

改进的投影孪生支持向量机

陈素根^{1,2}, 吴小俊¹

(1. 江南大学物联网工程学院, 江苏无锡 214122; 2. 安庆师范大学数学与计算科学学院, 安徽安庆 246133)

摘 要: 针对投影孪生支持向量机 (Projection Twin Support Vector Machine, PTSVM) 在训练和求解过程中存在的问题, 提出了一类改进的投影孪生支持向量机 (Improved PTSVM), 简称为 IPTSVM. 该文首先构造了改进的线性投影孪生支持向量机, 然后利用核技巧轻松将其推广到了非线性形式. 本文的主要贡献有: (1) 提出了投影孪生支持向量机的新模型, 克服了原始 PTSVM 在训练之前需要求解两个逆矩阵的问题; (2) 继承了传统 SVM (Support Vector Machine) 的精髓, 利用核技巧直接将线性 IPTSVM 推广到非线性形式; (3) 引入了一个新的参数, 可以调节模型的性能, 提高了 IPTSVM 的分类精度. 实验结果表明, 与 PTSVM 算法相比较, IPTSVM 不仅提高了分类精度, 而且克服了 PTSVM 的一些不足.

关键词: 支持向量机; 非平行平面支持向量机; 投影孪生支持向量机; 模式分类

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)02-0408-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.02.020

Improved Projection Twin Support Vector Machine

CHEN Su-gen^{1,2}, WU Xiao-jun¹

(1. School of Iot Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China;

2. School of Mathematics and Computational Science, Anqing Normal University, Anqing, Anhui 246133, China)

Abstract: An Improved Projection Twin Support Vector Machine (IPTSVM) is presented. The target of the proposed IPTSVM is to deal with a set of problems in the training and solving steps of PTSVM. We first propose a linear IPTSVM for binary classification. Then we extend it to the corresponding nonlinear version using kernel tricks. The paper has three main contributions to the community: (1) A new PTSVM-based method is proposed, in which we do not have to compute the inverse of a large matrix before the training step. (2) We design the nonlinear IPTSVM that is obtained by using kernel tricks. (3) A new parameter is introduced, which can adjust the performance of the model and improve the classification accuracy of IPTSVM. Experimental results obtained from several datasets demonstrate that, compared with PTSVM, IPTSVM not only improves the classification accuracy but also overcomes some deficiencies to a certain extent.

Key words: support vector machine; nonparallel hyperplane support vector machine; projection twin support vector machine; pattern classification

1 引言

支持向量机 (SVM) 是经典的分类算法之一^[1], 因为它具有坚实的理论基础和良好的泛化性能而得到广泛应用^[2,3]. 传统的 SVM 算法在解决小样本、非线性和高维模式问题中表现出了许多优势, 但由于在训练过程可能会求解大规模的逆矩阵问题, 就会表现出训练速度慢和效率低下等问题. 为了解决这些问题, 一方面, 诸多学者对如何设计高效的求解算法进行了深入研

究, 取得了许多优秀研究成果, 如 Chunking 算法^[4]、分解算法^[5]和 SMO (Sequential Minimal Optimization (序列最小优化算法)) 算法^[6]等; 另一方面, 构建新型 SVM 算法也逐渐引起了大家的关注, 非平行平面支持向量机算法就是代表性成果之一. 最早开始研究非平行平面支持向量机算法的是 Mangasarian 和 Wild, 2006 年, 他们提出了广义特征值中心支持向量机算法 (GEPSVM, Generalized Eigenvalues Proximal Support Vector Machine)^[7]来处理两类分类问题, 对每一类训练样本构造

一个分类超平面,将问题转化为求解两个规模较小的广义特征值问题,开辟了新的分类和决策思路. 2007 年, Khemchandani 等人在 GEPSVM 算法的基础上提出了孪生支持向量机算法 (Twin Support Vector Machine, TWSVM)^[8], TWSVM 将问题转化为求解两个规模较小的二次规划 (QP) 问题, 训练速度相当于传统 SVM 的四分之一. 2011 年, 陈小波等人提出了投影孪生支持向量机算法 (PTSVM)^[9], 寻找两个最优的投影方向, 使得本类样本投影之后尽可能聚集在该类的投影中心周围, 同时它类样本投影之后尽可能远离该投影中心. 于是, 诸多学者又纷纷提出了各种改进的投影孪生支持向量机算法, 如 RPTSVM (A Regularization For The Projection Twin Support Vector Machine)^[10]、LSPTSVM (Least Squares Projection Twin Support Vector Machine)^[11,12]. 从此, 非平行平面支持向量机算法得到了广泛和深入的研究, 其中代表性的成果有: TBSVM (Twin Bounded Support Vector Machine)^[13], TPMSVM (Twin Parametric-Margin Support Vector Machine)^[14], NPSVM (Nonparallel Support Vector Machine)^[15], L_1 -NPSVM^[16] 和 NHSVM (Nonparallel Hyperplane Support Vector Machine)^[17] 等. 为了克服 TWSVM 和 TBSVM 存在需计算逆矩阵和对线性与非线性问题分别构造不同模型的缺陷, 田英杰等人通过引入 ε 不敏感损失函数构造了 NPSVM 和 L_1 -NPSVM 算法. 然而, NPSVM 和 L_1 -NPSVM 模型的规模比 TWSVM 和 TBSVM 要复杂. 另外, 丁世飞等人对非平行平面支持向量机算法进行了综述^[18].

虽然 PTSVM 算法是非常优秀的非平行平面支持向量机算法之一, 但它依然有一些不足之处. 首先, PTSVM 算法在训练之前必须计算两个矩阵的逆矩阵, 当矩阵规模巨大时会增加模型的计算量. 甚至在有些时候会出现矩阵奇异的情况, 往往会利用近似逆矩阵来代替, 导致解的不稳定性; 其次, 原始 PTSVM 算法没有给出非线性的形式. 另外, 在目前已有的非平行平面支持向量机算法中, 很多算法在构造非线性模型时采用构造两个与线性模型不同的优化问题, 没有继承传统 SVM 基于核技巧推广到非线性模型的精髓. 针对以上问题, 受 NPSVM 和 L_1 -NPSVM 模型的启发, 本文首先提出了改进的投影孪生支持向量机的线性模型, 简称为 IPTSVM (Improved PTSVM), 然后利用核技巧直接将其推广到非线性形式. 与 PTSVM 算法或大多数已有的非平行平面支持向量机算法相比较, IPTSVM 算法有如下优点: (1) 在训练模型之前无需求解两个逆矩阵, 回避了可能的奇异和解不稳定的问题; (2) 继承了传统 SVM 的精髓, 利用核技巧直接将线性模型推广到非线性模型, 无需对线性与非线性问题分别构造两个不同的优化模型; (3) 引入了一个新的参数, 可以更好地调节模型的

性能, 提高了 IPTSVM 算法的分类精度. 最后, 在人工数据集和标准的 UCI 分类数据集上验证了本文算法的有效性.

