

## 基于不同 Margin 的人脸特征选择及识别方法

李伟红 陈伟民 杨利平 龚卫国

(重庆大学光电技术及系统教育部重点实验室 重庆 400044)

**摘要:** Margin 在机器学习中具有很重要的意义, 基于 margin 的特征选择方法就是从分类的角度对特征集各特征的权重进行分析。该文对不同的 margin 进行了分析, 提出将 sample-margin 和 hypothesis-margin 分别作为特征选择标准对 SBS 特征选择方法进行改进, 然后设计具有最佳超参数的 SVM 多项式分类器进行人脸识别。实验在 FRERT 人脸图像库上进行并与 Relief 特征选择方法进行了比较, 对 SVM 和 NN 分类器的实验结果也进行了分析。实验结果显示: 该文提出的人脸识别特征选择及识别方法是有效、适用的。

**关键词:** 人脸识别; Margin; 特征选择; 支持向量机(SVM); 顺序后退法(SBS)

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2007)07-1744-05

## Face Feature Selection and Recognition Based on Different Types of Margin

Li Wei-hong Chen Wei-min Yang Li-ping Gong Wei-guo

(Key Lab of Optoelectronic Technology and Systems of Education Ministry of China,  
Chongqing University, Chongqing 400044, China)

**Abstract:** Margin plays an important role in research of machine learning. Margin-based feature selection methods choose the weights of features from the view of classification. This paper analyzes different types of margin and proposed methods to improve the Sequential Backward Selection (SBS) method respectively using sample-margin and hypothesis-margin as feature selection criterion. A SVM polynomial classifier, which has optimal hyper-parameters, is then designed for face recognition. Experiments are conducted on FERET face database. Recognition accuracies between the proposed methods and relief feature selection method are compared. Experiments are also conducted by respectively using SVM and Nearest Neighbor (NN) classifier. Experimental results indicate that the proposed feature selection and recognition methods are efficient for face recognition.

**Key words:** Face recognition; Margin; Feature selection; Support Vector Machine (SVM); Sequential Backward Selection (SBS)

### 1 引言

特征选择是模式识别领域一个极其重要的问题。特征选择可定义为<sup>[1]</sup>: 已知一特征集, 从中选择一个子集使评价标准最优。特征选择有两个需要解决的问题, 一是选择的标准, 二是要找到一种较好的算法, 以便在允许的时间内找出最优的特征子集。特征选择的最终的评价将取决于其分类识别的能力。一种好的特征选择方法应该使所选择的特征子集即使采用最普通的分类器也可以获得较好的分类结果<sup>[2]</sup>。基于 margin 的特征选择方法实际上就是从分类识别的角度出发对特征集的各特征进行分析, 通过一定的搜索算法获得各特征的权重和最佳特征组合, 且分类误差最小。这类分析方法在近十年已经成为机器学习的基本方法<sup>[3]</sup>, 根据不同的 margin 定义, 在机器学习领域已经发展出目前两个最强大有力

的研究方——Support Vector Machine(SVM)<sup>[4]</sup> 和 AdaBoost<sup>[5]</sup>。SVM利用一组特殊的称为支持向量的样本, 代替原始样本对整个训练数据集进行描述, 并寻找一个最优超平面使待分类的两类的分类间隔最大, 这个间隔即sample-margin。AdaBoost则是利用具有 $L_1$ 范式的hypothesis-margin作为各类别间的距离, 且使分类误差达到最小。目前根据不同的margin的定义, 人们进行了大量的研究<sup>[6-10]</sup>。本文作者针对人脸识别问题, 对最大化sample-margin也进行过分析、实验<sup>[11]</sup>。该文将探究不同margin在人脸特征选择及识别方面的作用, 提出分别将两种类型的margin作为顺序后退特征选择方法的特征选择标准, 然后设计具有最佳超参数的SVM多项式分类器, 实际用于人脸识别。在SVM设计中, 由于超参数是一个不容忽视的因素, 该文采用grid-search方法获取不同情况下的SVM多项式最佳超参数。实验在FRERT人脸图像库上将该文提出的方法与Relief特征选择方法进行了比较。并进行了SVM及NN分类器识别的对比实验。综合实验结果表明: 该文提出的方法是一种有效、实用

2005-12-05 收到, 2006-04-03 改回

国家教育部科学研究重点项目(02057), 重庆市自然科学基金重点项目(CSTC2005BA2002)和重庆市自然科学基金(CSTC2005BB2181)资助课题

