基于相似度差的大间隔快速学习模型

应文豪[1,2] 王士同[1]

(江南大学数字媒体学院 无锡 214122)[1] (常熟理工学院计算机科学与工程学院 常熟 215500)[2]

摘 要 许多模式分类方法比如支持向量机和 L_2 核分类器等都会利用核方法并转化为二次规划问题进行求解,而计算核矩阵需要 $O(m^2)$ 的空间复杂度,求解 QP 问题则需要 $O(m^3)$ 的时间复杂度,这就使得此类方法在大样本数据上的学习性能非常低下。对此,首次提出了相似度差支持向量机算法 DSSVM。算法旨在寻求样本与某类相似度的一个最佳线性表示,并从线性表示的稀疏性以及相似度差意义上的间隔最大化角度构造了新的最优化问题。同时,证明了该算法等价于中心约束型最小包含球问题,这样就可以通过引入最小包含球的快速学习理论将相似度差支持向量机扩展为相似度差核支持向量机 DSCVM,从而较好地解决了大规模数据集的分类问题。实验证明了相似度差支持向量机和相似度差核支持向量机的有效性。

关键词 相似度差,稀疏,核心集,最小包含球,支持向量机

中图法分类号 TP181 **文献标识码** A

Large Margin and Fast Learning Model Based on Difference of Similarity

YING Wen-hao[1,2] WANG Shi-tong[1]

(School of Digital Media, Jiangnan University, Wuxi 214122, China)[1]

(School of Computer Science and Engineering, Changshu Institute of Technology, Changshu 215500, China)[2]

Abstract Many pattern classification methods, such as support vector machine and L_2 kernel classification, often use the kernel methods and are formulated as quadratic programming problems, but computing kernel matrix would require $O(m^2)$ computation, and solving QP may take up to $O(m^3)$, which limits these methods to train on large datasets. In this paper, a new classification method called difference of similarity support vector machine (DSSVM) was proposed. DSSVM pursues a best linear representation of a total similarity between any sample and a particular class. According to the sparsity of the linear representation and the max margin of the difference of similarity, a new optimization problem is obtained. Meanwhile, the difference of similarity support vector machine can be equivalently formulated as the center constrained minimal enclosing ball, and thus difference of similarity support vector machine can be extended to difference of similarity core vector machine (DSCVM) by introducing fast learning theory of minimal enclosing ball, to solve the classification for large datasets. This is confirmed in the experimental studies.

Keywords Difference of similarity, Sparsity, Core set, Minimum enclosing ball, Support vector machine

1 引言

模式分类问题是机器学习领域重要的研究方向之一。对于一个模式分类机,通常会引入训练样本并赋予不同权重来进行决策函数的构造。然而,随着训练样本集规模的增大,大量的训练样本被引入用来构造决策函数势必将降低该决策函数的决策效率。为提高决策效率,一种有效的方法是选用具有代表性的样本子集取代整个训练样本集来构造决策函数。如由 Vapnik 等人提出了支持向量机(Support Vector Machine, SVM)[1-4],SVM 通过寻找一个最优超平面来将两类样本分开,而该分类超平面的构造仅仅依赖于支持向量;David 等人在 2004 年提出支持向量数据描述(Support Vector Data Description, SVDD)[5],SVDD 旨在寻找一个最小超球来包含所有的目标样本,该超球的构造也仅依赖于支持向量;JooSeuk Kim 等人于 2009 年提出了 L_2 核分类器(L_2 kernel classification)[6],L_2 核分类器通过最小化正负类样本概率密度差和真实的密度差之间的 L_2 距离(ISE)来获得决策函数的参数 α。同样,α 的稀疏性使得 L_2 核分类器的决策函数也只依赖于部分训练样本。这种稀疏特性对于解决大规模样本问题具有重要的意义。

最近,Allen Y. Yang 等人提出了稀疏分类器(SC)[7],与 SVM,SVDD 等方法不同的是,SC 直接从参数稀疏的角度出发定义优化问题的目标函数。SC 假设一个新的测试样本可以近似地由该类训练样本的线性组合来表示,并且这一线性组合体现了稀疏性,即每类训练样本都具有某些子集,测试样本可由子集中的样本线性表示。基于 SC, Majumdar 等于 2009 年提出了组稀疏分类器(GSC)[8,9]。SC 与 GSC 算法的不同之处在于对优化问题目标函数的选择。SC 算法假设权

到稿日期:2012-10-31 返修日期:2013-03-13 本文受国家自然科学基金项目(61272210,61202311)资助。

应文豪(1979-),男,博士生,讲师,主要研究方向为模式识别、人工智能,E-mail:cslgywh@163.com;王士同(1964-),男,硕士,教授,主要研究方向为模式识别、模糊系统等。

向量 α 是稀疏的，使用权重向量的 l_1 范数（$\|\alpha\|_1$）来定义目标函数[10,11]；而 GSC 则定义每类权重向量的 l_2 范数为目标函数。

受 SC 及 GSC 的启发，本文提出了一种全新的分类器算法，即相似度差支持向量机（Difference of Similarity Support Vector Machine，DSSVM）。DSSVM 用一个样本与某类内所有样本的相似度的线性组合来表示该样本与该类的相似度，或称该样本隶属于该类的隶属度。DSSVM 的目标是使这个线性组合最优化，并尽可能要求正负类样本在相似度差意义上的 Margin 最大化。由于 DSSVM 本质上也是求解 QP 问题，同 SVM 和 SVDD 一样，DSSVM 对求解较大规模样本问题没有优势，但本文通过研究 DSSVM 原问题的对偶问题，发现其等价于 CC-MEB 问题，因而利用最小包含球的逼近理论和算法可以处理大样本数据集，从而得到 DSSVM 的另一个版本，即相似度差核心集向量机（Difference of Similarity Core Vector Machine，DSCVM）。

本文第 2 节介绍 DSSVM 算法及其核化形式；第 3 节回顾 MEB 和 CC-MEB；第 4 节揭示核化的 DSSVM 与 MEB 间的关系；第 5 节描述 DSSVM 及 DSCVM 的实验过程；最后总结全文。