2 投影孪生支持向量机

考虑二类分类问题, 假设给定的训练集为

$$T = \{(\mathbf{x}_1, +1), (\mathbf{x}_2, +1), \dots, (\mathbf{x}_p, +1), (\mathbf{x}_{p+1}, -1), (\mathbf{x}_{p+2}, -1), \dots, (\mathbf{x}_{p+q}, -1)\} \quad (1)$$

其中: $\mathbf{x}_i \in \mathbf{R}^n, i = 1, 2, \dots, p+q$. 对于线性分类问题, PTSVM^[9] 寻找两个最优的投影方向 $\mathbf{w}_1$ 和 $\mathbf{w}_2$, 使得本类训练样本投影后聚集在本类投影中心的周围, 同时它类样本投影后尽可能分散. 具体问题可归结为求解以下两个二次规划问题:

$$\begin{aligned} \min_{\mathbf{w}_1, \xi} & \frac{1}{2} \sum_{i=1}^p (\mathbf{w}_1^T \cdot \mathbf{x}_i - \mathbf{w}_1^T \cdot \frac{1}{p} \sum_{j=1}^p \mathbf{x}_j)^2 + C_1 \sum_{k=p+1}^{p+q} \xi_k \\ \text{s. t. } & \mathbf{w}_1^T \cdot \mathbf{x}_k - \mathbf{w}_1^T \cdot \frac{1}{p} \sum_{j=1}^p \mathbf{x}_j + \xi_k \geq 1, \\ & k = p+1, p+2, \dots, p+q \\ & \xi_k \geq 0, k = p+1, p+2, \dots, p+q \end{aligned} \quad (2)$$

$$\begin{aligned} \min_{\mathbf{w}_2, \xi} & \frac{1}{2} \sum_{i=p+1}^{p+q} (\mathbf{w}_2^T \cdot \mathbf{x}_i - \mathbf{w}_2^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} \mathbf{x}_j)^2 + C_2 \sum_{k=1}^p \xi_k \\ \text{s. t. } & \mathbf{w}_2^T \cdot \mathbf{x}_k - \mathbf{w}_2^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} \mathbf{x}_j + \xi_k \geq 1, k = 1, 2, \dots, p \\ & \xi_k \geq 0, k = 1, 2, \dots, p \end{aligned} \quad (3)$$

其中: $C_1 > 0, C_2 > 0$ 是惩罚参数, $\xi_+ = (\xi_1, \xi_2, \dots, \xi_p)^T$, $\xi_- = (\xi_{p+1}, \xi_{p+2}, \dots, \xi_{p+q})^T$, ξ_k 是非负松弛变量.

为了求解优化问题式 (2) 和式 (3), 作者通过引入拉格朗日函数并结合 KKT (Karush-Kuhn-Tucker) 条件, 将优化问题转化为其对偶形式:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha^T (\mathbf{B} - \frac{1}{p} \mathbf{e}_2 \mathbf{e}_1^T \mathbf{A}) \mathbf{S}_1^{-1} (\mathbf{B} - \frac{1}{p} \mathbf{e}_2 \mathbf{e}_1^T \mathbf{A})^T \alpha - \mathbf{e}_2^T \alpha \\ \text{s. t. } & 0 \leq \alpha \leq C_1 \mathbf{e}_2 \end{aligned} \quad (4)$$

$$\begin{aligned} \min_{\beta} & \frac{1}{2} \beta^T (\mathbf{A} - \frac{1}{q} \mathbf{e}_1 \mathbf{e}_2^T \mathbf{B}) \mathbf{S}_2^{-1} (\mathbf{A} - \frac{1}{q} \mathbf{e}_1 \mathbf{e}_2^T \mathbf{B})^T \beta - \mathbf{e}_1^T \beta \\ \text{s. t. } & 0 \leq \beta \leq C_2 \mathbf{e}_1 \end{aligned} \quad (5)$$

其中: α 和 β 是拉格朗日乘子, $\mathbf{e}_1 = (1, 1, \dots, 1)^T \in \mathbf{R}^p$, $\mathbf{e}_2 = (1, 1, \dots, 1)^T \in \mathbf{R}^q$, $\mathbf{A}$ 和 $\mathbf{B}$ 分别表示正类训练样本和负类训练样本的集合, $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 分别表示类内协方差矩阵, 且:

$$\mathbf{S}_1 = \sum_{i=1}^p (\mathbf{x}_i - \frac{1}{p} \sum_{j=1}^p \mathbf{x}_j) (\mathbf{x}_i - \frac{1}{p} \sum_{j=1}^p \mathbf{x}_j)^T \text{ 和}$$

$$\mathbf{S}_2 = \sum_{i=p+1}^{p+q} (\mathbf{x}_i - \frac{1}{q} \sum_{j=p+1}^{p+q} \mathbf{x}_j) (\mathbf{x}_i - \frac{1}{q} \sum_{j=p+1}^{p+q} \mathbf{x}_j)^T$$

通过求解对偶问题式 (4) 和式 (5), 可以最终获得最优

的投影方向 w_1 和 w_2 . 对于新的测试样本 $x \in R^n$, 根据分类规则式(6)将测试样本进行分类.

$$\text{label}(x) = \underset{i=1,2}{\operatorname{argmin}} \{d_i\} \quad (6)$$

$$\text{其中: } d_1 = \left| w_1^T \cdot x - w_1^T \cdot \frac{1}{p} \sum_{j=1}^p x_j \right|,$$

$$d_2 = \left| w_2^T \cdot x - w_2^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} x_j \right|.$$

3 改进的投影孪生支持向量机

PTSVM 是一类重要的非平行平面支持向量机, 然而, PTSVM 依然存在引言中所指出的一些问题. 为此, 本文提出一类改进的投影孪生支持向量机, 简称为 IPTSVM.

3.1 线性 IPTSVM

对于给定的训练样本集式(1), 线性 IPTSVM 模型求解如下两个优化问题:

$$\begin{aligned} \min_{w, \xi, \eta, \eta^*} & \frac{1}{2} \|w_+\|^2 + C_1 \sum_{i=1}^p (\eta_i + \eta_i^*) + C_2 \sum_{k=p+1}^{p+q} \xi_k \\ \text{s. t. } & w_+^T \cdot x_i - w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j \leq \varepsilon + \eta_i, i = 1, 2, \dots, p \\ & - (w_+^T \cdot x_i - w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j) \leq \varepsilon + \eta_i^*, i = 1, 2, \dots, p \\ & w_+^T \cdot x_k - w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j \geq 1 - \xi_k, k = p+1, p+2, \dots, p+q \\ & \eta_i, \eta_i^* \geq 0, \xi_k \geq 0, i = 1, 2, \dots, p; \\ & k = p+1, p+2, \dots, p+q \end{aligned} \quad (7)$$

$$\begin{aligned} \min_{w, \xi, \eta, \eta^*} & \frac{1}{2} \|w_-\|^2 + C_3 \sum_{i=1}^p (\eta_i + \eta_i^*) + C_4 \sum_{k=1}^p \xi_k \\ \text{s. t. } & w_-^T \cdot x_i - w_-^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} x_j \leq \varepsilon + \eta_i, \\ & i = p+1, p+2, \dots, p+q \\ & - (w_-^T \cdot x_i - w_-^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} x_j) \leq \varepsilon + \eta_i^*, \\ & i = p+1, p+2, \dots, p+q \\ & - (w_-^T \cdot x_k - w_-^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} x_j) \geq 1 - \xi_k, k = 1, 2, \dots, p \\ & \eta_i, \eta_i^* \geq 0, \xi_k \geq 0, i = p+1, p+2, \dots, p+q; k = 1, 2, \dots, p \end{aligned} \quad (8)$$

其中: $C_i > 0, i = 1, 2, 3, 4$ 是惩罚参数, $\xi_+ = (\xi_1, \xi_2, \dots, \xi_p)$, $\eta_+ = (\eta_1, \eta_2, \dots, \eta_p)$, $\eta_+^* = (\eta_1^*, \eta_2^*, \dots, \eta_p^*)$, $\xi_- = (\xi_{p+1}, \xi_{p+2}, \dots, \xi_{p+q})$, $\eta_- = (\eta_{p+1}, \eta_{p+2}, \dots, \eta_{p+q})$, $\eta_-^* = (\eta_{p+1}^*, \eta_{p+2}^*, \dots, \eta_{p+q}^*)$, ξ_k, η_i, η_i^* 是非负松弛变量, ε 是一个新引入正常数.