的人脸特征选择及识别方法。

## 2 间隔(Margin)

Margin在机器学习中具有很重要的意义。它可以度量一个分类器预测的可信程度。Crammer<sup>[3]</sup>对margin的描述为: margin是分界面与被预测样本间的距离测量, 它可以用于学习算法的设计以及误差界的计算。

### 2.1 样本间隔(sample-margin)

Sample-margin是从一个样本到由分类规则推导出的决策边界的距离, 如图 1(a)所示。SVM需要寻找的就是使待分类的两类的 sample-margin 最大的那个最优超平面。Bartlett<sup>[12]</sup>研究了样本到分界面之间的距离与决策分界面问题, 且导出了其范化误差。

在SVM中, 设给定样本点:  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ ,  $\mathbf{x}_i \in R^d$ ,  $y_i \in \{+1, -1\}$  是类别标记。在特征空间中分类面为:  $(\mathbf{w} \cdot \mathbf{x}) + b = 0$ , 使得

$$y_i[(\mathbf{w} \cdot \mathbf{x}) + b] \geq 1, \quad i = 1, 2, \dots, n \quad (1)$$

满足上述条件, 且使  $M_s = 2/\|\mathbf{w}\|$  最大的分类面就是最优分类面, 该文取  $M_s$  表示 sample-margin。

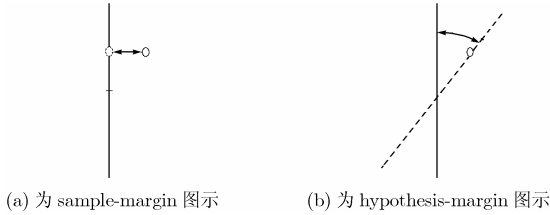


图 1

### 2.2 假设间隔(hypothesis-margin)

Hypothesis-margin 是在不改变任何样本点分类结果的前提下分类器可以移动的距离。需要注意的是, 这种 margin 要求对分类器之间的距离进行度量。它被应用在 AdaBoost 方法中, 如图 1(b)所示, 该文取  $M_\theta$  表示 hypothesis-margin。

设有  $\mu = \{\mu_j\}_{j=1}^k$ ,  $\hat{\mu} = \{\hat{\mu}_j\}_{j=1}^k$  定义了两个分类器, 其距离则可以表示为

$$\rho(\mu, \hat{\mu}) = \max_{i=1}^k \|\mu_i - \hat{\mu}_i\|_2 \quad (2)$$

**引理 1** 设  $\mu = \{\mu_j\}_{j=1}^k$  是一个类别集,  $\mathbf{x}$  是其中一个样本。  $\mu$  相对于  $\mathbf{x}$  的 hypothesis-margin

$$M_\theta(\mathbf{x}) = \frac{1}{2} (\|\mu_{\text{nearmiss}} - \mathbf{x}\| - \|\mu_{\text{nearhit}} - \mathbf{x}\|) \quad (3)$$

这里  $\mu_{\text{nearhit}}, \mu_{\text{nearmiss}}$  为与  $\mathbf{x}$  具有相同和不同类别的最近的样本点<sup>[2]</sup>。

**引理 2** 设  $S = \{\mathbf{x}_i\}_{i=1}^m$  是一个样本,  $\mu = (\mu_1, \dots, \mu_k)$  是一个假设集合<sup>[2]</sup>, 则

$$\text{sample-margin}_s(\mu) \geq \text{hypothesis-margin}_s(\mu) \quad (4)$$

Crammer<sup>[3]</sup>的研究结果表明, 如果我们能找到一个具有最大 hypothesis-margin 的类别集, 那么它同样具有最大的 sample-margin。sample-margin 在分类识别中具有最自然的特点, 为此它使 SVM 具有很好的分类性能, 而 hypothesis-

margin 则在计算上更容易, 具有比 sample-margin 更低的误差界。

## 3 基于不同 margin 的 SBS 特征选择方法

顺序后退法(Sequential Backward Selection, SBS)是一种特征选择的基本方法, 它是从全体特征中开始每次剔除一个特征, 剔除特征后仍然使保留的特征组的可分性判据最大。

由于 SBS 方法在特征剔除前并未考虑被剔除特征本身的特性, 而是考虑特征剔除后可分性判据的变化。一旦一个特征被剔除将不能再次返回特征集。一般改进方法采用顺序前进法(SFS)与 SBS 结合, 预先用 SFS 方法将特征一个个加入特征集, 然后用 SBS 方法将特征一个个剔除。但这种方法计算量明显很大, 对于人脸特征的选择就显得更加困难。该文基于 sample-margin 和 hypothesis-margin 获得各特征权重, 在 SBS 算法中逐一剔除权重较小的特征, 直到满足所需的特征数或其分类误差达到一定的阈值。

