227-1234.
- [10] YE Q L, ZHAO C X, YE N, et al. Multi-Weight Vector Projection Support vector machines[J]. Pattern Recognition Letters, 2010, 31: 2006-2011.
- [11] YE Q L, YE N, YIN T M. Enhanced multi-weight vector projection support vector machine [J]. Pattern Recognition Letters, 2014, 42: 91-100.
- [12] JAYADEVA, KHEMCHANDANI R, CHANDRA S. Twin support vector machines for pattern classification[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910.
- [13] TIAN Y, QI Z, JU X, et al. Nonparallel support vector machines for pattern classification[J]. IEEE Transactions on Cybernetics, 2014, 44(7): 1067.
- [14] CEVIKALP H. Best Fitting Hyperplanes for Classification[J]. IEEE Transactions on Pattern Analysis & Machine, 2017, 39(6): 1076-1088.
- [15] QI Z, TIAN Y, SHI Y. Structural twin support vector machine for classification[J]. Knowledge-Based Syst., 2013, 43: 74-81.

(下转第 130 页)

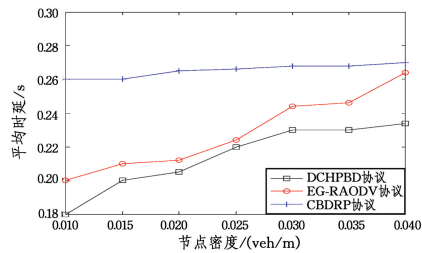


图8 节点密度与平均时延的关系曲线

Fig. 8 Relation curves between node density and average delay

结束语 本文针对城市环境下的数据分组转发问题,提出一种车载自组织网络中基于车辆密度的双簇头路由协议。该路由协议主要考虑两个重要问题:簇的形成和簇头的选择。簇的形成降低了对单个车辆节点的运动感知。选择的簇头作为中继节点转发数据分组,簇头的生存期越大,则网络重新选择中继节点的次数越少,网络越稳定。仿真结果表明,DCHPBD协议延长了簇头生存期,同时减少了数据传输的平均跳数并降低了时延,通信性能更佳。

参考文献

- [1] CAI R,ZHANG G A,JI Y C. Adaptive Routing Protocol Based on Directional Transmission in VANETs[J]. Computer Science, 2016,43(5):62-66. (in Chinese)
蔡蓉,章国安,季彦呈. 车载自组织网络中基于定向传播的自适应路由协议[J]. 计算机科学,2016,43(5):62-66.
- [2] CHUANG M C,CHENM C. DEEP:density-aware emergency message extension protocol for VANETs[J]. IEEE Transactions on Wireless Communications,2013,12(10):4983-4993.
- [3] MILDRED M,HONG C S,BANG Y. An adaptable mobility-aware clustering algorithm in vehicular networks[C]//Network Operations and Management Symposium (APNOMS). IEEE, 2011:1-6.
- [4] LUO Y,ZHANG W,HU Y. A new cluster based routing protocol for VANET[C]//2010 Second International Conference on Networks Security,Wireless Communications and Trusted Computing. Wuhan,China,2010:176-180.
- [5] SINGHA J P,BALI R S. A hybrid backbone based clustering

algorithm for vehicular Ad Hoc networks[C]//IEEE International Conference on Information and Communication Technologies (ICICT),Procedia Computer Science. Kochi,India,2014:1005-1013.

- [6] IDJMAYYEL H,QAZI B R. Energy efficient double cluster head routing scheme in a city vehicular network[C]//2013 27th International Conference on Advanced Information Networking and Applications Workshops. Barcelona, Spain, 2013: 1594-1599.
- [7] SOUZA E,NIKOLAIDIS I,GBURZYNSKI P. A new aggregate local mobility(ALM) clustering algorithm for VANETs[C]//IEEE 2010 International Conference on Communications(ICC). Cape Town,South Africa. 2010:1-5.
- [8] SALEET H,BASIR O,LANGAR R,et al. Region-based location service-management protocol for VANETs [J]. IEEE Transaction on Vehicular Technology,2010,59(2):917-931.
- [9] LI G,BOUKHATEM L. An Intersection-based delay sensitive routing for VANETs using ACO algorithm[C]//IEEE International Conference on Computer Communication and Networks. Shanghai,China,2014:1-8.
- [10] RAW S R,DAS S. Performance Analysis of P-GEDIR Protocol for Vehicular Ad Hoc Network in Urban Traffic Environments [J]. Wireless Personal Communications,2013,68(1):65-78.
- [11] SONG T,XIA W,SHEN L,et al. A cluster-based directional routing protocol in VANET[C]//2010 IEEE 12th International Conference on Communication Technology (ICCT). Nanjing, China,2010:1172-1175.
- [12] EIZA M H,NI Q. An evolving graph-based reliable routing scheme for VANETs[J]. IEEE Transaction on Vehicular Technology,2013,62(4):1493-1504.
- [13] MEHRA R,BALI R S,KAUR P. Efficient Clustering based OLSR Routing Protocol for VANET[C]//2016 Symposium on Colossal Data Analysis and Networking(CDAN). Indore,India, 2016:1-7.
- [14] CHEN P J,TANG L,CHEN Q B. Clustering Protocol Based on Motion State in VANETs[J]. Computer Engineering, 2013, 39(3):87-98. (in Chinese)
陈培菊,唐伦,陈前斌. 车载自组织网络中基于运动状态的簇路由协议[J]. 计算机工程,2013,39(3):87-98.

(上接第105页)

- [16] 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,2014,73(1):276-288.
- [17] KHEMCHANDANI R,SAIGAL P,CHANDRA S. Improvements on ν -Twin Support Vector Machine[J]. Neural Netw., 2016,7(79):97-107.
- [18] KUMAR M A,GOPA M. Application of smoothing technique on twin support vector machines[J]. Pattern Recognition Letters,2008,29:1842-1848.
- [19] YE Q L,ZHAO C X. A Feature Selection Method for TWSVM via a Regularization Technique[J]. Journal of Computer Research and Development,2011,48(6):1029-1037. (in Chinese)
业巧林,赵春霞. 基于正则化的 TWSVM 特征选择算法[J]. 计算机研究与发展,2011,48(6):1029-1037.

- [20] KUMAR M A,GOPAL M. Least squares twin support vector machines for pattern classification[J]. Expert Systems with Applications,2009,36(4):7535-7543.
- [21] WANG H X,LU X S,HU Z L,et al. Fisher discriminant analysis with L1-norm[J]. IEEE Trans. Cybern., 2014,6(44):828-842.
- [22] LI C N,SHAO Y H,DENG N Y. Robust L1-norm non-parallel proximal support vector machine[J]. Optimization,2016,65(1):1-15.
- [23] ZHENG W M,LIN Z C,WANG H X. L1-norm distance Kernel Discriminant Analysis via Bayes Error Bound Optimization for Robust Feature Extraction [J]. IEEE Trans. Neural Netw., 2014,4(24):793-805.
- [24] WANG H,TANG Q,ZHENG W M. L1-Norm-Based Common Spatial Patterns [J]. IEEE Trans. Biomed. Engineering, 2012, 59(3):653-662.