2 相似度差支持向量机（DSSVM）

本文所提到的相似度有两种含义，一种是样本与样本间的相似度，它具有多种定义方式，比如向量空间余弦相似度（Cosine Similarity）、皮尔森相关系数（Pearson Correlation Coefficient）、Jaccard 相似系数（Jaccard Coefficient）等，用 $s(x_i,x_j)$ 来表示；另一种是样本与类的相似度（即一个样本隶属于某类的隶属度），用 $S(x)$ 表示。为了区分两种不同含义的相似度，下文将第二种相似度表达为隶属度。

图 1 从相似度的角度描绘了本文的核心思想。从图 1 可以看到，样本点 $p1$ 隶属于正类的隶属度可用 $p1$ 和正类内的样本点的相似度来表示；然而，并不是所有的样本点都应对隶属度的表示具有贡献，$p2$，$p3$，$p4$，$p5$ 由于离 $p1$ 很近，应对 $p1$ 的隶属度表示做较大贡献，而 $p6$，$p7$，$p8$，$p9$ 离 $p1$ 较远，应对 $p1$ 隶属度表示做较小贡献或不做贡献；从另外的角度来观察，对于负类样本点 $n1$，如果要描绘其隶属于正类的隶属度，则 $p2$，$p3$，$p4$，$p5$ 反而应该做较小的贡献或不做贡献，而 $p6$，$p7$，$p8$，$p9$ 应该做较大贡献。因此可得如下结论：(1)样本点隶属于某类的隶属度可由该样本点与类内样本点的相似度来表示，且可用不同的权值来刻画各样本点对隶属度的贡献程度；(2)这种隶属度的表示方法具有稀疏性，即并不是所有的样本点对隶属度的表示都应具有贡献。

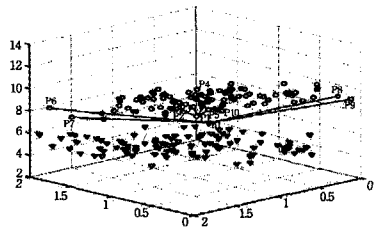

图 1 DSSVM 示意图

2.1 DSSVM 原始形式

对于两类分类问题，假设已知一个包含两类的样本集，$X\subseteq R^d$，$X=\{x_i|x_i\in R^d\}$，其中 $x_i=(x_{i1},x_{i2},\cdots,x_{id})^T$ 为列向量。样本类标签为 $y_i\in\{+1,-1\}$，当 $i\in I^+$ 时，$y_i=+1$，当 $i\in I^-$ 时，$y_i=-1$；并且已知正类样本间的相似度函数为 $s_+(x_i,x_j)\geqslant 0(i,j\in I^+)$，负类样本间的相似度函数为 $s_-(x_i,x_j)\geqslant 0(i,j\in I^-)$。

定义 $S_+(x)$ 表示样本 x 隶属于正类的隶属度，$S_-(x)$ 表示样本 x 隶属于负类的隶属度。为简单起见，根据上述思想，$S_+(x)$ 用样本 x 与 $x_i(i\in I^+)$ 间的相似度来线性表示，$S_-(x)$ 用 x 与 $x_i(i\in I^-)$ 间的相似度来线性表示，即

$$S_+(x;\alpha)=\sum_{i\in I^+}\alpha_i s_+(x,x_i) \tag{1}$$

$$S_-(x;\beta)=\sum_{i\in I^-}\beta_i s_-(x,x_i) \tag{2}$$

权值 $\alpha_i(i\in I^+)$ 表示第 i 个正类样本对 $S_+(x)$ 的贡献程度；权值 $\beta_i(i\in I^-)$ 表示第 i 个负类样本对 $S_-(x)$ 的贡献程度；定义 α 为权值 α_i 所构成的向量，β 为权值 β_i 所构成的向量。根据 GSC 的稀疏性思想，将大间隔的基本方法[12]融入本文的核心思想，构造如下最优化问题：

$$\begin{aligned}&\min\ \|\alpha\|_2^2+\|\beta\|_2^2-v\rho\\&\text{s.t.}\\&S_+(x_i)-\lambda S_-(x_i)>\rho,i\in I^+\\&S_+(x_i)-\lambda S_-(x_i)<-\rho,i\in I^-\\&\sum_{i\in I^+}\alpha_i+\sum_{i\in I^-}\beta_i=1\end{aligned} \tag{3}$$

式中，λ 为人为设定的参数，用来反映类别的先验概率；ρ 称为相似度差（Difference of Similarity）意义上的间隔 Margin；v 为调节参数。最优化问题(3)中的目标函数最小化可以使参数 α,β 获得稀疏化[8]，同时使正负类样本在相似度差意义上的间隔最大化；约束条件的含义为当训练样本属于正类时，该样本隶属于正类的隶属度应该大于该样本隶属于负类的隶属度，且其间隔应大于 ρ。当训练样本属于负类时，该样本隶属于正类的隶属度应该小于该样本隶属于负类的隶属度，且其间隔也应大于 ρ。此时，最大化间隔 ρ 可使正负类样本尽量分开。根据决策理论，最优分类器具有如下形式：

$$g(x)=\text{sgn}\{S_+(x)-\lambda S_-(x)\} \tag{4}$$

为了求解上述最优化问题，现给出定理 1。

定理 1 DSSVM 的原始优化形式为：

$$\begin{aligned}&\min\frac{1}{2}\|w\|_2^2-v\rho\\&\text{s.t.}\\&y_i(w^TX_i)\geqslant\rho,i=1,\cdots,N\\&w^T\mathbf{1}=1,\mathbf{1}=(1,1,\cdots,1)^T_{1\times N}\end{aligned} \tag{5}$$

证明：首先根据相似度函数定义一个新的空间，样本 x_i 映射到新空间后得到一个新的向量 X_i，即

$$\begin{aligned}X_i=&[s_+(x_i,x_1),\cdots,s_+(x_i,x_{N^+}),\\&-\lambda s_-(x_i,x_1),\cdots,-\lambda s_-(x_i,x_{N^-})]\end{aligned} \tag{6}$$