### 3.1 基于 sample-Margin 的 SBS 特征选择算法

设输入训练样本  $\mathbf{X}_0 = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_l]^T$ , 类别标签  $y = [y_1, y_2, \dots, y_k, \dots, y_l]^T$ 。初始化原始特征集  $\mathbf{s} = [1, 2, \dots, n]$ , 按特征标准排列后得到的特征集  $\mathbf{r} = []$ , 特征选择循环直到  $\mathbf{s} = []$ 。

(1) 训练样本划分为单特征:  $\mathbf{X} = \mathbf{X}_0(:, \mathbf{s})$

(2) 训练 SVM 分类器, 获得各单特征的  $M_s$ 。

(3) 剔除在排列标准中最后的特征, 即权重最小的特征:  $f = \text{argmin} M_s$ ,

(4) 更新特征集:  $\mathbf{r} = [\mathbf{s}(f), \mathbf{r}]$

(5) 排除特征后剩余特征集:  $\mathbf{s} = (1 : f-1, f+1 : \text{length}(\mathbf{s}))$

(6) 输出特征集:  $\mathbf{r}$

### 3.2 基于 Hypothesis-Margin 的 SBS 特征选择算法

该文采用基于 margin 的循环搜索方法(Iterative Search Margin Based Algorithm, ISMBA)<sup>[2]</sup>。

Samba 特征权重搜索算法:

(1) 初始化权重  $\mathbf{w} = (1, 1, \dots, 1)$

(2) 循环  $T$  次, 计算每个单特征的权重: for  $t=1$  to  $T$

(a) 从训练样本集  $S$  中随机选出一个样本  $\mathbf{x}_i, y_i$

(b) 计算与  $\mathbf{x}_i$  具有不同和相同类别的最近的点  $\mu_{\text{nearmiss}}, \mu_{\text{nearhit}}$  及  $M_\theta$ 。

(c) for  $i=1$  to  $N$ , 计算

$$\Delta_i = \frac{1}{2} \left( \frac{(\mathbf{x}_i - \mu_{\text{nearmiss}(i)})^2}{\|\mathbf{x} - \mu_{\text{nearmiss}}\|_w} - \frac{(\mathbf{x}_i - \mu_{\text{nearhit}(i)})^2}{\|\mathbf{x} - \mu_{\text{nearhit}}\|_w} \right) \mathbf{w}_i \quad (5)$$

(d)  $\mathbf{w} = \mathbf{w} + \Delta$

(3) 输出权重  $\mathbf{w} \leftarrow \mathbf{w}^2 / \|\mathbf{w}^2\|_\infty$ , 其中  $(\mathbf{w}^2)_i := (\mathbf{w}^2)_i$ 。

这样就可以计算出每个特征的权重。

## 4 Relief 特征选择算法

Relief 是 Kira<sup>[13]</sup>提出的被公认的性能较好的特征选择方

法,其基本思想是,一个好的特征应该使最近邻的同类之间特征值相同或相近,而使最近邻的不同样本之间的特征值不同和差别很大。据此对每个特征赋予不同的权值进行特征排序,通过设定特征权值的阈值或特征子集的数目进行特征选择。

Relief 特征选择算法:

(1)初始化权重:  $w=0$

(2)循环  $T$  次, for  $t=1$  to  $T$

(a)从训练样本  $S$  中随机选出一个实例

(b)循环  $N$  次, for  $i=1$  to  $N$

$$w_{(i)} = w_{(i)} + (\mathbf{x}_{(i)} - \mu_{\text{nearmiss}(i)})^2 - (\mathbf{x}_{(i)} - \mu_{\text{nearhit}(i)})^2 \quad (6)$$

