软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2013,24(11):2548−2557 [doi: 10.3724/SP.J.1001.2013.04475]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

加权光滑 CHKS 孪生支持向量机*

丁世飞[1,2], 黄华娟[1,2], 史忠植[2]

[1](中国矿业大学 计算机科学与技术学院,江苏 徐州 221116)

[2](中国科学院 计算技术研究所 智能信息处理重点实验室,北京 100190)

通讯作者: 黄华娟, E-mail: hhj-025@163.com

摘 要: 针对光滑孪生支持向量机(smooth twin support vector machines,简称 STWSVM)采用的 Sigmoid 光滑函数逼近精度低和 STWSVM 对异常点敏感的问题,引入一种性能更好的光滑函数——CHKS 函数,提出了光滑 CHKS 孪生支持向量机模型(smooth CHKS twin support vector machines,简称 SCTWSVM).在此基础上,根据样本点的位置为每个训练样本赋予不同的重要性,以降低异常点对非平行超平面的影响,提出了加权光滑 CHKS 孪生支持向量机(weighted smooth CHKS twin support vector machines,简称 WSCTWSVM).不仅从理论上证明了 SCTWSVM 具有严凸性和任意阶光滑的性能,而且在数据集上的实验结果表明,相对于 STWSVM,SCTWSVM 可以在更短的时间内获得更高的分类精度,同时验证了 WSCTWSVM 的有效性和可行性.

关键词: 孪生支持向量机;光滑孪生支持向量机;CHKS 函数;光滑;加权

中图法分类号: TP181 **文献标识码**: A

中文引用格式: 丁世飞,黄华娟,史忠植.加权光滑 CHKS 孪生支持向量机.软件学报,2013,24(11):2548−2557. http://www.jos.org.cn/1000-9825/4475.htm

英文引用格式: Ding SF, Huang HJ, Shi ZZ. Weighted smooth CHKS twin support vector machines. Ruan Jian Xue Bao/Journal of Software, 2013,24(11):2548−2557 (in Chinese). http://www.jos.org.cn/1000-9825/4475.htm

Weighted Smooth CHKS Twin Support Vector Machines

DING Shi-Fei[1,2], HUANG Hua-Juan[1,2], SHI Zhong-Zhi[2]

[1](School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

[2](Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

Corresponding author: HUANG Hua-Juan, E-mail: hhj-025@163.com

Abstract: Smooth twin support vector machines (STWSVM) uses Sigmoid function to transform the unsmooth twin support vector machines (TWSVM) into smooth ones. However, because of the low approximation ability of Sigmoid function, the classification accuracy of STWSVM is unsatisfactory. Furthermore, similar to TWSVM, STWSVM is sensitive to the abnormal samples. In order to address the above problems, this paper introduces CHKS function, and proposes a smooth twin support vector machines, smooth CHKS twin support vector machines (SCTWSVM). In order to reduce the influence of abnormal samples on SCTWSVM, different importance are given for each training sample according to the sample point positions for SCTWSVM, resulting in weighted smooth CHKS twin support vector machines (WSCTWSVM). The study proves that SCTWSVM is not only strictly convex, but also can meet the arbitrary order smooth performance. Meanwhile, the experimental results show that SCTWSVM has better performance than STWSVM. Furthermore, the experimental results also show that WSCTWSVM is effective and feasible relative to SCTWSVM.

Key words: twin support vector machines; smooth twin support vector machines; CHKS function; smooth; weight

支持向量机(support vector machine,简称 SVM)自 1995 年由 Vapnik 提出以来就受到理论研究和工程应用

* 基金项目: 国家自然科学基金(61379101); 国家重点基础研究发展计划(973)(2013CB329502)
收稿时间: 2013-04-24; 修改时间: 2013-07-12, 2013-08-02; 定稿时间: 2013-08-27

© 中国科学院软件研究所 http://www.jos.org.cn

两方面的重视,目前已成为机器学习的研究热点,并在很多领域如模式识别、时间序列预测、图像处理等得到了成功的应用[1−4].与其他机器学习方法相比,比如与人工神经网络相比[5,6],SVM 成功地解决了高维问题和局部极小值问题,因此具有更好的泛化能力.虽然 SVM 已经取得了较好的学习性能,但是它的训练时间非常高,达到 $O(l^3)$.其中,l 是整个训练集的样本数.为了提高 SVM 的训练速度,探寻关于 SVM 优化问题合理且高效的算法已成为 SVM 研究领域的一个重要研究课题.目前,大多数改进算法的一个共同思想就是用分解的方法将大规模二次规划问题转化为小规模问题求解.常用的算法包括选块算法(chunking algorithm)[7]、分解算法(decomposition algorithm)[8]、序列最小优化算法(sequential minimal optimization,简称 SMO)[9]等等.这些经典的改进方法在一定程度上节省了计算机内存,提高了算法的计算效率,但算法的设计和实现比较复杂.近年来,很多学者开始研究基于标准 SVM 的变形算法.例如,2001 年,Fung[10]等人提出了近似支持向量机分类器(proximal support vector machine,简称 PSVM).与 SVM 不同,PSVM 的基本思想是在每类样本集中设置与样本点邻近的平行超平面,并且使两个平行超平面之间的距离达到最大.该算法的时间复杂度为 $O(n^3)$,其中,n 为样本的维数.基于 PSVM 的思想,2006 年,Mangasarian 等人[11]提出了广义特征近似支持向量机(proximal support vector machines via generalized eigenvalues,简称 GEPSVM).该算法舍弃了 PSVM 平行约束的条件,优化目标要求超平面离本类样本尽可能地近,离它类样本尽可能地远,问题归结为求解两个广义特征值问题.与 PSVM 相比,GEPSVM 能够更好地处理异或(XOR)问题.