式中，N^+ 和 N^- 分别为正类训练样本和负类训练样本的数目。对样本 x_i 和类中其它样本的相似度进行归一化处理，即

$$\begin{aligned}s_+^*(x_i,x_1)&=\frac{s_+(x_i,x_1)}{s_+(x_i,x_1)+,\cdots,+s_+(x_i,x_{N^+})}\\&\vdots\\s_+^*(x_i,x_{N^+})&=\frac{s_+(x_i,x_{N^+})}{s_+(x_i,x_1)+,\cdots,+s_+(x_i,x_{N^+})}\end{aligned} \tag{7}$$

$$s_-^*(x_i,x_1)=\frac{s_-(x_i,x_1)}{s_-(x_i,x_1)+,\cdots,+s_-(x_i,x_{N^-})}$$

$$\vdots \tag{8}$$

$$s_-^*(x_i,x_{N^-})=\frac{s_-(x,x_{N^+})}{s_-(x_i,x_1)+,\cdots,+s_-(x_i,x_{N^-})}$$

归一化处理后 x_i 映射到新空间的样本可表示为：

$$X_i=[s_+^*(x_i,x_1),\cdots,s_+^*(x_i,x_{N^+}),-\lambda s_-^*(x_i,x_1),\cdots,-\lambda s_-^*(x_i,x_{N^-})] \tag{9}$$

另定义 $w=[\alpha_1,\cdots,\alpha_{N^+},\beta_1,\cdots,\beta_{N^-}]^T$，类标签 $y_i\in\{-1,1\}$，正负类训练样本总数 $N=N^++N^-$，则定理得证。

显然形式(5)是一个对偶问题，可得定理 2。

定理 2 DSSVM 原始问题的对偶问题为：

$$\max_{\gamma}-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_jX_i^TX_j+\frac{(1-\lambda)^2}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_j-\frac{\lambda}{N}\sum_{i=1}^{N}\gamma_iy_i$$

s. t. (10)

$$\sum_{i=1}^{N}\gamma_i=v$$

$$\gamma_i\geqslant 0,i=1,\cdots,N$$

证明：构造上述模型的拉格朗日方程：

$$L(w,\rho,\gamma,\mu)=\frac{1}{2}\|w\|^2-v\rho+\sum_{i=1}^{N}\gamma_i(\rho-y_iw^TX_i)+\mu(1-w^T1) \tag{11}$$

其中，$\gamma=(\gamma_1,\gamma_2,\cdots,\gamma_N)^T\geqslant 0$ 是拉格朗日乘子，方程 $L(w,\rho,\gamma,\mu)$ 分别对原始问题变量 ρ 和 w 求偏导数，可得：

$$\frac{\partial L}{\partial\rho}=\sum_{i=1}^{N}\gamma_i-v=0 \tag{12}$$

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{N}\gamma_iy_iX_i-\mu 1=0 \tag{13}$$

由式(12)、式(13)可得：

$$v=\sum_{i=1}^{N}\gamma_i \tag{14}$$

$$w=\sum_{i=1}^{N}\gamma_iy_iX_i+\mu 1 \tag{15}$$

由式(5)中的第二个条件可得：

$$(\sum_{i=1}^{N}\gamma_iy_iX_i+\mu 1)^T1=1 \tag{16}$$

$$\mu=\frac{1-\sum_{i=1}^{N}\gamma_iy_iX_i^T1}{N} \tag{17}$$

将式(14)、式(15)、式(17)代入拉格朗日方程(11)，可得原始问题的拉格朗日对偶形式：

$$\max_{\gamma}-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_jX_i^TX_j+\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_j1^TX_iX_j^T1-\frac{1}{N}\sum_{i=1}^{N}\gamma_iy_iX_i^T1+\frac{1}{2N} \tag{18}$$

s. t.

$$\sum_{i=1}^{N}\gamma_i=v$$

$$\gamma_i\geqslant 0,i=1,\cdots,N$$

根据式(6)－式(8)可得：

$$1^TX_i=X_i^T1=1-\lambda \tag{19}$$

将式(19)代入式(18)，则定理可得。

2.2 核化形式

一般样本在低维空间很难做到准确划分，因此需要进行核化，实质上是找到一个合适的映射 $\varphi:X_i\in R^d\to\varphi(X_i)\in R^D$，一般的 $d<<D$，并用核函数 $K(u,v)$ 表示映射后的点积 $\varphi(u)\cdot\varphi(v)$，核化后对偶问题的目标函数为：

$$\max_{\gamma}-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_j\varphi(X_i)^T\varphi(X_j)+\frac{(1-\lambda)^2}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_j-\frac{\lambda}{N}\sum_{i=1}^{N}\gamma_iy_i$$
$$=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_jK(X_i,X_j)+\frac{(1-\lambda)^2}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}y_iy_j\gamma_i\gamma_j-\frac{\lambda}{N}\sum_{i=1}^{N}\gamma_iy_i \tag{20}$$

s. t.

$$\sum_{i=1}^{N}\gamma_i=v$$

$$\gamma_i\geqslant 0,i=1,\cdots,N$$

2.3 决策函数

显然问题式(20)是一个 QP 问题，设参数 γ^* 为该问题的解，则根据式(15)、式(17)可得：

$$w^*=\sum_{i=1}^{N}\gamma_i^*y_iX_i+\frac{1-\sum_{i=1}^{N}\gamma_iy_iX_i^T1}{N}1$$
$$=\sum_{i=1}^{N}\gamma_i^*y_iX_i+\frac{1-(1-\lambda)\sum_{i=1}^{N}\gamma_i^*y_i}{N}1 \tag{21}$$

结合式(4)可得决策函数：

$$g(X)=\mathrm{sgn}(\sum_{i=1}^{N}\gamma_i^*y_iK(X_i,X)+\frac{1-(1-\lambda)\sum_{i=1}^{N}\gamma_i^*y_i}{N}1^TX) \tag{22}$$

