

# 基于 SSVM 的递归统计不相关特征抽取算法

任世锦<sup>1),2)</sup> 王小林<sup>1)</sup> 吕俊怀<sup>1)</sup> 张晓光<sup>2)</sup>

<sup>1)</sup>(徐州师范大学计算机学院 江苏 徐州 221116)

<sup>2)</sup>(中国矿业大学机电工程学院 江苏 徐州 221008)

**摘 要** 文章旨在研究数据分布未知的高维、小样本问题的特征抽取算法. 基于支持向量机原理和特征统计不相关思想, 提出基于散度支持向量机(SSVM)的递归统计不相关特征抽取算法, 解决现有算法抽取特征之间存在相关性、算法受到样本分布影响等问题. 针对高维小样本问题, 使用 PCA 把 SSVM 优化问题变换到同构低维空间; 给出边界鉴别向量集的递归求取方法, 把模式高维特征投影到边界鉴别向量集, 实现了统计不相关特征的抽取; 分析了算法的收敛性和终止条件. 文中使用核方法把线性 SSVM 推广到非线性 SSVM, 通过 KPCA 方法把非线性 SSVM 优化问题转换到低维空间中的等价优化问题, 在低维空间抽取不相关非线性特征. 仿真结果证明了文中算法的有效性.

**关键词** 散度支持向量机(SSVM); 分类; 特征抽取; 统计不相关边界鉴别向量; 主元分析(PCA)

**中图法分类号** TP391 **DOI 号**: 10.3724/SP.J.1016.2011.00443

## A SSVM-Based Recursive Uncorrelated Feature Extraction Algorithm

REN Shi-Jin<sup>1),2)</sup> WANG Xiao-Lin<sup>1)</sup> LV Jun-Huai<sup>1)</sup> ZHANG Xiao-Guang<sup>2)</sup>

<sup>1)</sup>(School of Computer Science and Technology, Xuzhou Normal University, Xuzhou, Jiangsu 221116)

<sup>2)</sup>(School of Mechanical and Electrical Engineering, China University of Mining & Technology, Xuzhou, Jiangsu 221008)

**Abstract** A feature extraction algorithm for high dimensional data with unknown distribution and small sample size problem is discussed in this paper. Based on support vector machines and the idea of uncorrelated features, a scatter support vector machine (SSVM)-based recursive uncorrelated feature extraction algorithm is presented to deal with drawbacks of existing algorithms, such as correlations among extracted features, performance decrease from distribution of samples etc. To cope with small sample size problem, the optimization problem of SSVM is transformed into that in isomorphic lower dimension space through PCA; Then the method of recursively extracting margin discriminant vectors is proposed, and the uncorrelated features can be yielded by projecting the data in margin discriminant vectors; Finally, the convergence and termination condition of the proposed algorithm are analyzed. The algorithm can be generalized into nonlinear cases through kernel methods, the optimization problem of nonlinear SSVM can be transformed into equivalent optimization problem in lower dimension through KPCA, and then uncorrelated nonlinear features can be extracted. The simulation results demonstrate the efficiencies of the proposed algorithm.

**Keywords** scatter support vector machine (SSVM); classification; feature extraction; uncorrelated margin discriminant vectors; principal component analysis (PCA)

收稿日期: 2008-08-31; 最终修改稿收到日期: 2011-01-02. 本课题得到徐州师范大学培育项目(08XLY10)、中国博士后科学基金(20060390277)和江苏省“六大人才高峰”计划(06-E-05)资助. 任世锦, 男, 1971 年生, 博士, 副教授, 研究兴趣为机器学习、模式识别、图像处理. E-mail: sjren\_phd@163.com. 王小林, 女, 1980 年生, 副教授, 研究兴趣为数据挖掘、人工智能和模式识别. 吕俊怀, 男, 1959 年生, 副教授, 研究兴趣为机器学习、多尺度分析和系统辨识. 张晓光, 男, 1963 年生, 教授, 博士生导师, 研究兴趣为图像处理、机器学习、故障诊断和测试技术.

## 1 引言

特征抽取方法能够降低数据存储空间以及消除样本中冗余信息和噪声,便于人们对数据进行分析 and 理解,解决了因样本维数过高而导致分类器训练代价过大、泛化性能低等问题<sup>[1-3]</sup>. 常见的特征抽取算法有主元分析(PCA)、线性鉴别分析(LDA)、独立主元分析(ICA)和基于支持向量机(SVM)方法等<sup>[2-6]</sup>,其中,LDA 和 SVM 在机器学习领域中引起了人们极大关注. LDA 能够求取反映样本集全局信息的鉴别向量,具有较高的分类性能,被认为是目前最为有效的特征抽取算法,LDA 及其改进算法的基本思想就是根据 Fisher 鉴别准则求取多个最佳鉴别向量,然后再将模式高维特征向量投影到该最佳鉴别向量集上,构成低维鉴别特征空间<sup>[6]</sup>. 鉴于 LDA 算法不支持高维小样本数据以及 F-S LDA 算法抽取的特征分量之间存在统计相关等问题,人们提出了很多改进算法,如广义 LDA、零子空间(Null-subspace) LDA、基于广义 SVD 的 LDA、统计不相关 LDA 等<sup>[6-9]</sup>. 然而 LDA 存在单类样本为高斯分布时才能取得最优解、不能准确区分线性可分样本等问题<sup>[10-11]</sup>.

SVM 是基于结构风险最小化原理的有监督分类方法,具有支持小样本数据、无需假定样本分布、计算量与维数无关等优点<sup>[5,12]</sup>. 然而 SVM 只能抽取样本集局部信息,存在超平面法线方向与样本分布不一致的问题,影响了泛化性能的提高<sup>[11]</sup>. 对二分类问题,SVM 超平面法向量与 LDA 最佳鉴别向量具有相同的物理意义. 基于上述考虑,文献[4]通过把样本高维特征投影到 SVM 分类超平面法线方向上,构成一维特征空间,取得了较好的分类性能;类似 F-S LDA 算法<sup>[6]</sup>,文献[5]提出基于 SVM 的递归提取正交边界鉴别向量方法,其性能优于 SVM 算法. 然而上述算法性能受到 SVM 性能和抽取特征之间存在相关性的影响. 其原因是:(1) SVM 只利用了样本集的局部信息,影响了其泛化性能,因此与常见的特征抽取算法相比,文献[4]算法性能并没有显著的提高. 理论分析和实验证明,利用样本集全局信息能够有效提高 SVM 的性能<sup>[11,13]</sup>; (2) 基于统计不相关的鉴别向量集优于正交鉴别向量集的模式识别理论<sup>[6-7,14]</sup>,文献[5]抽取的特征之间存在相关性,影响了算法性能的提高.

基于上述分析,本文提出基于散度支持向量机(Scatter Support Vector Machine, SSVM)的统计

不相关特征抽取算法. 首先提出了 SSVM,该方法实现了 SVM 与 LDA 之间的折中,保证了 SSVM 性能的优越性;给出基于 SSVM 的递归抽取统计不相关边界鉴别向量集算法,证明了算法的收敛性结论以及算法终止条件;并通过核方法把线性抽取算法推广到非线性抽取算法. 本文算法继承了 SVM 和 LDA 的优点,通过使用 PCA/KPCA 把高维空间的 SSVM 优化问题变换到同构低维空间中等价优化问题,解决了小样本情况下 SSVM 优化问题求解的问题,有效地克服现有算法的缺点. 使用 UCI 数据集和 XM2VTS 人脸图像进行了仿真实验,结果说明本文算法是可行有效的. 下面首先介绍 SSVM 原理.

