DOI:10.3969/j.issn.1673-4785.201312040
网络出版地址:http://www.cnki.net/kcms/doi/10.3969/j.issn.1673-4785.201312040.html

# 支持向量机的多观测样本二分类算法

李欢,王士同
(江南大学 数字媒体学院,江苏 无锡 214000)

**摘 要**:针对多观测样本的分类问题,提出基于 SVM 的多观测样本二分类算法。每次分类时,首先限制组成多观测样本的所有单观测样本属于同一类别,对多观测样本的类别做 2 次假设,通过比较不同类别假设下的目标函数最优解来确定多观测样本的类别。该方法无需对分类器进行训练或提前对训练集进行特征表示,而是将已知标签样本集和多观测样本作为一个整体,充分利用特征空间中同类样本连续分布这一特点,使得分类更加准确。结果表明所提方法的有效性。

**关键词**:模式识别;多观测;同类样本;SVM;二分类
**中图分类号**: TP391.4 **文献标志码**:A **文章编号**:1673-4785(2014)04-392-09

**中文引用格式**:李欢,王士同. 支持向量机的多观测样本二分类算法[J]. 智能系统学报, 2014, 9(4): 392-400.
**英文引用格式**:LI Huan, WANG Shitong. Binary-class classification algorithm with multiple-access acquired objects based on the SVM[J]. CAAI Transactions on Intelligent Systems, 2014, 9(4): 392-400.

# Binary-class classification algorithm with multiple-access acquired objects based on the SVM

LI Huan, WANG Shitong
(1. School of Digital Media, Jiangnan University, Wuxi 214000, China; 2. School of Digital Media, Jiangnan University, Wuxi 214000, China)

**Abstract**: The binary-class classification algorithm with multiple-access acquired objects based on the SVM is proposed for the purpose of classification of an object given with multiple observations in this paper. In each classification, initially all single observation samples in the multiple observation sample set are restricted to a same class. Two hypotheses are made for the class of the multiple observation sample set, and the class is determined by comparing the optimal values of the different objective functions under different class hypotheses. This method does not require training the classifier or early feature representation of the training set, instead, it takes advantage of the continuity law of the feature space of similar samples with the labeled samples and multiple observation samples as a whole, making the algorithm more accurate for classifications. Experiments show that the proposed method is valid and efficient.

**Keywords**: pattern recognition; multiple observations; similar samples; SVM; binary-class classification

传统模式识别主要针对测试模式为单观测样本的情况。然而,随着人工智能技术的飞速发展,数据采集工作变得越来越容易,人们常常可以获取某特定模式在不同时刻或不同条件下的多个观测样本。例如,日常生活中,可以用摄像头获取一个物体或一个人在不同时刻、不同光照条件下的图像数据,也可以借用多个摄像头从不同的角度获取图像数据。此外,即使是相同的观测数据,若用不同的方法进行数据转换,得到的特征值也不一样,这些就构成了同一模式的多观测样本。多观测样本相对于单观测样本

**收稿日期**:2013-12-20. **网络出版日期**:2014-06-21.
**基金项目**:国家自然科学基金资助项目(61272210);江苏省自然科学基金资助项目(BK2011417, BK2011003);江苏省“333”工程基金资助项目(BRA2011142).
**通信作者**:李欢. E-mail:huanli1130@126.com.

能提供更多关于测试模式的信息,从而提高分类精度[1]。由此可以预见,多观测样本分类问题将得到国内外研究学者的广泛关注。

目前,多观测样本的分类方法主要有 2 类:一类是基于参数模型的方法。例如,文献[2]提出了基于概率密度的 KLD(KL-divergence),该方法把所有样本集看作是独立的,并且服从高斯分布,然后通过计算测试样本集和各个训练样本集间的 KL 散度来确定多观测样本的类别。但是此方法仅仅对那些服从单高斯分布的样本集比较适用,难以精确地描述数据呈非线性分布的情况。针对这一情况,O.Arandjelovic 等[3]提出了半参数混合高斯模型,并将其应用在 KL 散度的计算中,从而解决了非线性分布的多观测样本分类问题。然而,此方法的计算复杂度相对较大。F.Cardinaux 等[4]通过嵌入局部特征来扩展 GMM(Gaussian mixture model),在保证低复杂度的同时进一步提高了分类性能。文献[5]提出了一种基于核函数的分类方法,该方法利用信息论的相关知识,把 RAD(resistor-average distance)看作是多观测样本间的相似度来完成多观测样本的分类。以上这些方法的不足在于它们不但要解决复杂的参数估计问题,而且当多观测样本和测试样本集之间的统计相关性较弱时,它们的性能会有大的波动。另一类是基于非参数模型的方法,其中最具代表性的是基于子空间的方法,此类方法把子空间的相似度作为多观测样的分类依据,例如,文献[6]提出的 MSM (mutual subspace method),首先用 PCA 特征子空间来表示每一类的训练样本集和多观测样本,再利用子空间之间的主成分角作为相似性度量,最后用子空间的典型相关性(canonical correlation)来实现多观测样本的分类,但该算法对数据的变化较为敏感。为此,K.Fukui 等[7]又提出 CMSM(constraint mutual subspace method)来消除 MSM 的数据敏感性,将原空间的所有样本集都映射到同一约束子空间,在此约束空间中计算样本集间的主成分角,再用子空间的典型相关性完成多观测样本的分类。但上述 2 种方法并没有考虑到数据的非线性分布问题,针对这一问题,H.Sakano 等[8]提出 KMSM(kernel mutual subspace method)算法,L.Wolf 等[9]提出 KPA(kernel principal angles)算法,使用核函数来解决数据的非线性问题,进而完成多观测样本的分类。虽然 KMSM 和 KPA 考虑了数据的非线性分布,但是这 2 种方法用到的核函数对参数的依赖性较大。以上这些方法都没有考虑到通过转换数据可以提取到更多的判别信息,T.K.Kim 等[10]提出 DCC(discriminant canonical correlation)算法,其首先通过训练获得一个能使类内典型相关性最大而类间典型相关性最小的判别转换矩阵,然后把原空间数据映射到新的子空间上,在此基础上把典型差分相关性作为相似度量进行分类,此方法存在未考虑数据非线性分布的缺点。一些研究者曾认为所有典型相关性对分类的贡献是相同的,即权值相等。但后来 T.K.Kim 等[11]发现在分类中不同的典型相关性所起的作用是不同的,继而提出了 BoMPA(boosted manifold principal angles)算法,该算法首先通过 PPCA(probabilistic PCA)搜索局部线性模块,并将得到的所有模块表示成 PCA 子空间的形式,进而计算子空间之间的典型相关性,然后把训练集表示为正负样本特征的形式,同时采用 AdaBoost 算法得到相应的权值,最后用加权后的主成分角来度量子空间的相似性,实现多观测样本的分类。在此基础上,X.Li 等[12]提出 Boosted 全局和局部主成分角联合的分类算法。文献[13]提出 MMD (manifold-manifold distance)方法,该方法将典型相关性和局部线性模块结合起来,首先用联合局部线性模型的集合来表示子空间所描述的流形,从而把 MMD 转换为线性模块的组合,最终通过 MMD 的计算来对观测样本进行分类,但该方法的计算量和复杂度相对较大。W.S.Chu [14]提出 KDT(kernel discriminant transformation)来解决多观测样本的分类问题,该方法用核子空间来表示每个样本集,同时定义一个能使类内核子空间相似性最大而类间核子空间相似性最小的 KDT 矩阵,从而把多观测样本的分类问题转换为寻求 KDT 矩阵的最优解问题。近来,E.Kokiopoulou 等[15]在标记传播算法的基础上提出了 MASC(mAniflod-based smoothing under constrain)算法,该算法将 k-近邻图运用到多观测样本的分类问题中,但是 k-近邻图的边权值的计算采用了欧式距离下的高斯核函数,而基于欧式距离的测度无法全面反映数据的空间分布特性。

