

# 控制与决策

Control and Decision

## 定点孪生支持向量机

刘峤, 方佳艳

引用本文:

刘峤, 方佳艳. 定点孪生支持向量机[J]. 控制与决策, 2020, 35(2): 272–284.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.0531>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于核正交流形角不相似度的非线性动态过程监测方法

Kernel orthogonal manifold angle based dissimilarity for nonlinear dynamic process monitoring

控制与决策. 2018, 33(6): 1141–1146 <https://doi.org/10.13195/j.kzyjc.2017.0425>

### 基于多代表点的大规模数据模糊聚类算法

Fuzzy clustering algorithm based on multiple medoids for large-scale data

控制与决策. 2016, 31(12): 2122–2130 <https://doi.org/10.13195/j.kzyjc.2015.1488>

### 基于图像特征的钼浮选矿浆pH值预测控制

Predictive control of slurry pH based on froth characteristics for antimony flotation process

控制与决策. 2016, 31(11): 1973–1978 <https://doi.org/10.13195/j.kzyjc.2015.1335>

### 经验模式分解与时间序列分析在网络流量预测中的应用

Network traffic prediction based on empirical mode decomposition and time series analysis

控制与决策. 2015, 30(5): 905–910 <https://doi.org/10.13195/j.kzyjc.2014.0453>

### 带反向预测及斥力因子的改进粒子群优化算法

Improved particle swarm optimization algorithm with reverse forecast and repulsion

控制与决策. 2015(2): 311–315 <https://doi.org/10.13195/j.kzyjc.2013.1669>

### 基于具有自适应分段损失函数支持向量机的产品销售预测模型

Product sale forecasting method based on support vector machine with adaptive segmented loss function

控制与决策. 2015(10): 1803–1809 <https://doi.org/10.13195/j.kzyjc.2013.1541>

### 一种基于FCOWA-ER的SVM多分类方法

A multi-class SVM based on FCOWA-ER

控制与决策. 2015(10): 1773–1778 <https://doi.org/10.13195/j.kzyjc.2014.1215>

### 基于MultiBoost的集成支持向量机分类方法及其应用

MultiBoost with SVM-based ensemble classification method and application

控制与决策. 2015(1): 81–85 <https://doi.org/10.13195/j.kzyjc.2013.1562>

# 定点孪生支持向量机

刘 峤, 方佳艳<sup>†</sup>

(1. 电子科技大学 信息与软件工程学院, 成都 611731; 2. 电子科技大学  
网络与数据安全省级重点实验室, 成都 611731)

**摘 要:** 孪生支持向量机(TWSVM)以及最近提出的各种变体模型均是在高维空间内独立求解两个带有约束条件的对偶二次规划问题(QPP). 然而, 由于每个对偶的QPP所需求解的对偶变量的数量由他类样本的数量决定, 当需要处理大规模数据集时, 这种直接求解标准QPP的方法将会导致非常高的计算复杂度. 为此, 提出一种改进的孪生支持向量机模型, 称为定点孪生支持向量机(FP-TWSVM). 所提模型将传统的TWSVM及其变体模型中处在高维空间内的对偶QPP转化成一系列有限个一维空间内的单峰函数优化问题. 可以采用高效的线性搜索方法求解这些一维的单峰函数优化问题, 例如斐波那契算法、黄金分割法. 在标准数据集包括大规模数据集上的数值实验验证了FP-TWSVM算法的有效性. 实验结果表明, FP-TWSVM在保持与其他模型相当的分精度的同时, 具有更快的训练速度, 消耗更少的内存空间.

**关键词:** 孪生支持向量机; 定点; 单峰函数; 线性搜索方法; 二次规划

**中图分类号:** TP181      **文献标志码:** A

## Fixed-point twin support vector machine

LIU Qiao, FANG Jia-yan<sup>†</sup>

(1. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. Provincial Key Laboratory of Network and Data Security, University of Electronic Science and Technology of China, Chengdu 611731, China)

**Abstract:** The twin support vector machines (TWSVMs) and the recently proposed variant models are all designed to solve two dual quadratic programming problems (QPPs) with constraint conditions independently in high dimensional space. However, since each dual QPP involves a set of dual variables with its size determined by the number of samples of other classes, when we need to cope with large-scaled datasets, the method of directly solving QPP will lead to very high computational complexity. Therefore, this paper proposes an improved twin support vector machine model, known as a fixed-point TWSVM (FP-TWSVM). This model transforms the traditional TWSVM and its variant models into a series of unimodal function optimization problems in one-dimensional space. Efficient linear search methods are used such as Fibonacci algorithm and golden section method to solve these one-dimensional unimodal function optimization problems. The validity of FP-TWSVM algorithm is verified by numerical experiments on several datasets including large datasets. The experimental results show that the FP-TWSVM has faster training speed and consumes less memory space while maintaining vertically the same classification accuracy as other models.

**Keywords:** twin support vector machine; fixed-point; unimodal function; line search methods; quadratic programming

## 0 引 言

近年来,支持向量机(SVM)<sup>[1-3]</sup>在机器学习领域尤其是非监督式学习中逐渐发展成为一种具有高效计算性能的方法. 由于它自身一些独特的数学特性, 在很多应用领域中<sup>[4-6]</sup>, 其性能表现更加优于其他已有的机器学习技术. 第一, 在训练模型中, 并不像传统

的监督式学习算法那样仅仅最小化经验风险, 还采用了结构风险最小化(SRM)原则, 使得算法产生的二元分类器具有“最大间隔”的特性, 因此提升了模型的泛化能力. 第二, SVM模型是通过求解一个二次凸优化问题产生最优分类超平面的, 因此保证了模型求出的最优解一定是全局最优解. 第三, 由于模型公

收稿日期: 2018-04-25; 修回日期: 2018-08-27.

基金项目: 国家自然科学基金重点项目(61133016); 国家自然科学基金面上项目(61772117); 四川省科技厅科技支撑项目(2017GZ0308); 十三五装备预研领域基金项目(61403120102).

责任编辑: 阳春华.

<sup>†</sup>通讯作者. E-mail: fyk80@163.com.

式的数学特性,使得核技巧<sup>[7]</sup>可以很容易地被合并到 SVM 模型中,从而实现高维空间中任意两个向量内积的隐式计算。

为了加速 SVM 模型的计算速度, Khewehandani 等<sup>[8]</sup>提出了孪生支持向量机(TWSVM),该模型被视为稀疏核学习的一个里程碑。TWSVM旨在求解两个最优的非平行超平面,每一个超平面离本类样本集最近并且尽可能地离他类样本集远。尽管TWSVM的公式构造与SVM的公式构造在思想上很相似,但它们在基础的公式构造上还是有很大不同。在TWSVM模型中,需要求解两个规模较小的QPP而不是SVM中单一的规模较大的QPP,并且所有样本点相对于分类器的分布遵循这样的规则(以二元分类为例):来自某一类别的样本集决定着另一个类别所对应的QPP中的约束条件,反之亦然。因此,TWSVM比传统的SVM模型具有更快的训练速度。

由于采用了分解思想,将一个较大规模的优化问题分解成两个独立的规模较小的优化问题,TWSVM模型的计算速度得到了提升。因此,近年来,TWSVM备受研究者的关注,并在此基础上提出很多TWSVM的变体模型,例如双子边界支持向量机(TBSVM)<sup>[9-10]</sup>、 $v$ -孪生支持向量机( $v$ -TWSVM)<sup>[11]</sup>、基于球体的孪生支持向量机(sphere-based TWSVM)<sup>[12-14]</sup>、基于结构化信息的孪生支持向量机(structure information based TWSVM)<sup>[15-16]</sup>、粗糙的基于间隔的 $v$ -TWSVM(rough margin-based  $v$ -TWSVM)<sup>[17-18]</sup>以及其他相关变体模型<sup>[19-20]</sup>。Shao等<sup>[9]</sup>提出了双子边界支持向量机(TBSVM),通过在模型的目标函数中加入一个正则项,实现了基于TWSVM模型的结构风险最小化(SRM)原则。Peng等<sup>[11]</sup>提出了 $v$ -TWSVM,引入了一对先验参数( $v$ )来分别控制边界错误所占比例的上界和支持向量所占比例的下界。另外,在TWSVM中的单位距离被松弛化为一个可以在原始问题中被优化的变量 $\rho$ 。在粗糙的基于间隔的 $v$ -TWSVM(rough margin-based  $v$ -TWSVM)模型<sup>[17]</sup>中,对于不同的违反边界的样本点,根据它们距离分类器的相对位置赋予它们不同的惩罚值,因此在一定程度上提高了模型的分类精度。Wang等<sup>[18]</sup>在粗糙的基于间隔的 $v$ -TWSVM的基础上引入结构风险最小化原则来最大化分类间隔,不仅提高了测试数据集的分类精度,而且避免了大矩阵的求逆运算,降低了模型的计算复杂度。

这些基于TWSVM的变体模型,包括TWSVM本

身,其模型公式构造的结果均是求解高维空间里的二次规划问题的对偶形式。然而,当需要处理大规模数据集时,求解这些二次规划问题将面临巨大的计算复杂度和时间开支。为了使得TWSVM及其变体模型能够更加适应于大规模数据集的分类学习,本文提出定点孪生支持向量机(fixed-point TWSVM, FP-TWSVM)来加速模型的训练速度,从而降低时间消耗。此外,该算法的思想还可以用于TWSVM的变体模型中。FP-TWSVM的主要特性如下:

1) FP-TWSVM的公式构造思想继承了TWSVM,将传统SVM模型中单一的规模较大的QPP分解成两个独立的规模较小的QPP,从而产生两个最邻近的分割超平面,每一个超平面都离本类样本集最近,离他类样本集尽可能地远。