为了求解模型式(7)和式(8), 我们引入拉格朗日函数将问题转化为对偶问题. 为此, 对于模型式(7), 构造拉格朗日函数如下:

$$\begin{aligned} L(w_+, \eta_+, \eta_+^*, \xi_-, \alpha_+, \alpha_+^*, \beta_-, \gamma_+, \gamma_+^*, \lambda_-) \\ = \frac{1}{2} \|w_+\|^2 + C_1 \sum_{i=1}^p (\eta_i + \eta_i^*) + C_2 \sum_{k=p+1}^{p+q} \xi_k \\ + \sum_{i=1}^p \alpha_i (w_+^T \cdot x_i - w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j - \varepsilon - \eta_i) \\ + \sum_{i=1}^p \alpha_i^* (-(w_+^T \cdot x_i - w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j) - \varepsilon - \eta_i^*) \\ + \sum_{k=p+1}^{p+q} \beta_k (1 - \xi_k - w_+^T \cdot x_k + w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j) \\ - \sum_{i=1}^p \gamma_i \eta_i - \sum_{i=1}^p \gamma_i^* \eta_i^* - \sum_{k=p+1}^{p+q} \lambda_k \xi_k \end{aligned} \quad (9)$$

根据 KKT 条件, 可得:

$$\begin{aligned} \frac{\partial L}{\partial w_+} &= w_+ + \sum_{i=1}^p (\alpha_i - \alpha_i^*) (x_i - \frac{1}{p} \sum_{j=1}^p x_j) \\ &\quad - \sum_{k=p+1}^{p+q} \beta_k (x_k - \frac{1}{p} \sum_{j=1}^p x_j) = 0 \end{aligned} \quad (10)$$

$$\frac{\partial L}{\partial \eta_+} = C_1 e_+ - \alpha_+ - \gamma_+ = 0 \quad (11)$$

$$\frac{\partial L}{\partial \eta_+^*} = C_1 e_+ - \alpha_+^* - \gamma_+^* = 0 \quad (12)$$

$$\frac{\partial L}{\partial \xi_-} = C_2 e_- - \beta_- - \lambda_- = 0 \quad (13)$$

其中: $e_+ = (1, 1, \dots, 1)^T \in R^p$, $e_- = (1, 1, \dots, 1)^T \in R^q$, $\alpha_+ = (\alpha_1, \alpha_2, \dots, \alpha_p)^T \in R^p$, $\alpha_+^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_p^*)^T \in R^p$, $\beta_- = (\beta_{p+1}, \beta_{p+2}, \dots, \beta_{p+q})^T \in R^q$, $\gamma_+ = (\gamma_1, \gamma_2, \dots, \gamma_p)^T \in R^p$, $\gamma_+^* = (\gamma_1^*, \gamma_2^*, \dots, \gamma_p^*)^T \in R^p$, $\lambda_- = (\lambda_{p+1}, \lambda_{p+2}, \dots, \lambda_{p+q})^T \in R^q$.

因为 $\gamma_+ \geq 0, \gamma_+^* \geq 0, \lambda_- \geq 0$, 根据式(11), 式(12)和式(13), 我们可以得到:

$$0 \leq \alpha_+ \leq C_1 e_+, 0 \leq \alpha_+^* \leq C_1 e_+, 0 \leq \beta_- \leq C_2 e_- \quad (14)$$

根据式(10), 我们可以得到:

$$\begin{aligned} w_+ &= \sum_{i=1}^p (\alpha_i^* - \alpha_i) (x_i - \frac{1}{p} \sum_{j=1}^p x_j) \\ &\quad + \sum_{k=p+1}^{p+q} \beta_k (x_k - \frac{1}{p} \sum_{j=1}^p x_j) \end{aligned} \quad (15)$$

将式(15)代入式(9), 结合式(11) ~ (13), 我们可以得到模型式(8)的对偶问题如下:

$$\begin{aligned} \min_{\alpha_+, \alpha_+^*, \beta_-} & \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) \\ & \cdot ((x_i - \frac{1}{p} \sum_{s=1}^p x_s) \cdot (x_j - \frac{1}{p} \sum_{s=1}^p x_s)) \\ & - \sum_{i=1}^p \sum_{j=p+1}^{p+q} (\alpha_i^* - \alpha_i) \beta_j ((x_i - \frac{1}{p} \sum_{s=1}^p x_s) \\ & \cdot (x_j - \frac{1}{p} \sum_{s=1}^p x_s)) + \sum_{j=p+1}^{p+q} \beta_j \\ & + \frac{1}{2} \sum_{i=p+1}^{p+q} \sum_{j=p+1}^{p+q} \beta_i \beta_j ((x_i - \frac{1}{p} \sum_{s=1}^p x_s) \end{aligned}$$

$$\begin{aligned} & \cdot (x_j - \frac{1}{p} \sum_{s=1}^p x_s) - \varepsilon \sum_{i=1}^p (\alpha_i^* + \alpha_i) \\ \text{s. t. } & 0 \leq \alpha_+ \leq C_1 e_+, 0 \leq \alpha_+^* \leq C_1 e_+, \\ & 0 \leq \beta_- \leq C_2 e_- \end{aligned} \quad (16)$$

为了表示方便,我们记:

$$A = (x_1, x_2, \dots, x_p)^T \in R^{p \times n}, B = (x_{p+1}, x_{p+2}, \dots, x_{p+q})^T \in R^{q \times n},$$

$$H = A - \frac{1}{p} e_+ e_+^T A \in R^{p \times n}, G = B - \frac{1}{p} e_- e_-^T B \in R^{q \times n},$$

$$\tilde{\pi} = (\alpha_+, \alpha_+^*, \beta_-) \in R^{(2p+q)},$$

$$\tilde{f} = (\varepsilon e_+, \varepsilon e_+, -e_-) \in R^{(2p+q)},$$

$$\tilde{C} = (C_1 e_+, C_1 e_+, C_2 e_-) \in R^{(2p+q)},$$

$$\tilde{F} = \begin{pmatrix} -H \\ H \\ G \end{pmatrix} \in R^{(2p+q) \times n},$$

$$\tilde{Q} = \begin{pmatrix} HH^T & -HH^T & -HG^T \\ -HH^T & HH^T & HG^T \\ -GH^T & GH^T & GG^T \end{pmatrix} \in R^{(2p+q) \times (2p+q)}.$$

具体写出矩阵 $\tilde{Q}$ 的元素构成,如下:

$$\begin{aligned} HH^T &= (A - \frac{1}{p} e_+ e_+^T A) (A - \frac{1}{p} e_+ e_+^T A)^T \\ &= AA^T - \frac{1}{p} e_+ e_+^T AA^T - \frac{1}{p} AA^T e_+ e_+^T + \frac{1}{p^2} e_+ e_+^T AA^T e_+ e_+^T, \\ HG^T &= (A - \frac{1}{p} e_+ e_+^T A) (B - \frac{1}{p} e_- e_-^T B)^T \\ &= AB^T - \frac{1}{p} e_+ e_+^T AB^T - \frac{1}{p} AA^T e_+ e_-^T + \frac{1}{p^2} e_+ e_+^T AA^T e_- e_-^T, \\ GH^T &= (B - \frac{1}{p} e_- e_-^T B) (A - \frac{1}{p} e_+ e_+^T A)^T \\ &= BA^T - \frac{1}{p} e_- e_-^T BA^T - \frac{1}{p} BA^T e_+ e_+^T + \frac{1}{p^2} e_- e_-^T BA^T e_+ e_+^T, \\ GG^T &= (B - \frac{1}{p} e_- e_-^T B) (B - \frac{1}{p} e_- e_-^T B)^T \\ &= BB^T - \frac{1}{p} e_- e_-^T BB^T - \frac{1}{p} BB^T e_- e_-^T + \frac{1}{p^2} e_- e_-^T BB^T e_- e_-^T. \end{aligned}$$