2007 年,Jayadeva 等人[12]在深入研究了标准 SVM 形式的基础上,借鉴 GEPSVM 的思想提出了孪生支持向量机(twin support vector machines,简称 TWSVM).TWSVM 将 GEPSVM 问题转换为两个较小的形如 SVM 的凸二次规划问题,因此时间开销缩减到 SVM 的 1/4.鉴于其明显的分类优势,TWSVM 已被应用于说话人识别[13]、医学检测[14,15]等领域.然而,TWSVM 的标准算法是在对偶空间求解两个二次规划问题,这种求解方法对于样本数目较大的问题将受到时间和内存的制约.为了解决这个问题,2008 年,Kumar 等人[16]引入正号函数,把 TWSVM 原始模型转化成两个不光滑的无约束优化问题,力求在原始空间求解 TWSVM 模型.为了可以使用具有快速收敛能力的梯度方法求解无约束优化问题,Kumar 等人采用 Sigmoid 光滑函数,对不光滑的无约束优化问题做光滑处理,提出了可以直接在原空间求解 TWSVM 模型的光滑孪生支持向量机(smooth twin support vector machines,简称 STWSVM).相对于 TWSVM,在 NDC 大数据集上的实验结果表明,STWSVM 的效率得到了很大的提高.然而,Sigmoid 函数对正号函数的逼近精度比较低,导致 STWSVM 的分类精度不够理想.并且和 SVM 一样,STWSVM 对异常点非常敏感.然而,STWSVM 并没有考虑输入样本点对最优分类超平面产生的不同影响.CHKS 函数是一种比 Sigmoid 函数具有更好逼近能力的光滑函数,已经被成功地应用到光滑支持向量机中[17].在本文中,我们引入 CHKS 函数,采用 CHKS 函数逼近孪生支持向量机的不可微项,提出了光滑 CHKS 孪生支持向量机模型(smooth CHKS twin support vector machines,简称 SCTWSVM).在此基础上,通过为每个训练样本赋予不同的样本重要性来减弱异常样本对超平面的影响,提出了加权 CHKS 孪生支持向量机学习算法(weighted smooth CHKS twin support vector machines,简称 WSCTWSVM).我们不仅从理论上证明了 SCTWSVM 具有严凸性、任意阶光滑和全局收敛的性能,而且在 UCI 数据集和 NDC 大数据集上的实验结果表明 SCTWSVM 能够在更短的时间内获得比 STWSVM 更高的分类精度,并且,WSCTWSVM 比 SCTWSVM 具有更好的分类性能.

1 光滑 CHKS 孪生支持向量机

1.1 SCTWSVM原理及其性质

给定两类 n 维的 m 个训练点,分别用 $m_1\times n$ 的矩阵 A 和 $m_2\times n$ 的矩阵 B 表示+1 类和−1 类,这里,m_1 和 m_2 分别代表两类样本的数目.TWSVM 的目标是在 n 维空间中寻找两个非平行的超平面:$x^Tw_1+b_1=0$,$x^Tw_2+b_2=0$,要求每一个超平面离本类样本尽可能地近,离它类样本距离尽可能地远.

TWSVM 可以归结为求解下面两个二次规划问题:

© 中国科学院软件研究所 http://www.jos.org.cn

$$\begin{cases}\min\limits_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\|Aw^{(1)}+e_1b^{(1)}\|^2+c_1e_2^T\xi^{(2)} \\ \text{s.t.}\quad -(Bw^{(1)}+e_2b^{(1)})\geqslant e_2-\xi^{(2)} \\ \xi^{(2)}\geqslant 0\end{cases} \tag{1}$$

$$\begin{cases}\min\limits_{w^{(2)},b^{(2)},\xi^{(1)}} \frac{1}{2}\|Bw^{(2)}+e_2b^{(2)}\|^2+c_2e_1^T\xi^{(1)} \\ \text{s.t.}\quad (Aw^{(2)}+e_1b^{(2)})\geqslant e_1-\xi^{(1)} \\ \xi^{(1)}\geqslant 0\end{cases} \tag{2}$$

其中,c_1,c_2 是两个惩罚参数,e_1,e_2 是两个全由 1 组成的列向量, $A=[x_1^{(1)},x_2^{(1)},...,x_{m1}^{(1)}]^T, B=[x_1^{(2)},x_2^{(2)},...,x_{m2}^{(2)}]^T, x_j^{(i)}$ 表示第 i 类的第 j 个样本.

在 TWSVM 算法中,TWSVM 丧失了稀疏性,当训练样本集较大时,算法效率将会比较低.为了提高 TWSVM 的训练速度,本文在求解公式(1)和公式(2)时不是在对偶空间寻找原始问题的最优解,而是直接在原空间求解原始问题.

由 KKT 条件,公式(1)和公式(2)的最优解有:

$$\xi^{(2)}=\max\{0,e_2+(Bw^{(1)}+e_2b^{(1)})\} \tag{3}$$

$$\xi^{(1)}=\max\{0,e_1-(Aw^{(2)}+e_1b^{(2)})\} \tag{4}$$

将公式(3)、公式(4)代入公式(1)和公式(2),得:

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\|Aw^{(1)}+e_1b^{(1)}\|^2+c_1e_2^T\max\{0,(e_2+Bw^{(1)}+e_2b^{(1)})\} \tag{5}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}} \frac{1}{2}\|Bw^{(2)}+e_2b^{(2)}\|^2+c_2e_1^T\max\{0,(e_1-Aw^{(2)}-e_1b^{(2)})\} \tag{6}$$

令 $\phi_1(w^{(1)},b^{(1)})=(u_1)_+=\max\{0,e_2+(Bw^{(1)}+e_2b^{(1)})\}$,$\phi_2(w^{(2)},b^{(2)})=(u_2)_+=\max\{0,(e_1-Aw^{(2)}-e_1b^{(2)})\}$,其中,$(u_1)_+$和$(u_2)_+$称为正号函数.

定理 1(不光滑性). 无约束的孪生支持向量机模型由公式(5)和公式(6)给出,该模型连续但不光滑.