## 2 散度支持向量机(SSVM)

LDA 通过如下形式的 Fisher 准则求取鉴别向量,即

$$\max_w \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \quad (1)$$

其中, $\mathbf{S}_b$ 和 $\mathbf{S}_w$ 分别为类间、类内散度矩阵. 设样本总体散度矩阵为 $\mathbf{S}_t$ ,满足关系: $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w$ ;  $\mathbf{w}$ 为鉴别向量. 上面优化问题具有如下等价形式<sup>[15]</sup>:

$$\max_{\mathbf{w}, b} \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_t \mathbf{w}} \quad (2)$$

由于 LDA 充分利用样本集全局信息以及类别信息,鉴别向量方向与样本分布一致. 受 LDA 和文献[16]启发,文献[11]提出 MCSVM (Minimum Class variance, SVM),其优化问题通过 $\mathbf{w}^T \mathbf{S}_w \mathbf{w}$ 代替 SVM 优化问题 $\mathbf{w}^T \mathbf{w}$ 项得到. MCSVM 考虑了样本集局部和全局信息,实现 FLDA 与 SVM 之间的折衷,具有更优的分类性能<sup>[11]</sup>. 对两类样本集 $\{(\mathbf{x}_j, y_j)\}_{j=1}^N$ , $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jd}]^T$ ;  $y_j \in \{-1, +1\}$ ,假设分类超平面模型具有如下形式:

$$y = \mathbf{w}^T \mathbf{x} + b.$$

受文献[11,13]的启发,本文通过求解如下改进 SVM 优化问题 $P_1$ 得到 $\mathbf{w}$ 和 $b$ 的最优解,即

$$P_1: \min_{\mathbf{w}, b, \xi_j} \frac{1}{2} \mathbf{w}^T \mathbf{S}_t \mathbf{w} + C \sum_{j=1}^N \xi_j \quad (3)$$

$$\text{s. t. } \begin{cases} y_j (\mathbf{w}^T \mathbf{x}_j + b) \geq 1 - \xi_j \\ \xi_j \geq 0, \quad j=1, 2, \dots, N \end{cases}$$

这里, $\xi_1, \xi_2, \dots, \xi_N$ 是非负松弛变量; $C$ 为误差惩罚因子; $\mathbf{S}_t = \frac{1}{N} \sum_{j=1}^N (\mathbf{x}_j - \mathbf{m})(\mathbf{x}_j - \mathbf{m})^T$ 为样本总体散度矩阵,需要满足 $\mathbf{w}^T \mathbf{S}_t \mathbf{w} > 0$ . 从上式可以看出,优化问题

$P_1$  就是使用  $\mathbf{w}^T \mathbf{S}_i \mathbf{w}$  代替传统 SVM 优化问题的  $\mathbf{w}^T \mathbf{w}$ , 且满足约束条件  $\mathbf{w}^T \mathbf{S}_i \mathbf{w} > 0$ . 由于式(1)和(2)所示优化问题的等价关系, 优化问题  $P_1$  表示的改进 SVM 具有与 MCSVM 类似特性, 因而在理论上保证 SSVM 具有更优的分类性能. 由于  $\mathbf{S}_i$  为样本集的总体散度矩阵, 代表样本的全局信息, 因此, 本文把这种改进的 SVM 称作散度支持向量机(SSVM).

根据式(3)所示优化问题  $P_1$  的目标函数和约束条件构造 Lagrange 方程, 即

$$L = \min_{\mathbf{w}, b, \alpha_j, \beta_j} \frac{1}{2} \mathbf{w}^T \mathbf{S}_i \mathbf{w} + C \sum_{j=1}^N \xi_j - \sum_{j=1}^N \alpha_j (y_j (\mathbf{w}_i^T \mathbf{x}_j + b_i) - 1 + \xi_j) - \sum_{j=1}^N \beta_j \xi_j \quad (4)$$

其中,  $\alpha_j$  和  $\beta_j$  ( $j = 1, 2, \dots, N$ ) 是 Lagrange 乘子. 式(4)对  $\mathbf{w}, b, \alpha_j, \beta_j$  求导并等于 0 得

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{S}_i \mathbf{w} - \sum_{j=1}^N \alpha_j y_j \mathbf{x}_j = 0 \quad (5)$$

$$\frac{\partial L}{\partial b} = \sum_{j=1}^N \alpha_j y_j = 0 \quad (6)$$

$$\frac{\partial L}{\partial \xi_j} = C - \alpha_j - \beta_j = 0 \quad (7)$$

把式(5)~(7)代入优化问题  $P_1$ , 根据 KKT 条件, 优化问题  $P_1$  转换为如下形式, 即

$$P_2: \max_{\alpha_j} L = \sum_{j=1}^N \alpha_j - \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N y_j y_k \alpha_j \alpha_k \mathbf{x}_j^T \mathbf{S}_i^{-1} \mathbf{x}_k$$

$$\text{s. t.} \begin{cases} 0 \leq \alpha_j \leq C, j = 1, 2, \dots, N \\ \sum_{j=1}^N \alpha_j y_j = 0 \end{cases} \quad (8)$$

对于 SSVM, 其决策函数为

$$g(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x} + b) = \text{sign}\left(\frac{1}{2} \sum_{j=1}^N a_j y_j \mathbf{x}_j^T \mathbf{S}_i^{-1} \mathbf{x} + b\right).$$

根据 KKT 条件, 支持向量  $\mathbf{x}_k$  其对应的  $a_k \neq 0$ ,  $\xi_k = 0$ . 设支持向量集为  $\mathbf{SV}$ , 对支持向量  $\mathbf{x}_k \in \mathbf{SV}$  有下式成立:

$$y_k \left( \frac{1}{2} \sum_{j=1}^N y_j a_j \mathbf{x}_j^T \mathbf{S}_i^{-1} \mathbf{x}_k + b \right) = 1,$$

那么  $b$  的最优值可由下式求取, 即

$$b = \frac{1}{|\mathbf{SV}|} \sum_{\mathbf{x}_k \in \mathbf{SV}} \left( y_k - \frac{1}{2} \sum_{j=1}^N y_j a_j \mathbf{x}_j^T \mathbf{S}_i^{-1} \mathbf{x}_k \right).$$

需要注意的是, 对于高维小样本数据集, 即样本数据维数  $d \geq N$ ,  $\mathbf{S}_i$  是往往是奇异的, 这样就不能直接通过优化问题  $P_1$  求取向量  $\mathbf{w}_i$  的最优解. 然而, 对于高维小样本数据集, 在第 3 节将证明, 通过 PCA 对数据集进行维数约简, 把优化问题  $P_1$  转换为低维空间上等价的优化问题, 从而求取 SSVM 的最优

解. 下面讨论小样本情况下的 SSVM 优化问题求解问题.