由上述可知,目前的多观测样本分类算法都有一定的不足和局限性。本文在经典 SVM 算法的基础上,用 SVM 的相关理论来实现多观测样本的分类。与传统的 SVM 算法相同,本文方法适用于小样本情况,利用核函数解决了非线性问题和维数问题,其算法复杂度与样本维数无关。然而,与传统分类方法的不同在于,该方法无需对分类器进行训练或提前对训练集进行特征表示,而是将测试集和训练集作为一个整体,充分利用特征空间中同类样本连续分布这一特点,使得分类更加准确。

## 1 多观测样本二分类问题的描述

多观测样本形成示意图如图 1 所示。在多观测样本的二分类问题中,若假设测试模式为 $s$, 则该问题就是将测试模式的多观测样本确定为 2 种类别中的一类。

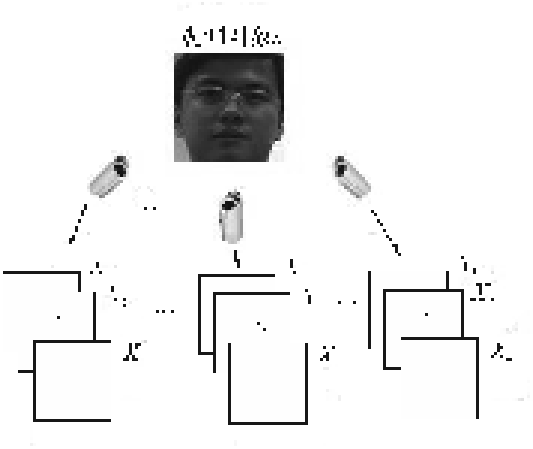


**图 1 多观测样本形成示意图**

**Fig.1 Schematic diagram of producing multiple observations**

假定测试模式 $s$ 的多测样本为

$$x_i^{(u)} = o_i(s), i = 1,2,\cdots,m \tag{1}$$

式中:上标 $(u)$ 表示各个观测样本是未标记的, $m$ 表示观测样本的数目, $o_i(s)$ 表示模式 $s$ 的第 $i$ 个单观测样本,它可能是模式 $s$ 经过平移、旋转、缩放或者是透视投影得到的,也可能是模式 $s$ 在某一特定时刻的观察记录。

多观测样本二分类问题的数据集可表示为 $X = \{X^{(l)}, X^{(u)}\}$ ,其中 $X^{(l)} = \{x_1, x_2, \cdots, x_l\} \subset R^d$ , $d$ 为样本维数, $X^{(l)}$ 表示已知标签的样本集,含有 $l$ 个样本, $X^{(l)}$ 涵盖了所有类别的数据。$X^{(u)} = \{x_{l+1}, x_{l+2}, \cdots, x_n\} \subset R^d$ , $n = l + m$ , $X^{(u)}$ 表示未知标签的样本集,含有 $m$ 个样本,并且所有样本属于同一类别,其对应于式(1)的多观测样本,即 $X^{(u)} = \{x_{l+1}, x_{l+2}, \cdots, x_n\} \triangleq \{x_1^{(u)}, x_2^{(u)}, \cdots, x_m^{(u)}\}$ 。因为二分类问题中的所用数据样本只属于 2 个类别,所以可以将数据的标签集表示为: $Y = \{-1, +1\}$ 。

综上所述,多观测样本二分类问题可正式定义为:给定已知标签的样本集 $X^{(l)}$ 和未知标签的样本集 $X^{(u)}$ ,而 $X^{(u)}$ 对应于模式 $s$ 的多观测样本, 即 $X^{(u)} \triangleq \{x_j^{(u)} = o_j(s), j = 1,2,\cdots,m\}$ ,问题就是确定未知标签的多观测样本的正确类别。其实,多观测样本二分类问题就是一种特殊的半监督学习,限制测试集中的所有样本属于同一类别,进而把多观测样本作为一个整体进行测试。而在一般的半监督学习所解决的分类问题中,测试集中的样本是属于多个类别的。因此,经典的半监督学习分类算法并不适合解决多观测样本二分类问题。同时,目前已有的多观测样本算法都存在着一定的不足。针对上问题,本文提出了一种新的算法,即基于 SVM 的多观测样本二分类算法。

