基于集成的年龄估计方法

张 宇 [1] 周志华 [1]

摘 要 近十年来, 由于广泛的应用前景, 关于人脸识别的研究得到了广泛的关注. 但目前有一种影响人脸识别技术的因素尚未被研究者所重视, 那就是年龄变化. 而在适用于年龄变化的人脸识别技术中有一个重要的问题, 即年龄估计. 本文基于典型相关分析和代价敏感学习提出了两种年龄估计算法, 并在此基础上利用集成技术来提高年龄估计的准确性. 最终实验结果验证了本文方法的有效性.

关键词 人脸识别, 年龄估计, 典型相关分析, 代价敏感学习, 集成学习
中图分类号 TP18

A New Age Estimation Method Based on Ensemble Learning

ZHANG Yu[1] ZHOU Zhi-Hua[1]

Abstract For the last decade, research on face recognition has attracted much attention due to its wide application. However, the factor that affects the performance of face recognition and has been neglected by researchers is the aging variation. Age estimation is an important problem in the face recognition technique. This paper proposes two algorithms for age estimation by utilizing canonical correlation analysis and cost-sensitive learning. Furthermore, these two algorithms are combined into the ensemble method to improve the accuracy of age estimation. Finally, the experimental results show the effectiveness of the proposed method.

Key words Face recognition, age estimation, canonical correlation analysis, cost-sensitive learning, ensemble learning

由于具有广泛的应用前景, 人脸识别技术在近十年来得到了长足的发展. 目前大多数研究者认为人脸识别技术可能面对的主要挑战在于姿态、面部特征、面部表情、遮挡、图像方位及成像条件等六个方面[1]. 研究者提出了很多方法来处理这六种问题. 但另外有一种对人脸识别性能有很大影响的因素, 即年龄变化, 却没有受到足够的重视. 在一个人从小孩长成青年人的过程中, 人脸部骨骼不断生长, 从而脸型不断变长变大, 相貌也不断变化; 在由青年人不断衰老变成老年人的过程中, 皮肤逐渐衰老, 皱纹增多, 相貌也在不断变化. 由此可见年龄变化所带来的人脸相貌变化与上述六方面问题给人脸相貌带来的变化是截然不同的, 这就需要研究者进行研究. 因此适用于年龄变化的人脸识别技术是一个重要的研究方向. 而这方面存在一个重要的问题, 即年龄估计. 如果能准确地根据人脸图像估计年龄, 则可以有助于进行后继的识别. 此外年龄估计在人机交互领域有着很多应用, 例如网页浏览器可以根据用户的年龄来决定是否允许用户查看某些网页; 自动售货机可以拒绝向未成年人出售香烟和酒精性饮料等.

收稿日期 2007-05-10 收修改稿日期 2007-09-25
Received May 10, 2007; in revised form September 25, 2007
国家自然科学基金 (60635030), 国家高技术研究发展计划 (863 计划) (2007AA01Z169) 资助
Supported by National Natural Science Foundation of China (60635030), National High Technology Research and Development Program of China (863 Program) (2007AA01Z169)
1. 南京大学计算机软件新技术国家重点实验室 南京 210093
1. National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093
DOI: 10.3724/SP.J.1004.2008.00997