于是,式(15)和式(16)分别转化为:

$$w_+ = H^T (-\alpha_+ + \alpha_+^*) + G^T \beta_- = \tilde{F}^T \tilde{\pi} \quad (17)$$

$$\begin{aligned} & \min_{\tilde{\pi}} \frac{1}{2} \tilde{\pi}^T \tilde{Q} \tilde{\pi} + \tilde{f}^T \tilde{\pi} \\ \text{s. t. } & 0 \leq \tilde{\pi} \leq \tilde{C} \end{aligned} \quad (18)$$

很明显,式(18)与式(4)或式(5)具有完全一样的形式,都可以利用 SOR((Successive Overrelaxation(超松弛求解算法))^[13,17]技巧来求解. 类似地,模型式(9)的求解及其对偶形式可以转化为:

$$w_- = R^T (-\alpha_- + \alpha_-^*) - S^T \beta_+ = \hat{F}^T \hat{\pi} \quad (19)$$

$$\begin{aligned} & \min_{\hat{\pi}} \frac{1}{2} \hat{\pi}^T \hat{Q} \hat{\pi} + \hat{f}^T \hat{\pi} \\ \text{s. t. } & 0 \leq \hat{\pi} \leq \hat{C} \end{aligned} \quad (20)$$

$$\text{其中: } R = B - \frac{1}{q} e_- e_-^T B \in R^{q \times n}, S = A - \frac{1}{q} e_+ e_+^T A \in R^{p \times n},$$

$$\hat{\pi} = (\alpha_-, \alpha_-^*, \beta_+) \in R^{(2q+p)}, \hat{f} = (\varepsilon e_-, \varepsilon e_-, -e_+) \in R^{(2q+p)},$$

$$\hat{C} = (C_3 e_-, C_3 e_-, C_4 e_+) \in R^{(2q+p)},$$

$$\hat{F} = \begin{pmatrix} -R \\ R \\ -S \end{pmatrix} \in R^{(2q+p) \times n},$$

$$\hat{Q} = \begin{pmatrix} RR^T & -RR^T & RS^T \\ -RR^T & RR^T & -RS^T \\ SR^T & -SR^T & SS^T \end{pmatrix} \in R^{(2q+p) \times (2q+p)};$$

具体写出矩阵 $\hat{Q}$ 的元素构成,如下:

$$\begin{aligned} RR^T &= (B - \frac{1}{q} e_- e_-^T B) (B - \frac{1}{q} e_- e_-^T B)^T \\ &= BB^T - \frac{1}{q} e_- e_-^T BB^T - \frac{1}{q} BB^T e_- e_-^T + \frac{1}{q^2} e_- e_-^T BB^T e_- e_-^T, \\ RS^T &= (B - \frac{1}{q} e_- e_-^T B) (A - \frac{1}{q} e_+ e_+^T A)^T \\ &= BA^T - \frac{1}{q} e_- e_-^T BA^T - \frac{1}{q} BB^T e_- e_+^T + \frac{1}{q^2} e_- e_-^T BB^T e_+ e_+^T, \\ SR^T &= (A - \frac{1}{q} e_+ e_+^T A) (B - \frac{1}{q} e_- e_-^T B)^T \\ &= AB^T - \frac{1}{q} e_+ e_+^T AB^T - \frac{1}{q} AB^T e_- e_-^T + \frac{1}{q^2} e_+ e_+^T AB^T e_- e_-^T, \\ SS^T &= (A - \frac{1}{q} e_+ e_+^T A) (A - \frac{1}{q} e_+ e_+^T A)^T = AA^T - \frac{1}{q} e_+ e_+^T AA^T \\ &\quad - \frac{1}{q} AA^T e_+ e_+^T + \frac{1}{q^2} e_+ e_+^T AA^T e_+ e_+^T. \end{aligned}$$

经上述分析,我们发现通过求解出优化问题式(18)和式(20)之后,利用式(17)和式(19)式可以计算出两个最优的投影方向 w_+ 和 w_- . 对于新的测试样本 $x \in R^n$,利用式(21)对其进行分类.

$$\text{label}(x) = \arg \min_{i \in \{+, -\}} \{d_i\} \quad (21)$$

$$\text{其中: } d_+ = \left| w_+^T \cdot x - w_+^T \cdot \frac{1}{p} \sum_{j=1}^p x_j \right| = |\tilde{\pi}^T \cdot U_+|,$$

$$d_- = \left| w_-^T \cdot x - w_-^T \cdot \frac{1}{q} \sum_{j=p+1}^{p+q} x_j \right| = |\hat{\pi}^T \cdot U_-|,$$

$$U_+ = \begin{pmatrix} -Ax + \frac{1}{p} e_+ e_+^T Ax + \frac{1}{p} AA^T e_+ - \frac{1}{p^2} e_+ e_+^T AA^T e_+ \\ Ax - \frac{1}{p} e_+ e_+^T Ax - \frac{1}{p} AA^T e_+ + \frac{1}{p^2} e_+ e_+^T AA^T e_+ \\ Bx - \frac{1}{p} e_- e_-^T Bx - \frac{1}{p} BA^T e_+ + \frac{1}{p^2} e_- e_-^T BA^T e_+ \end{pmatrix};$$

$$U_- = \begin{pmatrix} -Bx + \frac{1}{q} e_- e_-^T Bx + \frac{1}{q} BB^T e_- - \frac{1}{q^2} e_- e_-^T BB^T e_- \\ Bx - \frac{1}{q} e_- e_-^T Bx - \frac{1}{q} BB^T e_- + \frac{1}{q^2} e_- e_-^T BB^T e_- \\ -Ax + \frac{1}{q} e_+ e_+^T Bx + \frac{1}{q} AB^T e_- - \frac{1}{q^2} e_+ e_+^T AB^T e_- \end{pmatrix}$$

综合上述分析,可得到定理 1.

定理 1 对于训练样本集式(1),线性 IPTSVM 模型式(7)和式(8)的对偶问题分别为模型式(18)和式(20),最优投影方向 $\mathbf{w}_+$ 和 $\mathbf{w}_-$ 分别由式(17)和式(19)计算确定. 对于新的测试样本 $\mathbf{x} \in R^n$,可以利用式(21)对其进行分类.

3.2 非线性 IPTSVM

观察优化模型式(18),式(20)和决策函数式(21),我们发现,这里面的运算仅仅涉及到向量之间的内积,容易利用核技巧推广到非线性情形. 于是,对于非线性分类问题,与 TWSVM^[8] 和 TBSVM^[13] 不同,我们不需要重新构造两个新的优化问题. 具体来说,我们引入核函数 $K(x, x') = (\Phi(x) \cdot \Phi(x'))$, 相应的变换为 $x^* = \Phi(x)$, 利用核技巧就可以直接将线性 IPTSVM 模型推广到非线性 IPTSVM 模型. 对于寻找最优投影方向 $\mathbf{w}_+$, 可以求解优化问题式(22),并得到最优解式(23).