## 2 基于 SVM 的多观测样本二分类

### 2.1 支持向量机

支持向量机(support vector machine, SVM)是一种基于结构风险最小化(structural risk minimization, SRM)原理,在统计学习理论的基础上发展起来的机器学习方法[16]。SVM 的基本实现方法就是在原空间或者经过投影后的高维空间中构造最优分类面,并将此分类面作为分类决策面进行数据分类。

SVM 最基本的理论是用来解决二分类问题的,SVM 的目标就是构造线性最优分类超平面,使其将 2 类样本完全正确地分开,同时使分类间隔最大。对于给定的样本集, $(\boldsymbol{x}_i, y_i), i = 1,2,\cdots,l$ , $\boldsymbol{x}_i \in \mathbf{R}^d, y_i = \pm 1$ ,当样本集线性可分时,对应的线性判别函数的一般形式为: $g(x) = (\boldsymbol{w}^{\mathrm{T}}x) + \boldsymbol{b}$ ,其中 $\boldsymbol{w}$、$\boldsymbol{b}$ 为 $n$ 维向量,对判别函数作归一化处理,使离分类面最近的样本满足 $|g(x)| = 1$, 则分类间隔等于 $2/\|\boldsymbol{w}\|$ , 使分类间隔最大等价于使 $\|\boldsymbol{w}\|^2$ 最小;要求分类面能将所有样本正确分类,也就是要求它满足:

$$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) \geqslant 1, i = 1,2,\cdots,n \tag{2}$$

且使 $\|\boldsymbol{w}\|^2$ 最小的分类面就是最优分类面。

综上所述,最优分类面的求解问题等价于在式(2)的约束下最小化式(3):

$$\Phi(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2 = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} \tag{3}$$

而这一问题可以通过定义拉格朗日函数(式(4))来求解:

$$L(\boldsymbol{w},\boldsymbol{b},a) = \|\boldsymbol{w}\|^2/2 - \sum_{i=1}^{n}\alpha_i[y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) - 1] \tag{4}$$

式中: $\alpha_i \geqslant 0$ 为 Lagrange 系数,则问题转换成对 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 求 Lagrange 函数的最小值。式(4)分别对 $\boldsymbol{w}$、$\boldsymbol{b}$ 求偏微分,并令结果为零,则有

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \displaystyle\sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_i \\ \dfrac{\partial L}{\partial \boldsymbol{b}} = 0 \Rightarrow \displaystyle\sum_{i=1}^{n}\alpha_i y_i = 0 \end{cases} \tag{5}$$

将式(5)代入式(4),则原问题可以进一步转化为凸

二次规划的对偶问题:

$$\begin{cases}\max\limits_{\alpha}\sum\limits_{i=1}^{n}\alpha_i-\sum\limits_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}x_j/2\\ \text{s.t.}\quad \alpha_i\geqslant 0,i=1,2,\cdots,n\\ \sum\limits_{i=1}^{n}\alpha_i y_i=0\end{cases}\tag{6}$$

在式(6)所得的结果中,只有少数的 $\alpha_i$ 不等于零,其对应的样本离最优分类面最近,这些样本被称为支持向量。上述问题存在惟一最优解,若最优解为 $\alpha_i^*$ ,则 $\boldsymbol{w}^*=\sum\limits_{i=1}^{n}\alpha_i^* y_i\boldsymbol{x}_i$, $\boldsymbol{b}^*$ 可由式(2)取等号时得到,因此,最终的最优分类函数为

$$f(x)=\operatorname{sgn}((\boldsymbol{w}^*)^{\mathrm{T}}x+\boldsymbol{b}^*)=$$
$$\operatorname{sgn}(\sum_{i=1}^{n}\alpha_i^* y_i\boldsymbol{x}_i^{\mathrm{T}}x+\boldsymbol{b}^*)$$

对于样本集非线性可分的情况,可以先把原始空间的样本集通过非线性变换 $\varphi$ 映射到一个高维的特征空间,使得样本集在新的空间线性可分,然后构造最优分类平面。这种非线性变换可以通过引入适当的核函数来实现,用 $k(\boldsymbol{x}_i,x_j)=\varphi(\boldsymbol{x}_i)\cdot\varphi(x_j)$ 代替线性可分情况下的点积 $(\boldsymbol{x}_i^{\mathrm{T}},x_j)$ ,式(6)中的优化函数变为

$$Q(\alpha)=\sum_{i=1}^{n}\alpha_i-\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k(\boldsymbol{x}_i,x_j)/2$$

最终的分类函数为

$$f(x)=\operatorname{sgn}(\sum_{i=1}^{n}\alpha_i^* y_i k(\boldsymbol{x}_i,x)+\boldsymbol{b}^*)$$

在线性不可分的问题中,SVM 还引入了惩罚因子 $C$ 和松弛变量 $\xi$ ,此时最优分类面的求解问题可描述为

$$\begin{cases}\min\ \|\boldsymbol{w}\|^2/2+C\sum\limits_{i=1}^{n}\xi_i\\ \text{s.t.}\quad y_i(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+\boldsymbol{b})\geqslant 1-\xi_i,i=1,2,\ldots,n\\ \xi_i\geqslant 0,i=1,2,\cdots,n\end{cases}$$

同样地,通过定义拉格朗日函数的方法可以将原问题转换为凸二次规划的对偶问题:

$$\begin{cases}\max\limits_{\alpha}\sum\limits_{i=1}^{n}\alpha_i-\dfrac{1}{2}\sum\limits_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k(\boldsymbol{x}_i,x_j)\\ \text{s.t.}\quad 0\leqslant\alpha_i\leqslant C,i=1,2,\ldots,n\\ \sum\limits_{i=1}^{n}\alpha_i y_i=0\end{cases}$$

对应的最优分类函数为

$$f(x)=\operatorname{sgn}(\sum_{i=1}^{n}\alpha_i^* y_i k(\boldsymbol{x}_i,x)+\boldsymbol{b}^*)$$

## 2.2　基于 SVM 的多观测样本二分类

由于支持向量机具有结构简单、推广性能好、优化求解时具有惟一最优解等优点,本文将用 SVM 的相关理论来解决多观测样本二分类问题,确定多观测样本的类别。根据 SVM 的原理可知,SVM 要解决的数学问题为

$$\begin{cases}\min\dfrac{1}{2}\|\boldsymbol{w}\|^2+C\sum\limits_{i=1}^{n}\xi_i\\ \text{s.t.}\quad y_i(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+\boldsymbol{b})\geqslant 1-\xi_i,i=1,2,\ldots,n\\ \xi_i\geqslant 0,i=1,2,\cdots,n\end{cases}\tag{7}$$