根据计算出的特征权重,对特征进行重新排队,根据需要的特征维数或阈值进行特征选择。

### 5 实验及分析

#### 5.1 实验设计

实验在FRERT人脸图像数据库上进行。它包括1199个人的14,126幅图像(256×384),其中同一个人的图像差异包括不同表情、不同光照、不同姿态以及不同时期(相隔18个月以上)拍摄差异等。我们选择其中在同一年拍摄的200人的1400幅图像作为实验的图像子集,原始图像实例如图2所示。实验首先对图像进行了简单的预处理:根据眼睛的位置提取每幅图像中的人脸部分;然后采用直方图均衡化对其进行光度归一化;进一步对其进行了标准化处理,使其具有零均值和单位方差;最后将所有图像的分辨率统一调整为211×222,如图3所示。由于此时图像尺寸用于特征选择和识别仍然较大,所以我们采用2阶小波分解对原始图像进行压缩。小波变换(wavelet transform)通过多分辨率分析可以获得有效的人脸低频特征,Nastar等<sup>[14]</sup>研究了人脸外观变化与频谱变化之间的关系,指出人脸的光照、少许遮罩、旋转扭曲和面部表情只影响图像中高频部分,人脸图像的低频部分仍然保持稳定。

**基本实验** (1)训练SVM多项式分类器(核函数为  $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d$ ), 获得sample-margin作为SBS算法



图2 FERET人脸图像库的子库中的两个人的变化像例



图3 预处理后的人脸图像例

的特征选择标准进行特征选择,再用线性 SVM 分类器进行分类实验。(2)基于 hypothesis-margin 的 simba 搜索算法获得各单特征的权重,利用 SBS 算法逐一剔除权重较小的特征,获得最佳特征组合。训练 SVM 多项式分类器进行分类实验。(3)用 Relief 特征选择方法获得最佳特征组合,同样采用 SVM 多项式分类器进行分类实验。

**对比实验** (1)对采用 SVM 与 NN 分类器进行人脸识别的结果进行比较。(2)采用 grid-search 搜索方法获取 SVM 多项式分类器的超参数( $C, d$ )。 $C$ 取值 10~10000,  $d$ 取值 0.1~3.0。

图 4,图 5 分别将 sample-margin 和 hypothesis-margin 作为 SBS 算法的特征选择标准进行特征选择,并与 Relief 特征选择方法进行比较。该文列出 SVM 多项式分类器超参数为  $C=1000, d=0.1, 0.5, 1.0, 2.0$  时的实验结果(其余类似略)。基于 hypothesis-margin 的方法采用 simba 循环搜索法(损失函数采用线性函数  $(L(\theta) = \sum_i L(\theta_i))$  和 sigmoid 函数  $L(\theta) = \sum_i [1/(1 + e^{-\beta\theta_i})]$ ,  $\beta$  为常数。图 6 为经过不同特征选择方法后, SVM 分类器和 NN 分类器的分类识别结果。图 7 表示 SVM 多项式分类器超参数在  $C$  固定( $C=1000$ ) $d$  变化时的分类识别率,图 7(a)为采用 Relief 作为特征选择标准的特征选择及识别结果,图 7(b)为 sample-margin 特征选择方法的识别结果。图 8 表示 SVM 多项式分类器超参数在  $C$  固定( $C=1000$ ) $d$  变化时的分类识别率,图 8(a),图 8(b)分别为采用 simba(线性损失函数-lin)和 simba(sigmoid 损失函数-sig)搜索方法的分类结果。

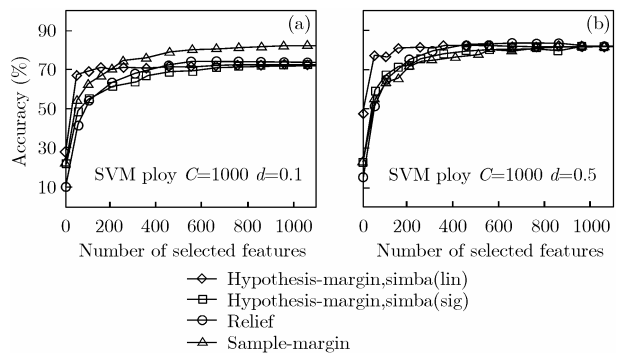


图 4 几种特征选择方法与 SVM 分类器在超参数不同时的分类识别率( $C=1000, d=0.1, d=0.5$ )

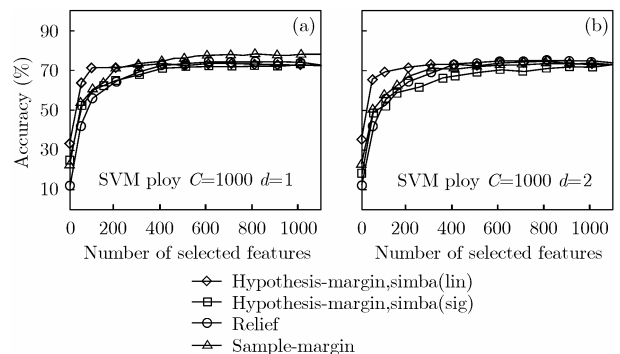


图 5 几种特征选择方法与 SVM 分类器在超参数不同时的分类识别率( $C=1000, d=1, d=2$ )