### 3 小样本情况下 SSVM 的求解

由函数分析理论可知,  $\mathbf{S}_i$  在空间  $R^d$  是有界、紧的、正定的算子, 根据 Hilbert-Schmidt 定理, 矩阵  $\mathbf{S}_i$  的特征向量是  $R^d$  上的正交基<sup>[17]</sup>. 令  $\mathbf{D}$  表示  $\mathbf{S}_i$  非零特征值对应特征向量张成的空间,  $\mathbf{D}_\perp$  表示  $\mathbf{S}_i$  的零特征值对应特征向量张成的空间, 则空间  $\mathbf{D}_\perp$  表示空间  $\mathbf{D}$  的补空间. 因此, 对任意向量  $\mathbf{w} \in R^d$  可以表示为<sup>[16]</sup>

$$\mathbf{w} = \mathbf{v} + \boldsymbol{\zeta}, \quad \mathbf{v} \in \mathbf{D}, \quad \boldsymbol{\zeta} \in \mathbf{D}_\perp \quad (9)$$

定义映射  $L: R^d \rightarrow \mathbf{D}$ , 对任意的  $\mathbf{w} \in R^d$ , 均存在唯一的映射  $\mathbf{v}$  使得下式成立, 即

$$\mathbf{v} = L(\mathbf{w}) \quad (10)$$

这样, 映射  $L$  把优化问题  $P_1$  从空间  $R^d$  转换为低维空间  $\mathbf{D}$  的优化问题  $P_3$ . 很显然, 如果优化问题  $P_1$  和  $P_3$  等价, 那么对高维空间  $P_1$  的求解就转化为对低维空间  $\mathbf{D}$  上  $P_3$  进行求解, 解决了小样本情况下因  $\mathbf{S}_i$  奇异而导致的 SSVM 难以求解的问题. 定理 1 给出优化问题  $P_1$  等价于  $P_3$  的结论.

**定理 1.** 设  $\mathbf{S}_i$  为奇异散度矩阵, 设  $\mathbf{D}$  和  $\mathbf{D}_\perp$  分别表示  $\mathbf{S}_i$  非零、零特征值对应特征向量张成的空间, 那么空间  $R^d$  上的优化问题  $P_1$  等价于如下所示空间  $\mathbf{D}$  上的优化问题  $P_3$ , 即

$$P_3: \min_{\mathbf{v}, \boldsymbol{\zeta}} \frac{1}{2} \mathbf{v}^T \mathbf{S}_i \mathbf{v} + C \sum_{j=1}^N \xi_j$$

$$\text{s. t.} \begin{cases} y_j (\mathbf{v}^T \mathbf{x}_j + b) \geq 1 - \xi_j \\ \xi_j \geq 0, j = 1, 2, \dots, N \\ \mathbf{v}_i \in \mathbf{D} \end{cases} \quad (11)$$

其中,  $\mathbf{v}^T \mathbf{S}_i \mathbf{v} > 0$ .

证明. 根据式(9), 对任意  $\mathbf{w} \in R^d$  可以表示为

$$\mathbf{w} = \mathbf{v} + \boldsymbol{\zeta}, \quad \mathbf{v} \in \mathbf{D}, \quad \boldsymbol{\zeta} \in \mathbf{D}_\perp.$$

根据函数理论<sup>[17]</sup>, 显然有下式成立, 即

$$\mathbf{v}^T \mathbf{S}_i \mathbf{v} > 0, \quad \boldsymbol{\zeta}^T \mathbf{S}_i \boldsymbol{\zeta} = 0, \quad \mathbf{v}^T \mathbf{S}_i \boldsymbol{\zeta} = 0.$$

因此有

$$\mathbf{w}^T \mathbf{S}_i \mathbf{w} = (\mathbf{v} + \boldsymbol{\zeta})^T \mathbf{S}_i (\mathbf{v} + \boldsymbol{\zeta}) = \mathbf{v}^T \mathbf{S}_i \mathbf{v} \quad (12)$$

则式(4)所示的 Lagrange 方程转化为

$$L = \min_{\mathbf{w}_i, b, \alpha_j, \beta_j} \frac{1}{2} \mathbf{w}^T \mathbf{S}_i \mathbf{w} + C \sum_{j=1}^N \xi_j - \sum_{j=1}^N \alpha_j (y_j (\mathbf{w}^T \mathbf{x}_j + b) - 1 + \xi_j) - \sum_{j=1}^N \beta_j \xi_j$$

$$= \mathbf{v}^T \mathbf{S}_i \mathbf{v} + C \sum_{j=1}^N \xi_j - \sum_{j=1}^N \beta_j \xi_j -$$

$$\sum_{j=1}^N \alpha_j (y_j (\mathbf{v}^T \mathbf{x}_j + \zeta_j^T \mathbf{x}_j + b) - 1 + \xi_j) \quad (13)$$

如果对  $\zeta \in R^d$  有  $\zeta^T \mathbf{S}_t \zeta = 0$  成立, 那么对任意不同样本  $\mathbf{x}_j, \mathbf{x}_k$  均有  $\zeta^T \mathbf{x}_j = \zeta^T \mathbf{x}_k$  成立, 即所有样本在  $\zeta$  上的投影均为常数. 令  $c = \zeta^T \mathbf{x}_j, \forall j$ , 则有下式成立:

$$\sum_{j=1}^N \alpha_j y_j \zeta^T \mathbf{x}_j = c \sum_{j=1}^N \alpha_j y_j = 0 \quad (14)$$

根据上式, 式(13)可以写为

$$L = \min_{\mathbf{w}, b, \alpha_j, \beta_j} \frac{1}{2} \mathbf{v}^T \mathbf{S}_t \mathbf{v} + C \sum_{j=1}^N \xi_j - \sum_{j=1}^N \alpha_j (y_j (\mathbf{v}^T \mathbf{x}_j + b) - 1 + \xi_j) - \sum_{j=1}^N \beta_j \xi_j \quad (15)$$

从式(14)可以看出, 上式与优化问题  $P_3$  对应的 Lagrange 方程完全一致, 而式(14)表示的 Lagrange 方程又与优化问题  $P_1$  等价. 因此, 优化问题  $P_3$  与优化问题  $P_1$  是等价的. 证毕.

令  $\mathbf{S}_t$  非零特征值对应的特征向量作为变换矩阵  $\mathbf{P}$ , 且特征向量的数目  $K \leq d-1$ . 在很多情况下, 可以假定训练样本是独立、同分布的, 此时  $K = d-1$ . 由于  $\mathbf{P}$  的列向量是空间  $R^d$  的正交基, 变换矩阵  $\mathbf{P}$  就是空间  $\mathbf{D}$  到  $R^K$  的一对一映射, 根据函数分析理论, 在 PCA 变换矩阵  $\mathbf{P}$  下空间  $\mathbf{D}$  与  $R^K$  同构<sup>[17]</sup>, 即

$$\mathbf{v} = \mathbf{P}\mathbf{u}, \quad \mathbf{u} \in R^K \quad (16)$$

根据式(16)、式(11)所示的优化问题  $P_3$  可以转换为如下所示的等价优化问题  $P_4$ , 即

$$P_4: \min_{\mathbf{v}, b, \xi} \frac{1}{2} \mathbf{u}^T \tilde{\mathbf{S}}_t \mathbf{u} + C \sum_{j=1}^N \xi_j, \quad \mathbf{u}^T \tilde{\mathbf{S}}_t \mathbf{u} > 0$$

$$\text{s. t. } \begin{cases} y_j (\mathbf{u}^T \tilde{\mathbf{x}}_j + b) \geq 1 - \xi_j \\ \xi_j \geq 0, \quad j=1, 2, \dots, N \\ \mathbf{u} \in R^{d-1} \end{cases} \quad (17)$$

其中,  $\tilde{\mathbf{S}}_t = \mathbf{P}^T \mathbf{S}_t \mathbf{P}$ ;  $\tilde{\mathbf{x}}_j = \mathbf{P}^T \mathbf{x}_j$  为样本  $\mathbf{x}_j$  在 PCA 变换矩阵上的投影, 且  $\tilde{\mathbf{x}}_j \in R^{N-1}$ .