而从多观测样本二分类问题的描述可知,二分类问题中的所有数据只属于 2 个类别,数据的标签集为 $\{-1,+1\}$ ,设多观测样本集 $X^{(u)}$ 的标签为 $y$ ,则 $y=-1$ 或 $y=+1$。因此可以通过假设多观测样本的标签来增加式(7)的约束条件:

$$\begin{cases}\min\dfrac{1}{2}\|\boldsymbol{w}\|^2+C\sum\limits_{i=1}^{l}\xi_i\\ \text{s.t.}\quad y_i(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+\boldsymbol{b})\geqslant 1-\xi_i,i=1,2,\ldots,l\\ y_j(\boldsymbol{w}^{\mathrm{T}}\varphi(x_j)+\boldsymbol{b})\geqslant 1,j=l+1,2,\ldots,n\\ y_{l+1}=y_{l+2}=\cdots=y_n=y\\ \xi_i\geqslant 0,i=1,2,\cdots,l\end{cases}\tag{8}$$

可以先假设 $y=-1$,求解得到目标函数值 $g_1$。再假设 $y=+1$,求解得到目标函数值 $g_2$。只有当假设的标签与多观测样本的实际标签相同时,相应得到的目标函数值才是最优解。因此,可以通过比较两次得到的目标函数值来确定待测试的多观测样本的标签。如式(9)所示:

$$\hat{y}=\begin{cases}-1,g_1>g_2\\ +1,g_1\leqslant g_2\end{cases}\tag{9}$$

为求解式(8)所述的优化问题,引入拉格朗日函数 $L$ :

$$L(\boldsymbol{w},\boldsymbol{b},\xi_i,\alpha_i,\beta_i,r_i)=\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{l}\xi_i-$$
$$\sum_{i=1}^{l}\alpha_i[y_i(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+\boldsymbol{b})-1+\xi_i]-$$
$$\sum_{i=1}^{l}r_i\xi_i-\sum_{i=l+1}^{n}\beta_i[y(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+\boldsymbol{b})-1]\tag{10}$$

式中:$\alpha_i$、$\beta_i$、$r_i$ 为 Lagrange 系数,$\alpha_i\geqslant 0$,$\beta_i\geqslant 0$,$r_i\geqslant 0$,$\xi_i\geqslant 0$。要使函数 $L$ 关于 $\boldsymbol{w}$、$\boldsymbol{b}$、$\xi_i$ 最小化,由极值存在的必要条件可知,函数 $L$ 的极值满足下列条件:

$$\begin{cases}\partial L/\partial\boldsymbol{w}=0\\ \partial L/\partial\boldsymbol{b}=0\\ \partial L/\partial\xi_i=0\end{cases}\tag{11}$$

解方程(11)可得

$$\begin{cases} \boldsymbol{w} = \sum_{i=1}^{l} \alpha_i y_i \varphi(\boldsymbol{x}_i) + \sum_{j=l+1}^{n} \beta_j y \varphi(x_j) \\ \sum_{i=1}^{l} \alpha_i y_i + \sum_{j=l+1}^{n} \beta_j y = 0 \\ C - \alpha_i - r_i = 0 \end{cases} \tag{12}$$

将式(12)代入式(10)得到优化问题式(8)的对偶形式,即关于 $\alpha_i$、$\beta_j$ 的最大化函数:

$$\begin{cases} \max \sum_{i=1}^{l} \alpha_i + \sum_{i=l+1}^{n} \beta_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j k(\boldsymbol{x}_i, x_j) - \\ \frac{1}{2} \sum_{i=l+1}^{n} \sum_{j=l+1}^{n} \beta_i \beta_j y_i y_j k(\boldsymbol{x}_i, x_j) - \sum_{i=1}^{l} \sum_{j=l+1}^{n} \alpha_i \beta_j y_i y_j k(\boldsymbol{x}_i, x_j) \\ \text{s.t.} \quad 0 \leqslant \alpha_i \leqslant C, i = 1, 2, \cdots, l \\ \beta_i \geqslant 0, i = l+1, 2, \cdots, n \\ \sum_{i=1}^{l} \alpha_i y_i + \sum_{j=l+1}^{n} \beta_j y = 0 \end{cases} \tag{13}$$

若设 $\boldsymbol{Y} = [y_1 \quad \cdots \quad y_l \quad y_{l+1} \quad \cdots \quad y_n]^{\mathrm{T}}$,因为 $y_{l+1} = y_{l+2} = \cdots = y_n = y$,所以 $\boldsymbol{Y} = [y_1 \cdots y_l \, y \cdots y]^{\mathrm{T}}$。令 $\boldsymbol{O} = [1 \quad 1 \quad \cdots \quad 1]$,$\boldsymbol{A} = [\alpha_1 \cdots \alpha_l \, \beta_{l+1} \cdots \beta_n]^{\mathrm{T}}$,则式(13)变为

$$\max \boldsymbol{OA} - \frac{1}{2} \boldsymbol{A}^{\mathrm{T}} ((\boldsymbol{YY}^{\mathrm{T}}) \cdot K) \boldsymbol{A} \tag{14}$$

可以看到,通过求解式(14)可以能得到两次标签假设对应的目标函数值 $g_1$ 和 $g_2$,从而根据式(9)确定待测试的多观测样本的标签。

### 2.3 基于 SVM 的多观测样本二分类的算法描述

基于 SVM 的多观测样本二分类的算法如下:

输入:

$X^{(l)}$、$Y^{(l)}$:已标记样本集和它的标签集;

$X^{(u)}$:多观测样本集;

$l$:已标记样本的数目;

$m$:多观测样本数目。

输出:

$\hat{y}$:多观测样本的类别。

处理:

1)由 $X^{(l)}$ 和 $X^{(u)}$ 得到样本矩阵 $\boldsymbol{X}, \boldsymbol{X} \subset \mathbf{R}^{n \times d}$,由 $Y^{(l)}$ 得到标签矩阵 $\boldsymbol{Y}$;

