

文章编号: 1000-5862(2019)01-0068-08

基于自适应多重多元回归的人脸年龄估计

曾雪强^{1,2}, 罗明珠¹, 陈素芬³, 吴水秀², 万中英²

(1. 南昌大学信息工程学院, 江西 南昌 330031; 2. 江西师范大学计算机信息工程学院, 江西 南昌 330022;

3. 南昌工程学院信息工程学院, 江西 南昌 330099)

摘要: 针对基于标记分布学习的多重多元回归模型不能生成和人脸老化趋势一致标记分布的问题, 提出自适应多重多元回归的人脸年龄估计方法. 在为不同年龄生成具有适合标准差的离散高斯分布的基础上, 采用偏最小二乘模型并有效地利用邻近年龄的人脸老化信息进行年龄估计. 在 MORPH 人脸数据库上的对比实验结果表明, 该文的人脸年龄估计模型具有更好的性能.

关键词: 人脸年龄估计; 自适应多重多元回归; 标记分布学习; 偏最小二乘

中图分类号: TP 391 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2019.01.12

0 引言

人脸图像包含许多与个人特性相关的生物学信息, 比如身份、性别、种族、年龄和情绪状态等. 在这些生物特征中, 年龄是一种具有重要实用价值的信息. 基于人脸图像的年龄估计在互联网接入控制、未成年香烟自动售货机使用、生物特征识别、安全控制和监控监测等领域具有广泛的应用前景^[1]. 基于人脸图像的年龄估计面临的困难主要来自于 2 个方面: (i) 人的遗传因子的多样性决定着不同的人具有不同的人脸特征变化模式; (ii) 人在成长过程中, 外在环境的多样性对人脸特征变化有着显著的影响. 因此, 如何开发出一个高性能的自动人脸年龄估计方法已经成为一个具有吸引力和挑战性的课题.

近年来学者们提出多种人脸年龄估计算法^[2-8], 但是依然没有出现一种精度足够高的人脸年龄估计模型. 人脸年龄图像的训练数据不足是导致人脸年龄估计模型性能不高的最主要原因之一^[3]. 这是因为很少有人有意识地在过去的任意年龄保留至少一张图像, 也没有人能够提前获取自己未来的图像. Geng Xin 等^[9]基于给每张人脸图像分配一个标记分布而不是单个的年龄标记的思想, 提出一种利用相近年龄信息的标记分布学习 (Label Distribution Learning, LDL) 方法来进行人脸年龄估

计. 每张人脸图像对应的标记分布覆盖其真实年龄附近的一些年龄, 每个标记值表示该年龄对真实年龄的描述度. 因此, 每张人脸图像不仅有助于其真实年龄的学习, 也有助于其附近年龄的学习. 在没有改变训练图像总数的情况下, 相当于每个年龄的训练样本得到增加, 从而缓解了训练数据不充足和不完整的问题. Geng Xin 等^[9]采用最大熵模型和神经网络提出了 2 种标记分布学习算法. 针对基于最大熵模型的标记分布学习为每个年龄建立单独模型的问题, 曾雪强等^[10]提出多重多元回归 (Multivariate Multiple Regression) 来解决标记分布学习问题, 并采用偏最小二乘 (Partial Least Squares, PLS)^[11-12]回归方法为所有年龄建立模型, 取得了优于基于最大熵模型的标记分布学习的实验效果.

然而, 在传统的标记分布学习中, 每张人脸图像的标记分布必须通过合理的假设和训练生成. 传统的标记分布学习, 假设人脸的老化速度在不同的年龄阶段是一致的, 为每张人脸图像生成统一标准差的离散高斯分布. 统一的标准差, 意味着对于任意一张人脸图像, 相同位置的附近年龄对于其真实年龄的学习贡献度是无差异的. 但是, 事实上人脸的老化速度在不同的阶段有着显著的差异. 从出生到成年, 主要的变化是颅面的生长^[13], 在这期间, 面部的尺寸越来越大, 颅骨的形状也逐渐改变. 从成年到老年, 主要的变化是皮肤的老化^[14], 这期间皮肤变暗

收稿日期: 2018-10-09

基金项目: 国家自然科学基金(61463033, 61866017), 江西省教育厅科学技术研究(GJJ150354)和江西省杰出青年人才资助计划(20171BCB23013)资助项目.