2) 与TWSVM模型不同的是,FP-TWSVM并不是直接求解高维空间内的标准QPP,而是设计了一种定点算法,将单个QPP转换成一系列有限多个一维空间里的单峰函数优化问题,因此一些高效的线性搜索方法,例如斐波那契算法和黄金分割法,都可以用来求解一维单峰函数优化问题,避免了大规模二次型矩阵的存储问题,极大地提高了模型的训练速度和存储效率。

3) 通过引入核生成曲面,核技巧能够与FP-TWSVM结合使用,从而实现FP-TWSVM算法在非线性情形的应用。当使用线性核函数时,FP-TWSVM的非线性形式退化成了线性形式。

## 1 孪生支持向量机(TWSVM)

这一部分简要回顾TWSVM的基本数学构造。考虑一个在 $d$ 维空间内的二元分类问题,有 $m$ 个正类别样本点和 $n$ 个负类别样本点。所有的正类别样本点存放在数据矩阵 $A \in R^{m \times d}$ 中,所有的负类别样本点存放在数据矩阵 $B \in R^{n \times d}$ 中, $d$ 是特征空间的维度。 $A_i(i = 1, 2, \dots, m)$ 表示矩阵 $A$ 的第 $i$ 行, $B_i(i = 1, 2, \dots, n)$ 表示矩阵 $B$ 的第 $i$ 行, $y \in \{-1, +1\}$ 表示样本的标签值。

### 1.1 线性孪生支持向量机(linear TWSVM)

TWSVM<sup>[8]</sup>是一种非常优秀的分类算法,可以将传统SVM模型中单一的规模较大的QPP分解成两个规模较小的独立的QPP,由此降低了计算复杂度和时间消耗。两个非平行分割超平面如下:

$$\mathbf{w}_1^T \mathbf{x} + b_1 = 0, \mathbf{w}_2^T \mathbf{x} + b_2 = 0. \quad (1)$$

其中: $\mathbf{w}_1$ 和 $\mathbf{w}_2$ 是权值向量, $b_1$ 和 $b_2$ 是超平面的截距。

TWSVM公式构造的思想与SVM很相似,然而它们的基本构造还是有很大区别的.在TWSVM模型中,旨在找到两个最优分割超平面(1),要求在这两个分割超平面的周围尽可能多地聚集着本类别的样本点.另外,在每一个QPP中的约束条件集合中,并不同时包含所有的数据点,只有那些他类的样本点的数量才决定着本类别对应求解的QPP的约束条件集合的大小.通过求解以下两个QPP,可以得到TWSVM的分类器:

$$\begin{aligned} \min_{\mathbf{w}_1, b_1, \xi_2} & \frac{1}{2} (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1)^T (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{e}_2^T \xi_2; \\ \text{s.t.} & -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) \geq \mathbf{e}_2 - \xi_2, \xi_2 \geq 0. \end{aligned} \quad (2)$$

$$\begin{aligned} \min_{\mathbf{w}_2, b_2, \xi_1} & \frac{1}{2} (\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2)^T (\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2) + c_2 \mathbf{e}_1^T \xi_1; \\ \text{s.t.} & (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) \geq \mathbf{e}_1 - \xi_1, \xi_1 \geq 0. \end{aligned} \quad (3)$$

其中:  $c_1, c_2 > 0$  是常数,分别用来权衡每个分类器到本类样本集的距离和对应的松弛变量的惩罚值;  $\xi_1$  和  $\xi_2$  是由松弛变量组成的向量;  $\mathbf{e}_1$  和  $\mathbf{e}_2$  是实数空间里具有合适维度的单位向量.以上优化问题(2)和(3)中的目标函数的第1项分别是各个非平行超平面到本类样本点的距离之和,因此最小化第1项就相当于使得每个超平面离本类的样本集最接近.问题(2)和问题(3)中的约束条件要求每个分割超平面到他类样本点的距离至少为单位1.  $\xi_1$  和  $\xi_2$  是针对每个超平面分别用来记录那些到本类分类器距离小于1的他类样本点的距离违反量.目标函数的第2项是惩罚函数,分别用来记录每个类别的分类器对应的他类样本违反量的总和,因此等同于最小化边界错误.

通过引入两组拉格朗日乘子,上述的原始优化问题可以被构造对偶形式<sup>[8]</sup>的优化问题.原始问题(2)对应的拉格朗日函数如下:

$$\begin{aligned} L(\mathbf{w}_1, b_1, \xi_2, \boldsymbol{\alpha}, \boldsymbol{\beta}) = & \\ & \frac{1}{2} (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1)^T (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + c_1 \mathbf{e}_2^T \xi_2 + \\ & \boldsymbol{\alpha}^T (\mathbf{e}_2 - \xi_2 + \mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \boldsymbol{\beta}^T (-\xi_2), \end{aligned} \quad (4)$$

其中  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m)^T$  和  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)^T$  是由拉格朗日乘子组成的向量.将KKT条件代入拉格朗日函数(4)中,所有的原始变量均被消去,原始问题(2)会被转换成一个关于对偶变量的对偶形式的优化问题,如下所示:

$$\max_{\boldsymbol{\alpha}} \mathbf{e}_2^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T G (H^T H)^{-1} G^T \boldsymbol{\alpha};$$

$$\text{s.t. } \mathbf{0} \leq \boldsymbol{\alpha} \leq C_1 \mathbf{e}. \quad (5)$$

其中:  $H = [A, \mathbf{e}_1]$ ,  $G = [B, \mathbf{e}_2]$ .  $\mathbf{u} = [\mathbf{w}_1, b_1]^T$  是一个扩展的权值向量,因此  $\mathbf{u}$  可以用对偶变量  $\alpha_i (i = 1, 2, \dots, n)$  和样本数据点表示如下:

$$\mathbf{u} = -(H^T H)^{-1} G^T \boldsymbol{\alpha}. \quad (6)$$

采用与原始问题(2)相同的推导思路,可以得到原始问题(3)对应的对偶问题

$$\begin{aligned} \max_{\boldsymbol{\gamma}} & \mathbf{e}_1^T \boldsymbol{\gamma} - \frac{1}{2} \boldsymbol{\gamma}^T P (Q^T Q)^{-1} P^T \boldsymbol{\gamma}; \\ \text{s.t.} & \mathbf{0} \leq \boldsymbol{\gamma} \leq C_2 \mathbf{e}. \end{aligned} \quad (7)$$

其中:  $P = [A, \mathbf{e}_1]$ ,  $Q = [B, \mathbf{e}_2]$ . 同样地,  $\mathbf{v} = [\mathbf{w}_2, b_2]^T$  是一个扩展的权值向量,可以用对偶变量  $\gamma_i (i = 1, 2, \dots, m)$  和样本数据点表示如下:

$$\mathbf{v} = (Q^T Q)^{-1} P^T \boldsymbol{\gamma}. \quad (8)$$

通过求解对偶的QPP (5)和(7)获得了最邻近分割超平面(1)后,可以根据如下判别式赋予任意新的数据点  $\mathbf{x} \in R^n$  的类别标签:

$$y = \arg \min_{k=1,2} \frac{|\mathbf{w}_k^T \mathbf{x} + b_k|}{\|\mathbf{w}_k^T\|}.$$

## 1.2 非线性孪生支持向量机(nonlinear TWSVM)

通过引入如下两个核生成曲面,可以将1.1节的线性TWSVM<sup>[8]</sup>拓展到非线性情形:

$$K(\mathbf{x}^T, C^T) \mathbf{u}_1 + b_1 = 0, K(\mathbf{x}^T, C^T) \mathbf{u}_2 + b_2 = 0. \quad (9)$$

其中:  $C^T = [A, B]^T$ ,  $K$  是一个选择适当的核函数.式(1)可以被看成是式(9)的一种特殊情况,因为在式(1)中相当于使用了线性核函数.可以用与线性情况相同的推导方法得出非线性情形中正类别对应的原始问题,如下所示:

$$\begin{aligned} \min_{\mathbf{u}_1, b_1, \xi_2} & \frac{1}{2} \|(K(A, C^T) \mathbf{u}_1 + \mathbf{e}_1 b_1)\|^2 + C_1 \mathbf{e}_2^T \xi_2; \\ \text{s.t.} & -(K(B, C^T) \mathbf{u}_1 + \mathbf{e}_2 b_1) \geq \mathbf{e}_2 - \xi_2, \xi_2 \geq 0. \end{aligned} \quad (10)$$

问题(10)对应的对偶问题如下:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} & \mathbf{e}_2^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T R (S^T S)^{-1} R^T \boldsymbol{\alpha}; \\ \text{s.t.} & \mathbf{0} \leq \boldsymbol{\alpha} \leq C_1 \mathbf{e}. \end{aligned} \quad (11)$$

其中:  $S = [K(A, C^T), \mathbf{e}_1]$ ,  $R = [K(B, C^T), \mathbf{e}_2]$ . 记  $\mathbf{z} = [\mathbf{u}_1, b_1]^T$ , 则  $\mathbf{z}$  可以通过下式求解:

$$\mathbf{z} = -(S^T S)^{-1} R^T \boldsymbol{\alpha}. \quad (12)$$

同理,非线性情形下的负类别对应的原始问题如下:

$$\begin{aligned} \min_{\mathbf{u}_2, b_2, \xi_1} & \frac{1}{2} \|K(B, C^T)\mathbf{u}_2 + \mathbf{e}_2 b_2\|^2 + C_2 \mathbf{e}_1^T \xi_1; \\ \text{s.t.} & K(A, C^T)\mathbf{u}_2 + \mathbf{e}_1 b_2 \geq \mathbf{e}_1 - \xi_1, \quad \xi_1 \geq \mathbf{0}. \end{aligned} \quad (13)$$

问题(13)的对偶问题如下:

$$\begin{aligned} \max_{\gamma} & \mathbf{e}_1^T \gamma - \frac{1}{2} \gamma^T L(N^T N)^{-1} L^T \gamma; \\ \text{s.t.} & \mathbf{0} \leq \gamma \leq C_2 \mathbf{e}. \end{aligned} \quad (14)$$

其中:  $L = [K(A, C^T), \mathbf{e}]$ ,  $N = [K(B, C^T), \mathbf{e}_2]$ . 记  $\mathbf{z}_2 = [\mathbf{u}_2, b_2]^T$ , 同样,  $\mathbf{z}_2$  可以由下式求出:

$$\mathbf{z}_2 = (N^T N)^{-1} L^T \gamma. \quad (15)$$

求解完对偶问题(11)和(14)得到两个核生成曲面(9)后, 可以使用与线性情形同样的判别规则来预测新的数据点  $\mathbf{x} \in R^n$  的标签值.