由定理 1 可知,公式(5)和公式(6)的第 2 项不光滑,无法用梯度方法进行求解,因为使用梯度方法求解无约束优化问题的前提条件是要求目标函数为一阶或二阶可微.为此,本文采用 CHKS 函数 $\phi(x,\varepsilon)=\frac{x+\sqrt{x^2+4\varepsilon^2}}{2}$ 来逼近$\phi(x)=\max\{0,x\}$,其中,ε是足够小的参数.可得$\phi_1(w^{(1)},b^{(1)})$和$\phi_2(w^{(2)},b^{(2)})$的光滑函数分别为

$$\phi_1(w^{(1)},b^{(1)},\varepsilon)=\frac{1}{2}(e_2+Bw^{(1)}+e_2b^{(1)})+\frac{1}{2}\sqrt{(e_2+Bw^{(1)}+e_2b^{(1)})^2+4\varepsilon^2} \tag{7}$$

$$\phi_2(w^{(2)},b^{(2)},\varepsilon)=\frac{1}{2}(e_1-Aw^{(2)}-e_1b^{(2)})+\frac{1}{2}\sqrt{(e_1-Aw^{(2)}-e_1b^{(2)})^2+4\varepsilon^2} \tag{8}$$

采用公式(7)和公式(8)作为光滑函数,可以得到光滑 CHKS 孪生支持向量机的模型为

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}} \Phi_1(w^{(1)},b^{(1)},\varepsilon)=\frac{1}{2}\|Aw^{(1)}+e_1b^{(1)}\|^2+c_1e_2^T\phi_1(w^{(1)},b^{(1)},\varepsilon) \tag{9}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}} \Phi_2(w^{(2)},b^{(2)},\varepsilon)=\frac{1}{2}\|Bw^{(2)}+e_2b^{(2)}\|^2+c_2e_1^T\phi_2(w^{(2)},b^{(2)},\varepsilon) \tag{10}$$

引理 1. $\Phi_1(w^{(1)},b^{(1)},\varepsilon)$和$\Phi_2(w^{(2)},b^{(2)},\varepsilon)$由公式(9)和公式(10)给出,则有:

(1) 对任意的 $w^{(1)}\in R^n,w^{(2)}\in R^n,b^{(1)}\in R,b^{(2)}\in R,\Phi_1(w^{(1)},b^{(1)},\varepsilon)$和$\Phi_2(w^{(2)},b^{(2)},\varepsilon)$分别关于 $w^{(1)},b^{(1)}$和 $w^{(2)},b^{(2)}$任意阶光滑.

(2) 对任意的 $w^{(1)}\in R^n,w^{(2)}\in R^n,b^{(1)}\in R,b^{(2)}\in R,\Phi_1(w^{(1)},b^{(1)},\varepsilon)$和$\Phi_2(w^{(2)},b^{(2)},\varepsilon)$单调递增.

(3) 对任意的 $w^{(1)}\in R^n,w^{(2)}\in R^n,b^{(1)}\in R,b^{(2)}\in R,\varepsilon>0$,有:

- $\Phi_1(w^{(1)},b^{(1)})\leqslant\Phi_1(w^{(1)},b^{(1)},\varepsilon)\leqslant\Phi_1(w^{(1)},b^{(1)})+\varepsilon$;
- $\Phi_2(w^{(2)},b^{(2)})\leqslant\Phi_2(w^{(2)},b^{(2)},\varepsilon)\leqslant\Phi_2(w^{(2)},b^{(2)})+\varepsilon$.

© 中国科学院软件研究所 http://www.jos.org.cn

(4) 对任意的$\varepsilon>0$,$\Phi_1(w^{(1)},b^{(1)},\varepsilon)$和$\Phi_2(w^{(2)},b^{(2)},\varepsilon)$是连续可微且严格凸的.

证明:

(1) 容易证明$\phi_1(w^{(1)},b^{(1)},\varepsilon)$和$\phi_2(w^{(2)},b^{(2)},\varepsilon)$是任意阶光滑的,因此结论(1)成立.证明略.

(2) 在此我们只证明$\Phi_1(w^{(1)},b^{(1)},\varepsilon)$是单调递增的,$\Phi_2(w^{(2)},b^{(2)},\varepsilon)$同理可证;

对任意的$x\in R^n$,$\varepsilon>0$,有:

$$\frac{\partial\phi_1(x,\varepsilon)}{\partial\varepsilon}=\frac{2\varepsilon}{\sqrt{x^2+4\varepsilon^2}}>0,\ \frac{\partial\phi_1(x,\varepsilon)}{\partial x}=\frac{1}{2}\left(1+\frac{x}{\sqrt{x^2+4\varepsilon^2}}\right)=\frac{\sqrt{x^2+4\varepsilon^2}+x}{2\sqrt{x^2+4\varepsilon^2}}>0.$$

因此,函数$\phi_1(x,\varepsilon)$关于ε是单调递增函数,即对任意的 $0<\varepsilon_1<\varepsilon_2$,有$\phi_1(x,\varepsilon_1)<\phi_1(x,\varepsilon_2)$;再由$\Phi_1(w^{(1)},b^{(1)},\varepsilon)$的定义,结论显然是成立的.

(3) $\phi_1(x,\varepsilon)-\max\{0,x\}=\dfrac{\left(\sqrt{x^2+4\varepsilon^2}+x\right)}{2}-\dfrac{x+\sqrt{x^2}}{2}=\dfrac{\sqrt{x^2+4\varepsilon^2}-\sqrt{x^2}}{2}.$

因此对任意的 $w^{(1)}\in R^n,b^{(1)}\in R,\varepsilon>0$,有 $0\leqslant\phi_1(w^{(1)},b^{(1)},\varepsilon)-\phi_1(w^{(1)},b^{(1)})\leqslant\varepsilon$.

由$\Phi_1(w^{(1)},b^{(1)},\varepsilon)$的定义,易得:$\Phi_1(w^{(1)},b^{(1)})\leqslant\Phi_1(w^{(1)},b^{(1)},\varepsilon)\leqslant\Phi_1(w^{(1)},b^{(1)})+\varepsilon$.

$\Phi_2(w^{(2)},b^{(2)})\leqslant\Phi_2(w^{(2)},b^{(2)},\varepsilon)\leqslant\Phi_2(w^{(2)},b^{(2)})+\varepsilon$,同理可证.

(4) 由定理 1 可知,对任意的$\varepsilon>0$,$\Phi_1(w^{(1)},b^{(1)},\varepsilon)$显然是连续可微的.下面证明它是严格凸函数.