$$\min_{\tilde{\pi}^*} \frac{1}{2} \tilde{\pi}^{*T} \tilde{Q}^* \tilde{\pi}^* + \tilde{f}^{*T} \tilde{\pi}^* \quad (22)$$

$$\text{s. t. } 0 \leq \tilde{\pi}^* \leq \tilde{C}^*$$

$$\mathbf{w}_+^* = \tilde{F}^{*T} \cdot \tilde{\pi}^* \quad (23)$$

$$\text{其中: } H^* = \Phi(A) - \frac{1}{p} \mathbf{e}_+ \mathbf{e}_+^T \Phi(A),$$

$$G^* = \Phi(B) - \frac{1}{p} \mathbf{e}_- \mathbf{e}_-^T \Phi(A),$$

$$\tilde{\pi}^* = (\alpha_+, \alpha_+, \beta_-) \in R^{(2p+q)},$$

$$\tilde{f}^* = (\varepsilon \mathbf{e}_+, \varepsilon \mathbf{e}_+, -\mathbf{e}_-) \in R^{(2p+q)},$$

$$\tilde{C}^* = (C_1 \mathbf{e}_+, C_1 \mathbf{e}_+, C_2 \mathbf{e}_-) \in R^{(2p+q)},$$

$$\tilde{F}^* = \begin{pmatrix} -H^* \\ H^* \\ G^* \end{pmatrix},$$

$$\tilde{Q}^* = \begin{pmatrix} H^* H^{*T} & -H^* H^{*T} & -H^* G^{*T} \\ -H^* H^{*T} & H^* H^{*T} & H^* G^{*T} \\ -G^* H^{*T} & G^* H^{*T} & G^* G^{*T} \end{pmatrix} \in R^{(2p+q) \times (2p+q)},$$

具体写出矩阵 $\tilde{Q}^*$ 的元素构成,如下:

$$H^* H^{*T} = K(A, A^T) - \frac{1}{p} \mathbf{e}_+ \mathbf{e}_+^T K(A, A^T) - \frac{1}{p} K(A, A^T) \mathbf{e}_+ \mathbf{e}_+^T + \frac{1}{p^2} \mathbf{e}_+ \mathbf{e}_+^T K(A, A^T) \mathbf{e}_+ \mathbf{e}_+^T,$$

$$H^* G^{*T} = K(A, B^T) - \frac{1}{p} \mathbf{e}_+ \mathbf{e}_+^T K(A, B^T) - \frac{1}{p} K(A, A^T) \mathbf{e}_+ \mathbf{e}_-^T + \frac{1}{p^2} \mathbf{e}_+ \mathbf{e}_+^T K(A, A^T) \mathbf{e}_+ \mathbf{e}_-^T,$$

$$G^* H^{*T} = K(B, A^T) - \frac{1}{p} \mathbf{e}_- \mathbf{e}_-^T K(A, A^T) - \frac{1}{p} K(B, A^T) \mathbf{e}_+ \mathbf{e}_+^T + \frac{1}{p^2} \mathbf{e}_- \mathbf{e}_-^T K(A, A^T) \mathbf{e}_+ \mathbf{e}_+^T,$$

$$G^* G^{*T} = K(B, B^T) - \frac{1}{p} \mathbf{e}_- \mathbf{e}_-^T K(A, B^T)$$

$$- \frac{1}{p} K(B, A^T) \mathbf{e}_+ \mathbf{e}_-^T + \frac{1}{p^2} \mathbf{e}_- \mathbf{e}_-^T K(A, A^T) \mathbf{e}_+ \mathbf{e}_-^T.$$

对于寻找最优投影方向 $\mathbf{w}_-$, 可以求解优化问题式(24),并得到最优解式(25).

$$\min_{\hat{\pi}^*} \frac{1}{2} \hat{\pi}^{*T} \hat{Q}^* \hat{\pi}^* + \hat{f}^{*T} \hat{\pi}^* \quad (24)$$

$$\text{s. t. } 0 \leq \hat{\pi}^* \leq \hat{C}^*$$

$$\mathbf{w}_-^* = \hat{F}^{*T} \cdot \hat{\pi}^* \quad (25)$$

其中:

$$R^* = \Phi(B) - \frac{1}{q} \mathbf{e}_- \mathbf{e}_-^T \Phi(B),$$

$$S^* = \Phi(A) - \frac{1}{q} \mathbf{e}_+ \mathbf{e}_+^T \Phi(B),$$

$$\hat{\pi}^* = (\alpha_-, \alpha_-, \beta_+) \in R^{(2q+p)},$$

$$\hat{f}^* = (\varepsilon \mathbf{e}_-, \varepsilon \mathbf{e}_-, -\mathbf{e}_+) \in R^{(2q+p)},$$

$$\hat{C}^* = (C_3 \mathbf{e}_-, C_3 \mathbf{e}_-, C_4 \mathbf{e}_+) \in R^{(2q+p)},$$

$$\hat{F}^* = \begin{pmatrix} -R^* \\ R^* \\ -S^* \end{pmatrix},$$

$$\hat{Q}^* = \begin{pmatrix} R^* R^{*T} & -R^* R^{*T} & R^* S^{*T} \\ -R^* R^{*T} & R^* R^{*T} & -R^* S^{*T} \\ S^* R^{*T} & -S^* R^{*T} & S^* S^{*T} \end{pmatrix} \in R^{(2q+p) \times (2q+p)},$$

具体写出矩阵 $\hat{Q}^*$ 的元素构成,如下:

$$R^* R^{*T} = K(B, B^T) - \frac{1}{q} \mathbf{e}_- \mathbf{e}_-^T K(B, B^T) - \frac{1}{q} K(B, B^T) \mathbf{e}_- \mathbf{e}_-^T + \frac{1}{q^2} \mathbf{e}_- \mathbf{e}_-^T K(B, B^T) \mathbf{e}_- \mathbf{e}_-^T,$$

$$R^* S^{*T} = K(B, A^T) - \frac{1}{q} \mathbf{e}_- \mathbf{e}_-^T K(B, A^T) - \frac{1}{q} K(B, B^T) \mathbf{e}_- \mathbf{e}_+^T + \frac{1}{q^2} \mathbf{e}_- \mathbf{e}_-^T K(B, B^T) \mathbf{e}_- \mathbf{e}_+^T,$$

$$S^* R^{*T} = K(A, B^T) - \frac{1}{q} \mathbf{e}_+ \mathbf{e}_+^T K(B, B^T) - \frac{1}{q} K(A, B^T) \mathbf{e}_- \mathbf{e}_-^T + \frac{1}{q^2} \mathbf{e}_+ \mathbf{e}_+^T K(B, B^T) \mathbf{e}_- \mathbf{e}_-^T,$$

$$S^* S^{*T} = K(A, A^T) - \frac{1}{q} \mathbf{e}_+ \mathbf{e}_+^T K(B, A^T) - \frac{1}{q} K(A, B^T) \mathbf{e}_- \mathbf{e}_+^T + \frac{1}{q^2} \mathbf{e}_+ \mathbf{e}_+^T K(B, B^T) \mathbf{e}_- \mathbf{e}_+^T.$$

于是,通过求解优化问题式(22)和式(24),并利用式(23)和式(25)分别可以计算出两个最优的投影方向 $\mathbf{w}_+$ 和 $\mathbf{w}_-$. 此时,对于新的测试样本 $\mathbf{x} \in R^n$,将其利用非线性映射映射到高维空间,得到 $\Phi(\mathbf{x})$, 利用式(26)对其进行分类.

$$\text{label}(\mathbf{x}) = \underset{i \in \{+, -\}}{\operatorname{argmin}} \{d_i^*\} \quad (26)$$