通常 SVM 模型的复杂度不超过  $N^3$ , 其中  $N$  是正、负类别训练样本的总数. 然而, 由于 TWSVM 将 SVM 中单一的规模较大的优化问题分解成了两个规模较小的独立的优化问题(2)和(3), 并且每个优化问题的规模约为  $N/2$ , 因此 SVM 模型和 TWSVM 模型的训练时间之比为  $[N^3 / (2 \times (N/2)^3)] = 4$ . 由此可得, TWSVM 的训练速度大约是传统 SVM 模型训练速度的 4 倍<sup>[8]</sup>.

在实际应用中, 若某个应用在正类别或者负类别需要大规模的训练集, 则可以使用矩形核技术<sup>[21-22]</sup>来降低问题(10)和(13)的数据维度. 此外, 在线性情形中, 需要引入一个正则项  $\varepsilon I$  来应对需要求逆的矩阵  $S^T S$  和  $N^T N$  可能存在的病态性. 这样, 可以进一步使用 Sherman-Morrison-Woodbury 公式求解矩阵的逆, 降低需要求逆的矩阵的阶数.

## 2 定点孪生支持向量机(FP-TWSVM)

无论是 TWSVM 还是它的变体模型, 最后均要求解对偶的标准二次规划问题(QPP)得到最优分割超平面. 在处理规模适中的数据时, 这些算法的训练时间可以控制在一个合理的范围. 但是, 当需要处理超大规模数据集时, 这些模型的计算复杂度就会相当高, 其时间开支已经超出了实际应用程序的可接受范围. 为解决这一问题, 该部分将提出定点孪生支持向量机 FP-TWSVM 算法. 与引言中的算法不同的是, FP-TWSVM 算法在求解优化问题的过程中引入了定点思想, 将 TWSVM 算法中的求解单个标准对偶 QPP 问题转换成求解一系列连续的一维单峰函数优化问题, 每一个优化问题都可以采用高效的线性搜索方法求得最佳解. 并且, 每个被分解的标准 QPP 问题中的目标函数一定随着每个一维单峰函数优化问题

的最优解的求出呈单调递减趋势. 与 TWSVM 相同的是, FP-TWSVM 也可以通过构造核生成曲面由线性分类拓展到非线性分类. 在 2.1 节中先将定点思想运用在线性情形中, 然后再扩展到非线性情况. 该定点思想在 TWSVM 的其他变体模型中同样适用.

为了方便本文中的后续数学推导, 先给出一些与 TWSVM 模型中不一样的数学符号. 考虑到计算的简便性, 去掉了偏移量  $b$ , 并将它隐式地加在权值向量  $\mathbf{w}$  的最后一个维度. 因此, 每个数据点要用  $n + 1$  维的向量来表示, 且每个样本向量的最后一个维度固定设为 1. 为了更好地体现本文新提出的算法的定点思想, 数学推导中将使用求和符号将所有的数据点以及原始变量和对偶变量显式地表示出来, 而不是像 TWSVM 中那样将所有的正类别和负类别样本点分别集中放在两个数据矩阵中, 将所有的原始变量和对偶变量表示成向量形式.

考虑一个二元分类问题. 假设正负类别的样本点分别有  $m$  个和  $n$  个;  $\{\mathbf{x}_i^{(1)}\}_{i=1}^m$  表示正类别样本数据集,  $\{\mathbf{x}_j^{(2)}\}_{j=1}^n$  表示负类别样本数据集,  $i$  和  $j$  分别表示正类别和负类别的样本索引;  $\mathbf{w}_1$  和  $\mathbf{w}_2$  分别表示正类别和负类别的维度扩展的权值向量.

### 2.1 线性孪生支持向量机(linear FP-TWSVM)

#### 2.1.1 构建 FP-TWSVM 的非平行分类器

在线性 FP-TWSVM 中, 目标是要求解如下两个分割超平面:

$$\mathbf{w}_1^T \mathbf{x} = 0, \quad \mathbf{w}_2^T \mathbf{x} = 0. \quad (16)$$

仿照 TWSVM 模型, 可以通过求解两个规模较小的 QPP 问题来得到上述两个最优非平行二元分类器. 在此, 只展示正类别对应的数学推导, 负类别的分类器的求解过程与正类别几乎一致. 正类别对应的最优分割超平面可以通过求解以下 QPP 得到:

$$\begin{aligned} \min_{\mathbf{w}_1, \xi} & \frac{1}{2} \sum_{i=1}^m (\mathbf{w}_1^T \mathbf{x}_i^{(1)})^2 + c_1 \sum_{j=1}^n \xi_j; \\ \text{s.t.} & -\mathbf{w}_1^T \mathbf{x}_j^{(2)} \geq 1 - \xi_j, \quad \xi_j \geq 0, \quad j = 1, 2, \dots, n. \end{aligned} \quad (17)$$

其中:  $c_1 > 0$  是平衡常数,  $\xi_j$  是松弛变量. 对应的拉格朗日函数如下:

$$\begin{aligned} L(\mathbf{w}_1, \xi, \alpha, \beta) = & \frac{1}{2} \sum_{i=1}^m (\mathbf{w}_1^T \mathbf{x}_i^{(1)})^2 + C_1 \sum_{j=1}^n \xi_j + \\ & \sum_{j=1}^n \alpha_j (1 - \xi_j + \mathbf{w}_1^T \mathbf{x}_j^{(2)}) + \sum_{j=1}^n \beta_j (-\xi_j), \end{aligned} \quad (18)$$

其中  $\alpha_j$  和  $\beta_j$  是拉格朗日乘子. 将函数(18)关于  $\mathbf{w}_1$  和  $\xi_j$  的偏导数分别设为零, 得到以下 Karush-Kuhn-Tucker (KKT)<sup>[23]</sup> 条件:

$$\frac{\partial L}{\partial \mathbf{w}_1} = \sum_{i=1}^m (\mathbf{x}_i^{(1)} \mathbf{x}_i^{(1)\top}) \mathbf{w}_1 + \sum_{j=1}^n \alpha_j \mathbf{x}_j^{(2)} = 0, \quad (19)$$

$$\frac{\partial L}{\partial \xi_j} = C_1 - \alpha_j - \beta_j = 0, \quad 0 \leq \alpha_j \leq c_1. \quad (20)$$

由互补松弛定理得到:

$$\alpha_j (1 - \xi_j + \mathbf{w}_1^\top \mathbf{x}_j^{(2)}) = 0, \quad (21)$$

$$\beta_j (-\xi_j) = 0. \quad (22)$$

由式(19)可得

$$\begin{aligned} \mathbf{w}_1 = & - \left( \sum_{i=1}^m (\mathbf{x}_i^{(1)} \mathbf{x}_i^{(1)\top}) \right)^{-1} \left( \sum_{j=1}^n \alpha_j \mathbf{x}_j^{(2)} \right) = \\ & - (\mathbf{A}^\top \mathbf{A})^{-1} \left( \sum_{j=1}^n \alpha_j \mathbf{x}_j^{(2)} \right), \end{aligned} \quad (23)$$

其中  $\mathbf{A}$  是正类别数据矩阵, 它的第  $i$  行表示第  $i$  个正类别的样本点. 将式(20)和(23)代入(18)中, 得到如下简化的拉格朗日函数:

$$L(\boldsymbol{\alpha}) = -\frac{1}{2} \sum_j \sum_k \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j=1}^n \alpha_j, \quad (24)$$

其中  $H$  被记为  $H = (\mathbf{A}^\top \mathbf{A})^{-1}$ . 因此, 可以得到正类别的对偶问题如下:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2} \sum_j \sum_k \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j=1}^n \alpha_j; \\ \text{s.t.} \quad & 0 \leq \alpha_j \leq C_1, \quad j = 1, 2, \dots, n. \end{aligned} \quad (25)$$

求解问题(25)之后, 可以得到正类别对应的最优权值向量.

### 2.1.2 分解对偶优化问题

对偶问题(5)包含了  $n$  个变量, 可以用标准的求解二次规划问题的方法来得到问题(25)的最佳解. 然而, 如果负类别包含的样本点规模很大, 则在正类别的原始问题中会包含大量的约束条件, 就需要大量的拉格朗日乘子来代表这些约束条件, 从而构成对偶问题. 将对偶的规划问题(25)写成标准的QP形式需要一个  $n \times n$  的数据矩阵, 当  $n$  很大时, 求解这样的QP问题的计算复杂度就会非常大. 同时, 由于有大规模的二次型矩阵的存在, 会消耗大量的存储空间. 因此, 需要寻找新的求得拉格朗日乘子最优解的方法, 而不是直接用标准的QP方法来求解问题(25). 接下来, 将提出一种“定点”思想来解决这一问题.

这种算法思想的核心在于将问题(25)中的约束条件集合划分成  $n$  个互不相交的子集  $\{\alpha_j | \alpha_j \geq$

$0, \alpha_j \leq c_1\}_{j=1}^n$ . 整个算法按照多个轮次运行, 每个轮次需要选择一个样本索引  $p$ , 在约束条件  $\alpha_j \geq 0$  和  $\alpha_j \leq c_1$  下, 通过更新对偶变量  $\alpha_p$  来提高目标函数值的大小. 这样的若干轮次构成了一系列连续的小规模优化问题.