类似第 3 节, 优化问题  $P_4$  的对偶问题为

$$P_5: \max_{a_j} -\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N a_j a_k y_j y_k \tilde{\mathbf{x}}_j^T \tilde{\mathbf{S}}_t^{-1} \tilde{\mathbf{x}}_k + \sum_{j=1}^N a_j$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^N a_j y_j = 0 \\ 0 \leq a_j \leq C, \quad j=1, 2, \dots, N \end{cases} \quad (18)$$

根据上面的讨论, 对小样本问题, 为求出优化问题  $P_1$  的最优解, 首先把样本通过 PCA 变换矩阵  $\mathbf{P}$  映射到空间  $R^K$  上, 然后在低维特征空间  $R^K$  中求取优化问题的最优解. 然而, 如何在空间  $R^K$  求取统计

不相关的边界鉴别向量集以及算法是否收敛, 是本文需要研究的关键问题. 下节给出算法实现以及收敛性分析.

## 4 递归统计不相关特征抽取算法分析

### 4.1 算法的实现

设训练样本集  $\{(\mathbf{x}_i, y_i)\}_{i=1}^N, \mathbf{x}_i \in R^d, y_i \in \{-1, 1\}$ , SSVM 的分类超平面为  $y = \mathbf{w}_1^T \mathbf{x} + b_1$ . 分类超平面法向量  $\mathbf{w}_1$  包含了样本类别边界信息, 具有区分不同类别数据的能力. 对二分类样本,  $\mathbf{w}_1$  与 LDA 鉴别向量具有相同的物理意义, 因此, 本文把  $\mathbf{w}_1$  称作边界鉴别向量. 设  $\mathbf{w}_1$  为规格化向量, 即  $\|\mathbf{w}_1\| = 1$ , 样本  $(\mathbf{x}_i, y_i)$  在向量  $\mathbf{w}_1$  方向上的投影为

$$\hat{y}_i = \mathbf{w}_1^T \mathbf{x}_i \quad (19)$$

对上式左边同乘  $\mathbf{w}_1$  可得

$$\hat{\mathbf{y}}_i \triangleq \mathbf{w}_1 \hat{y}_i = \mathbf{w}_1 \mathbf{w}_1^T \mathbf{x}_i \quad (20)$$

很显然,  $\mathbf{w}_1 \mathbf{w}_1^T \mathbf{x}_i$  为在向量  $\mathbf{w}_1$  方向上对  $\mathbf{x}_i$  的逼近. 根据 PCA 理论, 假设在向量  $\mathbf{w}_1$  方向上样本集  $\mathbf{X}$  的逼近  $\hat{\mathbf{X}}$  构成空间  $\mathbf{S}_{w_1}$ , 样本集  $\mathbf{X}$  与逼近  $\hat{\mathbf{X}}$  之间的误差  $\tilde{\mathbf{X}}$  构成残差空间  $\mathbf{S}_r$ , 满足下面等式:

$$\mathbf{X} = \hat{\mathbf{X}} + \tilde{\mathbf{X}},$$

它们之间的几何关系如图 1 所示.

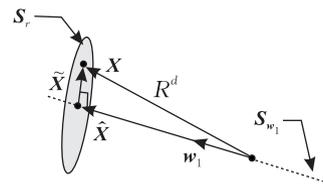


图 1  $\mathbf{X}$ 、逼近  $\hat{\mathbf{X}}$  和残差  $\tilde{\mathbf{X}}$  的几何关系

令  $\mathbf{w}_1 \mathbf{w}_1^T = \mathbf{I} + \mathbf{I}_{w_1}$ ,  $\mathbf{I}$  是  $d \times d$  单位矩阵,  $\mathbf{I}_{w_1}$  是非零矩阵, 那么式(20)可写为

$$\hat{\mathbf{x}}_i = \mathbf{w}_1 \mathbf{w}_1^T \mathbf{x}_i = \mathbf{x}_i + \mathbf{I}_{w_1}^T \mathbf{x}_i \quad (21)$$

根据图 1 所示, 样本  $\mathbf{x}_i$  与逼近  $\hat{\mathbf{x}}_i$  及其误差  $\tilde{\mathbf{x}}_i$  之间满足如下关系, 即

$$\mathbf{x}_i = \hat{\mathbf{x}}_i + \tilde{\mathbf{x}}_i \quad (22)$$

根据式(21)、(22),  $\tilde{\mathbf{x}}_i$  重新写为如下形式:

$$\tilde{\mathbf{x}}_i = (\mathbf{I} - \mathbf{w}_1 \mathbf{w}_1^T) \mathbf{x}_i \quad (23)$$

根据训练样本  $\{(\tilde{\mathbf{x}}_i, y_i)\}_{i=1}^N$ , 使用 SSVM 算法可求取边界鉴别向量  $\mathbf{w}_2$ ; 然后根据式(23)递归求取边界鉴别向量  $\mathbf{w}_3, \mathbf{w}_4, \dots, \mathbf{w}_l$ ; 最后把空间  $R^d$  模式高维特征投影到边界鉴别向量集  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_l\}$  上, 构成低维特征空间  $R^l$ . 在定理 2 给出低维空间中的特

征之间是否存在相关性的结论.

**定理 2.** 对训练样本集  $\{(\mathbf{x}_k, y_k)\}_{k=1}^N, \mathbf{x}_k \in R^d, y_k \in \{-1, 1\}$ , 基于上述方法使用 SSVM 递归求取边界鉴别向量  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_l$ , 且  $\|\mathbf{w}_i\| = 1, i = 1, 2, \dots, l$ . 对任意两个边界鉴别向量  $\mathbf{w}_i, \mathbf{w}_j$ , 样本  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$  在  $\mathbf{w}_i, \mathbf{w}_j$  方向上投影得到的特征是不相关的, 即

$$\text{cov}(\mathbf{w}_i^T \mathbf{X}, \mathbf{w}_j^T \mathbf{X}) = \mathbf{w}_i^T \mathbf{S}_t \mathbf{w}_j = 0, \quad i, j = 1, 2, \dots, l, \quad i \neq j \quad (24)$$

证明. 根据训练样本集, 由 SSVM 算法求取的  $\mathbf{w}_1$  作为第 1 个边界鉴别向量. 根据式(23),  $\mathbf{x}_k$  在  $\mathbf{w}_1$  方向上逼近残差为

$$\tilde{\mathbf{x}}_k^1 = (\mathbf{I} - \mathbf{w}_1 \mathbf{w}_1^T) \mathbf{x}_k \quad (25)$$

对残差样本集  $\{(\tilde{\mathbf{x}}_k^1, y_k)\}_{k=1}^N$ , 使用 SSVM 算法求取  $\mathbf{w}_2$  作为第 2 个边界鉴别向量. 然后基于式(25)递归求取边界鉴别向量  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_l, \|\mathbf{w}_i\| = 1, i = 1, 2, \dots, l$ . 对式(25)左边乘以  $\mathbf{w}_1^T$ , 有下面等式成立:

$$\mathbf{w}_1^T \tilde{\mathbf{x}}_k^1 = \mathbf{w}_1^T (\mathbf{I} - \mathbf{w}_1 \mathbf{w}_1^T) \mathbf{x}_k = 0 \quad (26)$$