作者简介: 曾雪强(1978-), 男, 江西南昌人, 教授, 博士, 博士生导师, 主研从事机器学习研究. E-mail: xqzeng@jxnu.edu.cn

和变得粗糙,并且会出现皱纹.一般而言,在幼年时期和老年时期容貌的改变比在中年时期更加明显.以 FG-NET^[15] 人脸图像数据库中某个人为例,图1展示了这名男子在3个年龄间隔的脸部图像,分别是0~18岁、18~46岁和46~54岁,这3个年龄段大致对应于幼年、中年和老年3个时期.从图1可以看出,这名男子的脸部变化在0~18岁和46~54岁这2个年龄段比在18~46岁年龄段更加明显,这表明他的脸部在幼年时期和老年时期的老化速度比他在中年时期的老化速度快.因此,传统的标记分布学习方法为所有年龄设定统一的标准差是不合理的,如何生成和脸部老化趋势一致的标记分布是提高人脸年龄估计模型精度的关键.

针对传统的标记分布学习方法不能生成和人脸老化趋势一致的标记分布的问题,在基于多重多元回归的标记分布学习基础上提出标准差自适应的人脸年龄估计方法,即自适应多重多元回归(Adaptive Multivariate Multiple Regression).基于人脸老化的速度随年龄改变而不同的思想,本文提出的方法将为每张人脸图像根据不同的年龄自适应地学习对应的标记分布,即生成不同标准差的离散高斯分布.在采

用偏最小二乘回归建立的多重多元回归模型的基础上,引入标准差的自适应,使该模型可以更有效地学习附近年龄信息来进行人脸年龄估计.将自适应多重多元回归应用于人脸年龄估计问题,有望大幅度提高人脸年龄估计的精度.

另外, Geng Xin 等^[16] 曾经提出2种自适应标记分布学习(Adaptive Label Distribution Learning, ALDL)的人脸年龄估计算法,即基于最大熵模型的 IIS-ALDL 算法和基于拟牛顿法的 BFGS-ALDL 算法.这2种算法在人脸年龄估计问题上取得了较好的效果.然而, IIS-ALDL 算法和 IIS-LDL 算法一样,采用最大熵模型假设且为每个年龄建立单独模型,并通过多次迭代进行求解,导致算法的计算复杂度较高; BFGS-ALDL 算法和 BFGS-LDL 算法类似,在迭代过程中只计算目标函数的一阶梯度,导致算法存在收敛快和容易过拟合的问题.本文将在实验中将提出的算法与 IIS-ALDL 和 BFGS-ALDL 进行对比分析.同时,本文还将与曾雪强等^[17] 提出的采用偏最小二乘(PLS)降维和支持向量回归(SVR)年龄估计算法进行对比分析.

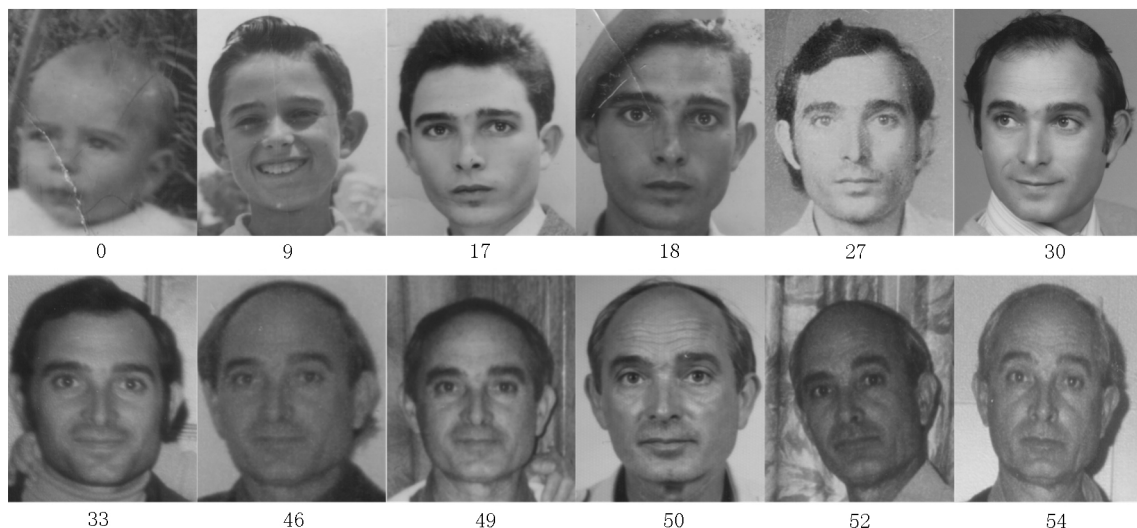


图1 FG-NET 人脸图像数据库中某个人在不同年龄的人脸图片(图片下的数字是其对应的年龄)