接下来仅仅关注样本索引  $p$ , 重新构造问题(25)中的目标函数, 将目标函数用对偶变量  $\alpha_p$  表示出来. 将那些关于  $\alpha_p$  的项从整个目标函数  $Q$  中独立出来, 如下所示:

$$\begin{aligned} Q(\boldsymbol{\alpha}) = & -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j=1}^n \alpha_j = \\ & -\frac{1}{2} \alpha_p^2 \mathbf{x}_p^{(2)\top} H \mathbf{x}_p^{(2)} - \sum_{j=1, j \neq p}^n \alpha_j \alpha_p \mathbf{x}_j^{(2)\top} H \mathbf{x}_p^{(2)} + \\ & \alpha_p - \frac{1}{2} \sum_{j \neq p} \sum_{k \neq p} \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j \neq p} \alpha_j = \\ & -\frac{1}{2} \alpha_p^2 \mathbf{x}_p^{(2)\top} H \mathbf{x}_p^{(2)} - \alpha_p \left( \sum_{j \neq p} \alpha_j \mathbf{x}_j^{(2)\top} H \mathbf{x}_p^{(2)} - 1 \right) - \\ & \frac{1}{2} \sum_{j \neq p} \sum_{k \neq p} \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j \neq p} \alpha_j. \end{aligned} \quad (26)$$

另外, 定义如下变量:

$$M_p = \mathbf{x}_p^{(2)\top} H \mathbf{x}_p^{(2)} > 0, \quad (27)$$

$$B_p = \sum_{j \neq p} \alpha_j \mathbf{x}_j^{(2)\top} H \mathbf{x}_p^{(2)} - 1, \quad (28)$$

$$C_p = -\frac{1}{2} \sum_{j \neq p} \sum_{k \neq p} \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j \neq p} \alpha_j. \quad (29)$$

将式(27)~(29)中定义的变量代入目标函数中, 得到

$$Q(\alpha_p) = -\frac{1}{2} M_p \alpha_p^2 - B_p \alpha_p + C_p. \quad (30)$$

为简化起见, 消去那些不影响求解过程的项. 因此, 经过上述的分解步骤以后, 每一个轮次需要求解的优化问题仅包含一个变量和两个条件, 如下所示:

$$\begin{aligned} \min_{\alpha_p} \quad & Q(\alpha_p) = \frac{1}{2} M_p \alpha_p^2 + B_p \alpha_p; \\ \text{s.t.} \quad & \alpha_p \geq 0, \quad \alpha_p \leq c_1. \end{aligned} \quad (31)$$

定义

$$v_p = \alpha_p - \frac{B_p}{M_p}, \quad (32)$$

完全平方之后, 问题(31)可以表示成如下更简洁的形式:

$$\begin{aligned} \min_{v_p} \quad & \frac{1}{2} v_p^2; \\ \text{s.t.} \quad & -\frac{B_p}{M_p} \leq v_p \leq c_1 - \frac{B_p}{M_p}. \end{aligned} \quad (33)$$

可以采用高效的线性搜索方法,如黄金分割法或者斐波那契算法,求得该一维单峰函数优化问题的最佳解.由式(23)可知,正类别对应的权值向量可以表示为

$$\mathbf{w}_1 = -H\left(\sum_{j=1}^m \alpha_j \mathbf{x}_j^{(2)}\right). \quad (34)$$

当求解完一系列与问题(33)类似的一维单峰函数优化问题得到一组最优解  $v_j^*(j = 1, 2, \dots, m)$  之后,可以求出对应的拉格朗日乘子  $\alpha_j^*(j = 1, 2, \dots, n)$  的最优解,最优权值向量为

$$\mathbf{w}_1^* = -H\left(\sum_{j=1}^m \alpha_j^* \mathbf{x}_j^{(2)}\right). \quad (35)$$

正类别对应的QPP问题求解过程的算法框架如下所示(用  $\alpha_j^* = 0$ ,对算法进行初始化,  $j = 1, 2, \dots, n$ ).

**算法1**

Input:  $\{(x_1^{(1)}, y_1^{(1)}), (x_2^{(1)}, y_2^{(1)}), \dots, (x_m^{(1)}, y_m^{(1)}), (x_1^{(2)}, y_1^{(2)}), (x_2^{(2)}, y_2^{(2)}), \dots, (x_n^{(2)}, y_n^{(2)})\}$ ;

Initialize:  $\alpha_1 = 0, \dots, \alpha_m = 0$ .

Repeat:

1) Choose a pattern  $p$  (vis KKT conditions).

2) Calculate the constants for the reduced optimization problem:

$$M_p = x_p^{(2)\top} H x_p^{(2)} > 0,$$

$$B_p = \sum_{j \neq p} \alpha_j x_j^{(2)\top} H x_p^{(2)} - 1.$$

3) Find optimal solution for the following reduced problem using a simple line search method in one dimension:

$$\min_{v_p} \frac{1}{2} v_p^2,$$

$$\text{s.t. } -\frac{B_p}{M_p} \leq v_p \leq c_1 - \frac{B_p}{M_p}.$$

4) Calculate  $\alpha_p$  via  $v_p = \alpha - \frac{M_p}{M_p}$ , until a certain stopping criterion is satisfied.

Output:  $\mathbf{w}_1^* = -H\left(\sum_{j=1}^m \alpha_j^* \mathbf{x}_j^{(2)}\right)$ .

为了完成该算法框架的细节问题,需要讨论以下几个问题:

1) 需要设计一种机制来选择每一轮次所需关注的样本点,得到如式(33)所示的优化问题.有两种常用的方法对每一轮的样本点进行选择,一种是随机抽样,另一种是按照某种序列对训练集进行扫描逐一选择样本点.本文则提出一种新的样本选择策略,这种

策略与常用的两种方法相比具有更好的表现.

2) 如何设置算法1框架中的主循环的终止条件.其中一种简单的方法是使得算法运行完固定的轮次后即可停止,这就意味着只需要求解一系列有限多个类似于式(33)的小规模优化问题.而在本文后续部分,将要介绍的一种更好的设置终止条件的策略是提前设置一个精度条件,当该精度条件满足时,才会终止算法1中的主循环.本文将在2.1.3节详细说明这两种策略.

求解负类别对应的最优权值向量的过程与上述过程基本类似.负类别的最优权值向量如下:

$$\mathbf{w}_2^* = -G\left(\sum_{j=1}^m \beta_j^* \mathbf{x}_j^{(2)}\right).$$

求解得到正负类别的权值向量以后,可以根据判别式  $y_i = \arg \min_{k=1,2} \frac{|\mathbf{w}_k^\top \mathbf{x}_i|}{\|\mathbf{w}_k\|}$  来确定新的样本点  $\mathbf{x}_i \in R^d$  的标签值  $k$ .其中:  $\mathbf{w}_k \in R^{d+1}$  是维度扩展的权值向量,偏移量  $b$  被隐式地添加在最后一个维度;  $\mathbf{w}_k^- \in R^d$  是不含偏移量的权值向量.

**2.1.3 样本选择策略**

在该部分,将以正类别求解分割超平面的过程为例,详细介绍如何进行样本选择.因此在TWSVM模型中,针对正类别,需要求解的对偶QPP问题如下:

$$\max_{\alpha} -\frac{1}{2} \sum_j \sum_k \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j=1}^n \alpha_j;$$

$$\text{s.t. } 0 \leq \alpha_j \leq C_1, j = 1, 2, \dots, n.$$

与问题(25)是一致的.

首先,需要使用Karush-Kuhn-Kucker定理<sup>[23]</sup>寻找使得  $\alpha_j$  成为问题(25)的最佳解的必要条件.因而,再次引入一组新的拉格朗日乘子,并构造问题(25)的对偶问题.问题(25)的拉格朗日函数如下所示:

$$Q(\alpha, \mathbf{u}, \mathbf{v}) =$$

$$-\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \mathbf{x}_j^{(2)\top} H \mathbf{x}_k^{(2)} + \sum_{j=1}^n \alpha_j +$$

$$\sum_{j=1}^n u_j (-\alpha_j) + \sum_{j=1}^n v_j (\alpha_j - c_1), \quad (36)$$

其中  $u_j$  和  $v_j$  分别是约束条件  $\alpha_j \geq 0$  和  $\alpha_j \leq c_1$  对应的拉格朗日乘子.因此,问题(25)的对偶问题如下所示:

$$\max_{\alpha, \mathbf{u}, \mathbf{v}} Q(\alpha, \mathbf{u}, \mathbf{v});$$

$$\text{s.t. } u_j \geq 0, v_j \geq 0, j = 1, 2, \dots, n. \quad (37)$$

设问题(37)中的目标函数关于  $\alpha_j$  的偏导数为零可得到如下等式:

$$\begin{aligned} \partial Q / \partial \alpha_j = & -2\alpha_j \mathbf{x}_j^{(2)T} H \mathbf{x}_j^{(2)} - \sum_{k=1, k \neq j}^n \alpha_k \mathbf{x}_j^{(2)T} H \mathbf{x}_k^{(2)} + \\ & 1 - u_j + v_j = 0. \end{aligned} \quad (38)$$

定义

$$F_j = -2\alpha_j \mathbf{x}_j^{(2)T} H \mathbf{x}_j^{(2)} - \sum_{k=1, k \neq j}^n \alpha_k \mathbf{x}_j^{(2)T} H \mathbf{x}_k^{(2)} + 1. \quad (39)$$

这样,式(38)可以写成

$$F_j + v_j = u_j. \quad (40)$$

根据互补松弛定理,有

$$u_j \alpha_j = 0 \quad (41)$$

和

$$v_j (\alpha_j - C_1) = 0. \quad (42)$$

第1种情况是  $\alpha_j - c_1 = 0$ , 这种情况下,等式(42)自然成立. 结合问题(37)中的条件  $v_j \geq 0$  和等式(40), 可得

$$u_j \geq \max_{j: \alpha_j = c_1} F_j. \quad (43)$$

第2种情况是  $\alpha_j - c_1 < 0$ , 为了使等式(42)成立, 必须有  $F_j = u_j$ . 可以用两个不等式条件  $F_j \geq u_j$  和  $F_j \leq u_j$  来代替单一的等式约束  $F_j = u_j$ , 由此可得

$$u_j \leq \min_{j: \alpha_j < c_1} F_j \text{ and } u_j \geq \max_{j: \alpha_j < c_1} F_j. \quad (44)$$