由公式(7)和公式(9)可得:

$$\nabla_{w^{(1)}}\Phi_1(w^{(1)},b^{(1)},\varepsilon)=A(Aw^{(1)}+e_1b^{(1)})+\frac{c_1e_2^T}{2}\left(1+\frac{e_2+Bw^{(1)}+e_2b^{(1)}}{\sqrt{(e_2+Bw^{(1)}+e_2b^{(1)})+4\varepsilon^2}}\right)B,$$

$$\nabla_{b^{(1)}}\Phi_1(w^{(1)},b^{(1)},\varepsilon)=e_1(Aw^{(1)}+e_1b^{(1)})+\frac{c_1e_2^T}{2}\left(1+\frac{e_2+Bw^{(1)}+e_2b^{(1)}}{\sqrt{(e_2+Bw^{(1)}+e_2b^{(1)})+4\varepsilon^2}}\right)e_1.$$

于是有:

$$\nabla^2\Phi_1(w^{(1)},b^{(1)},\varepsilon)=\begin{pmatrix}AA^T+BB^T\lambda_1(w^{(1)},b^{(1)},\varepsilon) & Ae_1+B\lambda_1(w^{(1)},b^{(1)},\varepsilon)\\ Ae_1+Be_1\lambda_1(w^{(1)},b^{(1)},\varepsilon) & e_1e_1^T+e_1e_1^T\lambda_1(w^{(1)},b^{(1)},\varepsilon)\end{pmatrix},$$

其中,$\lambda_1(w^{(1)},b^{(1)},\varepsilon)=\dfrac{2c_1e_2^T\varepsilon^2}{\left(\sqrt{(e_2+Bw^{(1)}+e_2b^{(1)})+4\varepsilon^2}\right)^3}$.

对任意的$\xi_1^T=(\xi^T,\xi_0)\in R^{n+1}$且$\xi_1\neq0,\xi\in R^n$,由于$\lambda_1(w^{(1)},b^{(1)},\varepsilon)>0$,有:

$$\begin{aligned}\xi_1^T\nabla^2\Phi_1(w^{(1)},b^{(1)},\varepsilon)\xi_1&=(A\xi^T)(A^T\xi)+(B\xi^T)(B^T\xi)\lambda_1(w^{(1)},b^{(1)},\varepsilon)+Ae_1\xi\xi_0+B\xi\xi_0\lambda_1(w^{(1)},b^{(1)},\varepsilon)+\\&\quad Ae_1\xi\xi_0+Be_1\xi\xi_0\lambda_1(w^{(1)},b^{(1)},\varepsilon)+\xi_0^2e_1e_1^T+\xi_0^2e_1e_1^T\lambda_1(w^{(1)},b^{(1)},\varepsilon)\\&=\|A\xi\|^2+(B\xi+\xi_0)^2\lambda_1(w^{(1)},b^{(1)},\varepsilon)>0.\end{aligned}$$

因此,对任意的$\varepsilon>0$,$\Phi_1(w^{(1)},b^{(1)},\varepsilon)$是严格凸函数.同理可证$\Phi_2(w^{(2)},b^{(2)},\varepsilon)$是连续可微且严格凸的. □

引理 2. 设 (w_1^k,b_1^k) 和 (w_2^k,b_2^k) 是目标函数公式(9)和公式(10)的极小点,(w_1^*,b_1^*) 和 (w_2^*,b_2^*) 是公式(5)和公式(6)的极小点,则当光滑因子ε趋于无穷小时,有:

$$0\leqslant\Phi_1(w_1^k,b_1^k,\varepsilon)-\Phi_1(w_1^*,b_1^*)\leqslant C\varepsilon \tag{11}$$

$$0\leqslant\Phi_2(w_2^k,b_2^k,\varepsilon)-\Phi_2(w_2^*,b_2^*)\leqslant C\varepsilon \tag{12}$$

其中,C 为常数.

由引理 1 容易得到引理 2 的结论,在此证明略.

定理 2. SCTWSVM 模型的解全局收敛于 TWSVM 原问题的解.即,设 (w_1^k,b_1^k) 和 (w_2^k,b_2^k) 是目标函数公式(9)和公式(10)的极小点,则存在公式(5)和公式(6)的极小点 (w_1^*,b_1^*) 和 (w_2^*,b_2^*),使得:

$$\lim_{k\to\infty}(w_1^k,b_1^k)=(w_1^*,b_1^*) \tag{13}$$

© 中国科学院软件研究所 http://www.jos.org.cn

$$\lim_{k\to\infty}(w_2^k,b_2^k)=(w_2^*,b_2^*) \tag{14}$$

证明:由于$\Phi_1(w_1,b_1,\varepsilon)$和$\Phi_2(w_2,b_2,\varepsilon)$是严格凸函数,可得问题公式(9)和公式(10)的极小值点(w_1^k,b_1^k)和(w_2^k,b_2^k)唯一.由引理 2 可知,$0\leqslant\Phi_1(w_1^k,b_1^k,\varepsilon)-\Phi_1(w_1^*,b_1^*)\leqslant\varepsilon$,$0\leqslant\Phi_2(w_2^k,b_2^k,\varepsilon)-\Phi_2(w_2^*,b_2^*)\leqslant\varepsilon$,因此,点列$\{(w_1^k,b_1^k)\}_{k=1}^{+\infty}$和$\{(w_2^k,b_2^k)\}_{k=1}^{+\infty}$收敛,再由$\Phi_1(w_1,b_1)$和$\Phi_2(w_2,b_2)$的连续性可证. □

1.2 非线性SCTWSVM

由前面的分析我们知道,SCTWSVM 可以解决线性分类问题,如果将前面的结论推广到非线性 SCTWSVM,就可以处理非线性问题.

基于核空间的 TWSVM 的两个超平面可以表示为

$$K(x^T,C^T)u_1+b_1=0,K(x^T,C^T)u_2+b_2=0 \tag{15}$$

其中,$C=[A^T,B^T]^T$,则非线性 TWSVM 的优化问题为

$$\begin{cases}\min\limits_{w^{(1)},b^{(1)},\xi^{(2)}}\dfrac{1}{2}\|K(A,C^T)w^{(1)}+e_1b^{(1)}\|^2+c_1e_2^T\xi^{(2)}\\ \text{s.t.}\quad -(K(B,C^T)w^{(1)}+e_2b^{(1)})\geqslant e_2-\xi^{(2)}\\ \xi^{(2)}\geqslant 0\end{cases} \tag{16}$$

$$\begin{cases}\min\limits_{w^{(1)},b^{(1)},\xi^{(2)}}\dfrac{1}{2}\|K(B,C^T)w^{(2)}+e_2b^{(1)}\|^2+c_2e_1^T\xi^{(1)}\\ \text{s.t.}\quad (K(A,C^T)w^{(2)}+e_1b^{(2)})\geqslant e_1-\xi^{(1)}\\ \xi^{(1)}\geqslant 0\end{cases} \tag{17}$$