2)计算样本矩阵 $\boldsymbol{X}$ 对应的核矩阵 $\boldsymbol{K}$;

3)设 $y = -1$,求解优化问题:$\max \boldsymbol{OA} - \boldsymbol{A}^{\mathrm{T}}((\boldsymbol{YY}^{\mathrm{T}}) \cdot K)\boldsymbol{A}/2$,得到 $g_1$;设 $y = +1$,求解优化问题:$\max \boldsymbol{OA} - \boldsymbol{A}^{\mathrm{T}}((\boldsymbol{YY}^{\mathrm{T}}) \cdot K)\boldsymbol{A}/2$,得到 $g_2$;

4)若 $g_1 > g_2$ 则 $\hat{y} = -1$,否则 $\hat{y} = +1$。

## 3 多图像样本集的分类

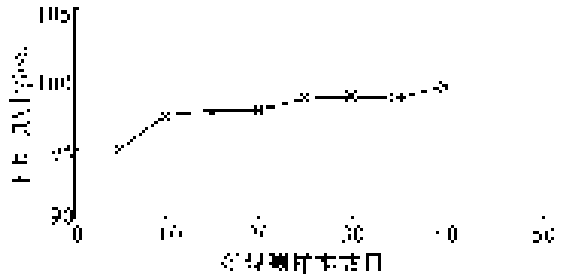
### 3.1 手写数字分类

为了验证基于 SVM 的多观测样本二分类算法的有效性,首先在手写数字数据库上进行实验。同类数字不同形式的手写图像组成多观测样本集,对此类样本集进行分类。实验中,使用 2 种不同的数据库:Binary 手写数字数据库和 USPS 手写数字数据库。Binary 数据库包含 0~9 共 10 类数字的手写图像,每类数字有 39 个样本,每个样本用大小为 20×16 的二值图像表示。USPS 数据库由 0~9 共 10 类手写数字组成,每类数字有 1 100 个样本,每个样本用大小为 16×16 的灰度图像表示。

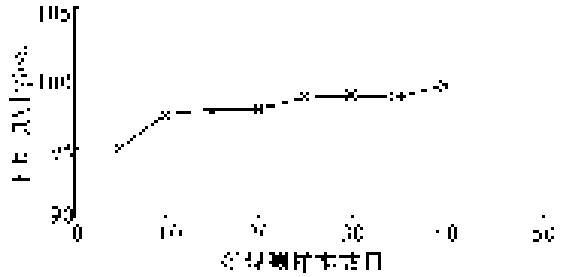
模式变换的鲁棒性是多观测样本分类的一种重要特性。可以使用虚拟样本来扩充已标记样本集,从而加强分类算法的抗变换性。虚拟样本一般通过原始样本的变换产生,虚拟样本的类别与原始样本相同,因此是已知标签的已标记样本。通过在数据集中添加虚拟样本,分类算法对测试样本的鲁棒性更强。因此,在本文所提的算法中使用这一方法,在原始数据集中添加大小为 $n^{vs}$ 的样本集 $X^{(vs)}$,数据集变为:$X = \{X^{(l)}, X^{(vs)}, X^{(u)}\}$。实验中,核函数选用高斯核函数,即:$k(x,y) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2/2\sigma^2)$。为计算参数 $\sigma$ 的大小,在数据集 $X$ 中随机选取 1 000 个样本,并计算两两样本之间的欧式距离,$\sigma$ 设置为所有距离的中值的 1/2。

对于每类数字,首先从对应样本中随机抽取 2 个样本组成训练集,剩下的样本组成测试集。再对训练集中的每个样本做连续的 4 次旋转变换,得到的样本放在训练集中,其中旋转角 $\theta$ 从 $[-40^\circ, 40^\circ]$ 的均匀采样序列中得到。这样的区间能避免“6”和“9”2 类数字的混淆。为了建立每类数字的多观测样本数据集 $X^{(u)}$,从每类数字的测试集中随机选取一个样本并对这个样本进行旋转变换,旋转角 $\theta \in [-40^\circ, 40^\circ]$。每次测试时,选取 2 类不同的数字进行实验,共有 45 种组合,即 (0,1), (0,2), …, (7,8), (7,9), (8,9)。再由这 2 类数字的训练样本共同组成算法的训练集 $X^{(t)}$,而对应的测试集作为算法的测试集,即多观测样本。该实验对不同大小的多观测样本进行了实验,样本数 $m = [5{:}5{:}40]$。对于不同大小的数据集 $X^{(u)}$,45 种组合中的每个组合进行 10 次随机实验,每个组合要对 2 个测试集进行测试,所以实验中的每个结果都是 900 次

随机实验的均值,如图 2 所示。



(a) 在 Binary 数据库上的平均识别率



(b) 在 USPS 数据库上的平均识别率

**图 2　在 2 种手写数字数据库上的识别率**

**Fig.2　Classification results measured on two different handwritten digit data sets**

从实验结果可以看出本文 SVM 算法在 Binary 数据库和 USPS 数据库上的识别率很高,尤其在 USPS 数据库上,当样本不少于 10 时识别率为 100%,这就说明基于 SVM 的多观测样本二分类算法的可行性。分析数据可得:算法的识别率随着多观测样本数目的增大而提高,因为增加多观测样本的数目能提供更多的某特定类别的信息,从而更加准确地判断类别。

### 3.2　物体图像分类

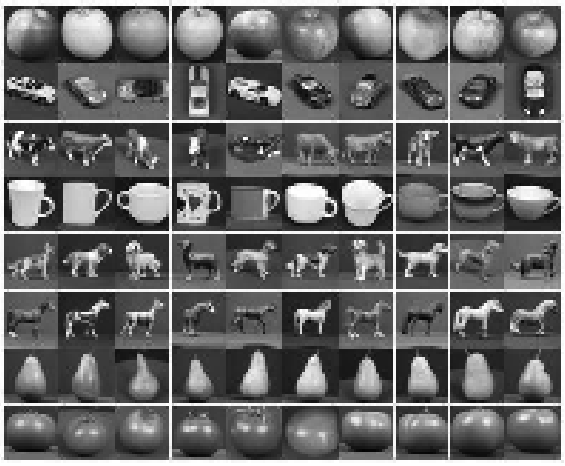
下面在物体图像数据库上验证基于 SVM 的多观测样本二分类算法的有效性,实验中同一物体的不同观测图像作为此类物体的多观测样本。并将本文算法与经典的多观测样本分类算法进行对比:

1) KLD[2] (KL-divergence):该方法是典型的基于密度估计的统计方法,把所有样本集看作是独立同分布的高斯随机变量,然后通过计算样本集间 KL 散度完成多观测样本的分类。实验中,协方差矩阵特征向量的长度按能量的 96%来选取。

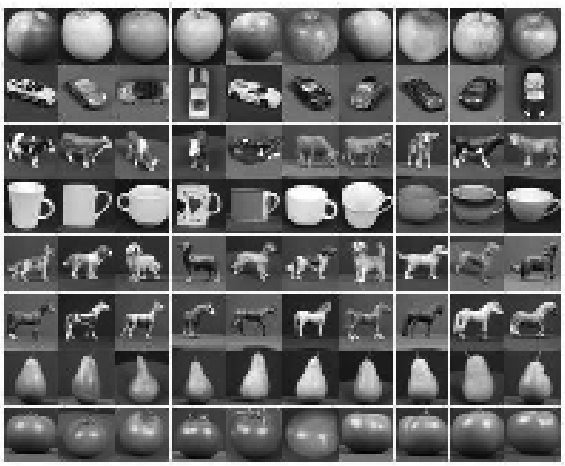
2) MSM[6] (mutual subspace method):MSM 是典型的子空间方法,该方法中的每个图像集用子空间来表示,而子空间通过主成分即协方差矩阵获得,把训练集与测试集之间的主成分角[17]作为相似性度量。实验中,当样本数目小于 9 时候,协方差矩阵的特征向量长度等于样本数目,否则设为 9。

3) KMSM[8] (kernel mutual subspace method):KMSM 是 MSM 在非线性空间的扩展,该方法考虑了图像集的非线性。与 MSM 不同的是,在用线性子空间建模之前,KMSM 需要先把图像样本非线性地映射到高维特征空间。也就是说,KMSM 用 KPCA 来取代 PCA,从而获得了数据的非线性。KMSM 方法中,使用高斯核函数 $k(x,y) = \exp(-\|\boldsymbol{x}-\boldsymbol{y}\|^2/2\sigma^2)$,其中 $\sigma$ 的选取与本文所提的算法相同。

实验选用 ETH-80 物体识别数据库,ETH-80 含有 8 个种类的图像:苹果、车子、牛、杯子、狗、马、梨和西红柿(如图 3(a)所示)。每个种类又含有 10 个物体类(例如,狗有 10 个不同的品种),每个物体类中包含该物体不同角度的 41 张图像,例如图 3(b)显示了狗这一种类中一个物体类的所有图像。数据库中的所有图像大小为 128×128,为了简化计算,对图像重新采样,使其大小为 32×32。



(a) ETH-80



(b) ETH-80 中一个物体类的 41 张图片

**图 3　ETH-80 数据库的样本图像**

**Fig.3　Sample images from the ETH-80 database**

每个种类的训练样本集由其 10 个物体类均随机抽取的 10 张图像组成,即每个种类的训练样本集含有 100 个样本,而剩余的 31 张图像组成每个物体类的测试集。实验中,从八大种类中选取 2 个不同的种类进行分类测试,共有 28 种组合,即 (1,2),(1,3),…,(6,7),(6,8),(7,8) 。由这 2 个种类的训练样本共同组成算法的训练集 $X^{(l)}$ ,再分别为这 2 个种类构建测试集:从每个种类对应的 10 个物体类中随机选取一个物体类,再从此物体类中选取 10 个样本组成多观测样本,即为该种类的测试集。对 28 种组合中的每个组合进行 10 次随机实验,每个组合要对 2 个测试集进行测试,所以实验中的每个结果都是 560 次随机实验的均值,如表 1 所示。

**表 1 在 ETH-80 数据库上的识别率**

**Table 1 Object recognition rate measured on the ETH-80 database** /%

| 算法 | KLD | MSM | KMSM | 本文 SVM |
|---|---|---|---|---|
| 识别率 | 85.714 | 97.500 | 91.071 | 98.036 |

实验结果表明,本文 SVM 算法在 ETH-80 数据库上的识别率很高,并且该算法优于其他 3 种算法,这就说明了基于 SVM 的多观测样本二分类算法的有效性。

## 4 基于视频的人脸识别

### 4.1 实验数据集

为了验证基于 SVM 的多观测样本二分类算法的有效性,在基于视频序列的人脸识别问题中进行实验,把视频中不同的视频帧作为同一个人的多观测样本。由于视频中人的头部姿势,人脸表情和光照都是变化的,所以本节是在真实的环境中验证所提算法的有效性。把所提算法与 4.2 节中描述的 KLD,MSM 和 KMSM 进行比较。由于实验中所用视频序列的视频帧在时间上是连续的,因此,该算法同样适用于基于图像集的人脸识别问题。

实验中,使用 2 个数据库:VidTIMIT 数据库[18] 和 Honda/UCSD[19] 数据库的第 1 部分。VidTIMIT 数据库包含了 43 个人在 3 个时间段的人脸视频序列,其中第 1 个和第 2 个时间段间隔 7 天,第 2 和第 3 个时间段间隔 6 天。每一个视频序列中,被拍摄者头部在不断运动:向左,向右,回到中间,再向下,向上,再回到中间位置。Honda/UCSD 数据库包含了 20 个人的 59 个视频序列,每个人有 2~5 个视频序列。与 VidTIMIT 数据库不同的是,Honda/UCSD 数据库的被拍摄者以不同的速度自由移动头部,同时脸部表情也在不断地变化。在两个数据库的预处理中,首先用 Viola P 的人脸检测方法[20] 从视频序列的视频帧中提取人脸区域。为了简化计算,把得到的人脸图像重新采样,使其大小为 32×32。