对于样本索引  $j$ , 将式(43)和(44)合成为一个不等式约束, 有

$$\max_{j: \alpha_j \leq c_1} F_j \leq u_j \leq \min_{j: \alpha_j < c_1} F_j. \quad (45)$$

最后, 去掉  $u_j$  可得

$$\max_{j: \alpha_j \leq c_1} F_j \leq \min_{j: \alpha_j < c_1} F_j. \quad (46)$$

定义

$$\psi_j = \max_{j: \alpha_j \leq c_1} F_j - \min_{j: \alpha_j < c_1} F_j, \quad (47)$$

由于  $\max_{j: \alpha_j \leq c_1} F_j \geq \min_{j: \alpha_j < c_1} F_j$ , 在可行域中使得  $\alpha_j$  成为问题(37)的最优解的充分必要条件是  $\psi_j = 0$ . 因此, 变量  $\psi_j$  可以帮助选择每一轮所要关注的样本点, 从而构建出特定的小规模优化问题. 在数值实验中, 需要保持存储空间的消耗尽可能小, 因此每个循环只选取一个样本点, 并针对该样本点对应的拉格朗日乘子作更新, 这种思想称为定点思想. 首先选择最大的  $\psi_p$  对应的样本点索引  $p$ , 然后通过求解一维单峰函数优化问题(33)得到最优解  $\alpha_p$  或  $v_p$ . 由于  $\alpha_p$  的改变, 需要

对所有的样本点的  $F_j$  和  $\psi_j$  值作更新. 在数值实验中, 只需要寻找满足  $\psi_j \leq \varepsilon$  条件的  $\alpha_j$  作更新即可,  $\varepsilon$  是预先设置好的精度参数. 因此, 只要还存在至少一个样本点  $(\mathbf{x}_j^{(2)}, y_j^{(2)})$ , 其对应的  $\psi_j$  值比预先设定的阈值  $\varepsilon$  大, 则继续执行算法1的主循环. 经过有限多轮后, 上述优化机制可以保证收敛到最优解, 这是因为每一轮中, 被分解的含有  $n$  个变量的目标函数均会单调递减. 由于空间限制, 读者可以参考文献[24]中提供的相关收敛性证明.

### 2.2 非线性孪生支持向量机(nonlinear FP-TWSVM)

通过构造以下两个核生成曲面, 可以很容易地将“定点”思想从线性TWSVM拓展到非线性情形:

$$K(\mathbf{x}^T, C^T) \mathbf{u}_1 = 0, K(\mathbf{x}^T, C^T) \mathbf{u}_2 = 0. \quad (48)$$

然后可以依照与线性情况类似的推导过程得到非线性分类问题的最优非平行分类器. 由于负类别分类器的求解过程与正类别相同, 不失一般性, 以正类别的求解过程为例. 对应的原始问题如下:

$$\begin{aligned} \min_{\mathbf{u}_1, \eta} & \frac{1}{2} \sum_{i=1}^m (\mathbf{u}_1^T k(\mathbf{x}_i^{(1)}, C^T))^2 + C_3 \sum_{j=1}^n \eta_j; \\ \text{s.t.} & -\mathbf{u}_1^T k(\mathbf{x}_j^{(2)}, C^T) \geq 1 - \eta_j, \\ & \eta_j \geq 0, j = 1, 2, \dots, n. \end{aligned} \quad (49)$$

其中  $c_3 > 0$  是平衡常数.  $K$  是选择适当的核函数, 当使用线性核函数时, 核生成曲面(48)就会退化成线性分类器. 构造拉格朗日函数如下:

$$\begin{aligned} L(\mathbf{u}_1, \boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\varphi}) = & \frac{1}{2} \sum_{i=1}^m \mathbf{u}_1^T k(\mathbf{x}_i^{(1)}, C^T)^2 + C_3 \sum_{j=1}^n \eta_j + \\ & \sum_{j=1}^n \theta_j (1 - \eta_j + \mathbf{u}_1^T k(\mathbf{x}_j^{(2)}, C^T)) + \sum_{j=1}^n \varphi_j (-\eta_j). \end{aligned} \quad (50)$$

与2.1.1节的推导类似, 由KKT条件可得

$$\begin{aligned} \mathbf{u}_1 = & - \left( \sum_{i=1}^m k(\mathbf{x}_i^{(1)}, C^T) k(\mathbf{x}_i^{(1)T}, C^T) \right)^{-1} \times \\ & \left( \sum_{j=1}^n \theta_j k(\mathbf{x}_j^{(2)}, C^T) \right) = \\ & - [k(A^T, C^T) k(A^T, C^T)]^{-1} \times \\ & \left( \sum_{j=1}^n \theta_j k(\mathbf{x}_j^{(2)}, C^T) \right). \end{aligned} \quad (51)$$

将KKT条件代入函数(50)中得到简化的拉格朗日函数(为防止推导过程冗余, 省略了KKT条件的推导过

程)

$$L(\theta) = -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \theta_j \theta_k k(\mathbf{x}_j^{(2)T}, C^T) Q k(\mathbf{x}_k^{(2)}, C^T) + \sum_{j=1}^n \theta_j, \quad (52)$$

其中  $Q = [k(A^T, C^T)k(A, C^T)]^{-1}$ .

综上,原始问题(49)可以被转化成如下的对偶问题:

$$\begin{aligned} \max_{\theta} \quad & -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \theta_j \theta_k k(\mathbf{x}_j^{(2)T}, C^T) \times \\ & Q k(\mathbf{x}_k^{(2)}, C^T) + \sum_{j=1}^n \theta_j; \\ \text{s.t.} \quad & 0 \leq \theta_j \leq C_3, j = 1, 2, \dots, n. \end{aligned} \quad (53)$$

其中  $\theta_j (j = 1, 2, \dots, n)$  是拉格朗日乘子.

非线性情况下,包含分类器的构造与分解策略的 FP-TWSVM 算法框架与算法1类似,只需要将  $\mathbf{x}_j^{(2)}$  替换成  $k(\mathbf{x}_j^{(2)T}, C^T)$ . 因此,为防止冗余,省略了详细的步骤.

非线性情况的分解策略如下. 首先,需要为问题(53)构造拉格朗日函数

$$\begin{aligned} Q(\theta, \delta, \mathbf{u}) = & -\frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \theta_j \theta_k k(\mathbf{x}_j^{(2)T}, C^T) Q k(\mathbf{x}_k^{(2)}, C^T) + \\ & \sum_{j=1}^n \theta_j + \sum_{j=1}^n \delta_j (-\theta_j) + \sum_{j=1}^n u_j (\theta_j - C_3), \end{aligned} \quad (54)$$

其中  $\delta_j$  和  $u_j (j = 1, 2, \dots, n)$  是为约束条件  $\theta_j \geq 0$  和  $\theta_j \leq c_3 (j = 1, 2, \dots, n)$  新引入的拉格朗日乘子. 因此,问题(53)的对偶问题如下:

$$\begin{aligned} \max_{\theta, \delta, \mu} \quad & Q(\theta, \delta, \mu); \\ \text{s.t.} \quad & \delta_j \geq 0, u_j \geq 0, j = 1, 2, \dots, n. \end{aligned} \quad (55)$$

设目标函数(54)关于  $\theta_j$  的偏导数为零,进而寻找使得  $\theta_j$  成为问题(53)的最优解的充分必要条件.

$$\begin{aligned} \frac{\partial Q}{\partial \theta_j} = & -2\theta_j k(\mathbf{x}_j^{(2)T}, C^T) Q k(\mathbf{x}_j^{(2)}, C^T) - \\ & \sum_{k=1, k \neq j}^n \theta_k k(\mathbf{x}_j^{(2)T}, C^T) Q k(\mathbf{x}_k^{(2)}, C^T) + \\ & 1 - \delta_j + u_j = 0. \end{aligned} \quad (56)$$

定义

$$\begin{aligned} E_j = & -2\theta_j k(\mathbf{x}_j^{(2)T}, C^T) Q k(\mathbf{x}_j^{(2)}, C^T) - \\ & \sum_{k=1, k \neq j}^n \theta_k k(\mathbf{x}_j^{(2)T}, C^T) Q k(\mathbf{x}_k^{(2)}, C^T) + 1. \end{aligned} \quad (57)$$

与2.1.3节的推导类似,定义变量  $\omega_j = \max_{j: \theta_j \leq C_3} E_j - \min_{j: \theta_j < C_3} E_j$ ,可以用来选择每一轮次的样本点,并对其拉格朗日乘子作更新. 求解完一系列小规模优化问题以后,可以得到一组最优的拉格朗日乘子  $\theta_j (j = 1, 2, \dots, n)$ ,进而得出最优的权值向量  $\mathbf{u}_1$  (式(51))和正类别对应的最邻近分割超平面. 负类别的最优分割超平面的推导过程与上述过程基本类似. 最后根据判别式  $y_i = \arg \min_{k=1,2} \frac{|k(\mathbf{x}_i^T, C^T) \mathbf{u}_k|}{\|\mathbf{u}_k^-\|}$  来预测新的样本点的标签值  $k$ . 其中:  $\mathbf{u}_k$  是扩展的权值向量,偏移量  $b$  被隐式地添加到该向量的最后一个维度;  $\mathbf{u}_k^-$  是不含偏移量  $b$  的权值向量.