尽管年龄估计非常重要, 但至今只有很少的工作[2−4] 关注于自动年龄估计, 这可能是由于年龄数据难以收集所导致的. FG-NET 年龄数据库[5] 的出现从一定程度上缓解了数据收集的困难. 文献 [2] 中定义了年龄函数, 即根据提取的特征来得到年龄的函数, 然后提出了四种年龄函数来进行年龄估计. 本文介绍最常用的年龄函数 — 加权年龄函数 (Weighted aging function). 本文假设外貌相似的人可能变老的方式也相似, 即年龄函数也相似. 基于该假设, 首先使用每个人的人脸图像建立每个人的年龄函数, 对于新的人脸图像, 根据每个人的年龄函数估计出新的人脸图像的年龄, 最后将估计年龄按新图像与每个人的外貌相似度加权得到最终的估计值, 这种方法称为 WAS 方法. 文献 [3] 中使用两种分类方式来进行年龄估计, 具体来说, 即单步和层次式年龄估计方法. 在单步年龄估计方法中, 本文使用自组织神经网络[6] 等方法来进行年龄分类. 层次式年龄估计方法使用层次式分类方法, 包括三种类型的分类器. 本文主要介绍特定年龄和外观分类器 (Appearance-and age-specific classifiers), 因为这是文献 [3] 中实验结果最好的方法. 由于作者发现相貌相似的人随年龄变化相貌变化方式也类似[2], 此方法首先对训练样本根据相貌进行聚类, 然后使用每个簇中的样本训练一个特定年龄分类器. 特定年龄分类器首先使用所有样本训练一个全局年龄分类器, 再将这个年龄段分成若干个子年龄段, 分别对每一年龄段的样本训练一个局部年龄分类器. 此方法同时训练了一个簇选择分类器, 用于在测试时为测试样本选择合适的簇进行年龄估计. 这种方法简称为 AAS 方法. 文献 [3] 采用了一种类似于非监督学习中的方法来研究年龄估计问题. 文献 [3] 首先定义了年龄模式, 即同一人的人脸图像按年龄排列的序列, 然后根据训练样本使用主成分分析方法[7] 来构建一个包含数据的大部分变化的子空间, 由于训练集每个人或多或少会缺失部分年龄的样本, 同时使用了 EM 算法[8] 来估计缺失的样本. 在进行年龄估计时, 首先假设新人脸的年龄为任何可能的值, 通过年龄子空间来重构此人脸, 最后选择重构误差最小的所对应年龄作为估计结果.

本文首先利用文献 [2] 中的 WAS (Weighted appearance specific) 方法, 构造每个人的年龄函数, 然后进行年龄估计. 与 WAS 方法不同的是, WAS 方法使用遗传算法[9] 进行参数估计, 而我们使用典型相关分析 (Canonical correlation analysis, CCA)[10] 提取特征, 再利用最小二乘法进行参数估计. 此外在实际应用中, 数据库中可能没有关于某些图像是属于同一个人的这样的信息, 因此目前常见的方法均会失效, 这时年龄估计问题可视为一个分类问题. 文献 [3] 将年龄估计问题看成一普通的分类问题, 使用各种分类器进行年龄估计. 但我们认为年龄估计问题并非普通的分类问题, 而是代价敏感的分类问题. 本文进一步借助代价敏感学习中的算法来研究年龄估计问题. 最终将这两种方法采用集成学习中的手段结合起来以期能获得更好的性能.

本文第 1 节介绍我们提出的方法, 第 2 节为实验结果, 最后是总结.

1 年龄估计方法

1.1 个体算法

1.1.1 基于典型相关分析的年龄估计方法

典型相关分析是研究两组不同维数的变量之间的线性关系的一种统计分析方法. 对于两组变量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$, 典型相关分析主要是分别为 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 寻找线性变换, 使得 $\bar{\boldsymbol{x}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{x}}$ 和

$\bar{\boldsymbol{y}} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{w}_{\boldsymbol{y}}$ 之间相关性最大.

相关性系数定义如下

$$\rho = \frac{\mathrm{E}[\bar{\boldsymbol{x}}\bar{\boldsymbol{y}}]}{\sqrt{\mathrm{E}[\bar{\boldsymbol{x}}^2]\mathrm{E}[\bar{\boldsymbol{y}}^2]}} = \frac{\mathrm{E}[\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{w}_y]}{\sqrt{\mathrm{E}[\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_x]\mathrm{E}[\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{w}_y]}} = \frac{\boldsymbol{w}_x^{\mathrm{T}}C_{xy}\boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}}C_{xx}\boldsymbol{w}_x\boldsymbol{w}_y^{\mathrm{T}}C_{yy}\boldsymbol{w}_y}} \tag{1}$$

其中 C_{xx} 和 C_{yy} 分别是 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的自协方差矩阵, $C_{xy} = C_{yx}^{\mathrm{T}}$ 是 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的协方差矩阵. 这里的目标是最大化式 (1).