公式(16)和公式(17)模型可转化为无约束规划问题:

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}}\frac{1}{2}\|K(A,C^T)w^{(1)}+e_1b^{(1)}\|^2+c_1e_2^T\max\{0,(e_2+K(B,C^T)w^{(1)}+e_2b^{(1)})\} \tag{18}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}}\frac{1}{2}\|K(B,C^T)w^{(2)}+e_2b^{(2)}\|^2+c_2e_1^T\max\{0,(e_1-K(A,C^T)w^{(2)}-e_1b^{(2)})\} \tag{19}$$

引入 CHKS 光滑函数:

$$\phi_{11}(w^{(1)},b^{(1)},\varepsilon)=\frac{1}{2}(e_2+K(B,C^T)w^{(1)}+e_2b^{(1)})+\frac{1}{2}\sqrt{(e_2+K(B,C^T)w^{(1)}+e_2b^{(1)})^2+4\varepsilon^2},$$

$$\phi_{12}(w^{(2)},b^{(2)},\varepsilon)=\frac{1}{2}(e_1-K(A,C^T)w^{(2)}-e_2b^{(2)})+\frac{1}{2}\sqrt{(e_1-K(A,C^T)w^{(2)}-e_1b^{(2)})^2+4\varepsilon^2}.$$

可得非线性 SCTWSVM 模型:

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}}\Phi_3(w^{(1)},b^{(1)},\varepsilon)=\frac{1}{2}\|K(A,C^T)w^{(1)}+e_1b^{(1)}\|^2+c_1e_2^T\phi_{11}(w^{(1)},b^{(1)},\varepsilon) \tag{20}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}}\Phi_4(w^{(2)},b^{(2)},\varepsilon)=\frac{1}{2}\|K(B,C^T)w^{(2)}+e_2b^{(2)}\|^2+c_2e_1^T\phi_{12}(w^{(2)},b^{(2)},\varepsilon) \tag{21}$$

经过推论易知,前面的定理也适用于非线性 SCTWSVM.

1.3 SCTWSVM算法

由定理 2 可知,SCTWSVM 的目标函数是任意阶光滑的,因此可以使用具有快速收敛能力的 Newton-Armijo 方法[18]进行求解.因为目标函数是严凸的,因此使用 Newton-Armijo 方法训练可以全局收敛,并且可以得到唯一的极小点.Newton-Armijo 求解 SCTWSVM 的过程如下:

算法 1.

Input:给定初始点$(w^0,\gamma^0)\in R^{n+1}$,$\eta$,令迭代步骤 $i=0$.

Output:目标函数的最优值.

Step 1. 计算$\Phi^i(w^i,b^i;\varepsilon)$和$g^i=\nabla\Phi(w^i,b^i;\varepsilon)$.

© 中国科学院软件研究所 http://www.jos.org.cn

Step 2. 如果$\|g^i\|\leqslant\eta$,则选取$(w^*,b^*)=(w^i,b^i)$,停机;否则,由方程$\nabla^2\Phi(w^i,b^i;\varepsilon)d^i=-g^i$计算下降方向 d^i.

Step 3. (Armijo 步)取 $\delta\in\left(0,\frac{1}{2}\right),\lambda_i=\max\left\{1,\frac{1}{2},\frac{1}{4},\ldots\right\}$,使得:$\Phi(w^i,b^i;\varepsilon)-\Phi((w^i,b^i)+\lambda_i d^i;\varepsilon)\geqslant-\delta\lambda_i g^i d^i$ 成立,

令$(w^{i+1},b^{i+1})=(w^i,b^i)+\lambda_i d^i$.

Step 4. 令 $i\leftarrow i+1$,转 Step 2.

2 加权光滑 CHKS 孪生支持向量机

2.1 加权光滑CHKS孪生支持向量机模型

和 STWSVM 一样,SCTWSVM 没有考虑到不同位置的训练样本对最优分类超平面产生的不同影响.当训练样本出现噪声等异常点时,给每个样本赋予相同的惩罚参数值将会影响到分类超平面的拓扑结构,进而降低算法的泛化能力.在这一节中,在 SCTWSVM 的基础上,我们将对处于不同位置的样本赋予不同的惩罚系数,提出加权光滑 CHKS 孪生支持向量机学习算法(WSCTWSVM).

对于线性情况,WSCTWSVM 的模型为

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}}\Phi_1(w^{(1)},b^{(1)},\varepsilon)=\frac{1}{2}\|Aw^{(1)}+e_1b^{(1)}\|^2+s_1c_1e_2^T\phi_1(w^{(1)},b^{(1)},\varepsilon)\tag{22}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}}\Phi_2(w^{(2)},b^{(2)},\varepsilon)=\frac{1}{2}\|Bw^{(2)}+e_2b^{(2)}\|^2+s_2c_2e_1^T\phi_2(w^{(2)},b^{(2)},\varepsilon)\tag{23}$$

对于非线性情况,WSCTWSVM 的模型为

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}}\Phi_3(w^{(1)},b^{(1)},\varepsilon)=\frac{1}{2}\|K(A,C^T)w^{(1)}+e_1b^{(1)}\|^2+s_1c_1e_2^T\phi_{11}(w^{(1)},b^{(1)},\varepsilon)\tag{24}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}}\Phi_4(w^{(2)},b^{(2)},\varepsilon)=\frac{1}{2}\|K(B,C^T)w^{(2)}+e_2b^{(2)}\|^2+s_2c_2e_1^T\phi_{12}(w^{(2)},b^{(2)},\varepsilon)\tag{25}$$

其中,$s_1,s_2\in(0,1]$分别表示正负类样本的加权系数值组成的向量.

2.2 加权系数的设计

和 TWSVM 一样,在用 SCTWSVM 求解最优分类超平面时,离每类样本集中心远的样本是少量的,这些样本点被称为异常点.分类超平面对这些样本点的分布是比较敏感的,若降低这些样本对分类超平面的影响,则可以提高算法的泛化能力.

