

文章编号: 1001- 2486(2009) 03- 0087- 05

# 判决特征选择的快速人脸识别算法\*

文益民<sup>1,2</sup>, 范志刚<sup>3</sup>

(1. 湖南大学 电气与信息工程学院, 湖南 长沙 410082; 2. 湖南工业职业技术学院 信息工程系, 湖南 长沙 410208;  
3. 上海交通大学 计算机系, 上海 200240)

**摘要:** 为了在确保识别精度的条件下提高人脸识别的速度, 提出了判决特征选择算法(SVM-DFS)。针对多类分类问题, 判决特征选择算法根据统计学习理论使用支持向量机来实现特征选择, 根据全概率定理把特征选择和多类分类集成到一个统一框架。在 UMIST 和 FERET 人脸数据库上的实验表明: SVM-DFS 算法可以用来挑选对分类最有用的特征, 这些挑选出来的特征具有明显的物理意义。使用判决特征选择方法不但可以加快分类器的响应速度, 而且不降低分类器的泛化能力。

**关键词:** 判决特征; 学习系统; 支持向量机; 人脸识别

中图分类号: TP391.4 文献标识码: A

## Discriminative Feature Selection for Fast Face Recognition

WEN Yi-min<sup>1,2</sup>, FAN Zhi-gang<sup>3</sup>

(1. College of Electrical and Information Engineering, Hunan Univ., Changsha 410082, China;  
2. Department of Information Engineering, Hunan Industry Polytechnic, Changsha 410208, China;  
3. Department of Computer, Shanghai Jiao Tong Univ., Shanghai 200240, China)

**Abstract:** In order to speed up the face recognition algorithm under the condition of keeping high accuracy, a discriminative feature selection algorithm is proposed to handle multi-class face recognition problems. It is supported by vector machine to select features and employ total probability rule to integrate feature selection and multi-class classification into an integral frame. The experiment on the face databases of UMIST and FERET displays that the proposed algorithm can effectively select the features which have obvious physical meanings, thus speeding up the response of classifier without degrading the generalization performance.

**Key words:** discriminative feature; learning systems; support vector machines; face recognition

人脸自动识别是一个热门研究方向<sup>[1-2]</sup>。随着网络技术、多媒体技术和个人 PC 的广泛应用, 由此衍生出的视频会议和基于内容的压缩和检索(MPEG-4 和 MPEG-7) 技术等都对人脸信息作深入处理。因此, 人脸自动识别算法具有重大研究价值和广阔应用前景。人脸自动识别通常有实时的要求, 因此人脸自动识别算法的响应速度是关系到人脸自动识别算法应用的重要因素。采用特征选择算法挑选对人脸识别最重要的特征, 同时实现对数据量的削减, 是提高人脸自动识别算法的响应速度的有效措施之一。

20 世纪 90 年代, 统计学习理论框架下产生了支持向量机(SVM) 这一新的通用机器学习方法。该方法在经验风险最小的前提下, 确保了最高的泛化能力。该方法已经在实践中表现出了优良特性。不仅是针对小样本问题, 更重要的是针对高维小样本问题, SVM 都表现出了很好的泛化能力<sup>[3]</sup>。基于支持向量机的理论和应用研究是一个很活跃的研究领域。最近又出现了基于支持向量机的特征选择方法<sup>[4-6]</sup>。Guyon 等在文献[4]中提出了使用支持向量机实现二类分类中的特征选择算法, 该算法通过挑选对应较大权重向量的特征来实现特征选择。文献[7]在此基础上也提出了一种利用支持向量机来实现特征挑选的算法。另外, 支持向量也被用作对特征进行排序的依据<sup>[8]</sup>。然而, 在多类分类问题中如何

\* 收稿日期: 2009-04-14

基金项目: 国家 863 计划资助项目(2007AA04Z244); 国家自然科学基金重点资助项目(60835004); 湖南省博士后科研资助专项计划资助项目(2008RS4005)

作者简介: 文益民(1969-), 男, 副教授, 博士后。

实现特征选择以加速多类分类过程呢?在文献[4]的基础上,本文提出了多类分类中的判决特征选择算法,该算法根据特征选择和和多类分类两个任务之间的关系,利用全概率定理将两者集成到一个统一框架,有效实现了多类分类中的特征选择。