其中:

$$\begin{aligned}
 d_+^* &= \left| \mathbf{w}_+^{*\top} \cdot \left(\Phi(\mathbf{x}) - \frac{1}{p} \sum_{j=1}^p \Phi(\mathbf{x}_j) \right) \right| \\
 &= |\hat{\pi}^{*\top} \cdot \mathbf{U}_+^*|; d_-^* = \left| \mathbf{w}_-^{*\top} \cdot \left(\Phi(\mathbf{x}) - \frac{1}{q} \sum_{j=p+1}^{p+q} \Phi(\mathbf{x}_j) \right) \right| = |\hat{\pi}^{*\top} \cdot \mathbf{U}_-^*|, \\
 \mathbf{U}_+^* &= \begin{pmatrix} -K(\mathbf{A}, \mathbf{x}) + \frac{1}{p} \mathbf{e}_+ \mathbf{e}_+^\top K(\mathbf{A}, \mathbf{x}) + \frac{1}{p} K(\mathbf{A}, \mathbf{A}^\top) \mathbf{e}_+ - \frac{1}{p^2} \mathbf{e}_+ \mathbf{e}_+^\top K(\mathbf{A}, \mathbf{A}^\top) \mathbf{e}_+ \\ K(\mathbf{A}, \mathbf{x}) - \frac{1}{p} \mathbf{e}_+ \mathbf{e}_+^\top K(\mathbf{A}, \mathbf{x}) - \frac{1}{p} K(\mathbf{A}, \mathbf{A}^\top) \mathbf{e}_+ + \frac{1}{p^2} \mathbf{e}_+ \mathbf{e}_+^\top K(\mathbf{A}, \mathbf{A}^\top) \mathbf{e}_+ \\ K(\mathbf{B}, \mathbf{x}) - \frac{1}{p} \mathbf{e}_- \mathbf{e}_+^\top K(\mathbf{A}, \mathbf{x}) - \frac{1}{p} K(\mathbf{B}, \mathbf{A}^\top) \mathbf{e}_+ + \frac{1}{p^2} \mathbf{e}_- \mathbf{e}_+^\top K(\mathbf{A}, \mathbf{A}^\top) \mathbf{e}_+ \end{pmatrix}, \\
 \mathbf{U}_-^* &= \begin{pmatrix} -K(\mathbf{B}, \mathbf{x}) + \frac{1}{q} \mathbf{e}_- \mathbf{e}_-^\top K(\mathbf{B}, \mathbf{x}) + \frac{1}{q} K(\mathbf{B}, \mathbf{B}^\top) \mathbf{e}_- - \frac{1}{q^2} \mathbf{e}_- \mathbf{e}_-^\top K(\mathbf{B}, \mathbf{B}^\top) \mathbf{e}_- \\ K(\mathbf{B}, \mathbf{x}) - \frac{1}{q} \mathbf{e}_- \mathbf{e}_-^\top K(\mathbf{B}, \mathbf{x}) - \frac{1}{p} K(\mathbf{B}, \mathbf{B}^\top) \mathbf{e}_- + \frac{1}{q^2} \mathbf{e}_- \mathbf{e}_-^\top K(\mathbf{B}, \mathbf{B}^\top) \mathbf{e}_- \\ -K(\mathbf{A}, \mathbf{x}) + \frac{1}{q} \mathbf{e}_+ \mathbf{e}_-^\top K(\mathbf{B}, \mathbf{x}) + \frac{1}{q} K(\mathbf{A}, \mathbf{B}^\top) \mathbf{e}_- - \frac{1}{q^2} \mathbf{e}_+ \mathbf{e}_-^\top K(\mathbf{B}, \mathbf{B}^\top) \mathbf{e}_- \end{pmatrix}.
 \end{aligned}$$

综合上述分析,可得到定理 2.

定理 2 对于训练样本集式(1),选定合适的核函数 K ,非线性 IPTSVM 模型的对偶问题分别为模型式(22)和式(24),最优投影方向 $\mathbf{w}_+^*$ 和 $\mathbf{w}_-^*$ 分别由式(23)和式(25)计算确定. 对于新的测试样本 $\mathbf{x} \in R^n$,可以利用式(26)对其进行分类.

仔细分析非线性 IPTSVM 模型,我们可以发现,当选择核函数为线性核函数时,非线性 IPTSVM 模型式(22)和式(24)等价于线性 PTSVM 模型式(18)和式(20). 总结上述过程,我们给出 IPTSVM 算法如下:

算法 1 改进的投影孪生支持向量机算法(IPTSVM 算法)

输入:训练样本集及相关参数;输出:分类精度及均方差.

Step1 给定训练样本集式(1);

Step2 选定合适的核函数 K 和参数 $\varepsilon > 0$,通过 10-折交叉验证选择合适的参数 $C_i > 0, i = 1, 2, 3, 4$;

Step3 构造优化模型式(22)和式(24),利用式(23)和式(25)分别求解最优的投影方向 $\mathbf{w}_+^*$ 和 $\mathbf{w}_-^*$;

Step4 利用决策函数式(26)进行分类.

3.3 IPTSVM 算法优缺点分析

综上所述,针对两类分类问题,我们提出了 IPTSVM 算法. IPTSVM 算法作为 PTSVM 算法的一种改进算法,与经典的非平行平面支持向量机算法(如:TWSVM、TBSVM 等)相比较而言,它具有一定的优势,也具有一些不足. 具体来说,有如下几个方面:

(1) 原始 PTSVM^[9] 仅仅给出了线性 PTSVM 算法,并没有给出非线性模型. RPTSVM^[10] 通过引入正则项,实现了对 PTSVM 算法的改进,同时给出了非线性形式. 然而,在文献[10]中线性 RPTSVM 和非线性 RPTSVM 是通过

构造两个不同的优化问题来实现的,这样的问题同样出现在 TWSVM 和 TBSVM 中. 也就是说,当核函数选取为线性核时,非线性 TWSVM、TBSVM 和 RPTSVM 都不等价于它们的线性模型. 然而,我们所提出的 IPTSVM 算法当核函数选取为线性核时,非线性 IPTSVM 模型是与线性 IPTSVM 等价的,因为我们的模型只需要构造线性形式,然后可以利用核技巧轻松推广到非线性情形,这真正意义上继承了传统 SVM 的精髓;(2) 在上述非平行平面支持向量机算法中,它们的求解过程中很多时候都需要在训练之前计算两个逆矩阵,这些逆矩阵甚至可能会出现奇异的情形. 这将带来两个方面的问题:(a) 当出现矩阵奇异的情况下,往往通过增加小的扰动实现近似求解,但这会导致解的不稳定性;(b) 当训练样本集规模较大时,求解逆矩阵会消耗很长的时间. RPTSVM 算法通过引入正则项虽然克服了 PTSVM 算法中可能出现的第一个问题,但依然存在第二个问题. 然而,我们的 IPTSVM 算法,在求解过程中不要求解任何逆矩阵,不会出现上述两个问题;(3) 我们所提出的 IPTSVM 算法中引入了一个新的参数 $\varepsilon > 0$,实际上相当于利用 ε 不敏感损失函数代替原始 PTSVM 模型中的 L_2 损失函数. 实验发现,适当选取参数 ε 的取值,可以提高模型的分类精度. 当然,如何选取最优的参数 ε 仍是个需要进一步研究的问题;(4) 从算法计算复杂度的角度进行考虑,我们提出的 IPTSVM 算法的优势并不明显. 事实上,假设我们的训练样本集中有 l 个训练样本,而且 $p \approx q \approx 0.5l$. 在这种情形下,当利用 SOR 技术求解 TBSVM、PTSVM、RPTSVM 和 IPTSVM 算法时,我们的提出的 IPTSVM 的计算复杂度大致为: $2 \times$ 迭代次数 $\times O(1.5l)$. 然而, TBSVM、PTSVM 和 RPTSVM 算法的计算复杂度大致为: $2 \times O(l^3) + 2 \times$ 迭代次数 $\times O$

(0.5I), 其中 $O(I^3)$ 为求解逆矩阵所消耗的时间. 因为逆矩阵的计算可以在训练之前完成, 如果不考虑这个计算时间的话, 那么我们的算法就相对要慢一点; 如果将计算逆矩阵的时间考虑在内的话, 我们的算法与上述算法效率差不多. 由于 IPTSVM 模型式 (22) 和式 (24) 二次规划问题的规模比 TBSVM、PTSVM 和 RPTSVM 的要大, 导致求解相对耗时, 特别是对大规模数据集表现更为明显. 这也是我们算法一个不足之处, 所以寻求更有效的求解算法也是有待深入研究的问题之一.