$\boldsymbol{w}_x$ 和 $\boldsymbol{w}_y$ 的求解过程如下: 首先对 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 中心化, 再定义 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的相关矩阵 C

$$C = \mathrm{E}[\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix}^{\mathrm{T}}] = \begin{bmatrix}C_{xx} & C_{xy}\\ C_{yx} & C_{yy}\end{bmatrix} \tag{2}$$

然后 $\boldsymbol{w}_x$ 和 $\boldsymbol{w}_y$ 可以通过对式 (1) 使用拉格朗日乘子法得到

$$\begin{cases} C_{xy}\boldsymbol{w}_y = \rho\lambda_x C_{xx}\boldsymbol{w}_x \\ C_{yx}\boldsymbol{w}_x = \rho\lambda_y C_{yy}\boldsymbol{w}_y \end{cases} \tag{3}$$

其中 $\lambda_x = \lambda_y^{-1} = \sqrt{\dfrac{\boldsymbol{w}_y^{\mathrm{T}}C_{yy}\boldsymbol{w}_y}{\boldsymbol{w}_x^{\mathrm{T}}C_{xx}\boldsymbol{w}_x}}$. 通过代入变换上述公式可以转换为式 (4)

$$\begin{cases} C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}\boldsymbol{w}_x = \rho^2\boldsymbol{w}_x \\ C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}\boldsymbol{w}_y = \rho^2\boldsymbol{w}_y \end{cases} \tag{4}$$

即 $\boldsymbol{w}_x$ 和 $\boldsymbol{w}_y$ 为相应矩阵中特征值为 ρ^2 时对应的特征向量. 求得 $\boldsymbol{w}_x$ 和 $\boldsymbol{w}_y$ 后, 用 $\bar{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_x$ 和 $\bar{y} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{w}_y$ 来进行处理.

在我们的方法中, 首先使用 CCA 对人脸表示进行处理, 其中人脸表示为 CCA 中的 $\boldsymbol{x}$ 变量, 年龄为 $\boldsymbol{y}$ 变量. 经过 CCA 处理后, 每幅人脸图像变成一维向量表示. 然后与文献 [2] 类似, 我们定义每个人的年龄方程 $f_i(x) = a_i x^2 + b_i x + c_i$, 其中 x 为人脸图像的一维表示, $f_i(x)$ 为此图像所对应的年龄, 接着我们使用最小二乘法来估计 a_i, b_i, c_i. 当对新的人脸图像 x_{new} 进行测试时, 估计的年龄为 $age_{new} = \sum_{i=1}^{n} p_i f_i(x_{new})$, 其中 n 为训练集中个体的数目, p_i 为 x_{new} 与第 i 个个体的相貌相似程度, 具体地, 根据文献 [11], 本文使用与第 i 个个体的所有图像之间的最短距离的倒数来定义 p_i.

这里使用 CCA 的好处有两点: 1) 可以提取和年龄相关的特征; 2) 由于每个个体的人脸图像数目比较少, 如果直接使用人脸特征来估计每个个体的年龄方程时, 由于未知数个数大于约束条件的个数, 所以这是个失定问题, 参数估计不太准确, 但经过 CCA 处理后, 每个人的年龄方程的估计变成过定问题, 参数估计更加准确, 这样更有利于年龄估计.

