

基于迭代特征选择的快速多角度人脸识别

范志刚 吕宝粮

(上海交通大学计算机科学与工程系 上海 200030)

(zgf@sjtu.edu.cn)

Fast Recognition of Multi-View Faces Based on Iterative Feature Selection

Fan Zhigang and Lü Baoliang

(Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030)

Abstract By utilizing support vector machines, an iterative feature selection method is proposed for multi-view face recognition. vector machines. According to the statistical relationship between the two tasks, feature selection and multi-class classification, the two tasks are integrated into a single consistent framework and the goal of feature selection is realized effectively. The classification process can be made more fast without degrading the generalization performance through this iterative feature selection method. The experiments conducted on the UMIST multi-view face database show that this iterative feature selection method can speed up the multi-view face recognition process without degrading the generalization performance.

Key words multi-view face recognition; fast classification; feature selection

摘要 针对快速多角度人脸识别问题,提出了一种利用支持向量机实现迭代特征选择的方法。根据特征选择和多类分类任务之间的统计关系,将这两个任务集成到一个统一的框架中,从而更有效地完成特征选择。这种迭代特征选择方法的特点是在不降低泛化能力的情况下大大加快分类速度。在 UMIST 多角度人脸数据库上的实验结果验证了所提方法的有效性。

关键词 多角度人脸识别;快速分类;特征选择

中图法分类号 TP391

1 引言

人脸识别技术在过去的几十年里一直是研究领域里的一个热点课题。其中多角度人脸识别是一个比正面人脸识别更加困难的识别任务。在多角度人脸识别问题中,数据空间中存在着非线性流形。所以现有的许多针对正面人脸识别的方法对多角度人

脸识别不是很有效。要更好地解决多角度人脸识别问题就必须有效地学习非线性流形结构。

子空间(subspace)方法是人脸识别的经典方法^[1],其中特征脸(eigenfaces)^[2,3]是最早的应用到人脸识别的经典方法。特征脸方法是采用主分量分析(PCA)来产生针对人脸识别的有效子空间。还有线性判决分析(LDA)^[4]、贝叶斯方法^[5]、独立分量分析(ICA)^[6]等人脸识别的子空间方法。因为人脸

模式的非线性分布,这些线性的人脸识别方法对于多角度人脸识别问题都不是很有效。近几年,针对人脸模式的非线性问题,人们提出了核子空间方法。Kernel PCA^[7]和 KDDA(kernel direct discriminant analysis)^[8]是可以用来解决多角度人脸识别的非线性核方法。这些方法是将核方法和子空间方法结合起来的产物。但是,由于核方法需要大量的计算,核子空间方法的分类速度特别慢。

为了加快多角度人脸识别的速度并保持识别系统的泛化能力,我们在前面的工作中提出了一种基于支持向量机的判决特征选择方法:SVM-DFS^[9]。本文在 SVM-DFS 的基础上,提出了利用支持向量机实现迭代特征选择的方法(SVM based discriminative feature selection in iterative way, SVM-DFS_I)。和子空间的投影方法不同,SVM-DFS_I是直接挑选对于分类最有用的特征。一种利用支持向量机实现递归特征消除的方法(SVM-RFE)^[10]已经被用来解决两类分类问题中的特征选择。本文提出的 SVM-DFS_I方法是用来解决人脸识别这样的多类分类问题中的特征选择。

2 基于支持向量机的迭代特征选择(SVM-DFS_I)

利用多类支持向量机,SVM-DFS_I能够完成多类分类问题中的特征选择任务。本文中多类支持向量机采用“一对其他”的问题分解方法^[11~14]。“一对其他”多类支持向量机是构造 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 类的样本看做为正类,把所有其他的样本看做为负类。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,它是用来判别第 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^[9,15,16]。 E 表示概率意义上的必然事件, ϕ 表示概率意义上的不可能事件。我们用 ω_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:

$$E = \omega_1 \cup \omega_2 \cup \dots \cup \omega_k, \quad (1)$$

并且

$$\phi = \omega_i \cap \omega_j, i \neq j. \quad (2)$$

$P(\omega_j)$ 是第 j 类为真的先验概率。定义一个随机事件 S_i 为“第 i 个特征被挑选出来作为在分类中使用的特征”。 $P(S_i | \omega_j)$ 表示在 ω_j 发生的条件下 S_i