## 1 多类分类中的判决特征选择算法

文中多类分类问题采用“一对其他”的策略实现。“一对其他”多类支持向量机构造  $k$  个二类支持向量机 ( $k$  是类别的数目)。第  $j$  个支持向量机  $f_j(X)$  的训练过程是把第  $j$  类的样本视为正类,把所有其他类别的样本视为负类。相对于整个多类分类问题,  $f_j(X)$  是一个二类分类的子模型,它用来判别第  $j$  类和所有其他类。 $r_{ij}$  表示根据二类分类子模型  $f_j(X)$  计算得到的第  $i$  个特征的排序指标<sup>[4]</sup>。 $E$  表示概率意义上的必然事件,  $f$  表示概率意义上的不可能事件。用  $\omega_j$  来表示第  $j$  类为真的事件。根据概率理论,事件  $\omega_1, \omega_2, \dots, \omega_k$  是样本空间的一个划分,也就是有:

$$E = \omega_1 \cup \omega_2 \cup \dots \cup \omega_k, \quad f = \omega \cup \bar{\omega}, \quad i \neq j \quad (1)$$

$P(\omega_j)$  是第  $j$  类为真的先验概率。定义一个随机事件  $S_i$  表示第  $i$  个特征被挑选出来作为对分类有用的特征。 $P(S_i | \omega_j)$  表示在  $\omega_j$  发生的条件下  $S_i$  的条件概率。当事件  $\omega_j$  发生时,第  $j$  个二类分类子模型  $f_j(X)$  对决定最终分类结果有效。根据第  $j$  个二类分类子模型  $f_j(X)$ , 通过计算特征排序指标  $r_{ij}$ <sup>[4]</sup> 可以得到

$$P(S_i | \omega_j) = \frac{r_{ij}}{\sum_{i=1}^L r_{ij}} \quad (2)$$

其中,  $L$  为特征的维数, 根据全概率定理有:

$$P(S_i) = \sum_{j=1}^k P(S_i | \omega_j) P(\omega_j) \quad (3)$$

于是  $P(S_i)$  可以用作针对整个多类分类问题的特征排序指标。

提出的多类分类中的判决特征选择算法的实现过程描述如下:

步骤 1 输入: 训练样本集合  $S = \{X_1, X_2, \dots, X_N\}^T, X_i \in R^L$ ;

迭代次数  $T$ ;

第  $t$  次迭代要选择特征的数量  $M_t$ 。

步骤 2 初始化: 为特征记上标号:  $s = [1, 2, \dots, L]$ ;

按照“一对其他”的策略训练  $k$  个二类支持向量机;

计算  $r_{ij}, 1 \leq i \leq L, 1 \leq j \leq k$ 。

步骤 3 For  $t = 1, 2, \dots, T$

(a) 根据式(2)和式(3)计算特征排序指标  $P(S_i)$ ;

(b) 按照  $P(S_i)$  值从大到小的次序对特征进行排序, 选择排在前面的  $M_t$  个特征并消除掉排在这  $M_t$  个特征后面的其他特征;

(c) 将消除掉的特征对应的标号去除, 然后更新  $s$ ;

(d) 根据  $s$  来更新训练样本, 即将训练数据通过投影限制在通过选择的特征上面:

$$S = S(:, s);$$

(e) 使用更新好的训练数据  $S$ , 按照“一对其他”的策略重新训练支持向量机并计算  $r_{ij}, 1 \leq i \leq L, 1 \leq j \leq k$ 。

步骤 3 输出: 经过挑选的特征和最终的支持向量机。

上述判决特征选择算法中的  $T$  和  $M_t$  是需要用户预先定义的常数。 $T$  是迭代的次数, 通常  $T$  的值不能取得太小, 但如果  $T$  值取得大则训练时间会变得很长。实际使用时,  $T$  值大小可以在精度和训练时间之间取折中值。 $M_t$  是在第  $t$  次迭代中挑选特征的数量。在  $T$  值确定的情况下,  $M_t$  的大小如何确