### 3 实现细节

至此,已经讨论了定点孪生支持向量机(FP-TWSVM)模型设计的基本原则和算法细节. 然而,为了使算法更加适用于大规模数据集,更加具有实用性,本文还做了一些实现技术上的改进. 尽管这些改进并不会改变算法的整体框架和设计原则,但是它们可以在实际操作中对算法的表现性能有很大的提升. 因此,该部分将着重介绍这些实现细节上的改进策略.

#### 3.1 通过KKT条件进行样本选择

不失一般性,同样以正类别为例. 对于每一个样本索引  $j$ ,计算出  $F_j$  的值,从而得到  $\psi_j$ . 如2.1.3节所述,在每一个轮次中,选择  $\psi_p$  最大的样本点  $p$  构造一个一维优化问题,然后反复这样的过程,直到所有的  $\psi_j$  值均小于预先设定的精度参数  $\epsilon$ . 事实上,  $\epsilon$  的选择

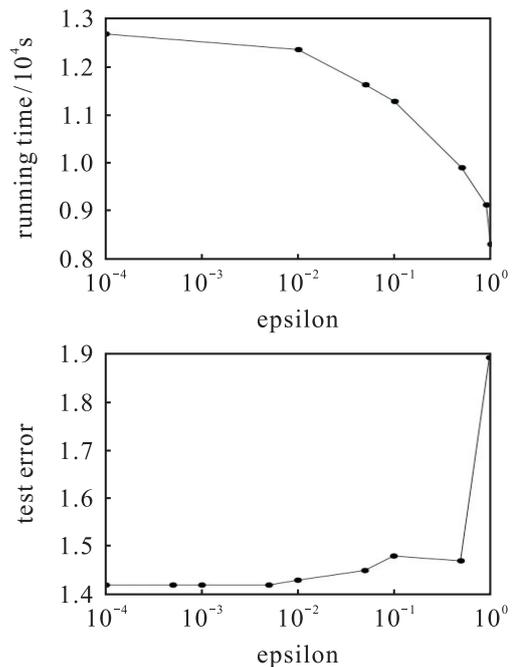


图1 运行时间和测试错误关于精度参数的函数图象

对于算法最终输出结果的影响并不是很大,在一个较大范围内选择 $\varepsilon$ 都可以使算法取得良好的分类表现。 $\varepsilon$ 越大,算法1的主循环终止越早。为了确保算法有良好的泛化能力,将 $\varepsilon$ 设成一个较大的值。图1分别绘制了运行时间(running time)和测试错误(testing error)关于 $\varepsilon$ 的函数图象。这两个函数图象说明, $\varepsilon$ 的中间值0.1具有很好的泛化能力。因此,只要选取的 $\varepsilon$ 值不是非常大,算法的表现性能对于实际的 $\varepsilon$ 取值是不太敏感的。

### 3.2 构建主动集(active set)

基于上述的实现,需要依次扫描负类别所有的训练样本,并计算出它们的 $\psi_j$ 。然而,如果只有一小部分样本对于最优分割超平面的求解有较大影响,也就是说,有很多样本对应的拉格朗日乘子 $\alpha_j$ 为零,将会花费很多时间在那些对最终解没有贡献的样本点上。考虑到这个问题,将负类别样本点划分成两个互不相交的子集。第1个子集记为 $M$ ,称为主动集,由那些对最优解有贡献的样本点组成,因此 $M = \{\mathbf{x}_j | \alpha_j \neq 0\}$ 。第2个集合是主动集 $M$ 的补集,也就是 $M^c = \{\mathbf{x}_j | \alpha_j = 0\}$ 。在执行主循环时,首先根据3.1节所述的利用KKT条件从主动集中选择一个样本点,对它的拉格朗日乘子作更新。只有当主动集中这样的样本点不存在时,意味着 $\forall j \in M, \psi_j < \varepsilon$ ,才会从补集中选择 $\psi_p$ 最大的样本点,将其移出补集,加到主动集中,用该样本点构造一个一维单峰函数优化问题,对 $\alpha_p$ 作一次更新。整个过程将大量的时间用于更新主动集中的样本点对应的拉格朗日乘子,并且只有当主动集中没有符合条件的样本点作更新时才会去访问补集,并将补集中 $\psi_p$ 最大的样本点加入到主动集中。这种构建主动集的策略可以大大降低运行时间。

### 3.3 调节精度参数

尽管主动集的引入可以显著降低时间消耗,但是这种机制仍然要迫使算法不断更新主动集中的 $\alpha_j$ 值,只要主动集中还存在至少一个样本点索引 $j$ ,满足 $\psi_j > \varepsilon$ 。这样的做法可能会导致目标函数 $Q$ 减小的速率很慢。图2绘制了当精度阈值为固定值时,拉格朗日函数 $Q$ 关于轮次总数的函数图象,如实线所示,该函数图象呈阶梯状。仔细观察目标函数 $Q$ 明显降低时对应的轮次可以发现,在这些轮次中,新的样本点刚好被加进了主动集中。当一个新的样本点被加入到主动集中后,又需要很多的轮次来更新主动集中的样本点对应的 $\alpha_j$ ,这就使得目标函数 $Q$ 在后续的一段时间内变化量不大。为了加速这一过程,逐渐改变

精度参数 $\varepsilon$ 的值而不是使 $\varepsilon$ 在整个算法过程中保持不变。在算法前阶段,将 $\varepsilon$ 设成一个较大的值,根据3.2节所述,可以频繁地向主动集中添加新的样本点。这样一来,在前阶段,算法就可以平衡对所有训练样本的关注度,而不会花费很多时间在初始化步骤所产生的主动集样本上。随着轮次的增加,逐渐降低阈值 $\varepsilon$ ,将算法的注意力从总体样本中转移到支持向量上,也就是主动集样本上。将这一缓慢减小精度参数的过程称为cooling,并选取对数函数 $\varepsilon(t) = \varepsilon_0 / \log_{10}(t + 10)$ 作为算法的cooling规则。该函数在 $t$ 较小时保持一个相对较大的值,继而缓慢减小。这种cooling策略使得目标函数 $Q$ 的值下降更平稳光滑,下降速率更快,如图2虚线所示。该策略将 $Q$ 的下降速率提高了一个数量级,使得整个优化算法可以更快地收敛。

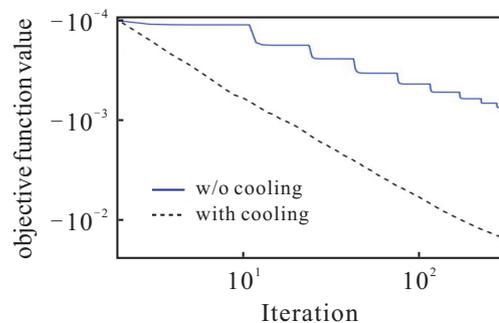


图2 在精度参数值固定和可变情况下,目标函数 $Q$ 关于轮次数量的函数

## 4 数值实验

为了验证本文算法的分类效果,将FP-TWSVM与TWSVM<sup>[8]</sup>、TBSVM<sup>[9]</sup>和 $v$ -TWSVM<sup>[11]</sup>作比较来验证新算法在训练速度(训练时间)和存储效率方面的优越性。本文将从分类精度和计算效率(反应在平均训练时间上)的角度对各个算法的性能进行衡量。实验使用的是具有3.40 GHz CPU和16 GB RAM的PC机,在Microsoft Windows环境中运行,使用的是Matlab 2012a版本。

### 4.1 参数设置

在进行实验之前,首先进行最优参数选择。在后续实验中,采用网格搜索法<sup>[25]</sup>对参数 $c_i$ ( $i = 1, 2, 3, 4$ )进行调节。其中: $c_1$ 和 $c_2$ 分别是正类别在线性和非线性情况下的平衡常数, $c_3$ 和 $c_4$ 分别是负类别在线性和非线性情况下的平衡常数。对于非线性情况,采用的是高斯核函数 $\exp(-\sigma \|\mathbf{x} - \mathbf{y}\|^2)$ ,其中 $\sigma$ 是核宽度参数,同样需要与 $c_i$ ( $i = 1, 2, 3, 4$ )一起进行参数调节。为了避免偏倚和过度拟合情况的发生,采用10折交叉验证技术<sup>[26-28]</sup>来选择 $c_i$ ( $i = 1, 2, 3, 3$ )和 $\sigma$ 的最优参数值。对于每一个数据集,随机抽取10%的样本点

构成验证集. 在实验中, 从  $\{10^i, i = -5 \text{ to } -1\}$  范围中选择  $c_1$  和  $c_3$  的最优参数值, 从  $\{0.1 - 1\}$  范围中调节参数  $c_2$  和  $c_4$  的数值. 对于每一个数据集, 在10个验证集上计算平均分类精度, 分类精度 accuracy 有如下定义:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (58)$$

其中 TP、TN、FP 和 FN 分别是分类正确的正类别点 (true positive)、分类正确的负类别点 (true negative)、分类错误的负类别点 (false positive) 和分类错误的正类别点 (false negative) 的样本总数.

### 4.2 标准数据集

为了验证新算法的有效性, 在一系列数据集上, 包括 UCI 和 NDC 这些经常用来作机器学习算法测试

的数据集上进行数值实验. 其中一些 NDC 数据集规模较大, 有利于凸显出新算法在训练速度 (由训练时间反应) 方面相对于其他算法的优越性. 此外, 样本点在训练之前进行了归一化, 使得每个维度的特征值都集中在  $[0, 1]$  范围内, 并分别采用线性核和高斯核获得最优非平行超平面.

#### 4.2.1 IUCI 数据集

选择 8 个用于二元分类的 UCI 数据集, 数据集特性如表 1 所示. 正类别和负类别样本总数的比值用  $(N_P : N_N)$  来表示. 数据集 Breast Cancer Wisconsin (prognostic) 简称为 WPBC, Australian Credit Approval 简称为 ACA, Contraceptive Method Choice 简称为 CMC, Connectionist Bench 简称为 Sonar, Congressional Voting Records 简称为 Votes.