的条件概率。当事件 ω_j 发生的时刻,第 j 个二类分类子模型 $f_j(x)$ 正好对决定最终分类结果是有效的。根据第 j 个二类分类子模型 $f_j(x)$,通过特征排序指标 r_{ij} 来计算概率 $P(S_i | \omega_j)$ 的值:

$$P(S_i | \omega_j) = \frac{r_{ij}}{\sum_{t=1}^n r_{tj}}. \quad (3)$$

根据全概率定理, $P(S_i)$ 可以通过 $P(S_i | \omega_j)$ 和 $P(\omega_j)$ 来计算得到:

$$P(S_i) = \sum_{j=1}^k P(S_i | \omega_j) P(\omega_j). \quad (4)$$

于是 $P(S_i)$ 可以用来作为针对整个多类分类问题的特征排序指标。SVM-DFS_I 迭代特征选择算法就是根据 $P(S_i)$ 来挑选特征的。SVM-DFS_I 迭代特征选择算法的实现过程描述如下。

(1) 输入训练样本:

$$X_0 = \{x_1, x_2, \dots, x_l\}. \quad (5)$$

(2) 初始化:

为特征记上标号: $s = [1, 2, \dots, n]$,利用训练样本 X_0 来训练支持向量机。

(3) For $t = 1, 2, \dots, T$:

① 根据训练好的支持向量机来计算特征排序指标 $P(S_i)$;

② 按照 $P(S_i)$ 的值从大到小的次序对特征进行排序,选择排在前面的特征并消除掉排在最后面的 M 个特征;

③ 通过把已经消除掉的特征的标号去除掉来更新 s ;

④ 根据经过选择的特征的标号 s 来更新训练样本,将训练数据限制在经过选择的特征上面:

$$X = X_0(:, s);$$

⑤ 用更新好的训练数据 X 来重新训练支持向量机。

(4) 输出:最终的支持向量机分类器和经过挑选的特征。

上述 SVM-DFS_I 算法中的 M 和 T 是需要用户定义的常数。 M 的值在不等于零的情况下应该定义得尽量小,这样特征选择的精度才会高。在下面的实验部分我们将说明 M 的不同取值对特征选择精度的影响。当然, M 值如果取得小则训练时间会变得很长。实际使用时, M 值的大小可以在精度和训练时间之间取折中。我们考虑在 M 值确定的情况下如何确定 T 的大小问题。 T 值的大小可以通过重新训练好的支持向量机模型的参数来确定。 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 Δ_i 能够足够大.

$$\Delta_i = \frac{1}{\|w^{(i)}\|}, \quad (6)$$

其中, $w^{(i)}$ 表示子模型 $f_i(x)$ 的权重向量. 又有等式^[11]:

$$\|w^{(i)}\|^2 = \sum_{\text{support vectors}} \alpha_j^{(i)}, \quad (7)$$

其中, $\alpha_j^{(i)}$ 表示子模型 $f_i(x)$ 的拉格朗日系数. 我们定义一个参数 L 为

$$L = \sum_{i=1}^k P(w_i) \left(\sum_{\text{support vectors}} \alpha_j^{(i)} \right). \quad (8)$$

我们可以根据参数 L 来确定 T 值的大小. T 的值如果选取得合理,则参数 L 的值能够足够小. 也可以用测试集验证的方法来确定 T 值的大小.

3 实验结果

为了验证本文提出的 SVM-DFSI 迭代特征选择算法在多角度人脸识别问题中的有效性,本节进行了一系列的实验. 所有的实验都是在 UMIST 人脸数据库^[17]上完成的. UMIST 数据库是一个多角度人脸数据库,整个库有 20 个人的 575 张灰度图像. 每个人的图像覆盖了从完全正面到完全侧面的人脸的各个角度. 所有输入图像的大小规格为 112×92 , 所以输入向量的维数是 $N = 10304$. 图 1 显示了 UMIST 人脸数据库中的一个几人脸图像样例. 在实验中,整个人脸数据库分为训练集和测试集两部分. 训练集包括 240 张图像,其中每个人根据人脸的不同角度仔细挑出 12 张图像. 剩余的 335 张图像作为测试集. 全部实验是在 Pentium IV PC 机上完成, CPU 主频为 3GHz, 内存为 1GB RAM. 所有的支持向量机都采用线性核函数,其中参数 $C = 10000$. 特征选择完成后,分类器就是使用



图 1 UMIST 数据库中一个人脸的不同角度的图像样例

SVM-DFSI 算法中输出的支持向量机模型.

为了测试经过特征选择以后不同维数对分类速度的影响,我们做了一系列实验. 让维数在区间 $(1, N)$ 上变化,人脸识别的速度随着维数的变化而改变. 这里说的维数等价于特征的数目,就是经过特征选择以后在分类中使用的特征个数. 图 2 给出了在参数 $M = 100$ 的情况下,不同维数与分类速度的关系. 图 2 中的分类测试时间(test time)表示整个测试集中所有测试样本的测试时间的总和. 从图 2 可以看出,经过特征选择以后,人脸识别的速度可以大大加快.