在本节中,计算每个样本点的权重值,采用的是基于样本点和其类中心的距离方法.离类中心近的样本点对最优分类超平面的贡献比较大,因此对此类样本的加权系数赋值为 1;对离中心较远的那部分样本点的加权系数赋值为足够小的正数,以降低这类样本对分类超平面拓扑结构的影响;中间的一部分样本根据距离的大小赋值为 0~1 之间的数.

正类的类中心定义为 $x_+=\frac{\sum_{y_i=1}x_i}{l_+}$,负类的类中心定义为 $x_-=\frac{\sum_{y_i=-1}x_i}{l_-}$,$i=1,2,\ldots,l$.其中,$l_+$和 l_-分别表示正、负类样本的样本总数.

计算正、负类样本到其对应类中心的最远距离,并将其定义为正、负类的半径.

正类的半径定义为$r_+=\max_{\{x_i:y=1\}}\|x_+-x_i\|,r_-=\max_{\{x_i:y=-1\}}\|x_--x_i\|$.根据每个样本到该类的距离,可以定义每个样本的加权系数:

$$s_1=\begin{cases}1, & r=x_+,y_i=1\\ 10^{-3}, & r=r_+,y_i=1\\ 1-\|x_+-x_i\|/(r_++\delta), & \text{其余的}y_i=1\end{cases},$$

© 中国科学院软件研究所 http://www.jos.org.cn

$$s_2=\begin{cases}1, & p=x_-, y_i=-1\\ 10^{-3}, & p=r_-, y_i=-1\\ 1-\|x_- - x_i\|/(r_- +\delta), & \text{其余的} y_i=-1\end{cases},$$

其中,$r=\|x_+-x_i\|$,$p=\|x_--x_i\|$,$\delta=10^{-6}$.δ的引入是为了避免加权系数为 0.

3 实验与分析

为了测试所提出的算法的有效性,在本节中,我们将做两组实验:在第 1 个实验中,为了测试 SCTWSVM 的性能,我们将对 NDC 大数据集进行测试,并与 STWSVM 和 TWSVM 的测试结果进行比较;在第 2 个实验中,为了测试 WSCTWSVM 的性能,我们将对 10 个 UCI 数据集进行测试,测试结果将与 SCTWSVM,STWSVM 和 TWSVM 进行比较.所有实验都在 Intel(R) Core(TM)2 Duo CPU E4500,2G 内存和 MATLAB7.11.0 的环境中进行.

3.1 NDC大数据集上的实验

为了测试 SCTWSVM 的分类性能,下面我们对大数据集 NDC 进行实验.NDC 数据集是由 David Musicant's NDC 数据产生器[19]产生的.表 1 描述了 NDC 数据集的内容.在这个实验中,我们设定 SCTWSVM,STWSVM 和 TWSVM 的惩罚参数 $c_1=c_2=1$,高斯核函数的参数$\sigma=2^{-17}$,Newton-Armijo 算法结束时,下降方向的模为ε_1=1.0E−3,CHKS 函数的参数ε=1.0E−5.表 2 是线性 SCTWSVM,STWSVM 和 TWSVM 处理几个 NDC 数据集的训练时间、迭代步数、训练正确率和测试正确率的比较结果.表 3 显示的是非线性 SCTWSVM,STWSVM 和 TWSVM 对几个 NDC 数据集的测试结果.

Table 1 NDC datasets

表 1 NDC 数据集

数据集	训练集	测试集	维度
NDC-500	500	50	32
NDC-700	700	70	32
NDC-900	900	90	32
NDC-1k	1 000	100	32
NDC-2k	2 000	200	32
NDC-3k	3 000	300	32
NDC-4k	4 000	400	32
NDC-5k	5 000	500	32
NDC-10k	10 000	1 000	32
NDC-1l	100 000	10 000	32
NDC-3l	300 000	30 000	32
NDC-5l	500 000	50 000	32
NDC-1m	1 000 000	100 000	32
NDC1	5 000	5 000	100
NDC2	5 000	5 000	1 000

Table 2 Comparison for linear algorithms on NDC datasets

表 2 线性算法对 NDC 数据集的测试结果对比

线性算法		数据集								
		NDC-3k	NDC-4k	NDC-5k	NDC-10k	NDC-1l	NDC-3l	NDC-5l	NDC1	NDC2
SCTWSVM	训练正确率(%)	82.14	79.98	82.12	86.49	87.62	82.12	79.65	89.68	90.67
	测试正确率(%)	77.85	74.73	80.27	83.29	86.89	78.56	78.08	86.31	86.95
	时间(s)	0.465	0.607	0.734	1.178	0.996	2.899	5.131 2	1.152	30.364
	迭代步数	7	6	7	8	7	10	11	7	8
STWSVM	训练正确率(%)	81.05	79.87	79.96	84.48	86.46	79.54	78.84	86.72	87.45
	测试正确率(%)	77.32	74.65	79.20	82.24	86.25	78.76	77.12	84.52	83.36
	时间(s)	0.578	0.666	0.761	1.239	1.014	3.103	5.350 5	1.574	33.585
	迭代步数	9	9	8	11	9	13	13	9	11
TWSVM	训练正确率(%)	82.96	79.82	82.15	87.45	–	–	–	–	–
	测试正确率(%)	77.68	74.75	79.25	85.35	–	–	–	–	–
	时间(s)	27.08	60.94	114.27	1 092.23	–	–	–	–	–

"–"表示训练时间过长,实验无法进行.