2.4 DSSVM 的优势与计算复杂度分析

DSSVM 是一种基于已知相似度信息的分类方法，算法首先根据训练样本的相似度信息构建新空间，如式(6)所示，然后在新空间下求解二次规划问题得到最优分类器，最后将测试样本映射到新空间进行分类。在第 5 节 DSSVM 实验部分可以看到 DSSVM 具有较好的分类精度和较快的速度；另外，从式(6)可以看到，新空间的维度为正负类样本个数之和，也就是说对于一些特殊样本(比如第 5 节 DSSVM 实验部分使用的 Colon、Leukemia、Metas 等基因表达数据集，此类数据采集成本高，因此样本量少，样本维度高)，DSSVM 将样本映射到新空间后起到了降维的作用，用降维后的样本进行训练势必会提高效率，这可以通过第 5 节 DSSVM 实验部分对 Colon、Leukemia、Metas 等数据集的实验得到证实。

DSSVM 的核化模型跟其他许多核化形式的分类方法一样，实质上都是求解 QP 问题。从式(20)可以看到，QP 一次项和二次项计算的空间复杂度均为 $O(m^2)$，而求解 QP 需要 $O(m^3)$ 的时间复杂度，所以 DSSVM 同 SVM、SVDD 一样很难用于较大样本训练。为了提高 DSSVM 对较大样本的分类能力，本文结合 CC-MEB 理论，提出了相似度差核心集向量机算法。

3 MEB 和 CC-MEB 最小球问题

最小包含球 MEB 问题在模式识别中得到了广泛研究[13-19]，特别是利用 MEB 的核心集 Core Set 解决大样本问题。假定一个球 $B(\boldsymbol{C},R)$，$\boldsymbol{C}$ 和 R 分别为球心和半径，给定 N 个样本数据集 $X=\{x_i|x_i\in R^d\}$，其中 x_i 为向量，$1\leqslant i\leqslant N$，原始样本空间通过函数 $\varphi:x_i\to\varphi(x_i)$ 映射到高维特征空间 Γ[20]。

3.1 MEB 最小球

最小包含球问题可用下列优化问题描述：

$$\min_{R,C} R^2$$
$$\text{s. t.} \quad \|x_i - C\|^2 \leqslant R^2 \tag{23}$$

引入核函数核化后，MEB 问题可描述为：

$$\min_{R,C} R^2$$
$$\text{s. t.} \quad \|\varphi(x_i) - C\|^2 \leqslant R^2 \tag{24}$$

通过构造拉格朗日函数，可以得到该优化问题的对偶形式：

$$\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}} diag(\boldsymbol{K}) - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}$$
$$\text{s. t.} \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1} = 1 \tag{25}$$

其中，$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_N)^{\mathrm{T}}_{1\times N} \geqslant 0$ 是拉格朗日乘子，$\mathbf{1} = (1, 1, \cdots, 1)^{\mathrm{T}}_{1\times m}$，$\boldsymbol{K} = [k(x_i, x_j)]_{N\times N}$ 为核矩阵。

Tsang 等人指出，如果引入的核函数满足条件 $k(x_i, x_j) = k$，那么原问题核化后对偶形式只要如同式(25)的优化问题均可视为 MEB 问题，如硬边界一类 SVDD 核化形式、软边界一类和二类 L_2-SVM 核化形式，这样用于求解 MEB 问题的一些优势方法就可以使用，如利用 MEB 核心集的快速逼近法等。

3.2 中心约束 CC-MEB 最小球

文献[15] Tsang 等人将 MEB 拓展到中心约束的最小球 CC-MEB 后，建立 MEB 和其他核化形式优化算法之间的联系，如 L_2-SVR，Ranking SVM 等，因而这些算法同样可以利用 MEB 技巧快速解决一些优化问题。

在 CC-MEB 中，将 Γ 特征空间每个样本点 $\varphi(x_i)$ 中引入一属性项 $\delta_i \in R$，从而构建一个拓展的 MEB 高 1 维特征空间 Γ'，其样本为 $\begin{bmatrix}\varphi(x_i)\\ \delta_i\end{bmatrix}$，且约束 Γ' 空间的中心点为 $\begin{bmatrix}\boldsymbol{C}\\ 0\end{bmatrix}$，其中 $\boldsymbol{C}$ 是 Γ 特征空间的中心点。因而 MEB 在 Γ' 空间中的优化问题可表述为式(26)，即为 CC-MEB 问题。

$$\min_{R,C} R^2$$
$$\text{s. t.} \quad \|\varphi(x_i) - C\|^2 + \delta_i \leqslant R^2 \tag{26}$$

其对偶形式如下：

$$\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}(diag(\boldsymbol{K}) + \Delta) - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}$$
$$\text{s. t.} \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1} = 1 \tag{27}$$

其中，

$$\Delta = [\delta_1^2, \delta_2^2, \cdots, \delta_m^2]^{\mathrm{T}} \geqslant 0 \tag{28}$$

此时，有

$$R = \sqrt{\boldsymbol{\alpha}^{\mathrm{T}}(\mathrm{diag}(\boldsymbol{K}) + \Delta) - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}} \tag{29}$$

$$\boldsymbol{C} = \sum_{i=1}^{m} \alpha_i \varphi(x_i) \tag{30}$$

Γ' 空间中任一点到球心的距离平方为：

$$\left\| \begin{bmatrix}\varphi(x_i)\\ \delta_i\end{bmatrix} - \begin{bmatrix}\boldsymbol{C}\\ 0\end{bmatrix} \right\|^2 = \|\varphi(x_i) - \boldsymbol{C}\|^2 + \delta_i^2 \tag{31}$$

由于约束条件 $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1} = 1$，因此将式(27)改成式(32)不影响优化问题的解。

$$\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}(diag(\boldsymbol{K}) + \Delta - \eta\mathbf{1}) - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}$$
$$\text{s. t.} \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1} = 1 \tag{32}$$

其中 $\eta \in R$ 为某一常数。

因此，只要是形如式(32)且满足条件(28)的优化问题就可视为 MEB 问题，只是此时为 CC-MEB 问题。

4 相似度差支持核向量机(DSCVM)

4.1 相似度差支持向量机和 CC-MEB 的关系

正如上节所述，很多优化问题在一定条件下可以转化为 MEB 问题或者 CC-MEB 问题，如 DSSVM 也可以转化为 MEB 问题或 CC-MEB 问题，则相关的快速逼近算法可以运用于 DSSVM。因此首先给出定理 3。