### 4.2 基于 VidTIMIT 数据库的人脸识别

首先在 VidTIMIT 数据库上对本文算法进行测试。图 4 显示了 VidTIMIT 数据库中一些样本图像。

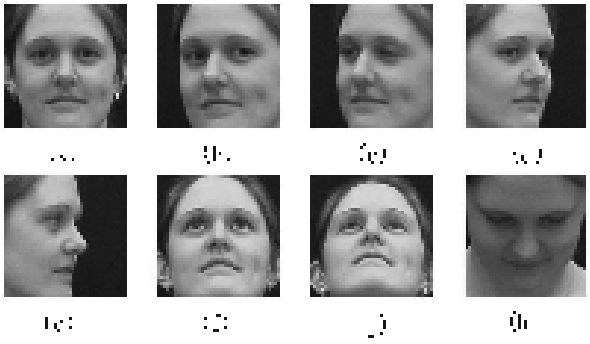


**图 4 VidTIMIT 数据库中的样本图像**

**Fig.4 Sample images in the VidTIMIT database**

由于数据库含有 3 个时间段的视频序列,因此采用下面的标准度量算法的性能:

$$\bar{e} = \frac{1}{6}\sum_{i=1}^{3}\sum_{j=1,j\neq i}^{3} e(i,j)$$

式中:$e(i,j)$ 表示第 $i$ 个时间段的数据作为训练集,第 $j$ 个时间段的数据作为测试集时的分类正确率。也就是说 $\bar{e}$ 是以下 6 次实验的平均正确率,即 (1,2),(1,3),(2,1),(2,3),(3,1) 和(3,2) 。在这 6 次实验的每次实验中,选取 2 种不同类别的数据进行实验,共有 45 种组合,即 (1,2),(1,3),…,(9,8),(9,10) 。再由这 2 类在第 $i$ 个时间段中的图像集共同组成算法的训练集 $X^{(l)}$ ,而这 2 类在第 j 个时间段的图像集作为算法的测试集,即多观测样本。该实验对不同大小的多观测样本进行了实验,样本数目 $m = [4:4:16]$ 。对于不同数目的多观测样本数据集 $X^{(u)}$ ,45 种组合中的每个组合都有 2 个测试集,因此每个组合要进行 2 次测试。所以实验中的每个 $e(i,j)$ 都是 90 次实验的均值,每个结果 $\bar{e}$ 是 540 次随机实验的结果,实验结果如表 2 所示。

**表 2 在 VidTIMIT 数据库上的识别率**

**Table 2 Recognition results on the VidTIMIT database/%**

| 样本数目 $m$ | KLD | MSM | KMSM | 本文 SVM |
|---|---|---|---|---|
| 4 | 50.600 | 72.965 | 86.157 | 97.287 |
| 8 | 88.410 | 89.277 | 92.229 | 98.828 |
| 12 | 94.630 | 95.995 | 95.496 | 99.148 |
| 16 | 96.630 | 96.742 | 96.244 | 99.003 |

图 5 用柱形图表示了实验结果,本文 SVM 算法在 VidTIMIT 数据库上的识别率很高。由图 5 可知,对不同数目的观测样本,KLD、MSM 和 KMSM 3 种算法的识别率变化较大,而本文 SVM 算法的识别率变化不大,这说明基于 SVM 的多观测样本二分类算法对不同数目的多观测样本具有更好的鲁棒性。

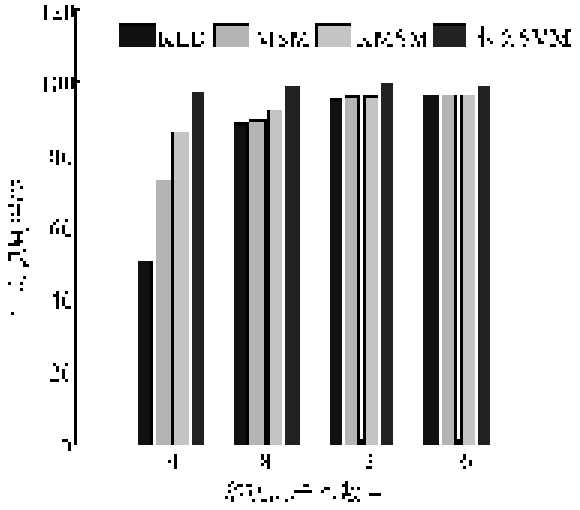


图 5　在 VidTIMIT 数据库上的识别率

**Fig.5　Recognition results on the VidTIMIT database**

### 4.3　基于 Honda/UCSD 数据库的人脸识别

在 Honda/UCSD 数据库上进一步验证本文 SVM 算法的有效性。图 6 显示了 Honda/UCSD 数据库中一些样本图像。实验中,选取 19 个人所对应的视频序列进行实验。实验中,选取 2 种不同类别的数据进行实验,共有 171 种组合,即 (1,2),(1,3),…,(17,18),(17,19),(18,19) 。由这 2 类数据的训练样本共同组成算法的训练集 $X^{(l)}$ ,而对应的测试集作为算法的测试集,即多观测样本。该实验对不同大小的多观测样本进行了实验, $m=[4:4:16]$ 。对于不同大小的数据集 $X^{(u)}$ ,171 种组合中的每个组合都有 2 个测试集,因此每个组合要进行 2 次测试。所以实验中的每个结果都是 342 次实验的均值,如表 3 所示,并用柱形图表出来(如图 7)。

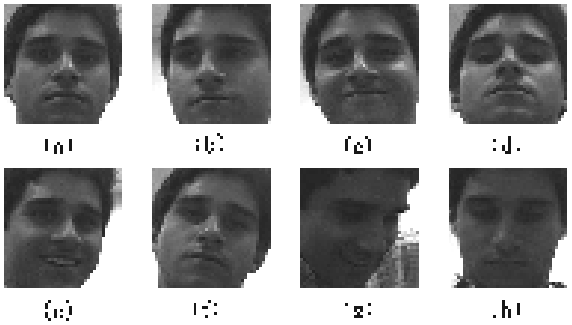


图 6　Honda/UCSD 数据库中的样本图像

**Fig.6　Sample images in the Honda/UCSD database**

实验结果表明,相比于以往的 KLD、MSM 和 KMSM 算法,本文 SVM 算法获得最高的识别率。这进一步说明了基于 SVM 的多观测样本二分类算法的有效性。

表 3　在 Honda/UCSD 数据库上的识别率

**Table 3　Recognition results on the Honda/UCSD database**

/%

| 样本数目 $m$ | KLD | MSM | KMSM | 本文 SVM |
|---|---|---|---|---|
| 4 | 54.678 | 63.158 | 79.825 | 88.596 |
| 8 | 85.371 | 85.088 | 82.164 | 94.444 |
| 12 | 92.690 | 90.936 | 83.333 | 95.322 |
| 16 | 93.860 | 92.105 | 85.673 | 95.906 |