© 中国科学院软件研究所 http://www.jos.org.cn

Table 3 Comparison for nonlinear algorithms on NDC datasets

表 3 非线性算法对 NDC 数据集的测试结果对比

非线性算法		数据集					
		NDC-500	NDC-700	NDC-900	NDC-1k	NDC-2k	NDC-3k
SCTWSVM	训练正确率(%)	100.00	99.28	99.67	98.85	100.00	100.00
	测试正确率(%)	82.19	84.29	81.43	85.36	88.34	90.36
	时间(s)	0.562 4	1.311 5	2.501 6	3.501 5	21.09	66.209
	迭代步数	6	8	9	11	7	8
STWSVM	训练正确率(%)	100.00	99.25	99.46	98.44	100.00	100.00
	测试正确率(%)	80.17	83.17	80.34	84.14	87.24	90.13
	时间(s)	0.568 5	1.312 6	2.532 4	3.519 4	21.25	66.612
	迭代步数	7	9	11	13	9	11
TWSVM	训练正确率(%)	99.25	99.27	99.56	98.89	99.69	99.57
	测试正确率(%)	82.21	84.29	80.58	84.56	88.29	90.45
	时间(s)	0.785	1.736 5	3.467 2	4.117 5	25.890 9	85.448

从表 2 中我们可以看出:当训练样本达到 100 000 时,采用 TWSVM 算法已经没有意义,训练时间过长,甚至出现死机现象;然而,SCTWSVM 和 STWSVM 在训练样本达到 500 000 时仍然可以在较短的时间内得到比较令人满意的分类精度,这表明了光滑孪生支持向量机在处理大数据集时具有优势.NDC1 和 NDC2 是两个高维的大数据集,从表 2 的实验结果也可以看出,SCTWSVM 和 STWSVM 处理高维数据集同样是有效的.并且,相对于 STWSVM,SCTWSVM 所需的迭代步数比较少,可以在更短的时间内获得更优的分类精度,这说明 SCTWSVM 的学习性能优于 STWSVM 的学习性能.从表 3 中我们也可以看出:SCTWSVM 的分类精度略高于 STWSVM 的分类精度,并且已经相当接近于 TWSVM 的精度.对于少数数据集,SCTWSVM 的分类精度甚至高于 TWSVM 的分类精度,而 SCTWSVM 所用的 CPU 时间是最少的.实验结果充分说明了 SCTWSVM 的学习能力比 STWSVM 要强,特别适用于处理大数据集.

3.2 UCI数据集上的实验

为了测试 WSCTWSVM 的性能,下面我们对 10 个 UCI 数据集进行实验.在这个实验中,非线性算法采用的核函数为高斯核函数,WSCTWSVM,SCTWSVM,STWSVM 和 TWSVM 的参数 c_1 和 c_2 以及核函数的参数 σ 都是采用网格划分方法进行确定,搜索范围为$[2^{-7},2^{12}]$.Newton-Armijo 算法结束时,下降方向的模为ε_1=1.0E−3,CHKS 函数的参数ε=1.0E−5.表 4 是线性 WSCTWSVM,SCTWSVM,STWSVM 和 TWSVM 处理 UCI 数据集时测试正确率、时间以及几种算法之间在置信水平为 0.05 下的配对 t 检验的比较结果.表 5 显示的是非线性 WSCTWSVM,SCTWSVM,STWSVM 和 TWSVM 对 UCI 数据集的测试结果.

Table 4 Comparison of classification accuracy for linear algorithms on UCI dataset

表 4 线性算法对 UCI 数据集的测试结果对比

线性算法		数据集						
		Hepatitis (155×19)	Housing (506×13)	Wdbc (596×30)	Glass6 (214×10)	Votes (435×16)	Pima (768×8)	Spect (267×22)
WSCTWSVM	测试正确率(%)	78.35±4.31	86.25±2.39	97.10±6.32	97.52±4.56	97.50±1.23	77.95±2.31	83.85±8.36
	时间(s)	0.212	3.521	9.845	0.327	1.282	16.51	1.124
SCTWSVM	测试正确率(%)	78.23±2.15	84.96±3.87	96.89±4.31	96.70±6.05	95.96±4.24	75.52±3.37	82.50±9.85
	时间(s)	0.206	3.378	8.525	0.307	1.115	15.97	0.985
	P-value	0.865	0.047	0.326	0.358	0.039	0.042	0.085
STWSVM	测试正确率(%)	78.08±2.16	84.42±4.53	96.22±6.67	96.55±2.40	95.25±2.24	75.40±9.68	81.98±2.17
	时间(s)	0.215	3.542	10.284	0.354	1.286	17.82	1.254
	P-value	0.474	0.038	0.129	0.295	0.034	0.036	0.052
TWSVM	测试正确率(%)	78.23±2.74	84.89±2.86	96.81±2.54	96.21±2.72	95.63±2.74	75.48±4.67	81.98±2.17
	时间(s)	0.798	10.122	25.642	0.9248	3.540	41.92	5.874
	P-value	0.887	0.044	0.285	0.185	0.035	0.039	0.052

© 中国科学院软件研究所 http://www.jos.org.cn

Table 5 Comparison of classification accuracy for nonlinear algorithms on UCI dataset

表 5 非线性算法对 UCI 数据集的测试结果对比

非线性算法		数据集			
		Wdbc (596×30)	Ionosphere (351×34)	Heart-statlog (270×14)	Hepatitis (155×19)
WSCTWSVM	测试正确率(%)	77.92±6.56	94.01±1.76	81.23±8.67	79.88±10.16
	时间(s)	0.518	0.597	0.125	0.142
SCTWSVM	测试正确率(%)	77.89±7.68	92.84±2.09	79.89±9.17	78.85±8.99
	时间(s)	0.498	0.526	0.114	0.129
	P-value	0.874	0.048	0.051	0.147
STWSVM	测试正确率(%)	77.27±8.62	92.68±2.74	79.26±6.34	78.32±9.18
	时间(s)	0.524	0.595	0.128	0.147
	P-value	0.652	0.042	0.045	0.098
TWSVM	测试正确率(%)	77.81±6.35	92.72±6.78	79.35±12.56	78.54±10.17
	时间(s)	0.862	1.181	0.478	0.352
	P-value	0.798	0.039	0.048	0.122

从表 4 和表 5 我们可以看出,在 0.05 的置信水平下,配对 t 检验计算出的 p 值表明:对于一部分数据集,其 P 值<0.05,表示该算法的分类精度和 WSCTWSVM 的分类精度有显著性差异.虽然 WSCTWSVM 所耗费的时间略多于 SCTWSVM,但少于 STWSVM.这一结果表明,根据样本点的位置为 SCTWSVM 的每一个样本赋予不同的重要性,可以降低异常样本对分类超平面拓扑结构的影响,进而可以提高算法的泛化能力.