定理 3 DSSVM 等价于 CC-MEB 问题。

证明：令 $v\lambda_i = \gamma_i$，由式(20)可得：

$$\max_{\gamma} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j v\lambda_i v\lambda_j K(X_i, X_j) + \frac{(1-\lambda)^2}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j v\lambda_i v\lambda_j - \frac{\lambda}{N}\sum_{i=1}^{N} v\lambda_i y_i$$
$$\text{s. t.} \quad \sum_{i=1}^{N} v\lambda_i = v \tag{33}$$
$$v\lambda_i \geqslant 0, i = 1, \cdots, N$$

将式(33)改为式(34)不影响 QP 问题的求解。

$$\max_{\gamma} -\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \lambda_i \lambda_j K(X_i, X_j) + \frac{(1-\lambda)^2}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \lambda_i \lambda_j - \frac{2\lambda}{vN}\sum_{i=1}^{N}\lambda_i y_i$$
$$\text{s. t.} \quad \sum_{i=1}^{N}\lambda_i = 1 \tag{34}$$
$$\lambda_i \geqslant 0, i = 1, \cdots, N$$

即

$$\max_{\gamma} -\frac{2\lambda}{vN}\sum_{i=1}^{N}\lambda_i y_i - \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \lambda_i \lambda_j \left(K(X_i, X_j) - \frac{(1-\lambda)^2}{N}\right)$$
$$\text{s. t.} \quad \sum_{i=1}^{N}\lambda_i = 1 \tag{35}$$
$$\lambda_i \geqslant 0, i = 1, \cdots, N$$

为与式(32)形式相同，令 $\lambda_i = \alpha_i$，则有

$$\max_{\alpha} \boldsymbol{\alpha}^{\mathrm{T}}(diag(\tilde{\boldsymbol{K}}) + \Delta - \eta\mathbf{1}) - \boldsymbol{\alpha}^{\mathrm{T}}\tilde{\boldsymbol{K}}\boldsymbol{\alpha}$$
$$\text{s. t.} \quad \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{1} = 1 \tag{36}$$

其中，

$$\tilde{\boldsymbol{K}} = [\tilde{k}(X_i, X_j)] = [y_i y_j (k(X_i, X_j) - \frac{(1-\lambda)^2}{N})]_{N\times N} \tag{37}$$

$$\Delta = -diag(\tilde{\boldsymbol{K}}) + \eta\mathbf{1} + \boldsymbol{T}_{linearterm} \tag{38}$$

$$\boldsymbol{T}_{linearterm} = [T_i] = [-\frac{2y_i\lambda}{vN}]_{N\times 1} \tag{39}$$

当取 $\eta = \max(diag(\tilde{\boldsymbol{K}}) - \boldsymbol{T}_{linearterm})$ 时，式(28)条件必定满足，定理得证。

至此，DSSVM 分类方法可以视为 CC-MEB 形式的最小球问题。因此，可以利用 $(1+\varepsilon)^2$ 或 $1+\varepsilon$ 逼近算法[15]快速找出大样本集的核心集 Core Set，其中 ε 是最小球的逼近参数，是最小球逼近的迭代终止条件。然后采用 DSSVM 对核心集进行训练，获得判决函数，因此得到了 DSSVM 求解大样本问

题的另一个版本 DSCVM。

4.2 DSCVM 算法的实现

实际上，对于 DSSVM 算法中的每一个训练样本 $Z_i=(X_i,y_i)$，在 DSCVM 中被映射到新空间的表示方法为式(40)，这个空间可以是复空间。

$$\begin{bmatrix}\tilde{\varphi}(Z_i)\\ \delta_i\end{bmatrix}=\begin{bmatrix}y_i\varphi(X_i)\\ y_i\dfrac{1-\lambda}{\sqrt{2N}}\mathbf{i}\\ \delta_i\end{bmatrix} \tag{40}$$

此时，采用 CC-CVM 的$(1+\varepsilon)^2$ 迭代算法需要计算最小球的半径和样本点到球心的距离(见下文中 DSCVM 算法)，半径可以采用式(29)计算；而在 DSCVM 算法中，样本点到球心的距离计算需要将式(30)、式(31)改成式(41)、式(42)，并通过式(42)计算获得。

$$\boldsymbol{C}=\sum_{i=1}^{m}\alpha_i\tilde{\varphi}(Z_i)=\sum_{i=1}^{N}\alpha_i y_i\begin{bmatrix}\varphi(x_i)\\ \dfrac{1-\lambda}{\sqrt{2N}}\mathbf{i}\end{bmatrix} \tag{41}$$

$$\begin{aligned}\left\|\begin{bmatrix}\tilde{\varphi}(Z_l)\\ \delta_l\end{bmatrix}-\begin{bmatrix}\boldsymbol{C}\\ 0\end{bmatrix}\right\|^2&=\|\tilde{\varphi}(Z_l)-\boldsymbol{C}\|^2+\delta_l^2\\&=\left\|\begin{bmatrix}y_l\varphi(X_l)\\ y_l\dfrac{1-\lambda}{\sqrt{2N}}\mathbf{i}\end{bmatrix}-\sum_{i=1}^{N}\alpha_i y_i\begin{bmatrix}\varphi(X_i)\\ \dfrac{1-\lambda}{\sqrt{2N}}\mathbf{i}\end{bmatrix}\right\|^2+\delta_l^2\\&=\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(k(X_i,X_j)-\frac{(1-\lambda)^2}{2N}\right)-2\sum_{i=1}^{m}\alpha_i y_i y_l\left(k(X_i,X_l)-\frac{(1-\lambda)^2}{2N}\right)+y_l y_l\left(k(X_l,X_l)-\frac{(1-\lambda)^2}{2N}\right)+\delta_l^2\end{aligned} \tag{42}$$

DSCVM 算法分 4 个阶段，第一阶段将训练样本映射到新空间，第二阶段采用 CC-CVM 的$(1+\varepsilon)^2$ 找出核心集，第三阶段利用 DSSVM 对核心集进行训练并输出判决函数，第四阶段根据核心集将测试样本映射到新空间并用决策函数进行分类。具体如下：