1.1.2 基于代价敏感学习的年龄估计方法

如果实际的年龄数据库中并没有包含哪些图像是属于同一个人的信息时, 则目前大部分年龄估计算法都会失效, 如文献 [2, 4] 中的算法. 这时年龄估计问题可以看成是一个分类问题[3]. 文献 [3] 中将年龄估计问题看成是一常规的分类问题, 使用各种分类算法来进行处理, 即认为将年龄为 i 岁的人预测成 j 岁和 k 岁 ($k \neq j$) 所引起的代价是一样, 实际上根据第 2.1 节的衡量标准可以看出年龄为 i 岁的人预测成 j 岁所引起的代价是 $|i-j|$, 而预测为 k 岁的代价是 $|i-k|$, 因此将年龄为 i 岁的人预测成 j 岁和 k 岁 ($k \neq j$) 所引起的代价是不一样的, 这说明了年龄估计问题不是常规的分类问题, 而是一种代价敏感的分类问题.

在标准的机器学习算法中通常关注的是错误率或正确率, 这隐含的假设就是不管在何种情况下错误分类带来的代价是相同的. 然而在实际应用中, 不同的错误分类情况通常有不同的代价. 比如在医疗诊断应用中, 将一个病人预测为健康者的代价远远大于反过来的情况, 因为前者的代价往往会是一个人的生命. 当分类问题具有代价敏感性时, 应当最小化总体代价而非错误率. 在代价敏感学习领域中, 错误分类代价是现实世界问题中最常见的, 这使得针对错误分类代价的代价敏感学习成为被最早研究并最受关注的一个子领域.

年龄估计问题中错误分类的代价很容易确定, 将 i 岁的样本错分成 j 岁的代价为 $|i-j|$, 如果写成代价矩阵的形式 $Cost_{C\times C}$, C 为类别数, $Cost(i,j)$ 表示将 j 岁的样本错分成 i 岁的代价, 则为一对称矩阵, 即年龄估计问题是一代价矩阵为一对称矩阵的多类代价敏感问题. 我们使用 LDA[12−13] + K-Nearest-Neighbour 分类器来进行处理, 这两种方法是人脸处理中最常用的方法之一. 但由于传统的 LDA 方法无法处理代价敏感问题, 而且我们实验中使用的数据集的样本分布不均衡 (这一点从图 2 中可以看出, 而不平衡的数据对 LDA 的性能有一定的影响[14]), 因此我们修改了类间散度和类内散度矩阵的定义, 用来改进 LDA 在处理代价敏感问题上的性能. 修改过的类间散度和类内散度矩阵的定义为

$$S_w = \sum_{k=1}^{C}\sum_{x_i\in\Pi_k} importance(k)(\boldsymbol{x}_i - \bar{\boldsymbol{m}}_k)(\boldsymbol{x}_i - \bar{\boldsymbol{m}}_k)^{\mathrm{T}} \tag{5}$$

$$S_b = \sum_{i=1}^{C}\sum_{i=1}^{C} Cost(i,j)(\bar{\boldsymbol{m}}_i - \bar{\boldsymbol{m}}_j)(\bar{\boldsymbol{m}}_i - \bar{\boldsymbol{m}}_j)^{\mathrm{T}} \tag{6}$$

其中数据集 $A = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, $\boldsymbol{x}_i \in \mathbf{R}^N (i = 1, \cdots, n)$, A 被分为 C 类, $A = \{\Pi_1, \cdots, \Pi_C\}$, $\bar{\boldsymbol{m}}_k = \dfrac{1}{n_k}\sum_{\boldsymbol{x}_i\in\Pi_k}\boldsymbol{x}_i$ 是第 k 类样本的均值, $importance(k) = \sum_{i=1}^{C} Cost(i,k)$ 表示第 k 类样本总的错分代价, 错分代价越高, 此类就越不能被错分, 其重要性也就越高. 最终用于降维的向量可由 $S_w^{-1}S_b$ 的特征向量得到. 这种方法称为 CSLDA 方法. 据我们所知, 这是第一个处理代价敏感学习问题的 LDA 方法.