定呢?  $M_t$  值可以通过重新训练的支持向量机的参数来确定。  $M_t$  值如果选取得合理, 则重新训练的各个支持向量机子模型  $f_i(X)$  的间隔  $\Delta$  将足够大:

$$\Delta_t = \frac{1}{\|w^{(i)}\|} \quad (4)$$

其中,  $w^{(i)}$  表示子模型  $f_i(X)$  的权重向量。根据文献[9]有:

$$\|w^{(i)}\|^2 = \sum_{\text{support vectors}} \alpha_j^{(i)} \quad (5)$$

其中,  $\alpha_j^{(i)}$  表示子模型  $f_i(X)$  的拉格朗日系数。因此, 为使得  $f_i(X)$  的间隔  $\Delta_t$  能够足够大, 定义参数  $M$  为:

$$M = \sum_{i=1}^k P(\omega_i) \left[ \sum_{\text{support vectors}} \alpha_j^{(i)} \right] \quad (6)$$

可以根据参数  $M$  来确定  $M_t$  值的大小。  $M_t$  值如果选取得合理, 则参数  $M$  的值就能够足够地小。为方便起见, 后面将提出的算法简称为 SVM-DFS。

## 2 实验

为了系统评测提出的多类分类中判决特征选择算法的性能, 使用 UMIST 人脸数据库<sup>[10]</sup>来进行快速人脸识别实验。UMIST 人脸数据库是一个多角度人脸数据库, 包含了 20 个人的 575 张灰度图像。每个人的图像都具有正面到侧面的各个角度的图像。所有输入图像的大小为  $92 \times 112$ , 所以输入向量的维数为  $L = 10\ 304$ 。整个数据库被分为两个部分: 训练数据集和测试数据集。训练数据集包含 240 张图像——每个人 12 张图像, 都是经过仔细挑选的包含各个角度的图像, 剩下的 335 张图像被用作测试数据集。SVM 训练算法中的参数  $C = 10\ 000$ , 参数  $\sigma$  的取值位于区间  $(10, 1000)$ 。支持向量机的训练使用 libSVM 算法<sup>[11]</sup>, 其缓存设置为 40MB。

在这一组实验里面, SVM-DFS 算法被设置为只有一个迭代步骤, 即参数  $T = 1$ 。为了比较 SVM-DFS 的性能, 在同等条件下使用 KDDA 算法。KDDA 算法的实验设置根据文献[12]来执行, 并且使用一个 RBF 核函数。实验中, 考虑到参数  $M$  和分类速度与精度的折中, 设置参数  $M_t$  的值为 2000。

在表 1 中, SVM-DFS-L 表示在特征选择以后使用线性核函数的 SVM 作为分类器; SVM-DFS-R 表示在特征选择以后使用 RBF 核函数的 SVM 模型作为分类器。测试时间为所有单个测试样本的测试时间的总和。

表 1 SVM-DFS 和 KDDA 对比的测试结果  
Tab. 1 The comparison of SVM-DFS with KDDA

Methods	Correct rate(%)	Test time(s)
KDDA	95.2239	2.9474
SVM-DFS-L	96.7164	0.0463
SVM-DFS-R	98.2090	1.0431

从表 1 中可以发现: SVM-DFS 在提高分类器泛化能力的情况下, 减少了特征维数, 同时还加速了测试过程。相对于 KDDA 特征选择算法, 使用 SVM-DFS 算法进行特征挑选后, 分类器的泛化能力更好, 这说明了 SVM-DFS 挑选出的特征更能体现分类信息。

为了探讨参数  $T > 1$  的情形下 SVM-DFS 算法的性能, SVM-DFS 算法被设置为使用多个迭代步。实验结果如图 1 和图 2 所示, 图 2 是图 1 的局部放大图。其中 SVM-DFS-1、SVM-DFS-2、SVM-DFS-3、SVM-DFS-4 和 SVM-DFS-5 分别表示 SVM-DFS 算法使用参数  $M_t = K_t - 1$ ,  $M_t = K_t - 5$ ,  $M_t = K_t - 10$ ,  $M_t = K_t - 20$  和  $M_t = K_t - 40$ 。常数  $K_t$  表示在当前  $t$  迭代步的特征空间的维数。在特征选择的过程中, SVM-DFS 算法使用线性核, 再使用线性 SVM 分类器来实现 SVM-DFS 特征选择后的分类。