算法 DSCVM

阶段 1　将训练样本映射到新空间，如式(8)所示

阶段 2　通过 CC-CVM 获得 Core Set

Step 1　初始化 t=0，S_t，$\mathbf{C}_t$ 和 R_t

Step 2　所有样本点均被 $B(\mathbf{C}_t,(1+\varepsilon)R_t)$包络，则 $S=S_t$ 转入阶段 2，否则转入 Step 3

Step 3　找出离 $\mathbf{C}_t$ 最远的点 z，将其加入核心集中，即 $S_{t+1}=S_t\cup\{\mathbf{z}\}$

Step 4　找出新的 MEB：$B(\mathbf{C}_{t+1},R_{t+1})$

Step 5　转 Step 1

阶段 3　获得决策函数

Step 6　采用 DSSVM 对核心集 S 训练

Step 7　根据式(18)输出决策函数

阶段 4　对测试样本分类

Step 8　根据得到的核心集，将测试样本映射到新空间

Step 9　对测试样本进行分类

在实现算法的 Step 3 时，不在球 $B(\boldsymbol{C}_t,(1+\varepsilon)R_t)$中的点假设有 n 个，其与 $\boldsymbol{C}_t$ 的距离需要进行计算，并找出最远点 z，由式(42)可知，此时计算复杂度为 $O(|S_t|^2+n|S_t|)$，当 n 非常大时，Step 3 是不可能实现的。为此，本算法采用文献[21]提出的概率加速方法(Probabilistic Speedup Method)，通过随机采样一个子集，当子集 $Size=59$ 时，所有最远点的 5%在该子集中的概率将会达到 95%，这样在保证性能的基础上使计算复杂度降为 $O(|S_t|^2)$，而一般的$|S_t|<<n$，从而提高了算法的效率。

5 实验及分析

为评价 DSSVM 及 DSCVM 算法的有效性，本文将实验分为两部分。第一部分实验使用 11 种二类小量样本数据集来验证 DSSVM 的有效性，其中一种是人造二维香蕉形数据，另外 10 种可从 http://www.ics.uci.edu/~mlearn/MLRepository.html 和 http://ict.ewi.tudelft.nl/~davidt/index.html 网站下载得到[22]。第二部分实验使用 5 种二类较大样本数据集，从 http://ict.ewi.tudelft.nl/~davidt/index.html 和 http://archive.ics.uci.edu/ml/datasets.html 下载得到，所有实验均在 Intel Core2，2.0GHz 主频，2G RAM，Windows XP 系统，Matlab2009a 平台实现。

本文所有实验均采用径向基(Radial Basis Function，RBF)核函数实现，另外，在实验时正负类的相似度均用向量空间余弦相似度来表示：

$$s(x_i,x_j)=\frac{x_i\cdot x_j}{\|x_i\|\,\|x_j\|} \tag{43}$$

5.1 DSSVM 实验

DSSVM 实验的参数设置如下：核函数中的 $2\sigma^2$ 选择以训练样本的平均 2 范数的平方 s 为基准，并在网格$\{s/32,s/16,s/8,s/4,s/2,s,2s,4s,8s,16s,32s\}$中搜索直至最优；参数 v 在网格$\{1/64,1/32,1/16,1/8,1/4,1/2,1,2,4,8,16,32,64\}$中搜索直至最优；$\lambda$ 为 N^+/N^-，此处 N^+ 和 N^- 分别为正类训练样本和负类训练样本的数目。

表 1　DSSVM 实验数据集

数据集	维数	样本数	+1 类	−1 类
Banana	2	300	150	150
Iris	4	150	50	100
Housing	13	506	458	48
Biomed	5	194	127	67
Balance-scale	4	625	288	337
Sonar	60	208	111	97
Ionosphere	34	351	225	126
Diabetes	8	768	500	268
Colon	1908	62	22	40
Leukemia	3571	72	25	47
Metas	4919	145	46	99

本文构造了一种人造香蕉形数据集，其均值为 0，标准差为 1，正负类各 150 个样本，如图 2 所示。此外，为验证算法的有效性，对 Iris、Housing 等 10 个真实数据集进行实验。11 类数据集属性如表 1 所列。实验时每次从正负类样本中各取 60%用于训练，其余样本用于测试，每个数据集进行 10 次随机交叉验证，以求得测试样本的分类精度及标准差。另外对此 11 类数据集使用 SVM 和 SVDD 进行实验，SVM 中的 $2\sigma^2$ 选择以训练样本的平均 2 范数的平方 s 为基准，并在网格$\{s/32,s/16,s/8,s/4,s/2,s,2s,4s,8s,16s,32s\}$中搜索直至最优；SVDD 实验中两个参数 C_1 和 C_2 均取 1，其余实验环境及参数配置与 SVM 相同。通过样本分类精度(单位%)、标准差(单位%)、样本训练时间 train-time(单位 sec)和测试样本的分类时间 class-time(单位 sec)比较了 3 种算法的运行效率。图 3 为 DSSVM 在香蕉集上的分类效果；表 2 为 3 类算

法在11个数据集上的分类时间、精度及标准差。

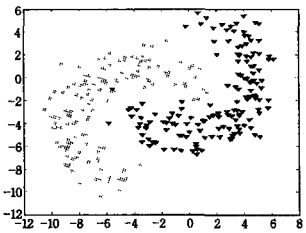

图2 人造香蕉集数据

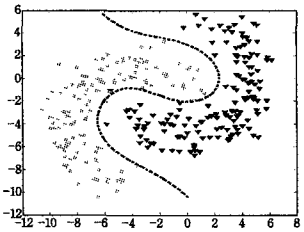

图3 DSSVM在香蕉集上的分类效果

从表2可以看出，除了数据集Biomed和Diabetes外，DSSVM的分类精度均优于SVM和SVDD；除了数据集Biomed外，训练时间均少于SVM和SVDD；除了数据集Housing、Balance-scale和Diabetes外，其分类速度均少于SVM和SVDD。另外，DSSVM将样本映射到新空间后起到了降维的作用，用降维后的样本进行训练时将缩短其所花费的时间，在Colon、Leukemia、Metas 3个基因数据集上的实验结果验证了这一结论。可见，本文提出的DSSVM算法可以有效解决分类问题。