即  $\mathbf{w}_1$  正交于样本残差空间. 根据式(5)可知,  $\mathbf{w}_2 = \mathbf{S}_t^{-1} \sum_{k=1}^N \alpha_k y_k \tilde{\mathbf{x}}_k^1$ , 对  $\mathbf{w}_2$  上式左边乘以  $\mathbf{w}_1^T \mathbf{S}_t$  并根据式(26)可得

$$\mathbf{w}_1^T \mathbf{S}_t \mathbf{w}_2 = \mathbf{w}_1^T \mathbf{S}_t \mathbf{S}_t^{-1} \sum_{k=1}^N \tilde{\alpha}_{2k} \tilde{\mathbf{x}}_k^1 = 0.$$

对任意的  $\mathbf{w}_i, \mathbf{w}_j, i \neq j, i, j = 1, 2, \dots, l$ , 同理可证明有下式成立:

$$\mathbf{w}_i^T \mathbf{S}_t \mathbf{w}_j = 0. \quad \text{证毕.}$$

定理 2 证明了本文算法能够从样本集抽取统计不相关的特征,

综合以上所述, 本文给出如下递归统计不相关特征抽取算法实现步骤:

假设有  $N$  个训练样本  $\{(\mathbf{x}_j, y_j)\}_{j=1}^N \in R^d \times R$ , 其中  $d$  是输入数据的维数, 根据 SSVM 原理, 本文提出的算法实现步骤如下:

1. 初始化. 给出算法终止条件, 迭代次数  $r=1$ , 令  $\mathbf{w}_r$  表示迭代计算得到的边界鉴别向量.

2. 确定本文 SSVM 模型参数. 基于第 2 节 SSVM 算法, 使用交叉校验方法确定 SSVM 的模型参数, 根据上述样本训练 SSVM, 并把求取得到的超平面法向量  $\mathbf{w}$  作为第 1 个边界鉴别向量  $\mathbf{w}_r; r \leftarrow r+1$ .

3. 求取边界鉴别向量  $\mathbf{w}_r$ . 使用式  $\tilde{\mathbf{x}}_i^r = (\mathbf{I} - \mathbf{w}_i \mathbf{w}_i^T) \tilde{\mathbf{x}}_i^{r-1}$  计算逼近残差样本集  $\{(\tilde{\mathbf{x}}_i^r, y_i)\}_{i=1}^N$ , 根据残差样本集对 SSVM 优化问题求解, 把求取的超平面法向量作为第  $r$  个边界鉴别向量  $\mathbf{w}_r$ .

4. 结束判断. 判断是否满足算法结束条件, 如果不满

足, 则  $r \leftarrow r+1$ , 转到步 3 继续递归求取统计不相关边界鉴别向量; 否则,  $r \leftarrow r-1$ , 转入步 5.

5. 抽取特征. 把样本投影到边界鉴别向量变量集  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r\}$ , 构成  $r$  维特征空间, 即抽取  $r$  个统计不相关特征.

6. 分类器训练: 在低维特征空间建立分类器模型.

7. 算法结束.

本文递归统计不相关特征抽取算法比较简单, 易于实现.

算法实现的关键是确定合适的终止条件, 而算法的收敛性决定算法终止条件的存在, 下面分析算法的收敛性.

## 4.2 算法的收敛性分析

SSVM 基于结构风险最小化原理, 其结构风险越小表示边界向量  $\mathbf{w}_l$  包含的分类信息越多. 如果 SSVM 的结构风险随着  $l$  的递增而增大, 则边界鉴别向量  $\mathbf{w}_l$  包含的分类信息越少, 即算法收敛. 基于以上分析, 定理 3 给出算法收敛的结论.

**定理 3.** 令  $\{\mathbf{w}_{l-1}, b_{l-1}, \xi_1^{l-1}, \xi_2^{l-1}, \dots, \xi_N^{l-1}\}$  和  $\{\mathbf{w}_l, b_l, \xi_1^l, \xi_2^l, \dots, \xi_N^l\}$  是对应 SSVM 优化问题的解, 如果

$$\frac{1}{2} \mathbf{w}_{l-1}^T \mathbf{S}_t \mathbf{w}_{l-1} + C \sum_{j=1}^N \xi_j^{l-1} \leq \frac{1}{2} \mathbf{w}_l^T \mathbf{S}_t \mathbf{w}_l + C \sum_{j=1}^N \xi_j^l \quad (27)$$

成立, 那么算法是收敛的.

证明. 由式(24)可知,  $\mathbf{x}_i^l$  和  $\mathbf{x}_i^{l-1}$  满足下面关系:

$$\mathbf{x}_i^l = (\mathbf{I} - \mathbf{w}_{l-1} \mathbf{w}_{l-1}^T) \mathbf{x}_i^{l-1} \quad (28)$$

对上式左边乘以  $\mathbf{w}_l^T \mathbf{S}_t$  得到如下形式:

$$\mathbf{w}_l^T \mathbf{S}_t \mathbf{x}_i^l = \mathbf{w}_l^T \mathbf{S}_t \mathbf{x}_i^{l-1}.$$

对式(28)进行迭代计算得到如下等式, 即

$$\mathbf{w}_l^T \mathbf{S}_t \mathbf{x}_i^l = \mathbf{w}_l^T \mathbf{S}_t \mathbf{x}_i^{l-1} = \dots = \mathbf{w}_l^T \mathbf{S}_t \mathbf{x}_i.$$

由于  $\mathbf{S}_t$  是已知正定矩阵, 显然下式成立:

$$\mathbf{w}_l^T \mathbf{x}_i^l = \mathbf{w}_l^T \mathbf{x}_i^{l-1} = \dots = \mathbf{w}_l^T \mathbf{x}_i \quad (29)$$

根据式(29), 有下面等式成立:

$$y_i (\mathbf{w}_l^T \mathbf{x}_i^l + b_l) = y_i (\mathbf{w}_l^T \mathbf{x}_i^{l-1} + b_l) = \dots = y_i (\mathbf{w}_l^T \mathbf{x}_i + b_l) \geq 1 - \xi_i^l, \quad \xi_i^l \geq 0, \quad i = 1, 2, \dots, N \quad (30)$$

由 SSVM 算法可知,  $\{\mathbf{w}_l, b_l, \xi_1^l, \xi_2^l, \dots, \xi_N^l\}$  必须满足以下约束条件:

$$\xi_j^l \geq 0, \quad y_i (\mathbf{w}_l^T \mathbf{x}_i^l + b_l) \geq 1 - \xi_i^l, \quad i = 1, 2, \dots, N.$$

根据式(30), SSVM 的最优解  $\{\mathbf{w}_l, b_l, \xi_1^l, \xi_2^l, \dots, \xi_N^l\}$  对样本集  $\{(\mathbf{x}_i^t, y_i)\}_{i=1,2,\dots,l}^t$  均满足如式(29)所示的约束条件,  $t = 1, 2, \dots, l$ . 同理,  $\{\mathbf{w}_{l-1}, b_{l-1}, \xi_1^{l-1}, \xi_2^{l-1}, \dots, \xi_N^{l-1}\}$  对  $\{(\mathbf{x}_i^t, y_i)\}_{i=1,2,\dots,l}^{t-1}$  满足约束条件, 而对样本集  $\{(\mathbf{x}_i^t, y_i)\}_{i=1}^N$  不满足约束条件, 因此下式

成立:

$$\frac{1}{2} \mathbf{w}_{l-1}^T \mathbf{S}_l \mathbf{w}_{l-1} + C \sum_{j=1}^N \xi_j^{l-1} \leq \frac{1}{2} \mathbf{w}_l^T \mathbf{S}_l \mathbf{w}_l + C \sum_{j=1}^N \xi_j^l.$$

上式表明,随着抽取特征数量  $l$  的递增,结构风险也随之增加,边界鉴别向量  $\mathbf{w}_l$  包含的分类信息也减少,算法逐渐收敛. 证毕.