目前多类代价敏感学习的通常做法是拆分多个两类代价敏感问题来解决[15], 但对于本文问题是不适用的, 这是因为本文问题如果拆成两类问题来解的话, 由于代价矩阵是对称的, 每个两类代价敏感问题均可等同于常规的分类问题, 这样就无法体现本文问题的代价敏感性. 因此我们提出了一种变体方法, 即每次随机选取 $t\,(t > 2)$ 个类的样本来使用 CSLDA 方法, 此过程重复 M 次 (这是由于全部列举 t 个类的组合数目过大, 计算存储开销过大), 最后将这 M 次方法的结果进行投票得到最终结果. 当然 CSLDA 方法也可以直接应用到多类代价敏感学习中, 但我们希望利用这种类似集成学习的方法来提高此方法的泛化能力. 本文称这种方法为 MCSLDA 方法. 在实验中取 t 为 5, M 为 2 000.

1.2 集成方法

集成学习 (Ensemble learning) 通过训练多个学习器来对一个问题进行求解, 可以显著地提高学习系统的泛化能力.

目前, 集成学习已成为当前机器学习领域的四个主要研究方向之一[16].

文献 [17] 指出当每个学习器的性能越好, 且学习器之间差异度越大, 则集成的性能就越好. 本文中的两种方法分别以回归和分类的角度来看待年龄估计问题, 所以这两种方法的差异度较大, 而且通过后面的实验结果可以看出这两种方法的性能较优, 所以我们通过平均法将两种方法结合起来, 希望能够得到性能更好的方法.

2　实验测试

2.1　实验设置

我们使用 FG-NET 年龄数据库进行实验. 此数据库中包含 82 个人的 1 002 幅人脸图像, 其中每个人有 6～18 幅标有年龄的图像. 数据库中的年龄变化范围在 0～69 岁之间, 一些典型的图像显示在图 1 中, 数据库中各年龄的样本数目见表 1, 从中可以看到数据库中仅有 91 幅年龄在 36～69 岁之间的图像, 所以我们使用年龄范围在 0～35 岁之间的数据上进行实验.



图 1　FG-NET 数据库中的典型图像

Fig. 1　Sample face images in the FG-NET database

表 1　FG-NET 数据库中各年龄样本数量分布

Table 1　The distributions of the sample amount belonging to each age in the FG-NET database

Age	Amount of samples	Age	Amount of samples	Age	Amount of samples	Age	Amount of samples
0	43	18	47	36	8	54	2
1	27	19	23	37	3	55	2
2	39	20	20	38	5	56	0
3	42	21	16	39	6	57	0
4	42	22	17	40	9	58	1
5	40	23	22	41	6	59	0
6	41	24	9	42	5	60	1
7	41	25	17	43	4	61	3
8	31	26	11	44	4	62	1
9	25	27	11	45	7	63	1
10	40	28	12	46	3	64	0
11	33	29	9	47	2	65	0
12	37	30	19	48	3	66	0
13	32	31	6	49	3	67	1
14	32	32	4	50	2	68	0
15	30	33	9	51	3	69	1
16	37	34	8	52	3		
17	28	35	11	53	2		

FG-NET 数据库中的人脸图像是在非受限的环境下拍摄得到的, 人脸图像的尺度、姿态及光照等因素变化较大, 这给年龄估计带来了很大的噪声. 仿照文献 [2−4], 实验使用主动外观模型 (Active appearance model)[18] 来提取人脸特征. 主动外观模型通过人脸图像上对应特征点对齐、人脸图像灰度的规范化等手段部分减少噪声, 从而使得年龄估计更加准确. 实验中主动外观模型需要 60 个模型参数来解释数据中 95 % 的变化.

所有方法采用 Leave-one-person-out (LOPO) 模式进行测试, 即每一轮, 同一个人的所有图像用来作为测试集, 其余作为训练集, 总共进行 82 轮测试. 之所以这样进行测试, 是因为这样更接近实际应用, 在实际应用中, 用来测试的图像的个体是数据库中所没有的.