表2 DSSVM测试精度比较

数据集	DSSVM			SVM			SVDD		
	精度及标准差	train-time	class-time	精度及标准差	train-time	class-time	精度及标准差	train-time	class-time
Banana	**99.20±0.24**	**5.67±0.14**	**0.43±0.05**	98.15±0.49	18.43±0.22	1.09±0.08	96.33±0.49	6.83±0.10	0.97±0.03
Iris	**97.87±0.85**	**0.54±0.09**	**0.12±0.02**	96.58±1.06	2.95±0.06	0.13±0.05	96.27±0.52	0.84±0.09	0.15±0.02
Housing	**88.74±1.21**	**3.40±0.25**	1.29±0.09	86.82±0.82	5.65±0.56	0.96±0.07	87.46±0.61	4.59±0.37	**0.85±0.03**
Biomed	81.77±2.14	**0.78±0.10**	**0.18±0.03**	**83.63±2.48**	3.65±0.12	0.18±0.02	83.12±1.98	0.87±0.11	0.37±0.05
Balance-scale	**91.35±0.35**	**6.15±0.17**	1.93±0.08	89.76±0.18	32.63±0.35	2.45±0.10	87.37±0.18	10.54±0.58	**1.89±0.07**
Sonar	**85.44±1.56**	**0.59±0.08**	0.32±0.03	84.12±0.87	1.26±0.11	0.45±0.03	84.86±1.08	0.95±0.07	**0.25±0.03**
Ionosphere	**90.57±0.28**	**3.31±0.11**	**0.43±0.04**	88.57±0.34	25.54±0.28	0.68±0.06	89.01±0.12	8.64±0.16	0.98±0.04
Diabetes	69.40±3.24	**21.32±0.58**	**1.85±0.15**	68.52±3.52	40.68±0.89	2.54±0.21	**69.94±4.26**	29.54±1.02	2.15±0.18
Colon	**85.48±2.31**	**0.12±0.02**	**0.05±0.01**	79.13±1.72	0.76±0.08	0.09±0.02	72.84±1.27	0.58±0.08	0.08±0.01
Leukemia	**98.57±0.56**	**0.20±0.05**	**0.06±0.02**	96.67±0.77	0.29±0.04	0.06±0.01	80.40±2.15	0.78±0.06	0.17±0.02
Metas	**68.06±0.71**	**0.58±0.06**	**0.29±0.08**	63.33±0.95	0.68±0.10	0.30±0.06	65.32±0.32	0.81±0.12	0.31±0.10

5.2 DSCVM实验

本节实验用于验证DSCVM在大样本上的分类性能。根据4.2节给出的DSCVM算法流程，对4个真实的UCI数据集进行实验研究。表3给出了4个数据集的属性。核函数中的$2\sigma^2$选择以训练样本的平均2范数的平方s为基准，并在$\{s/32, s/16, s/8, s/4, s/2, s, 2s, 4s, 8s, 16s, 32s\}$网格中搜索，直至最优；参数$v$在网格{1/64,1/32,1/16,1/8,1/4,1/2,1,2,4,8,16,32,64}中搜索，直至最优；λ为N^+/N^-。实验效率和分类精度分别采用train-time(sec)和g表示。

表3 DSCVM实验数据集

数据集	维数	样本数	+1类	−1类
Delft	64	1500	1124	376
Abalone	10	4177	1407	2770
Spambase	57	4601	1813	2788
Concordia	1024	4000	400	3600
Letter Recognition	16	20,000	8,387	11,613

实验1 利用Delft数据集分析逼近参数ε对DSCVM算法的影响。实验时从Delft数据集中随机抽取60%作为训练样本，其余作为测试样本。表4给出了$\sigma=16, v=12$时的实验结果。从表4可以看出，ε越小，算法的精度越高，但是此时算法的实现时间越长。

表4 逼近参数ε对DSCVM的影响

ε	精度及标准差	Core Set子集样本数 CVs	train-time (s)
1e−3	51.47±4.69	8	0.76±0.08
1e−4	61.62±3.17	12	1.26±0.10
1e−5	68.93±1.95	23	2.73±0.09
1e−6	84.02±2.38	48	9.21±0.17
1e−7	89.63±1.01	97	38.31±0.35
1e−8	96.60±0.82	161	142.78±1.75
1e−9	97.10±0.52	166	153.93±2.33

实验2 对表3中5个真实数据集进行实验，分别从4个数据集中随机抽取60%作为训练样本，其余作为测试样本。参数σ和v在上述规定范围内寻得最优。实验结果见表5。由表5可以看到，DSCVM在大样本上具有较好的分类效果。实验3将DSCVM和SVM、SVDD进行比较以说明其优势。

表5 DSCVM在5个样本上的分类精度及训练时间

数据集	ε	精度及标准差	Core Set子集样本数 CVs	train-time(s)
Delft	1e−8	96.91±0.32	158	140.28±3.28
Abalone	1e−6	85.42±0.87	356	1643.54±29.32
Spambase	1e−5	87.10±0.56	402	1956.67±34.12
Concordia	1e−5	94.18±0.44	341	1637.85±22.30
Letter Recognition	1e−4	85.02±0.21	628	3147.58±80.21

实验3 由于SVM及SVDD对大样本的训练速度较慢，为了能与DSCVM进行比较，实验3分别从4个数据集中随机抽取1000个作为训练样本，其余样本作为测试样本。DSCVM中参数σ和v在上述规定范围内寻得最优，参数ε取1e−6。SVM和SVDD的运行环境和参数设置同5.1节中所述。实验结果见表6。由表6可以看到，DSCVM可以在保证分类精度的前提下，大大缩短对样本的训练时间，因此在较大样本集的分类问题上具有一定优势。

表6 DSCVM性能比较

数据集	DSCVM		SVM		SVDD	
	g(%)	train-time	g(%)	train-time	g(%)	train-time
Delft	**91.63**	**34.6719**	90.40	400.2965	91.35	342.4839
Abalone	78.63	**28.4688**	**83.02**	236.2344	79.42	248.6485
Spambase	**87.96**	**40.4729**	87.12	407.8125	84.65	479.7549
Concordia	93.18	**26.5938**	**96.04**	680.6719	88.64	594.7840
Letter Recognition	83.32	**38.75**	85.54	902.5839	**85.93**	421.4732