定理 3 虽然给出了算法的收敛性结论,但不能直接作为算法终止条件. 定理 4 给出算法终止的条件.

**定理 4.** 对本文递归不相关特征抽取算法,设每次使用的训练样本集为  $\{(\mathbf{x}_i^l, y_i)\}_{i=1}^N$ , 其中,  $l=1, 2, \dots, m$  为抽取特征的数量,有如下关系成立:

$$\|\mathbf{x}_i^l\|^2 < \|\mathbf{x}_i^{l-1}\|^2 \quad (31)$$

对  $\forall \epsilon > 0$ , 必存在一个正整数  $m$  使得下式成立:

$$\max_i \|\mathbf{x}_i^m\| \leq \epsilon \quad (32)$$

证明. 根据式(28)有下式成立:

$$\|\mathbf{x}_i^{l-1}\|^2 - \|\mathbf{x}_i^l\|^2 = (\mathbf{w}_{l-1}^T \mathbf{x}_i^{l-1})^2 > 0,$$

即  $\|\mathbf{x}_i^l\|^2 < \|\mathbf{x}_i^{l-1}\|^2$ , 可以看出  $\|\mathbf{x}_i^l\|$  随着  $l$  的增大而单调递减.

注意到  $\mathbf{x}_i = \mathbf{x}_i^1$ , 根据 SSVM 算法可知

$$\mathbf{w}_1 = \mathbf{S}_1^{-1} \sum_{j=1}^N \alpha_j y_j \mathbf{x}_j.$$

由于  $\mathbf{S}_1^{-1}$  是常数矩阵, 那么  $\mathbf{w}_1 \in \mathbf{S} = \text{span}\{\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_N^1\}$ . 从式(28)可知:

$$\mathbf{x}_i^2 = (\mathbf{I} - \mathbf{w}_1 \mathbf{w}_1^T) \mathbf{x}_i,$$

很显然  $\mathbf{x}_i^2 \in \mathbf{S}, i=1, 2, \dots, N$ . 同理可以证明  $\mathbf{w}_2 \in \mathbf{S}$ .

由于  $\mathbf{S}$  是有限维空间, 一定存在一个正整数  $m$  使得下式成立:

$$\|\mathbf{x}_i^m\|^2 = \|\mathbf{x}_i\|^2 - \sum_{j=1}^{m-1} (\mathbf{w}_j^T \mathbf{x}_i)^2 = 0.$$

显然, 上式等价于式(32). 很明显, 式(32)成立时, 根据本节的分析可知, 算法是收敛的. 证毕.

本文提出 3 种算法终止条件: (1) 确定抽取的特征数量. 该方法难以确定合适的特征数量; (2) 根据定理 4, 设定阈值  $\epsilon > 0$ , 当  $\max_{i=1, 2, \dots, N} \|\mathbf{x}_i^l\| < \epsilon$  时算法终止, 其中,  $\mathbf{x}_i^l$  为第  $l-1$  次递归计算后求取的样本残差; (3) 基于分类精度不会随着统计不相关特征增加而下降的原理<sup>[5]</sup>, 首先设定分类精度变化阈值  $\theta_{s,pe}$ , 分类精度变换小于阈值  $\theta_{s,pe}$  时算法终止, 避免了抽取太多特征付出计算代价太大的问题.

## 5 非线性特征抽取算法

对大量的非线性模式数据, 需要把上述线性特

征抽取算法推广到非线性特征抽取算法. 本文首先通过核方法把上面线性 SSVM 推广到非线性 SSVM, 然后基于非线性 SSVM 递归抽取特征. 首先把上面训练样本  $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$  映射到 Hilbert 空间  $\{(\phi(\mathbf{x}_i), y_i)\}_{i=1}^N$ , 其中  $\phi(\cdot): R^d \rightarrow \mathbf{H}$  为高维非线性映射,  $\mathbf{H}$  空间表示 Hilbert 空间, 满足关系  $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = K_H(\mathbf{x}_i, \mathbf{x}_j)$ ,  $K_H(\mathbf{x}_i, \mathbf{x}_j)$  为满足 Mercer 条件的核函数. 在  $\mathbf{H}$  空间内散度函数定义为

$$\mathbf{S}_t^\phi = \frac{1}{N} \sum_{i=1}^N (\phi(\mathbf{x}_i) - \mathbf{m})(\phi(\mathbf{x}_i) - \mathbf{m})^T,$$

其中,  $\mathbf{m} = \frac{1}{N} \sum_{j=1}^N \phi(\mathbf{x}_j)$  是总体样本的均值.

在  $\mathbf{H}$  空间内, SSVM 的优化问题具有如下形式, 即

$$\begin{aligned} P_6: \min_{\mathbf{w}, b, \xi} & \frac{1}{2} \mathbf{w}^T \mathbf{S}_t^\phi \mathbf{w} + C \sum_{j=1}^N \xi_j, \quad \mathbf{w}^T \mathbf{S}_t^\phi \mathbf{w} > 0 \\ \text{s. t.} & \begin{cases} y_j (\mathbf{w}^T \phi(\mathbf{x}_j) + b) \geq 1 - \xi_j \\ \xi_j \geq 0, \quad j=1, 2, \dots, N \end{cases} \end{aligned} \quad (33)$$

众所周知,  $\mathbf{H}$  空间的维数很高,  $\mathbf{S}_t^\phi$  几乎总是奇异的, 不能满足条件  $\mathbf{w}_t^T \mathbf{S}_t^\phi \mathbf{w}_t > 0$ , 因此无法由优化问题  $P_6$  直接求取  $\mathbf{w}$  的最优解. 由于  $\mathbf{S}_t^\phi$  是  $\mathbf{H}$  空间中的有界、正定紧的自伴算子, 根据 Hilbert-Schmidt 定理<sup>[17]</sup>,  $\mathbf{S}_t^\phi$  的特征向量组成  $\mathbf{H}$  空间的正交基. 令  $\mathbf{D}^\phi$ 、 $\mathbf{D}_\perp^\phi$  分别为  $\mathbf{S}_t^\phi$  非零、零特征值对应特征向量张成的互补空间, 那么  $\mathbf{H}$  空间的向量  $\mathbf{w}$  可以表示为

$$\mathbf{w} = \mathbf{u} + \mathbf{v}, \quad \mathbf{u} \in \mathbf{D}^\phi, \quad \mathbf{v} \in \mathbf{D}_\perp^\phi \quad (34)$$

类似定理 2, 可以证明, 线性 SSVM 优化问题  $P_6$  可以转化为空间  $\mathbf{D}^\phi$  的等价优化问题. 设  $\mathbf{S}_t^\phi$  的非零特征向量的数目为  $K$ , 且  $K \leq N-1$ , 那么  $\mathbf{D}^\phi$  的维数也为  $K$ . 根据函数分析理论,  $\mathbf{D}^\phi$  与欧氏空间  $R^K$  同构, 存在同构映射

$$\mathbf{u}_i = \mathbf{P} \mathbf{q}_i, \quad \mathbf{q}_i \in R^{N-1} \quad (35)$$