表 1 在 UCI 数据集上的线性分类结果

dataset	$N_p : N_N$	TWSVM		TBSVM		$v$ -TWSVM		FP-TSVM	
		MA/%	TT/ $10^{-3}s$	MA/%	TT/ $10^{-3}s$	MA/%	TT/ $10^{-3}s$	MA/%	TT/ $10^{-3}s$
Heart-Statlog(270x13)	150:120	90.54 ± 4.83	5.78	92.23 ± 2.86	3.42	<b>94.48</b> ± 3.66	4.03	90.79 ± 4.72	<b>2.96</b>
WPBC(698x34)	458:240	93.60 ± 5.02	16.54	<b>97.40</b> ± 1.54	15.03	96.78 ± 2.47	16.28	93.97 ± 3.93	<b>10.56</b>
PIMA-Indians(768x8)	500:268	81.79 ± 3.69	8.95	<b>86.59</b> ± 3.62	8.44	84.43 ± 4.95	8.60	80.56 ± 4.62	<b>4.79</b>
CMC(1473x9)	844:629	77.54 ± 2.38	22.47	<b>77.72</b> ± 4.23	19.52	77.06 ± 3.48	21.48	77.48 ± 4.21	<b>15.66</b>
Heart-Cleveland (303x14)	164:139	91.92 ± 2.55	5.23	93.24 ± 2.75	4.99	92.64 ± 4.53	5.02	91.78 ± 3.27	<b>1.53</b>
Votes (435x16)	267:168	93.60 ± 3.81	7.69	95.29 ± 3.86	6.72	<b>95.39</b> ± 4.57	7.30	93.66 ± 3.98	<b>3.62</b>
Sonar(208x60)	111:97	85.98 ± 5.24	5.93	<b>86.54</b> ± 5.62	5.03	86.32 ± 6.55	5.51	85.78 ± 3.78	<b>2.36</b>
Lonosphere(351x34)	225:126	94.66 ± 3.39	9.22	96.43 ± 3.52	8.45	<b>97.02</b> ± 4.72	8.98	94.83 ± 5.22	<b>3.84</b>
avg.acc		88.70 ± 3.86		90.68 ± 3.50		90.51 ± 4.37		88.61 ± 4.22	
avg.time		10.25		8.95		9.65		5.67	

表 2 在 exp-NDC 数据集上的线性分类结果

Dataset	$(N_p : N_N)$	train-test	TWSVM		TBSVM		$v$ -TWSVM		FP-TWSVM	
			MA/%	TT/ $10^{-3}s$	MA/%	TT/ $10^{-3}s$	MA/%	TT/ $10^{-3}s$	MA/%	TT/ $10^{-3}s$
exp-NDC-500	(330:170)	500-50	87.79 ± 5.78	1.88	88.33 ± 4.55	1.65	88.04 ± 4.53	1.73	<b>87.63</b> ± 4.23	0.02
exp-NDC-700	(447:253)	700-70	86.57 ± 3.90	2.98	<b>88.14</b> ± 3.52	2.82	87.53 ± 3.48	2.9	86.32 ± 3.26	<b>0.03</b>
exp-NDC-900	(571:329)	900-90	85.59 ± 3.94	5.03	<b>87.44</b> ± 4.38	4.59	86.58 ± 3.21	4.88	85.61 ± 3.66	<b>2.53</b>
exp-NDC-1k	(627:373)	1k-100	86.40 ± 4.83	5.91	<b>88.90</b> ± 3.42	5.51	87.10 ± 3.74	5.81	86.54 ± 4.02	<b>2.29</b>
exp-NDC-2k	(1246:754)	2k-200	88.42 ± 2.95	5.83	88.53 ± 2.32	5.23	<b>88.72</b> ± 3.21	5.45	88.11 ± 3.25	<b>2.56</b>
exp-NDC-3k	(1860:1140)	3k-300	88.25 ± 2.41	9.33	<b>88.89</b> ± 4.13	8.41	88.44 ± 4.56	8.72	88.23 ± 3.24	<b>3.54</b>
exp-NDC-4k	(2474:1526)	4k-400	88.23 ± 1.88	12.21	<b>89.41</b> ± 1.25	10.25	88.35 ± 1.29	11.36	88.48 ± 1.01	<b>6.84</b>
exp-NDC-5k	(3086:1914)	5k-500	87.27 ± 2.45	18.54	<b>88.54</b> ± 3.64	16.32	87.45 ± 3.21	16.21	86.59 ± 2.58	<b>7.3</b>
exp-NDC-10k	(6138:3862)	10k-1k	94.29 ± 1.54	92.45	<b>96.33</b> ± 2.87	89.22	95.05 ± 1.78	92.63	94.31 ± 1.65	<b>41.5</b>
exp-NDC-50k	(30783:19217)	50k-5k							<b>87.41</b> ± 3.56	<b>314.15</b>
exp-NDC-100k	(61648:38352)	100k-10k							84.20 ± 2.15	774.42

#### 4.2.2 lexp-NDC datasets

NDC数据集是由David Musicants NDC数据生成器生成,数据集中的样本成正态分布.为了方便测试新算法的有效性,将对数正态分布作为转换函数,对NDC数据集作倾斜化处理.根据概率论,如果一个连续型随机变量的对数服从正态分布,则该随机变量本身服从对数正态分布.由此可得,如果随机变量 $X_1$ 服从对数正态分布,则 $X_2 = \ln|X_1|$ 服从正态分布;同样地,如果随机变量 $X_2$ 服从正态分布,则 $X_1 = \exp(X_2)$ 服从对数正态分布,该概率密度函数向右倾斜.因此,倾斜化的NDC数据集称为exp-NDC,每个样本点有32个特征维度,并且一个NDC数据集的规模可以在500~100 000范围内浮动.为了体现新算法在训练速度上的优越性,在一组规模递增的exp-NDC数据集上运行各个算法.关于exp-NDC数据集的其他详细特性见表2.

#### 4.3 分类结果:UCI数据集和exp-NDC数据集

##### 4.3.1 线性结果

算法FP-TWSVM、TWSVM、TBSVM和 $v$ -TWSVM在线性情形下的实验结果记录在表1中.黑色字体代表每一个数据集在分类精度或者训练时间方面的最优结果.对于分类精度,取的是每个数据集的10折验证集上的平均值,其结果和标准差一起记录在表1中,简记为MA,用百分数表示.对于表1中的大部分数据集,FP-TWSVM的平均分类精度与TWSVM相当,略低于TBSVM和 $v$ -TWSVM.FP-TWSVM算法在表1中所有的数据集上的平均精度为88.61%,而TWSVM、TBSVM和 $v$ -TWSVM在所有数据集上的平均分类精度分别是88.70%、90.68%和90.51%.由此可得,在大多数UCI数据集上,使用线性分类器时,相比于TWSVM和它的变体模型,FP-TWSVM同样可以取得理想的分类精度.在分类精度上,exp-NDC数据集上的实验结果同UCI数据集类似.

本文提出的算法改进了TWSVM模型,旨在加速训练时间,并减少内存和时间的消耗,尤其是在处理大规模数据集时,通过将标准QP分解成一系列连续的一维单峰函数优化问题,避免存储大规模的二次型矩阵,可以显著提升存储效率,加快训练时间,提高处理数据的能力.因此,从直觉上可以推断出,FP-TWSVM在训练时间、存储效率和处理数据的能力上明显要胜于TWSVM及其变体模型.为了验证这一观点,在表1的UCI数据集上记录了每

个算法的训练时间(简记为TT).在规模适中的UCI数据集上,训练时间由平均CPU时间来表示,用的是10折交叉验证技术来计算时间.表1记录的实验结果表明,对于每一个数据集,FP-TWSVM计算出最优分割超平面所需的时间明显少于其他算法.FP-TWSVM在所有UCI数据集上的平均训练时间为 $5.67 \times 10^{-3}$ s, TWSVM、TBSVM和 $v$ -TWSVM在所有UCI数据集上的平均训练时间分别为 $10.25 \times 10^{-3}$ s、 $8.95 \times 10^{-3}$ s和 $9.65 \times 10^{-3}$ s.

为了更好地体现新算法的时间高效性和存储高效性,同样在一组规模递增的exp-NDC数据集(包含了大规模数据集)上进行了线性分类实验,实验结果如表2所示.训练样本总数和测试样本总数记为“train-test”,正类别样本总数和负类别样本总数之比记为 $N_P : N_N$ .在保持与其他算法相当的分精度的条件下,FP-TWSVM的训练速度明显快于其他算法,并且具有比其他3种算法更高的存储效率,尤其是在处理大规模数据集(exp-NDC-10k, exp-NDC-50k, exp-NDC-100k)时,这种优势体现得更为明显.这些优越性归功于新算法中采用的分解策略、样本选择策略以及其他实现技术上的改进,这些改进加速了算法的收敛.随着exp-NDC数据集规模的逐渐增大,FP-TWSVM的训练时间也在逐渐增加,然而,训练时间的增加速率比其他算法要明显缓慢.值得注意的是,在一些超大规模的数据集(例如exp-NDC-50k和exp-NDC-100k数据集)上,其他3种算法由于内存溢出而无法得出最终分类结果,而FP-TWSVM仍然可以在一个合理的时间范围内产生最优分割超平面,这无疑证实了FP-TWSVM在面对大规模数据集时的存储高效性,反应了它强大的数据处理能力.

##### 4.3.2 非线性结果

通过引入两个非平行核生成曲面,新算法可以很容易地被推广到非线性分类问题中.非线性情况下,4种算法在UCI数据集和exp-NDC数据集上的分类精度和训练时间记录在表3和表4中.