结束语 相似度差支持向量机在已知样本相似度信息的前提下，将样本隶属于某类的隶属度表示成该样本与该类所有样本相似度的线性组合，并根据线性组合的稀疏性定义优化问题的目标函数，通过判断样本与各类的总相似度差来进

(下转第257页)

687-697

[3] Karaboga D, Akay B. A comparative study of artificial bee colony algorithm[J]. Applied Mathematics and Computation, 2009, 214(1): 108-132

[4] Li C Q, Niu P F, Xiao X J. Development and investigation of efcient articial bee colony algorithm for function optimization numerical function optimization [J]. Applied Soft Computing, 2012, 12: 320-332

[5] 罗钧，王强，付丽. 改进蜂群算法在平面误差评定中的应用[J]. 光学精密工程，2012，20(2)：422-430

[6] Horng M H. Multilevel thresholding selection based on the artificial bee colony algorithm for image segmentation[J]. Expert Syst Appl, 2011, 38(11): 13785-13791

[7] Öztürk C, Karabo˘ga D, Görkemli B. Artificial bee colony algorithm for dynamic deployment of wireless sensor networks[J]. Turk J Electr Eng Comput Sci, 2012, 20(2): 1-8

[8] 暴励，曾建潮. 一种双种群差分蜂群算法[J]. 控制理论与应用，2011，28(2)：266-272

[9] Wu B, Fan S H. Improved artificial bee colony algorithm with chaos[M]. Computer science for environmental engineering and ecoinformatics, Berlin: Springer, 2011: 51-56

[10] 罗钧，肖向海，付丽，等. 基于分段搜索策略的改进蜂群算法[J]. 控制与决策，2012，27(9)：1402-1405

[11] 张超群，郑建国，王翔. 蜂群算法研究综述[J]. 计算机应用研究，2011，28(9)：3201-3205

[12] Chen S H, Chen M C, Chang P C, et al. Guidelines for developing effective estimation of distribution algorithms in solving single machine scheduling problems[J]. Expert Systems with Applications, 2010, 37(9): 6441-6451

[13] Akay B, Karaboga D. Parameter tuning for the artificial bee colony algorithm[C]// International Conference on Computer and Computational Intelligence. 2009, 5796: 608-619

(上接第 244 页)

行分类。本文还通过最小包含球的核心集技巧建立了 DSSVM 与 MEB 之间的联系，在此基础上提出的 DSCVM 算法在解决较大样本分类问题上显现了一定的优势。

参 考 文 献

[1] Cortes C, Vapnik V N. Support vector networks [J]. Machine Learning, 1995, 20(3): 273-297

[2] Chang C-C, L in C-J. Training v-Support Vector Classifiers: Theory and Algorithms[J]. Neural Computation, 2002, 14: 1959-1977

[3] Hu M, Chen Y, Kwok J T. Building sparse multi-kernel SVM classifiers[J]. IEEE Transactions on Neural Networks, 2009, 20(5): 827-839

[4] 丁晓剑，赵银亮，李远成. 基于 SVM 的二次下降有效集算法[J]. 电子学报，2011，39(8)：1766-1770

[5] Tax D M J, Duin R P W. Support Vector Data Description[J]. Machine Learning, 2004, 54(1): 45-66

[6] Kim J S, Scott C D. L2 kernel classification[J]. IEEE Trans. PAMI, 2010, 32(10): 1822-1831

[7] Yang Y, Wright J, Ma Y, et al. Feature Selection in Face Recognition: A Sparse Representation Perspective[J]. IEEE Trans. PAMI, 2009, 31(2): 210-227

[8] Majumdar A, Ward R K. Classification via group sparsity promoting regularization[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. 2009: 861-864

[9] Majumdar A, Ward R K. Improved Group Sparse Classifier[J]. Pattern Recognition Letters, 2010, 31(13): 1959-1964

[10] Tibshirani R. Regression shrinkage and selection via the lasso [J]. J. Royal. Statist. Soc B., 1996, 58(1): 267-288

[11] Donoho D L. For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution[J]. Commun. Pure Appl. Math., 2006(6): 797-829

[12] 宋枫溪，程科，杨静宇，等. 最大散度差和大间距线性投影与支持向量机[J]. 自动化学报，2004，30(6)：890-896

[13] Fine S, Scheinberg K. Efficient SVM training using low-rank kernel representations[J]. J. Mach. Learn. Res., 2001, 2: 243-264

[14] Deng Z H, Chung F L, Wang S T. FRSDE: Fast Reduced Set Density Estimator using Minimal Enclosing Ball Approximation [J]. Pattern Recognition, 2008, 41: 1363-1372

[15] Tsang I W, Kwok J T, Zurada J M. Generalized core vector machines[J]. IEEE Transactions on Neural Networks, 2006, 17(5): 1126-1140

[16] Chung F L, Deng Z H, Wang S T. From Minimum Enclosing Ball to Fast Fuzzy Inference System Training on Large Datasets [J]. IEEE Transactions on Fuzzy Systems, 2009, 17(1): 173-184

[17] Bădoiu M, Clarkson K L. Optimal core sets for balls[J]. Computational Geometry: Theory and Applications, 2008, 40(1): 14-22

[18] Bădoiu M, Har-Peled S, Indyk P. Approximate clustering via core sets[C]// Proc. 34th Annu. ACM Symposium Theory Comput., Montreal, QC, Canada, 2002: 250-257

[19] Asharaf S, Murty M N, Shevade S K. Multiclass core vector machine[C]// Proc. 24th ICML. Corvalis, OR, 2007: 41-48

[20] 胡文军，王士同，邓赵红. 适合大样本快速训练的最大夹角间隔核心集向量机[J]. 电子学报，2011，39(5)：1178-1184

[21] Smola A, Schölkopf B. Sparse greedy matrix approximation for machine learning[C]// ICML'00 Proceedings of the 7th International Conference on Mechine Learning Stanford, CA, June 2000: 911-918

[22] Asuncion A, Newman D J. UCI Machine Learning Repository [OL]. http://www.ics.uci.edu/ ~mlearn/MLRepository.html, Irvine, CA: University of California, School of Information and Computer Science