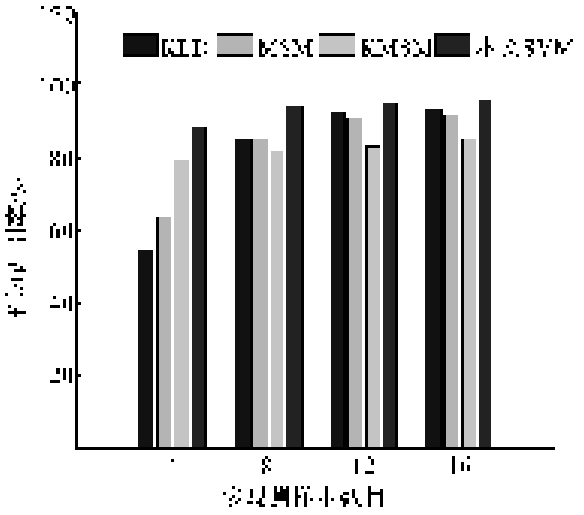


图 7　在 Honda/UCSD 数据库上的识别率

**Fig.7　Recognition results on the Honda/UCSD database**

## 5　结束语

本文提出基于 SVM 的多观测样本分类算法,该算法首先进行类别假设,然后求解优化问题得到相应的目标函数值,把目标函数值作为分类依据来实现多观测样本的二分类。实验结果表明本文算法在手写数字识别、物体识别和人脸识别中都能取得较好的分类效果,为模式识别问题提供了一种新的方法。但是本文针对的是二分类问题,如何在该算法的基础上实现多分类仍是需要进一步的研究。

## 参考文献:

[1] KIM T K, KITTLER J, CIPOLLA R. On-line learning of mutually orthogonal subspaces for face recognition by image sets[J]. IEEE Transactions on Signal Processing, 2010, 19(4): 1067-1074.

[2] SHAKHNAROVICH G, FISHER J W, DARREL T. Face recognition from long-term observations [C]//European

Conference on Computer Vision(ECCV). San Diego, USA, 2002, 3: 851-868.

[3] ARANDJELOVIC O, SHAKHNAROVICH G, FISHER J, et al. Face recognition with image sets using manifold density divergence[C]//IEEE International Conference on Computer Vision and Pattern Recognition(CVPR). San Diego, USA, 2005, 1: 581-588.

[4] CARDINAUX F, SANDERSON C, BENGIO S. User authentication via adapted statistical models of face images [J]. IEEE Transactions on Signal Processing, 2006, 54 (1): 361-373.

[5] ARANDJELOVIC O, CIPOLLA R. Face recognition from face motion manifolds using robust kernel resistor-average distance[C]//IEEE Workshop on Face Processing in Video. Washington D C, USA, 2004, 5: 88-93.

[6] YAMAGUCHI O, FUKUI K, MAEDA K, et al. Face recognition using temporal image sequence[C]//IEEE International Conference on Automatic Face and Gesture Recognition. Nara, Japan, 1998: 318-323.

[7] FUKUI K, YAMAGUCHI O. Face recognition using multi-viewpoint patterns for robot vision[C]//International Symposium on Robotics Research. Siena, Italy, 2005, 15: 192-201.

[8] SAKANO H, MUKAWA N. Kernel mutual subspace method for robust facial image recognition[C]//Fourth International Conference on Knowledge-based Intelligent Engineering Systems and Allied Technologies. [S.l.], 2000, 1: 245-248.

[9] WOLF L, SHASHUA A. Learning over sets using kernel principal angles[J]. Machine Learning Research, 2003, 4 (10): 913-931.

[10] KIM T K, KITTLER J, CIPOLLA R. Discriminative learning and recognition of image set classes using canonical correlations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1005-1018.

[11] KIM T K, ARANDJELOVIC O, CIPOLLA R. Boosted manifold principal angles for image set-based recognition [J]. Pattern Recognition, 2007, 40(9): 2475-2484.

[12] LI X, FUKUI K, ZHENG N N. Image-set based face recognition using boosted global and local principal angles [C]//9th Asian Conference on Computer Vision(ACCV). Xi'an, 2010: 323-332.

[13] WANG R P, SHAN S G, CHEN X L, et al. Manifold-manifold distance with application to face recognition based on image Set[C]//IEEE International Conference on Computer Vision and Pattern Recognition(CVPR). Anchorage, Alaska, USA, 2008: 1-8.

[14] CHU W S, CHEN J C, LIEN J. Kernel discriminant transformation for image set-based face recognition[J]. Pattern Recognition, 2011, 44(8): 1567-1580.

[15] KOKIOPOULOU E, PIRILLOS S, BROSSARD P. Graph-based classification of multiple observation sets[J]. Pattern Recognition, 2010, 43(12): 3988-3997.

[16] VAPNIK V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 21-32.

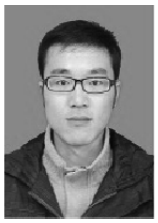
[17] GOLUB G H, LOAN C V. Matrix computations[M].3rd ed. Baltimore: The John Hopkins University Press, 1996: 15-16.

[18] SANDERSON C. Biometric person recognition: face, speech and fusion, VDM-Verlag, 2008.

[19] LEE K C, HO J, YANG M H, et al. Video-based face recognition using probabilistic appearance manifolds[C]// IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Madison, USA, 2003: 313-320.

[20] VIOLA P, JONES M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57 (2): 137-154.

**作者简介：**



李欢，男，1990 年生，硕士研究生，主要研究方向为人工智能、模式识别。

王士同，男，1964 年生，教授，博士生导师，主要研究方向为人工智能、模式识别和生物信息。