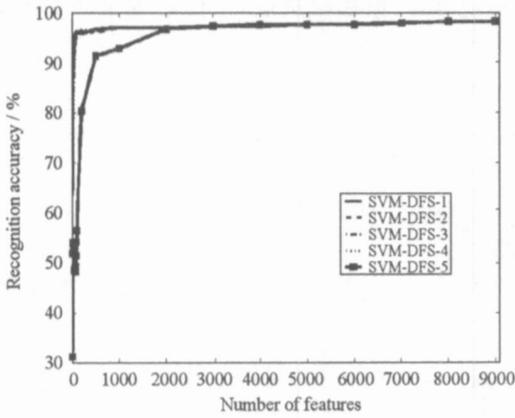


图1 在使用多个迭代步骤在不同维数情况下识别精度的变化情况

Fig.1 Recognition accuracy under various dimensionality with multiple iterative steps

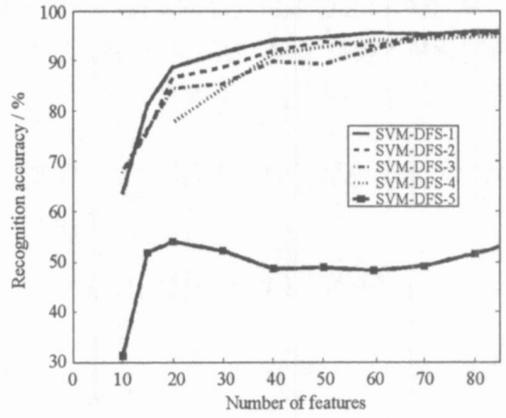


图2 图1的局部放大图

Fig.2 Enlarged figure of Fig.1

通过图1和图2可以看出:在使用更多的迭代步骤的情形下,使用SVM-DFS算法进行特征选择可以得到更高的识别精度。

### 3 分析与讨论

为什么SVM-DFS算法在减少特征维数的情形下还能确保分类器的泛化能力呢?使用FERET人脸数据库<sup>[3]</sup>来探讨SVM-DFS挑选出来的特征的性质。所有的图像都使用zero-mean-unit-variance操作来预处理,并且根据眼睛的位置来对齐。所有的图像都使用眼睛的位置来归一化,并且尺寸大小都统一为130×150。用一个mask模板去除背景和人物头发,使用直方图均衡化改善图像的质量。从FERET FA/FB数据集里面挑选100个人的图像,每个人两张图像,一张用来训练,一张用来测试。SVM-DFS算法中使用线性核函数,并且设置参数T=1。实验结果如图3和图4所示。

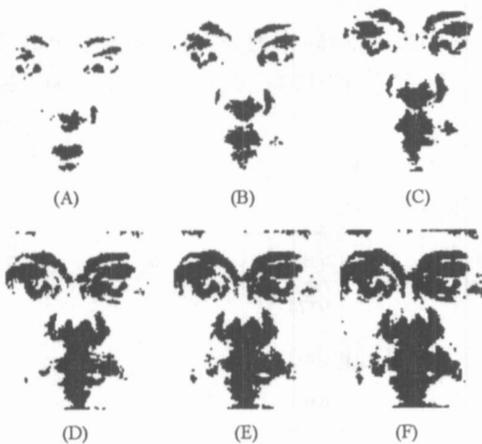


图3 选择出来的重要像素特征被显示为黑色像素  
Fig.3 The selected important pixel features are showed as black pixels

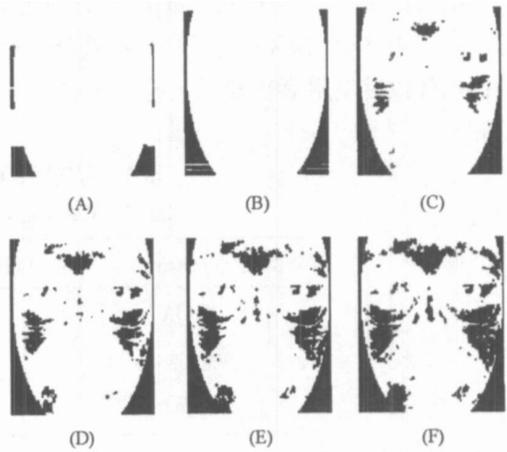


图4 选择出来的不重要像素特征被显示为黑色像素  
Fig.4 The selected unimportant pixel features are showed as black pixels

在图3中,黑色像素是被选择出来的重要像素特征:(A)表示被选择出来的最重要的1000个像素特征;(B)表示被选择出来的最重要的2000个像素特征;(C)表示被选择出来的最重要的3000个像素特征;(D)表示被选择出来的最重要的4000个像素特征;(E)表示被选择出来的最重要的5000个像素特征;(F)表示被选择出来的最重要的6000个像素特征。