本文采用文献 [4] 中的评价准则, 一种是平均绝对误差 (Mean absolute error, MAE), 即计算测试集中样本的估计年龄与真实年龄之差的绝对值的均值, 平均绝对误差越小, 算法的性能就越好; 另一种是累积指数 (Cum-score), 其定义为 $Cum\text{-}score\,(l) = M_{e\leq l}/M \times 100\,\%$, 其中 M 为测试样本总量, e 是误差绝对值, l 是指定的整型参数, $M_{e\leq l}$ 表示误差绝对值小于指定值的测试样本数目. 所以累积指数表示了估计误差在规定范围内的测试样本的比重, 在同一 l 下, 累积指数越大, 算法的性能就越好. l 的取值范围为 0～10. 之所以不显示更大的误差水平时各方法的性能, 是因为年龄估计中偏差 10 岁的估计是没有实际意义的.

本文用作对比的算法包括 WAS[2]、AAS[3] 和 AGES[4]. 对于 AGES, 年龄模式子空间的维数与文献 [4] 相同, 均设为 20. 注意通常年龄都是使用非负整数来表示, 而以上算法估计出的年龄往往是非负实数, 所以我们在实验中对这些算法的结果使用四舍五入方法进行处理.

2.2　实验结果

实验结果如表 2 和图 2 所示. 从表 2 可以看出本文方法的平均绝对误差要远远小于 WAS 和 AAS 的平均绝对误差, 也比 AGES 方法的平均绝对误差小. 从图 2 中可以看出本文方法在所有误差水平上均好于 WAS 和 AAS, 而且本文方法在除误差水平为 0 的所有其他误差水平上均优于 AGES 方法. 因此我们认为本文方法的性能比现有年龄估计方法略优.

表 2　四种年龄估计算法的平均绝对误差

Table 2　Mean absolutes error of four algorithms in age estimation

Method	MAE
WAS	6.5137
AAS	7.7622
AGES	4.2700
Our Method	**3.4863**

此外为了验证集成方法的有效性, 我们比较了两种个体方法 (分别记为 CCA 和 MCSLDA) 和集成后的方法的性能, 实验结果如表 3 和图 3 所示. 从表 3 可以看出集成后的方法的 MAE 均要小于 CCA 和 MCSLDA 方法的 MAE, 从图 3 中可以看出集成方法的累积指数在大部分误差水平上均优于两种个体方法 (由于比例的问题在图中可能看不出集成方法和 CCA 方法的差别, 但实际上集成方法的性能要好于 CCA 方法), 因此我们认为集成方法比较有效.

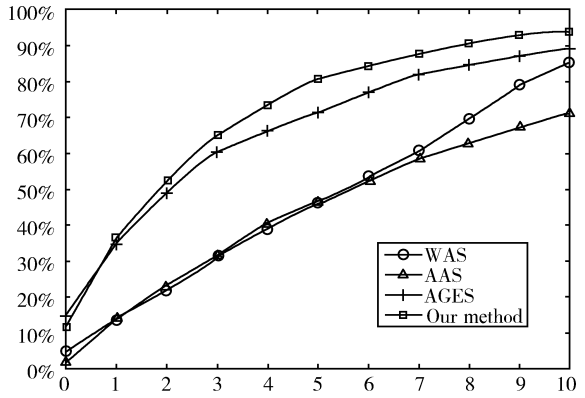


图 2 当误差水平在 0～10 之间, 四种年龄估计算法累积指数的比较

Fig. 2 Cumulative score of four algorithms in age estimation at error levels from 0 to 10 years

表 3 三种年龄估计算法的平均绝对误差

Table 3 Mean absolute errors of three algorithms in age estimation

Method	MAE
CCA	3.7168
MCSLDA	3.8891
Ensemble	**3.4863**

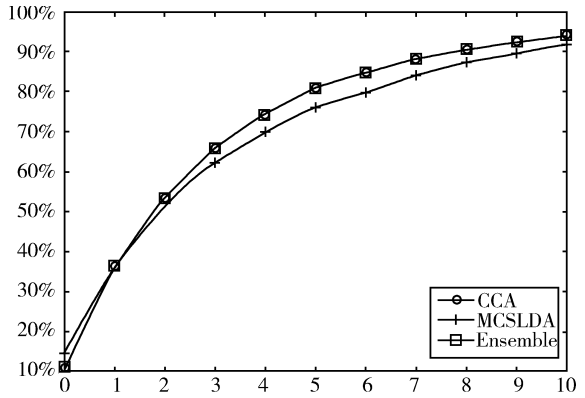


图 3 当误差水平在 0～10 之间, 三种年龄估计算法累积指数的比较

Fig. 3 Cumulative scores of three algorithms in age estimation at error levels from 0 to 10 years