其中, 变换矩阵  $\mathbf{P}$  的列向量由  $\mathbf{S}_t^\phi$  的非零特征向量组成. 优化问题  $P_6$  转化到  $R^K$  等价优化问题  $P_7$ , 即

$$\begin{aligned} P_7: \min_{\mathbf{w}, b, \xi} & \frac{1}{2} \mathbf{q}^T \tilde{\mathbf{S}}_t \mathbf{q} + C \sum_{j=1}^N \xi_j, \quad \mathbf{q}^T \tilde{\mathbf{S}}_t \mathbf{q} > 0, \quad \mathbf{q} \in R^K \\ \text{s. t.} & \begin{cases} y_j (\mathbf{q}^T \mathbf{z}_j + b) \geq 1 - \xi_j \\ \xi_j \geq 0, \quad j=1, 2, \dots, N \end{cases} \end{aligned} \quad (36)$$

这里,  $\tilde{\mathbf{S}}_t = \mathbf{P}^T \mathbf{S}_t^\phi \mathbf{P}$ ;  $\mathbf{z}_j = \mathbf{P}^T \phi(\mathbf{x}_j)$  是通过 KPCA 变换矩阵把  $\phi(\mathbf{x}_j)$  投影到  $R^K$ , 有关 KPCA 的详细介绍可参考文献<sup>[18]</sup>. 这样对非线性 SSVM 优化问题  $P_6$  的求解问题可转化为如式(36)所示  $R^K$  中优化问题  $P_7$  进行求解. 由于  $\tilde{\mathbf{S}}_t^\phi$  非奇异, 最优分类超平面法向量为

$$\mathbf{q} = \sum_{j=1}^N \alpha_j \mathbf{y}_j \tilde{\mathbf{S}}^{-1} \mathbf{P}^T \phi(\mathbf{x}_j).$$

非线性特征抽取方法为, 首先把样本  $\{(\mathbf{x}_j, \mathbf{y}_j)\}_{j=1}^N$  映射到高维特征空间  $\{(\phi(\mathbf{x}_j), \mathbf{y}_j)\}_{j=1}^N$ ; 然后通过 KPCA 非零特征向量组成的变换矩阵  $\mathbf{P}$  把  $\mathbf{H}$  空间样本投影到  $K$  维特征空间  $\{(\mathbf{z}_j = \mathbf{P}^T \phi(\mathbf{x}_j), \mathbf{y}_j)\}_{j=1}^N$ . 类似 4.1 节, 在  $K$  维特征空间中通过优化问题  $P_7$  递归求取统计不相关边界鉴别向量  $\mathbf{q}_i$ ,  $i=1, 2, \dots, L$ . 算法终止条件可与 4.2 节提出的方法类似.

根据以上分析, 可以通过核方法把线性特征抽取算法推广到非线性特征抽取问题.

## 6 仿真实验

为说明本文算法的有效性, 首先以 UCI 中 Heart-disease 和 Breast cancer 数据集为例, 研究算法抽取特征数量与分类精度之间关系. Heart-disease 数据集包含 270 个样本, 每个样本有 13 个属性; Breast cancer 数据集包括 277 个样本, 每个样本有 9 个属性. 由于 SVM 与递归支持向量集维数约简算法 (RSVM) 在文献[5]做了详细的比较, 因此, 本文算法只与 RSVM 进行比较, 在低维特征空间中使用最近邻分类器对测试样本分类. 本文 SSVM 算法使用高斯核函数, 使用 5-fold 交叉校验方法确定模型参数  $\sigma$  和  $C$ . 在仿真实验中, 从所有的数据集随机选取 50% 作为训练样本, 其余作为测试样本, 此过程重复 10 次, 计算分类精度的平均值作为最终分类精度. 仿真程序在 matlab7.0 环境下运行, 并使用其自带的优化工具包, 仿真结果如图 2、图 3 所示. 从图中可以看出, (1) RSVM 分类精度随着抽取特征数量的增加会出现下降的现象, 给确定最佳特征数量带来了困难. 本文算法分类精度在开始时会随着抽取特征数量递增快速增加, 当达到或接近最佳特征数时分类精度增加得非常缓慢, 甚至是不变, 从而说明本文提出的判断算法终止条件的合理性; (2) 本文算法优于 RSSVM 算法.

为进一步说明算法的有效性, 本文算法与 KFDDA<sup>[2]</sup>、KRDA<sup>[19]</sup>、SVM 和 RSSVM 算法对比. 实验选用 UCI 数据集 Breast-cancer、Heart-disease、Sonar、Cancer、Ionosphere、Waste plant 和 Waveform 以及 XM2VTS 人脸数据库进行镜片检测<sup>[19]</sup>. Waste plant 数据集有 13 个类别. 为简便起见, 将其合并成正常情况和非正常情况两类, 正常情况有

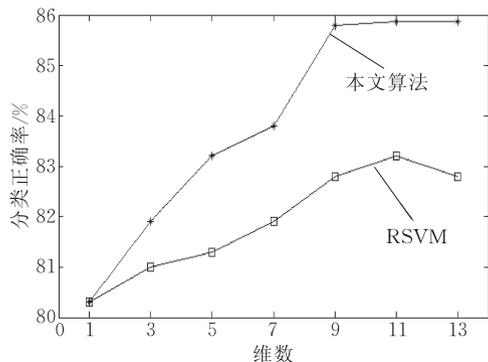


图 2 Heart-disease 抽取特征数量与分类精度

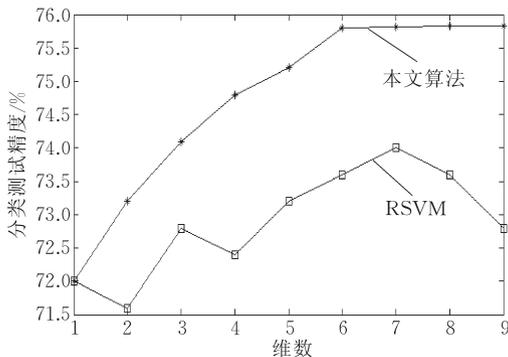


图 3 Breast-cancer 抽取特征数量与分类精度

332 个样本, 非正常情况包括性能超过平均值的正常情况、低输入的正常情况、二沉池故、暴雨以及固体溶度过负荷, 分别对应的样本数量为 116、65、7、3、4. UCI 数据集统计特性如表 1 所示. XM2VTS 数据库共有 2360 幅图像, 其中 1518 幅人脸图像有镜片, 842 幅无镜片. 人脸图像的分辨率  $720 \times 576$ , 每个图像被旋转和缩放, 使得所有眼睛中心放置在特定像素上, 然后修剪为  $85 \times 156$ , 通过直方图均衡化技术把图像强度规格化到均值为 0、方差为 1, 部分图像如图 4 所示. 实验中, 如果某个人出现在训练样本集中, 则从测试数据中剔除此人图片. 所有算法采用高斯核函数, 使用交叉验证法确定 SVM、RSSVM 以及 SSVM 正则化参数  $\sigma$  和  $C$ , KRDA 与 KFDDA 参数确定方法参考文献[20]. 实验中, UCI 数据集所有的数据属性被规格化为  $[-1, 1]$ , 在低维特征空间中使用最近邻分类器估计精度. 样本选取方法同上, 测试样本的平均精度如表 2 所示.