4 实验结果与分析

为了验证 IPTSVM 算法的有效性, 分别在人工数据集 Ripley^[19] 和 7 个两类分类 UCI 数据集上^[20] 进行实验, 并分别与 TBSVM、PTSVM 和 RPTSVM 方法进行对比. 所有实验均以 Matlab R2013a 为实验环境, 以 Intel P4(3.40GHz) 处理器、4GB 内存的 PC 为硬件平台. 在实验过程中, 使用逐次超松弛技术 (SOR)^[13,17] 来求解各算法中的 QP (Quadratic Programming (二次规划)) 问题. 对于非线性情形, 选择 RBF (Radial Basis Function

(径向基函数)) 核 $K(x, y) = e^{-\mu \cdot \|x-y\|^2}$ 为核函数, 其中 μ 为核参数. 在我们的实验过程中, 均采用 10-折交叉验证方法为各算法选择最优参数. 除了 IPTSVM 算法中的参数 ε 在 $\{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$ 中选取以外, 其它的参数选取范围均为 $\{2^i | i = -8, -7, \dots, 7, 8\}$. 在所有参数选定之后, 每个实验重复进行 10 次, 记录平均实验结果和均方差.

对于人工数据集上的实验, 我们选择了 Ripley 数据集进行验证. Ripley 数据集含有 250 个训练样本和 1000 个测试样本, 在这个数据集上, 我们分别研究了 IPTSVM 算法的性能以及参数 ε 对 IPTSVM 算法性能的影响. 表 1 给出了非线性情形下 IPTSVM 算法与 TBSVM、PTSVM 和 RPTSVM 算法性能上的比较, 分别给出了训练精度和测试精度. 从表 1 中发现, 我们的 IPTSVM 算法取得了相对较好的训练精度和测试精度. 图 1 描述了参数 ε 对 IPTSVM 算法性能的调节作用. 图 1 中的曲线变化说明了参数 ε 对 IPTSVM 算法性能的调节作用, 因此, 选择合适的参数 ε 可以提升 IPTSVM 算法的性能. 对于 UCI 数据集上的实验, 我们选择了 Australian、House votes、

表 1 非线性 IPTSVM 与几种算法在 Ripley 数据集上性能比较

数据集	TBSVM	PTSVM	RPTSVM	IPTSVM
	训练精度 $\pm$ 方差 (%) 测试精度 (%)	训练精度 $\pm$ 方差 (%) 测试精度 (%)	训练精度 $\pm$ 方差 (%) 测试精度 (%)	训练精度 $\pm$ 方差 (%) 测试精度 (%)
Ripley	88.40 $\pm$ 5.80 85.10	88.40 $\pm$ 7.41 84.50	88.80 $\pm$ 4.13 85.20	89.20 $\pm$ 6.81 85.50

表 2 线性 IPTSVM 与几种算法在两类分类 UCI 数据集上性能比较

数据集	TBSVM	PTSVM	RPTSVM	IPTSVM
	精度 $\pm$ 方差 (%) 时间 (s) / p 值	精度 $\pm$ 方差 (%) 时间 (s) / p 值	精度 $\pm$ 方差 (%) 时间 (s) / p 值	精度 $\pm$ 方差 (%) 时间 (s)
Australian (690 $\times$ 14)	86.10 $\pm$ 4.55 0.2308/0.0078	86.81 $\pm$ 2.96 0.2396/0.2690	86.67 $\pm$ 3.91 0.7044/0.5752	87.10 $\pm$ 3.45 1.1966
House votes (435 $\times$ 16)	95.47 $\pm$ 3.21 0.0734/0.2266	95.17 $\pm$ 1.50 0.0562/0.3620	95.56 $\pm$ 3.25 0.0570/0.0026	95.64 $\pm$ 2.96 0.1435
Heart-c (303 $\times$ 14)	85.13 $\pm$ 5.07 0.0527/0.3486	84.50 $\pm$ 4.22 0.0585/0.0279	85.36 $\pm$ 4.26 0.0518/0.0707	85.16 $\pm$ 5.20 0.0819
Heart-Statlog (270 $\times$ 13)	84.81 $\pm$ 4.77 0.0472/0.0018	83.70 $\pm$ 4.79 0.0615/0.2463	85.19 $\pm$ 4.62 0.0620/0.0714	85.19 $\pm$ 4.72 0.0752
Monk3 (432 $\times$ 7)	84.91 $\pm$ 3.96 0.0658/0.0066	83.88 $\pm$ 4.53 0.0433/0.0064	83.97 $\pm$ 5.75 0.0602/0.0383	84.33 $\pm$ 5.32 0.0916
Sonar (208 $\times$ 60)	80.13 $\pm$ 4.28 0.0579/0.9890	79.33 $\pm$ 4.96 0.0532/0.0090	80.54 $\pm$ 5.26 0.0355/0.6918	80.33 $\pm$ 6.00 0.0453
Spect (267 $\times$ 44)	80.19 $\pm$ 6.21 0.0527/0.0303	80.16 $\pm$ 5.97 0.0720/0.0150	80.33 $\pm$ 5.46 0.0386/0.3030	80.46 $\pm$ 5.89 0.0810
平均精度	85.25 $\pm$ 4.58	84.80 $\pm$ 4.10	85.37 $\pm$ 4.64	85.46 $\pm$ 4.79
W-T-L	6-0-1	7-0-0	4-1-2	

表 3 非线性 IPTSVM 与几种算法在两类分类 UCI 数据集上性能比较

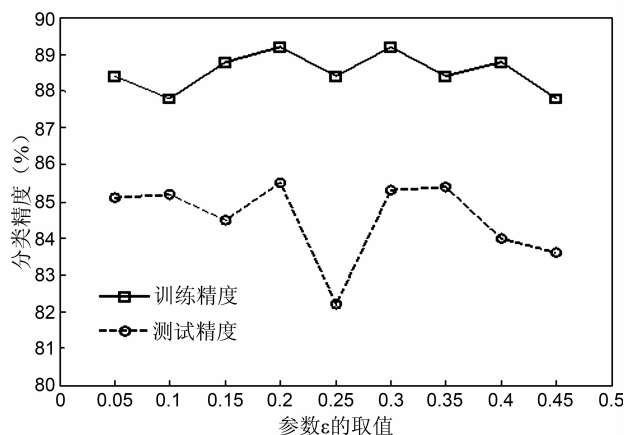
数据集	TBSVM	PTSVM	RPTSVM	IPTSVM
	精度 $\pm$ 方差 (%) 时间(s)/ p 值	精度 $\pm$ 方差 (%) 时间(s)/ p 值	精度 $\pm$ 方差 (%) 时间(s)/ p 值	精度 $\pm$ 方差 (%) 时间(s)
Australian(690 $\times$ 14)	86.96 $\pm$ 4.27 0.6735/0.0076	86.81 $\pm$ 2.68 0.1805/0.0211	87.10 $\pm$ 2.84 0.7057/0.0029	87.25 $\pm$ 3.54 2.4617
House votes(435 $\times$ 16)	95.87 $\pm$ 3.17 0.0498/0.0023	95.40 $\pm$ 1.82 0.0873/0.0390	95.87 $\pm$ 2.58 0.0600/0.0051	96.09 $\pm$ 2.29 1.1908
Heart-c(303 $\times$ 14)	85.13 $\pm$ 2.47 0.0544/0.0023	84.81 $\pm$ 4.01 0.0403/0.0045	85.19 $\pm$ 4.15 0.0759/0.6959	85.19 $\pm$ 3.73 0.0701
Heart-Statlog(270 $\times$ 13)	84.44 $\pm$ 3.08 0.0342/0.0165	84.48 $\pm$ 2.62 0.0289/0.8752	84.81 $\pm$ 5.91 0.0868/0.7052	85.56 $\pm$ 4.480.0852
Monk3(432 $\times$ 7)	98.26 $\pm$ 1.32 0.1095/0.0033	98.06 $\pm$ 3.13 0.0590/0.0029	98.14 $\pm$ 3.06 0.1434/0.0059	98.23 $\pm$ 3.01 0.1672
Sonar(208 $\times$ 60)	89.26 $\pm$ 7.19 0.0215/0.2217	88.94 $\pm$ 7.64 0.0131/0.0013	90.38 $\pm$ 7.25 0.0162/0.0536	89.62 $\pm$ 6.20 0.0266
Spect(267 $\times$ 44)	82.37 $\pm$ 5.91 0.0162/0.0430	82.25 $\pm$ 4.52 0.0256/0.5643	82.85 $\pm$ 5.98 0.0138/0.0021	82.64 $\pm$ 5.22 0.0603
平均精度	88.90 $\pm$ 3.92	88.67 $\pm$ 3.77	89.19 $\pm$ 4.54	89.23 $\pm$ 4.07
W-T-L	6-0-1	7-0-0	4-1-2	