4 结束语

光滑孪生支持向量机(STWSVM)的分类精度比较低,为了进一步地提高其学习性能,采用 CHKS 函数逼近孪生支持向量回归机的不可微项,并用 Newton-Armijo 算法求解相应的模型,提出了一种光滑孪生支持向量机模型——光滑 CHKS 孪生支持向量机(SCTWSVM).和 STWSVM 一样,SCTWSVM 对异常样本点比较敏感.为了减弱异常样本点对分类超平面拓扑结构的影响,通过为每个训练样本赋予不同的样本重要性,提出了加权 SCTWSVM(WSCTWSVM).我们不仅从理论上证明了 SCTWSVM 具有严凸性和满足任意阶光滑的性能,而且在 DNC 大数据集和 UCI 数据集上的实验表明了 SCTWSVM 比 STWSVM 具有更好的学习性能.实验结果也表明了 WSCTWSVM 的有效性和可行性.鉴于 SCTWSVM 和 WSCTWSVM 的良好性能,把它们应用到其他领域是下一步的工作.

致谢 在此,我们向对本文的工作给予支持和建议的同行表示感谢.

References:

[1] Zeng ZQ, Gao J. Simplified support vector machine based on reduced vector set method. Ruan Jian Xue Bao/Journal of Software, 2007,18(11):2719−2727 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/18/2719.htm [doi: 10.1360/jos182719]

[2] Ding SF, Qi BJ. Research of granular support vector machine. Artificial Intelligence Review, 2012,38(1):1−7. [doi: 10.1007/s10462-011-9235-9]

[3] Shi FQ, Xu J. Emotional cellular-based multi-class fuzzy support vector machines on product's KANSEI extraction. Applied Mathematics & Information Sciences, 2012,6(1):41−49.

[4] Ding SF, Jin FX, Zhao XW. Modern Data Analysis and Information Pattern Recognition. Beijing: Science Press, 2013. 35−38 (in Chinese).

[5] Ding SF, Xu L, Su CY, Jin FX. An optimizing method of RBF neural network based on genetic algorithm. Neural Computing and Applications, 2012,21(2):333−336. [doi: 10.1007/s00521-011-0702-7]

[6] Ding SF, Jia WK, Su CY, Zhang LW, Liu LL. Research on neural network algorithm based on fact or analysis and cluster analysis. Neural Computing and Applications, 2011,20(2):297−302. [doi: 10.1007/s00521-010-0416-2]

[7] Cortes C, Vapnik VN. Support vector networks. Machine Learning, 1995,20:273−297. [doi: 10.1007/BF00994018]

© 中国科学院软件研究所 http://www.jos.org.cn

[8] Osuna E, Freund R. An improved training algorithm for support vector machines. In: Proc. of the 1997 IEEE Workshop on Neural Networks for Signal Processing. New York: IEEE Press, 1997. 276−285. [doi: 10.1109/NNSP.1997.622408]

[9] Platt JC. Using analytic QP and sparseness to speed training of support vector machines. In: Kearns M, Solla S, Cohn D, eds. Advances in Neural Information Processing Systems 11. Cambridge: MIT Press, 1999. 557−563.

[10] Fung G, Mangasarian OL. Proximal support vector machine classifiers. In: Proc. of the 7th ACM SIFKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: IEEE Press, 2001. 77−86. [doi: 10.1145/502512.502527]

[11] Mangasarian OL, Wild EW. Multi-Surface proximal support vector machine classification via generalized eigenvalues. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(1):69−74. [doi: 10.1109/TPAMI.2006.17]

[12] Jayadeva KR, Suresh C. Twin support vector machines for pattern classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(5):905−910. [doi: 10.1109/TPAMI.2007.1068]

[13] Cong HH, Yang CF, Pu XR. Efficient speaker recognition based on multi-class twin support vector machines and GMMs. In: Proc. of the 2008 IEEE Conf. on Robotics, Automation and Mechatronics. Piscataway: IEEE Press, 2008. 348−352. [doi: 10.1109/RAMECH.2008.4681433]

[14] Zhang XS, Gao XB, Wang Y. Twin support tensor machine for MCS detection. Journal of Electronics (China), 2009,26(3): 318−325. [doi: 10.1007/s11767-007-0211-0]

[15] Zhang XS, Gao XB. MCs detection approach using bagging and boosting based twin support vector machine. In: Proc. of the 2009 IEEE Int'l Conf. on Systems, Man, and Cybernetics. Piscataway: IEEE Press, 2009. 5000−5005. [doi: 10.1109/ICSMC.2009.5346375]

[16] Kumar MA, Gopal M. Application of smoothing technique on twin support vector machines. Pattern Recognition Letters, 2008, 29(13):1842−1848. [doi: 10.1016/j.patrec.2008.05.016]

[17] Wu Q. Research on learning algorithms for support vector machines based on optimization theory [Ph.D. Thesis]. Xi'an: Xidian University in Candidacy, 2009 (in Chinese).

[18] Yuan YX, Sun WY. Optimization Theory and Methods. Beijing: Science Press, 1997. 45−49 (in Chinese).

[19] Kumar MA, Gopal M. Least squares twin support vector machines for pattern classification. Expert Systems with Applications, 2009,36(4):7535−7543. [doi: 10.1016/j.eswa.2008.09.066]

附中文参考文献:

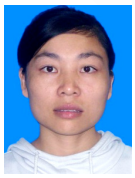
[1] 曾志强,高济.基于向量集约简的精简支持向量机.软件学报,2007,18(11):2719−2727. http://www.jos.org.cn/1000-9825/18/2719.htm [doi: 10.1360/jos182719]

[4] 丁世飞,靳奉祥,赵相伟.现代数据分析与信息模式识别.北京:科学出版社,2013.35−38.

[17] 吴青.基于最优化理论的支持向量机学习算法研究[博士学位论文].西安:西安电子科技大学,2009.

[18] 袁亚湘,孙文瑜.最优化理论与方法.北京:科学出版社,1997.45−49.

丁世飞(1963－),男,山东青岛人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为模式识别与人工智能,机器学习,数据挖掘.
E-mail: dingsf@cumt.edu.cn



黄华娟(1984－),女,博士生,主要研究领域为数据挖掘,模式识别.
E-mail: hhj-025@163.com

史忠植(1941－),男,研究员,博士生导师,CCF 高级会员,主要研究领域为智能科学,人工智能,机器学习.
E-mail: shizz@ics.ict.ac.cn

© 中国科学院软件研究所 http://www.jos.org.cn