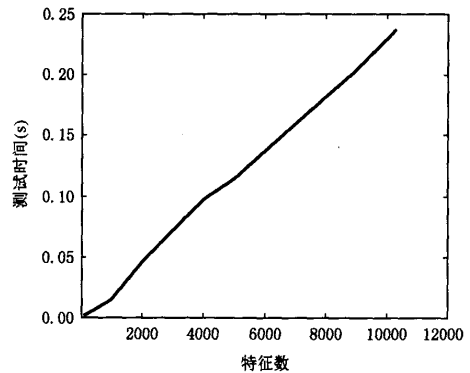


图 2 不同维数下测试时间的变化

图 3 显示了参数 M 分别取 1 和 0 的情况下人脸识别的准确率随维数变化的情况. 其中 SVM-DFSI-1 表示参数 M 取值为 1, SVM-DFSI-0 表示参数 M 取值为 0. 图 4 是图 3 的局部放大图. 从图 3 和 4 可以看出,参数 M 的不同取值对分类准确率的影响. 需要特别说明的是当 M 取值为 0 时是 SVM-DFSI 算法的一种特殊情况,此时 SVM-DFSI 算法实际上不能进行迭代而是要去直接一次性挑选排序在最前面的一定数量的特征. 从这些实验结果可以看出,在低维情况下,迭代的特征选择要比不迭代的特征选择的分类准确率高很多.

表 1 给出了在参数 M 分别取不同值的情况下人脸识别的测试准确率随维数改变而变化的情况. 从表 1 可以看出,在低维的情况下 $M = 1$ 的分类准确率最高. 随后 M 值从 5~1000 逐渐变大,则分类的准确率在降低,最后 $M = 0$ 时的分类准确率最低.

这些实验结果表明 SVM-DFSI 迭代特征选择算法的迭代次数应该尽量多,这样在低维情况下的分类准确率就会高. 当然迭代次数太多会使训练时间变得非常长,在实际使用时可以在精度和训练时间之间取折中.

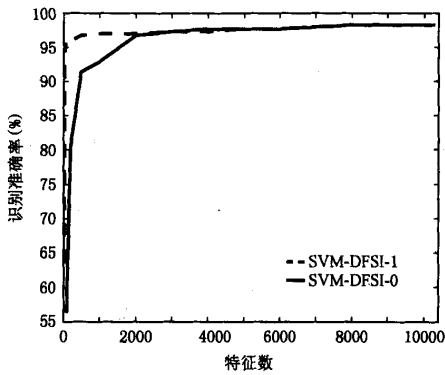


图3 不同维数下测试准确率的变化(M分别取1和0)

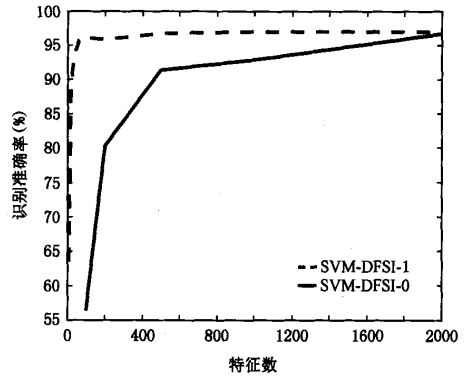


图4 图3的局部放大图

表1 不同维数和不同M值下人脸识别的测试准确率

维数	测试准确率(%)								
	无迭代	有迭代							
		M=0	M=1	M=5	M=10	M=20	M=50	M=100	M=500
9000	98.209	98.209	98.209	98.209	98.209	98.209	98.209	98.209	98.209
8000	98.209	98.209	98.209	98.209	98.209	98.209	97.910	97.910	98.209
7000	97.910	97.910	97.910	97.910	97.910	97.910	97.910	97.910	97.910
6000	97.612	97.612	97.612	97.612	97.612	97.910	97.612	97.612	97.612
5000	97.612	97.612	97.612	97.612	97.612	97.612	97.612	97.612	97.612
4000	97.612	97.313	97.313	97.313	97.313	97.313	97.313	97.313	97.313
3000	97.313	97.313	97.313	97.313	97.313	97.313	97.313	97.313	97.015
2000	96.716	97.015	97.015	97.015	97.015	97.015	97.015	97.313	97.313
1000	92.836	97.015	97.015	97.015	97.015	97.015	96.716	96.716	94.925
500	91.343	96.716	96.119	96.418	96.418	96.418	96.418	95.522	-
200	80.299	95.821	96.418	96.119	96.418	94.328	95.821	-	-
100	56.418	96.119	95.821	95.821	94.925	96.119	94.328	-	-
90	54.328	96.119	95.522	95.522	-	-	-	-	-
80	51.642	95.821	95.224	95.522	94.627	-	-	-	-
70	49.254	95.224	95.224	94.627	-	-	-	-	-
60	48.258	95.524	92.836	92.239	94.030	-	-	-	-
50	48.955	94.627	93.731	89.254	-	93.433	-	-	-
40	48.657	94.030	91.940	89.851	91.343	-	-	-	-
30	52.239	91.642	88.657	85.373	-	-	-	-	-
20	54.030	88.657	86.567	84.478	77.612	-	-	-	-
15	51.940	81.194	75.821	-	-	-	-	-	-
10	31.343	63.582	67.761	68.060	-	-	-	-	-