所有算法均采用高斯核函数 $K(x, x') = \exp(-\sigma \|x - x'\|^2)$ ,并且选取的是相同的核宽度参数.实验结果同样证实了非线性的FP-TWSVM消耗的CPU时间最少,占有的存储空间最小,同时又可以保持与其他算法相当的分精度.在UCI和exp-NDC数据集上最好的平均训练时间分别是 $8.13 \times 10^{-3}$ s和 $3.53 \times 10^{-2}$ s,均为FP-TWSVM所达到的训练时间,再次说明了FP-TWSVM强大的数据处理能力.

表 3 在 UCI 数据集上的非线性分类结果

dataset	TWSVM		TBSVM		v-TWSVM		FP-TWSVM	
	MA/%	TT/10 <sup>-3</sup> s	MA/%	TT/10 <sup>-3</sup> s	MA/%	TT/10 <sup>-3</sup> s	MA/%	TT/10 <sup>-3</sup> s
Heart-Statlog	85.75 ± 1.46	9.81	<b>87.23</b> ± 4.25	8.32	86.54 ± 3.84	8.56	85.98 ± 3.78	<b>3.19</b>
WPBC	96.12 ± 1.38	26.19	<b>97.59</b> ± 3.54	24.36	96.98 ± 2.57	25.85	96.38 ± 2.07	<b>12.34</b>
PIMA-Indians	86.25 ± 3.54	12.36	88.52 ± 4.23	10.54	<b>88.83</b> ± 1.54	12.01	86.41 ± 3.58	<b>5.32</b>
CMC	80.96 ± 2.31	48.41	<b>84.21</b> ± 4.56	36.78	81.44 ± 2.54	40.63	82.81 ± 2.55	<b>25.22</b>
Heart-Cleveland	93.45 ± 2.50	9.87	94.77 ± 3.21	8.56	<b>94.83</b> ± 1.25	8.99	94.02 ± 4.46	<b>3.58</b>
Votes	95.21 ± 2.76	13.45	<b>96.27</b> ± 1.57	10.08	96.04 ± 2.12	11.2	95.25 ± 4.89	<b>3.96</b>
Sonar	93.45 ± 4.33	15.17	<b>96.42</b> ± 5.21	12.44	95.25 ± 3.62	12.75	93.39 ± 4.29	<b>5.43</b>
Ionosphere	94.39 ± 3.24	13.98	<b>96.88</b> ± 4.22	11.54	96.01 ± 3.89	12.35	94.26 ± 4.32	<b>5.98</b>
accuracy(mean)	90.70 ± 2.69		92.74 ± 3.85		91.99 ± 2.67		91.06 ± 3.74	
time(mean)	18.66		15.33		16.54		8.13	

表 4 在 exp-NDC 数据集上的非线性分类结果

dataset	TWSVM		TBSVM		v-TWSVM		FP-TWSVM	
	MA/%	TT/10 <sup>-2</sup> s	MA/%	TT/10 <sup>-2</sup> s	MA/%	TT/10 <sup>-2</sup> s	MA/%	TT/10 <sup>-2</sup> s
exp-NDC-500	87.45 ± 6.03	4.69	<b>92.98</b> ± 5.98	4.24	88.75 ± 4.47	4.54	87.56 ± 4.32	<b>1.54</b>
exp-NDC-700	88.29 ± 5.33	5.98	<b>91.86</b> ± 4.21	5.21	89.86 ± 3.66	5.78	88.59 ± 4.16	<b>2.56</b>
exp-NDC-900	89.52 ± 5.22	8.55	<b>93.89</b> ± 5.63	7.12	90.78 ± 3.21	7.67	89.65 ± 3.87	<b>3.56</b>
exp-NDC-1k	93.26 ± 3.96	9.78	<b>95.11</b> ± 3.02	9.12	94.69 ± 3.74	9.56	93.01 ± 3.92	<b>4.29</b>
exp-NDC-2k	94.12 ± 2.63	11.3	98.05 ± 2.02	10.08	<b>98.41</b> ± 1.35	10.41	94.01 ± 2.69	<b>5.7</b>
accuracy(mean)	90.53 ± 4.63		94.38 ± 4.17		92.50 ± 3.29		90.56 ± 3.79	
time(mean)	8.06		7.15		7.59		3.53	

### 5 结 论

本文提出了一种孪生支持向量机的改进模型,称为定点孪生支持向量机(fixed-point twin support vector machine, FP-TWSVM). 通过将每个QPP分解成一系列有限多个连续的一维优化问题,显著提高了训练速度,减少了存储空间. 其中,每个一维的优化问题都是一个单峰函数最小化问题,可以使用高效的线性搜索方法(例如斐波那契算法或者黄金分割法)求出问题的最佳解. 另外,本文提出了一些实现技术上的改进,例如样本选择策略、构建主动集以及调节精度参数的cooling技术,均有利于加快算法的收敛. 此外,通过引入两个非平行核生成曲面,可以实现FP-TWSVM从线性分类情形到非线性分类情形的推广. 实验结果表明,在保持与其他算法相当的分精度的同时,FP-TWSVM相比于其他算法具有明显的时间高效性和存储高效性,因此更加适用于处理大规模数据集.

本文提出的这种“定点”思想以及相关的建立在特定算法设计上的实现技术改进策略,是否同样适用于TWSVM其他的变体模型将是下一步的研究工作.

### 参考文献(References)

- [1] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [2] Bertsekas D P. Nonlinear programming[J]. Journal of the Operational Research Society, 1997, 48(3): 334-334.
- [3] Abe S. Support vector machines for pattern classification[M]. London: Springer, 2005.
- [4] Nobel W S, Scholkopf B, Tsuda K, et al. Support vector machine applications in computational biology[J]. Kernel Methods in Computational Biology, 2004: 71-92.
- [5] Lal T N, Schroder M, Hinterberger T, et al. Support vector channel selection in BCI[J]. IEEE Transactions on Biomedical Engineering, 2004, 51(6): 1003-1010.
- [6] Trafalis T B, Ince H. Support vector machine for regression and applications to financial forecasting[C]. Proceedings of the IEEE-INNS-ENNS International Joint

- Conference on Neural Networks. Como: IEEE, 2000: 348-353.
- [7] Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis[M]. Cambridge: Cambridge University Press, 2004.
- [8] Khemchandani R, Chandra S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910.
- [9] 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.
- [10] Tian Y, Ju X, Qi Z, et al. Improved twin support vector machine[J]. Science China Mathematics, 2014, 57(2): 417-432.
- [11] Peng X J. A  $v$ -twin support vector machine ( $v$ -TWSVM) classifier and its geometric algorithms[J]. Information Sciences, 2010, 180(20): 3863-3875.
- [12] Peng X. Least squares twin support vector hypersphere (LS-TSVH) for pattern recognition[J]. Expert Systems with Applications, 2010, 37(12): 8371-8378.
- [13] Peng X, Xu D. A twin-hypersphere support vector machine classifier and the fast learning algorithm[J]. Information Sciences, 2013, 221: 12-27.
- [14] Xu Y, Yang Z, Zhang Y, et al. A maximum margin and minimum volume hyper-spheres machine with pinball loss for imbalanced data classification[J]. Knowledge-Based Systems, 2016, 95: 75-85.
- [15] Peng X, Kong L, Chen D. A structural information-based twin-hypersphere support vector machine classifier[J]. International Journal of Machine Learning and Cybernetics, 2017, 8(1): 295-308.
- [16] Pan X, Luo Y, Xu Y. K-nearest neighbor based structural twin support vector machine[J]. Knowledge-Based Systems, 2015, 88: 34-44.
- [17] Xu Y, Wang L, Zhong P. A rough margin-based  $v$ -twin support vector machine[J]. Neural Computing and Applications, 2012, 21(6): 1307-1317.
- [18] Wang H, Zhou Z. An improved rough margin-based  $v$ -twin bounded support vector machine[J]. Knowledge-Based Systems, 2017, 128: 125-138.
- [19] Cao L, Shen H. Combining Re-sampling with twin support vector machine for imbalanced data classification[C]. The 17th International Conference on Parallel and Distributed Computing, Applications and Technologies. IEEE, 2016: 325-329.
- [20] Cao Q, Fu X, Guo Y. Fuzzy chance constrained twin support vector machine for uncertain classification[C]. International Conference on Management Science and Engineering Management. Cham: Springer, 2017: 1508-1521.
- [21] Khemchandani R. Mathematical programming applications in machine learning[D]. New Delhi: Department of Computer Science and Engineering, Indian Institute of Technology Delhi, 2008.
- [22] Fung G M, Mangasarian O L. Multicategory proximal support vector machine classifiers[J]. Machine learning, 2005, 59(1/2): 77-97.
- [23] Cristianini N, John S T. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge: Cambridge University Press, 2000.
- [24] Lin C J. A formal analysis of stopping criteria of decomposition methods for support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(5): 1045-1052.
- [25] Lin C J. A practical guide to support vector classification[R]. Freiburg: University of Freiburg, 2003.
- [26] Duda R O, Hart P E, Stork D G. Pattern classification and scene analysis[M]. New York: Wiley, 1973.
- [27] 储茂祥, 王安娜, 巩荣芬. 一种改进的最小二乘孪生支持向量机分类算法[J]. 电子学报, 2014, 42(5): 998-1003.  
(Chu M X, Wang A N, Gong R F. An improved least square twin support vector machine classification algorithm[J]. Chinese Journal of Electronics, 2014, 42(5): 998-1003.)
- [28] 丁世飞, 黄华娟, 史忠植. 加权光滑CHKS孪生支持向量机[J]. 软件学报, 2013, 11: 2548-2557.  
(Ding S F, Huang H J, Shi Z Z. Weighted smooth CHKS twin support vector machines[J]. Journal of Software, 2013, 11: 2548-2557.)

### 作者简介

刘峤(1974—), 男, 教授, 博士生导师, 从事机器学习和知识图谱等研究, E-mail: 693083139@qq.com;

方佳艳(1997—), 女, 本科生, 从事机器学习和人工智能的研究, E-mail: fyk80@163.com.

(责任编辑: 齐 霖)