表 1 数据集的统计

数据集	样本数	维数
Sonar	208	60
Ionosphere	351	34
Waste plant	527	38
Cancer	683	10
Wave form	300	40



图 4 XM2VTS 数据库部分人脸图像

表 2 分类器测试精度率比较

数据集	精度/%				
	SVM	RSVM	KFDA	KRDA	本文算法
Breast cancer	73.3±4.7	74.0±4.2	71.2±4.6	74.5±4.2	75.7±4.3
Heart-disease	82.8±3.3	83.2±3.4	82.5±3.4	83.3±3.1	86.5±3.2
Sonar	83.5±4.4	84.9±6.1	83.3±4.3	83.5±4.2	85.9±4.3
Waste plant	74.8±3.5	76.3±3.2	71.6±4.7	75.1±3.1	78.2±3.2
Cancer	95.7±1.8	96.2±2.0	95.2±2.8	96.1±1.3	96.8±1.9
Waveform	79.9±2.4	81.2±2.5	77.1±3.1	92.3±2.2	89.3±2.3
Ionosphere	92.7±2.4	93.3±2.3	92.7±4.5	94.4±2.1	95.2±2.6
XM2VTS	97.2±2.8	97.6±2.6	93.0±5.1	97.4±2.3	98.6±2.5

从表 2 可以看出,本文算法明显优于 SVM、RSVM 和 KFDA 算法,对绝大部分数据集优于 KRDA 算法,只有个别数据集分类精度低于 KRDA 算法.出现上述现象的原因是,一方面 SSVM 具有类似 MCSVM 的特点,因而具有更高的分类性能;其次,本文算法抽取的特征之间不存在冗余信息,可以有效提高分类样本精度.

## 7 结 语

本文提出了一种基于 SSVM 的递归统计不相关特征抽取算法,基于核方法把线性特征抽取方法推广到非线性特征抽取,通过 PCA/KPCA 变换解决了高维特征空间中因散度矩阵奇异性导致 SSVM 优化问题无法求解的问题,证明了算法的收敛性以及抽取的特征具有统计不相关性,给出了算法终止条件.文中还给出了递归非线性特征抽取算法的实现方法.本文给出的线性、非线性特征抽取方法能够抽取不相关特征,有效地提高分类器的性能,克服了现有算法存在的问题,具有重要的理论和应用价值.需要指出的是,本文算法计算代价较大,如何使用现有的快捷优化算法,提高算法的速度,对于算法的实际应用具有重要的意义.

## 参 考 文 献

[1] Li Junhong, Cui Peiling. Improved kernel fisher discriminant analysis for fault diagnosis. *Expert Systems with Applications*, 2009, 36: 1423-1432

[2] Yang J, Frangi A F, Zhang D, Jin Z. KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(2): 230-244

[3] Sergios Theodoridis, Konstantinos Koutroumbas. *Pattern Recognition*. 3rd Edition. Beijing: Publishing House of Electronic Industry, 2006

[4] Ying Zi-Lu, Tang Jing-Hai, Li Jing-Wen, Zhang You-Wei. Support vector discriminant analysis and its application to facial expression recognition. *Acta Electronica Sinica*, 2008, 36(4): 725-730(in Chinese)

(应自炉, 唐京海, 李景文, 张有为. 支持向量鉴别分析及在人脸表情识别中的应用. *电子学报*, 2008, 36(4): 725-730)

[5] Qing Tao, Chu Dejun, Wang Jue. Recursive support vector machines for dimensionality reduction. *IEEE Transactions on Neural Networks*, 2008, 19(1): 189-193

[6] Yang Jing-Yu, Jin Zhong, Hu Zhong-Shan. A theorem on dimensionality of the uncorrelated optimal discriminant feature space. *Chinese Journal of Computers*, 2003, 26(1): 110-115(in Chinese)

(杨静宇, 金忠, 胡钟山. 具有统计不相关性的最佳鉴别特征空间的维数定理. *计算机学报*, 2003, 26(1): 110-115)

[7] Ye Jieping, Janardan Ravi, Qi Li et al. Feature reduction via generalized uncorrelated linear discriminant analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2006, 18(10): 1312-1322

[8] Xiong J, Ye T. Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *Journal of Machine Learning Research*, 2006, 7: 1183-1204

[9] Howland P, Park H. Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(8): 995-1006

- [10] Ye Jieping. Least squares linear discriminant analysis//Proceedings of the 24th International Conference on Machine Learning, 2007; 1087-1093
- [11] Stefanos Zafeiriou, Anastasios Tefas, Ioannis Pitas. Minimum class variance support vector machines. *IEEE Transactions on Image Processing*, 2007, 16(10): 2551-2564
- [12] Widodo Achmad, Yang Bo-Su. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 2007, 21: 2560-2574
- [13] Huang Kaizhu, Yang Haiqin, King Irwin et al. Maxi-Min margin machine: Learning large margin classifiers locally and globally. *IEEE Transactions on Neural Networks*, 2008, 19(2): 260-272
- [14] Jin Zhong, Yang Jingyu, Hu Zhongshan. Face recognition based on the uncorrelated discriminant transformation. *Pattern Recognition*, 2001, 34: 1405-1416
- [15] Liu K, Yang J Y et al. An efficient algorithm for Foley-sammon optimal set of discriminant vectors by algebraic method. *International Journal of Pattern Recognition and Artificial Intelligence*, 1992, 6(5): 817-829
- [16] Cevikalp H, Neamtu M, Wilkes M Barkana et al. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(1): 4-13
- [17] Kreyszig E. *Introductory Functional Analysis with Applications*. New York: Wiley, 1978
- [18] Smola A J. *Learning with kernels* [Ph.D. dissertation]. Berlin, Germany: Technischen Universitat, 1998
- [19] Moghaddam B, Yang Ming-Hsuan. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 707-711
- [20] Ji Shuiwang, Ye Jieping. Kernel uncorrelated and regularized discriminant analysis: A theoretical and computational study. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(10): 1131-1142



**REN Shi-Jin**, born in 1971, Ph. D., associate professor. His research interests include machine learning, pattern recognition, image processing.

**WANG Xiao-Lin**, born in 1980, associate professor. Her research interests include data mining, artificial intelli-

gent and pattern recognition.

**LV Jun-Huai**, born in 1959, associate professor. His research interests include machine learning, multi-scale analysis and system identification.

**ZHANG Xiao-Guang**, born in 1963, professor, Ph. D. supervisor. His research interests include image processing, machine learning, fault diagnosis and measurement technology.

## Background

High-dimensional data and small sample size problem occur in many applications, such as machine learning, face recognition, gene expression data analysis, and fault detection. To overcome the curse of dimensionality, efficient and robust feature extraction is necessary to be applied as a pre-processing step and has received wide attention in research fields. Since most of the existing methods can only work well on data with specified distribution, and commonly lack of robustness, it is challenging to guarantee the efficiencies of extracted dimensions.

This paper proposes a new feature extraction algorithm. Compared with the state-art of methods, the proposed method is based on the principle of support vector machines, and the extracted uncorrelated features are not only desirable in many applications because they contain minimum redundancy, but also can simultaneously keep classification margin maximal in the reduced feature space, guaranteeing the generalization performance of classifier in theory. The algorithm

can work well with respect to corrupted data with unknown distribution and its computation complexity is insensitive to the number of dimensions.

The work is supported by the Natural Science Foundation Training Program of Xuzhou Normal University (grant No.08XLY10), China Postdoctoral Science Foundation (grant No.20060390277) and Six Talent Flood Tide Project of Jiangsu Province (grant No.06-E-05). The first program aims to on-line monitor conditions of coal boiler combustion in power plant based on data-driven fault detection and diagnosis theory. The second foundation devotes to detect weld defects effectively and automatically through features extracted from weld images. The last project is to construct a platform to test the proposed methods. The authors of this paper have been engaged in image processing, feature extraction and fault monitoring. These studies have laid an important theoretical and application basis for future work.