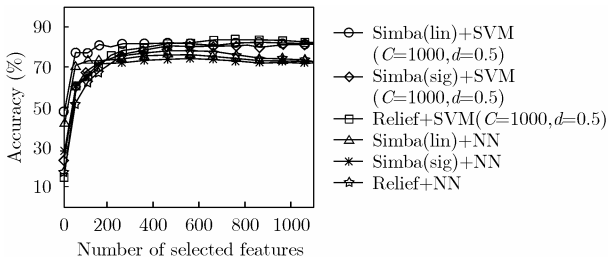


图6 采用不同特征选择方法后, SVM 或 NN 分类器的分类识别结果

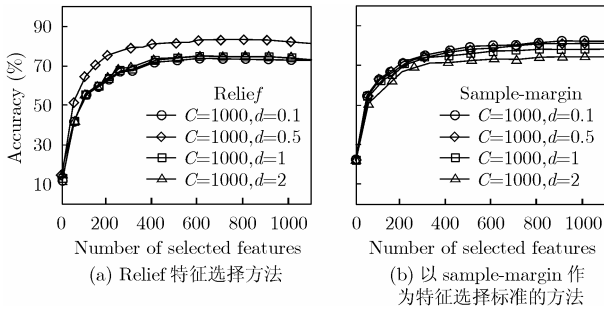


图7 SVM 多项式分类器超参数  $C$  固定( $C=1000$ )  $d$  变化时的分类识别率

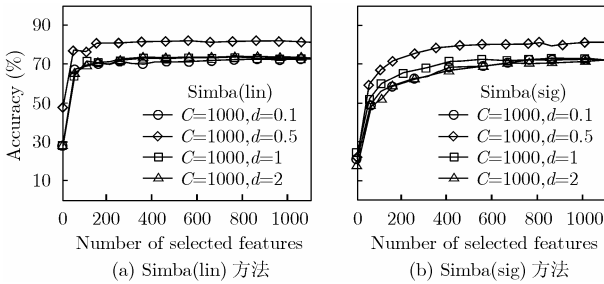


图8 SVM 多项式分类器超参数  $C$  固定( $C=1000$ )  $d$  变化的分类识别率

5.2 实验分析

从图3, 图4 我们可以得到几种特征选择方法以及 SVM 分类器在不同的最佳超参数情况下的分类识别率。最佳超参数通过 grid-search 搜索方法获取,  $C$  取值  $10 \sim 10000$ ,  $d$  取值  $0.1 \sim 3.0$ , 图示为其中部分结果(其余略)。从图4(a)可以得到: SVM 超参数  $C=1000$ ,  $d=0.1$  时, simba(lin)方法在特征维数小于 200 时分类识别效果比其他方法更好, 但特征维数大于 200 以后, sample-margin 方法明显优于其他方法。同样从图4(b)可以得到, SVM 多项式超参数为  $C=1000$ ,  $d=0.5$  时, 特征维数小于 500 时, simba(lin)方法为最优, 特征维数在 500 以后, 4 种方法效果相近。从图5(a)可以得到, 当 SVM 超参数为  $C=1000$ ,  $d=1$  时, simba(lin)方法在特征维数小于 200 时分类识别效果比其他方法好, 但特征维数在 200 以后, sample-margin 方法比其他方法略好。从图5(b)可以得到, 当 SVM 超参数为  $C=1000$ ,  $d=2$  时, 在整个特征维数范围内, simba(lin)方法为最好。从以上实验结果可以得到: 基于不同 margin 的特征选择方法, 都可以在特征维数较小时获得较好的分类效果。从总体上基于 hypothesis-

margin 的 simba(lin)方法较好。而对于通过不同特征选择方法获得的特征子集, SVM 分类器最佳超参数有所不同。

为了比较这几种特征选择方法对于分类器的敏感程度, 该文采用 SVM 多项式分类器与 NN 分类器进行对比实验。从图6可以得到: 特征维数小于 300 时几种特征选择方法效果有较大差异, 但当特征维数大于 300 以后, 分类结果几乎趋于一致。可以看到当经过几种特征选择后, NN 分类器与 SVM 分类器的人脸识别分类效果差异较小, 仅为 5~6% 左右。这证明该文提出的特征选择方法的确是一种较好的方法。