Heart-c、Heart-Statlog、Monk3、Sonar 和 Spect 这 7 个数据集进行验证. 表 2 和表 3 分别给出了 4 种算法在线性和非线性情形下的 10-折交叉验证结果和训练时间. 为了更好的体现算法的优越性, 在表 1 ~ 表 3 中, 我们用黑体数字代表特定数据集下获得的最高分类精度. 同时, 我们在表 2 和表 3 的倒数第 2 行给出了所有算法的平均分类精度, 在最后一行用“W-T-L”(win-tie-loss)比较了 IPTSVM 算法与其它算法的性能. 为了比较我们所提出的方法与已有方法的显著性差异, 我们采用配对 t -检验方法, 将显著性水平设置为 0.05, 表 2 和表 3 中的有很多小于 0.05 的 p 值, 进一步说明了我们的方法与已

有方法的差异性. 总而言之, 实验结果说明了 IPTSVM 算法在绝大部分数据上都取得了比 PTSVM 算法较好的实验结果, 所以我们的算法确实比 PTSVM 算法有一定的改进. 同时, IPTSVM 算法在大多数数据集上也取得的与 RPTSVM 相当或者较好的效果. 稍微不足的是, 我们 IPTSVM 算法的训练时间要相对长一点.

5 结论

本文针对投影孪生支持向量机 (PTSVM) 算法在训练和求解过程中存在的一些问题, 提出了一种改进的投影孪生支持向量机 (Improved PTSVM, IPTSVM) 算法. 与 PTSVM 算法相比较, 本文的主要贡献在于: 首先, 提出了 IPTSVM 新模型, 克服了原始 PTSVM 在训练之前需要求解两个逆矩阵的问题; 其次, 轻松利用核技巧将线性 IPTSVM 推广到非线性形式, 继承了传统 SVM 的精髓, 同时也弥补了原始 PTSVM 没有给出非线性形式的不足; 最后, 通过引入新的参数 ε , 可以方便调节 IPTSVM 模型的性能, 提高了分类精度. 在人工数据集和标准 UCI 数据集上的实验结果表明, 本文算法与 PTSVM 算法比较具有一定的优势. 然而, 我们所提出的 IPTSVM 算法也存在一些不足之处, 如: (1) 优化问题的规模约是 PTSVM 优化问题的 1.5 倍, 求解速度相当较慢; (2) 有多个参数需要选择. 这些问题都有待进一步研究. 另外, 将 IPTSVM 算法向多重递归投影方向或多

图1 参数 ε 对IPTSVM算法性能的调节作用

类分类问题推广将是我们下一步研究的重点。

参考文献

- [1] Cortes C, Vapnik V N. Support vector machine[J]. Machine Learning, 1995, 20(3): 273 – 297.
- [2] Osuna E, Freund R, Girosi F. Training support vector machines: An application to face detection[A]. Proceedings of Computer Vision and Pattern Recognition [C]. San Juan: IEEE Computer Society Conference on, 1997. 130 – 136.
- [3] Isa D, Lee L H, Kallimani V P, et al. Text document pre-processing with the Bayes formula for classification using the support vector machine [J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20 (9): 1264 – 1272.
- [4] Boser B, Guyon I, Vapnik V N. A training algorithm for optimal margin classifiers[A]. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory [C]. New York: ACM, 1992. 144 – 152.
- [5] Osuna E, Freund R, Girosi F. An improved training algorithm for support vector machines [A]. Proceedings of IEEE Workshop on Neural Networks for Signal Processing [C]. New York, USA: IEEE, 1997. 276 – 285.
- [6] Platt J. Fast Training of Support Vector Machines Using Sequential Minimal Optimization[M]. Advances in Kernel Methods; Support Vector Machines, Cambridge, MA: MIT Press, 1998.
- [7] Mangasarian O L, Wild E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 69 – 74.
- [8] Jayadeva, Khemchandai R, Chandra S. Twin support vector machine classification for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905 – 910.
- [9] Chen X B, Yang J, Ye Q L, et al. Recursive projection twin support vector machine via within-class variance minimization[J]. Pattern Recognition, 2011, 44(10): 2643 – 2655.
- [10] Shao Y H, Wang Z, Chen W J, et al. A regularization for the projection twin support vector machine[J]. Knowledge-Based Systems, 2013, 37: 203 – 210.
- [11] Shao Y H, Deng N Y, Yang Z M. Least squares recursive projection twin support vector machine for classification[J]. Pattern Recognition, 2012, 45(6): 2299 – 2307.
- [12] Ding S F, Hua X P. Recursive least squares projection twin support vector machines for nonlinear classification [J]. Neurocomputing, 2014, 130(4): 3 – 9.
- [13] Shao Y H, Zhang C H, Wang X B, et al. Improvements on twin support vector machines[J]. IEEE Transactions on Neural Networks, 2011, 22(6): 962 – 968.
- [14] Peng X J. TPMSVM: a novel twin parametric-margin support vector machine for pattern recognition[J]. Pattern Recognition, 2011, 44(10): 2678 – 2692.
- [15] Tian Y J, Qi Z Q, Ju X C, et al. Nonparallel support vector machines for pattern classification[J]. IEEE Transaction on Cybernetics, 2014, 44(7): 1067 – 1079.
- [16] Tian Y J, Ping Y. Large-scale linear nonparallel support vector machine solver [J]. Neural Networks, 2014, 50(2): 166 – 174.
- [17] Shao Y H, Chen W J, Deng N Y. Nonparallel hyperplane support vector machine for binary classification problems [J]. Information Sciences, 2014, 263(4): 22 – 35.
- [18] Ding S F, Hua X P, Yu J Z. An overview on nonparallel hyperplane support vector machine algorithms[J]. Neural Computing and Applications, 2014, 25(5): 975 – 982.
- [19] Ripley B D. Pattern Recognition and Neural Networks [M]. Cambridge University Press, 2008.
- [20] Blake C L, Merz C J. UCI repository of machine learning databases[R]. Irvine, CA: Department of Information and Computers Science, University of California, 1998.

作者简介



陈素根 男, 1982 生, 江南大学物联网工程学院博士研究生, 安庆师范大学副教授. 研究方向为模式识别与智能系统、机器学习等.
E-mail: chensugen@126.com



吴小俊 (通信作者) 男, 1967 年生, 江南大学物联网工程学院教授, 博士生导师. 研究方向为模式识别、智能计算、计算机视觉等.
E-mail: wu_xiaojun@jiangnan.edu.cn