3 小结

本文使用典型相关分析和代价敏感学习分别从回归及分类的角度来处理年龄估计问题, 并使用集成学习的方法将上述两种方法加以结合, 实验显示本文方法比现有的技术具有更好的性能. 本文中仅仅使用了简单的集成技术将两种算法进行结合, 能否采用其他的集成技术来达到更好的性能是一个值得研究的问题.

References

1 Zhao W Y, Chellappa R, Phillips P J, Rosenfeld A. Face recognition: a literature survey. *ACM Computing Survey*, 2003, **35**(4): 399−458

2 Lanitis A, Taylor C J, Cootes T. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(4): 442−455

3 Lanitis A, Draganova C, Christodoulou C. Comparing diffrent classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2004, **34**(1): 621−628

4 Geng X, Zhou Z H, Zhang Y, Li G, Dai H H. Learning from facial aging patterns for automatic age estimation. In: Proceedings of the 14th Annual ACM International Conference on Multimedia. Santa Barbara, USA: ACM, 2006. 307−316

5 Cootes T. The FG-NET aging database [Online], available: http://www.fgnet.rsunit.com, July 3, 2008

6 Kohonen T. *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988

7 Jolliffe I T. *Principle Component Analysis (Second Edition)*. New York: Springer-Verlag, 2002

8 Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, **39**(1): 1−38

9 Goldderg D E. *Genetic Algorithms in Search Optimization and Machine Learning*. Boston: Addison-Wesley Professional, 1989

10 Hotelling H. Relation between two sets of variants. *Biometrika*, 1936, **28**(12): 322−377

11 Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71−86

12 Fisher R A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, **7**: 179−188

13 Rao C R. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B*, 1948, **10**(2): 159−203

14 Xie J, Qiu Z D. The effect of imbalanced data sets on LDA: a theoretical and empirical analysis. *Pattern Recognition*, 2007, **40**(2): 557- 562

15 Zhou Z H, Liu X Y. On multi-class cost-sensitive learning. In: Proceedings of the 21st National Conference on Artificial Intelligence. Boston, USA: AAAI, 2006. 567−572

16 Dietterich T G. Machine learning research: four current directions. *AI Magazine*, 1998, **18**(4): 97−136

17 Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning. *Advances in Neural Information Processing Systems 7*. Cambridge: MIT Press, 1995. 231−238

18 Edwards G J, Taylor C J, Cootes T F. Interpreting face images using active appearance models. In: Proceedings of the 3rd International Conference on Face and Gesture Recognition. Freiburg, Germany: IEEE, 1998. 300

张 宇 香港科技大学计算机科学与工程系博士研究生. 2007 年于南京大学计算机科学与技术系获得硕士学位. 主要研究方向为机器学习, 数据挖掘和模式识别. E-mail: zhangyu@cse.ust.hk

(**ZHANG Yu** Ph. D. candidate in the Department of Computer Science and Engineering at Hong Kong University of Science and Technology. He received his master degree in the Department of Computer Science and Technology, Nanjing University in 2007. His research interest covers machine learning, data mining, and pattern recognition.)

周志华 博士, 南京大学教授. 主要研究方向为机器学习, 数据挖掘, 信息检索, 模式识别, 神经计算, 进化计算. 本文通信作者.
E-mail: zhouzh@nju.edu.cn

(**ZHOU Zhi-Hua** Ph. D., professor at Nanjing University. His research interest covers machine learning, data mining, information retrieve, pattern recognition, neural computation, and evolution computing. Corresponding author of this paper.)