由于 SVM 的超参数对分类识别结果有一定程度的影响, 该文进行了一系列相应的实验以确定 SVM 的最优超参数。从图7(a), 图8(a), 图8(b)可以得到: 经过 Relief 方法和基于 hypothesis-margin 的 simba 搜索方法进行特征选择后, 再设计 SVM 多项式进行人脸识别时, 都在  $C=1000$ ,  $d=0.5$  处获得最佳的分类结果。而图7(b)可以得到: 基于 sample-margin 的 SVM 多项式分类器的超参数在  $C$  固定时, 分类识别率随  $d$  的减小而提高。

6 结束语

该文分别对 sample-margin 和 hypothesis-margin 进行了分析, 提出基于不同 margin 对特征集中的各特征进行权重分析, 在 SBS 算法中剔除权重较小的特征, 当达到需要的特征维数或分类误差达到一定的阈值时, 输入 SVM 多项式分类器进行分类实验。基于 sample-margin 的特征选择方法将 SVM 与 SBS 进行结合可直接得到实验结果。基于 hypothesis-margin 方法的 simba 循环搜索方法(采用线性损失函数和 sigmoid 函数)进行特征选择, 再设计 SVM 多项式分类器进行分类实验。实验结果表明: 本文提出的特征选择方法均能有效地降低特征维数, 获得较好的分类结果, 为此本文提出的方法是实用、有效的人脸特征选择及识别方法。特别值得注意, 超参数对 SVM 分类器的分类能力有一定的影响。实验显示: 在选择合理的特征选择方法的同时设计具有最佳超参数的 SVM 分类器将会提高分类识别的能力。

参考文献

- [1] Langley P. Selection of relevant features in machine learning. In: Proc. AAAI Fall Symposium on Relevance, New Orleans, Louisiana, 1994: 140-144.
- [2] Gilad-Bachrach R, Navot A, and Tishby N. Margin based feature selection - theory and algorithms. In proceedings of the 21<sup>st</sup> International Conference on Machine Learning (ICML), Banff, Alberta, Canada, July 4-8, 2004: 43.
- [3] Crammer K, Gilad-Bachrach R, Navot A, and Tishby N. Margin analysis of the lvq algorithm. Proc.17<sup>th</sup> Conference on Neural Information Processing Systems, Banff, Alberta, Canada, 2002: 462-469.

- [4] Vapnik V. The Nature of Statistical Learning Theory. Springer-Verlag, 1995.
- [5] Freund Y and Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, 55(1):119-139.
- [6] Buckingham L and Geva L. L<sub>v</sub>q is a maximum margin algorithm. In Pacific Knowledge Acquisition Workshop PKAW'2000, Sydney, Australia, 2000.
- [7] Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, and Vapnik V. Feature selection for SVMs. In *Advances in Neural Information Processing Systems*. MIT Press, 2000, (13): 668-674.
- [8] Grandvalet Y and Canu S. Adaptive scaling for feature selection in svms. In *Advances in Neural Information Processing Systems*, MIT Press, 2003, (15): 553-560.
- [9] Guyon I, Weston J, Barnhill S, and Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning*, 2002, 46(1/3): 389-422.
- [10] Ahmad A and Dey L. A feature selection technique for classificatory analysis. *Pattern Recognition Letters*, 2005, 26(1): 43-56.
- [11] Weihong L and Weiguo G. Feature selection based KPCA, SVM and GSFS for face recognition. *Proceeding of 3rd International Conference on Advances in Pattern Recognition*, London, 2005, 2: 344-350.
- [12] Bartlett P. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Trans. on Information Theory*, 1998, 44(2): 525-536.
- [13] Kira K and Rendell L A. The feature selection problem: traditional methods and a new algorithm. In: *Proc. of the Ninth National Conf. on Artificial Intelligence*, Menlo Park, CA, USA, 1992: 129-134.
- [14] Nastar C and Ayache N. Frequency-based non-rigid motion analysis. *IEEE Trans. on Pattern Anal. Machine Intell.*, 1996, 18(11): 1067-1079.
- 李伟红: 女, 1962 年生, 博士生, 研究方向为模式识别。  
陈伟民: 男, 1955 年生, 教授, 博士生导师, 研究方向为仪器科学与技术。  
杨利平: 男, 1981 年生, 博士生, 研究方向为模式识别。  
龚卫国: 男, 1957 年生, 教授, 博士生导师, 研究方向为智能化信息技术及系统。