相对地,在图4中,黑色像素是被选择出来的不重要像素特征:(A)表示被选择出来的最不重要的

1000 个像素特征; (B) 表示被选择出来的最不重要的 2000 个像素特征; (C) 表示被选择出来的最不重要的 3000 个像素特征; (D) 表示被选择出来的最不重要的 4000 个像素特征; (E) 表示被选择出来的最不重要的 5000 个像素特征; (F) 表示被选择出来的最不重要的 6000 个像素特征。

从图 3 和图 4 可以知道:最重要的像素特征能够显示为一个人脸的框架图,而最不重要的像素特征则显示为一个 mask 模板的框架图。可以认为 SVM-DFS 算法能够有效地实现特征选择,挑选出来的特征具有明显的物理意义。因此,经过 SVM-DFS 算法挑选的特征仍然能够描述样本的类别信息,因而在一定范围内能够在降低特征维数的同时确保分类器的泛化能力不会下降。

## 4 结论

本文提出了多类分类中的判决特征选择算法。针对多类分类问题,根据全概率定理把特征选择和多类分类集成到一个统一框架,判决特征选择方法使用支持向量机实现特征选择。在 UMIST 人脸数据库上的实验表明:使用判决特征选择方法可以加快分类器的响应速度,同时不降低分类器的泛化能力。在 FERET 人脸数据库上的实验则表明:SVM-DFS 算法可以选择到对分类最有用的特征,选出来的特征具有明显的物理意义。本文的工作反映了使用对象的本质的物理特征能够确保分类成功,因此,如何从认知心理学的角度深入探讨人脸识别的特征选择问题,对于提高人脸识别的精度是一个非常有趣的问题。

## 参考文献:

- [1] Pentland A. Looking at People: Sensing for Ubiquitous and Wearable Computing [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2000, 22(1):107- 119.
- [2] 李江,郁文贤,匡刚要,等. 红外图像人脸识别方法[J]. 国防科技大学学报, 2006, 28(2): 73- 76.
- [3] Vapnik V N. Statistical Learning Theory[M]. New York: Wiley Interscience, 1998.
- [4] Guyon I, Weston J, Barhill S, et al. Gene Selection for Cancer Classification Using Support Vector Machines[J]. Machine Learning, 2002, 46: 389 - 422.
- [5] Mao K Z. Feature Subset Selection for Support Vector Machines through Discriminative Function Pruning Analysis[J]. IEEE Trans. Systems, Man, and Cybernetics, 2004, 34(1): 60- 67.
- [6] Chang Y W, Lin C J. Fe JMLR Workshop and Conference Proceedings: Causation and Prediction Challenge (WCCI 2008), 2008, 3: 53- 64.
- [7] Fan Z G, Lu B L. Fast Recognition of Multi-view Faces with Feature Selection[C]//Proceedings of the 10<sup>th</sup> IEEE International Conference on Computer Vision, 2005: 76- 81.
- [8] Heisele B, Serre T, Prntice S, et al. Hierarchical Classification and Feature Reduction for Fast Face Detection with Support Vector Machines[J]. Pattern Recognition, 2003, 36: 2007- 2017.
- [9] Vapnik V. An Overview of Statistical Learning Theory[J]. IEEE Trans. Neural Networks, 1999, 10(5): 988- 999.
- [10] Graham D B, Allinson N M. Characterizing Virtual Eigensignatures for General Purpose Face Recognition[J]. Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences, 1998, 163: 446- 456.
- [11] Fan R E, Chen P H, Lin C J. Working Set Selection Using Second Order Information for Training SVM[M]. Journal of Machine Learning Research, 2005, 6: 1889- 1918.
- [12] Lu J, Plataniotis K, Venetsanopoulos A. Face Recognition Using Kernel Direct Discriminant Analysis Algorithms[J]. IEEE Trans. Neural Networks, 2003, 14(1): 117- 126.
- [13] Phillips P J, Moon H, Rizvi S A, et al. The Feret Evaluation Methodology for Face Recognition Algorithms[J]. IEEE Trans. Pattern Analysis and Machine Learning, 2000, 22(10): 1090- 1104.