4 结论和进一步研究

本文提出了一种利用支持向量机实现迭代特征选择方法 SVM-DFS1. 该方法将特征选择和多类分

类两个任务集成到一个统一的框架中,实现了有效的特征选择. 我们在 UMIST 多角度人脸数据库上的一系列实验表明,这种迭代特征选择方法可以在不降低泛化能力的条件下加快多角度人脸识别的速度. 同时从实验结果看出, SVM-DFS1 算法的迭代

次数越多低维情况下的分类准确率越高。但是迭代次数太多会使训练时间变长。在实际使用时可以在精度和训练时间之间取折中。作为进一步研究工作,我们将使用小波变换作为 SVM-DFSI 算法的预处理。

参 考 文 献

- 1 X. Wang, X. Tang. Unified subspace analysis for face recognition. ICCV 2003, Nice, 2003
- 2 M. Turk, A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991, 3(1): 71~86
- 3 B. Moghaddam, A. Pentland. Probabilistic visual learning for object representation. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997, 19(7): 696~710
- 4 P. Belhumeur, J. Hespanha, D. Kriegeman, *et al.* Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997, 19(7): 711~720
- 5 B. Moghaddam, T. Jebara, A. Pentland. Bayesian face recognition. Pattern Recognition, 2000, 33(11): 1771~1782
- 6 M. S. Bartlett, J. R. Movellan, T. J. Sejnowski. Face recognition by independent component analysis. IEEE Trans. Neural Networks, 2002, 13(6): 1450~1464
- 7 B. Scholkopf, A. Smola, K. R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 1998, 10(5): 1299~1319
- 8 J. Lu, K. Plataniotis, A. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. IEEE Trans. Neural Networks, 2003, 14(1): 117~126
- 9 Z. G. Fan, B. L. Lu. Fast recognition of multi-view faces with feature selection. ICCV05, Beijing, 2005
- 10 I. Guyon, J. Weston, S. Barnhill, *et al.* Gene selection for cancer classification using support vector machines. Machine Learning, 2002, 46(3): 389~422
- 11 V. N. Vapnik. Statistical Learning Theory. New York: John Wiley and Sons, 1998
- 12 R. Rifkin, A. Klautau. In defense of one-vs-all classification. Journal of Machine Learning Research, 2004, 5(1): 101~141
- 13 C. Hsu, C. Lin. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Networks, 2002, 13(2): 415~425
- 14 B. L. Lu, K. A. Wang, M. Utiyama, *et al.* A part-versus-part method for massively parallel training of support vector machines. IJCNN'04, Budapest, 2004
- 15 Z. G. Fan, K. A. Wang, B. L. Lu. Feature selection for fast image classification with support vector machines. ICONIP 2004, Calcutta, 2004
- 16 B. Heisele, T. Serre, S. Prentice, *et al.* Hierarchical classification and feature reduction for fast face detection with support vector machine. Pattern Recognition, 2003, 36(9): 2007~2017
- 17 D. B. Graham, N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences, 1998, 163(1): 446~456



范志刚, 1978年生, 博士研究生, 主要研究方向为机器学习和模式识别。



吕宝粮, 1960年生, 博士, 教授, 博士生导师, 主要研究方向为仿脑计算理论与模型、神经网络、并行机器学习、脑-计算机接口、人脸识别、生物信息学和自然语言处理。

基于迭代特征选择的快速多角度人脸识别

作者: 范志刚, 吕宝粮

作者单位: 上海交通大学计算机科学与工程系, 上海, 200030

本文链接: http://d.g.wanfangdata.com.cn/Conference_6194392.aspx