1 标记分布学习

1.1 标记分布

在实例 x 的标记分布^[19] 中,为 x 的每一个可能的标记 w 赋予一个实数 $y_{x,w} \in [0, 1]$,称为标记 w 对实例 x 的描述度.描述度表示用 w 描述 x 的程度,且实例 x 的所有可能标记的描述度之和为1,即 $\sum_w y_{x,w} = 1$,表示使用实例 x 的标记集合中的所有标记能够完

整描述该实例.

对于一幅真实年龄为 α 的人脸图像,一个合理的标记分布应该满足如下2个性质:(i) 真实年龄 α 的描述度应该是标记分布中最大的,确保真实年龄在类别描述中的主导地位;(ii) 其他年龄的描述度随着离 α 的距离增大而减小,确保距离真实年龄越近的年龄对类别描述的贡献越大.

标记分布需要满足的约束条件和概率分布一致,因此统计学中的理论和方法也适用于标记分布.根据 Geng Xin 等^[9] 的实验结果,本文选取以真实年

龄 α 为中心, 标准差为 σ 的离散高斯分布 $N(\alpha, \sigma^2)$ 作为人脸图像 x 的先验标记分布, 其函数形式表示为 $y_{x, w} = \frac{1}{\sigma \sqrt{\pi Z}} \exp(- (w - \alpha)^2 / (2\sigma^2))$, 其中 Z 是归一化因数, 即 $Z = \frac{1}{\sigma \sqrt{2\pi}} \sum_w (- (w - \alpha)^2 / (2\sigma^2))$.

图 2 给出了一个真实年龄为 α 的离散高斯分布示意图.

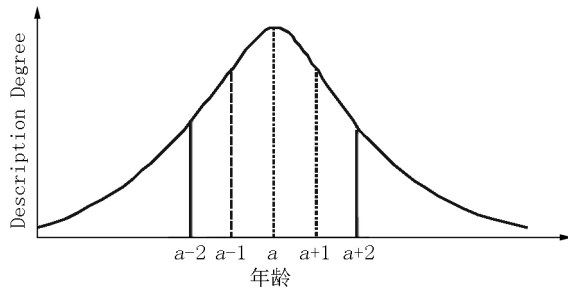


图 2 真实年龄为 α 的年龄标记高斯分布示意图

1.2 基于多重多元回归的标记分布学习

对年龄估计问题进行形式化定义. 令 $X \in \mathbf{R}^{n \times d}$ 表示图像特征空间 $w = \{w_1, w_2, \dots, w_c\}$ 表示可能年龄的有限集. 给定训练集 $S = \{(x_1, \alpha_1), (x_2, \alpha_2), \dots, (x_n, \alpha_n)\}$, 其中 x_i 是第 i 幅人脸图像的特征向量 $\alpha_i \in w$ 是 x_i 的真实年龄 $y_i = N(\alpha_i, \sigma^2)$ 是 x_i 的标记分布. 具体而言, 标记分布 y_i 由

$$y_{ij} = \frac{1}{\sigma \sqrt{2\pi Z}} \exp(- (w_j - \alpha_i)^2 / (2\sigma^2)) \quad (1)$$

生成, 其中 $j \in \{1, 2, \dots, c\}$.

将训练集 S 看作由 n 组 d 维自变量与 c 维因变量的观测数据组成, 曾雪强等^[10] 将标记分布学习问题转化为建立多组自变量和因变量之间的线性依赖关系, 即多重多元回归分析问题. 再采用多因变量偏最小二乘回归方法为所有年龄建立统一的年龄估计模型. 自变量矩阵 X 和因变量矩阵 Y 分别表示如下:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix},$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1c} \\ y_{21} & y_{22} & \cdots & y_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nc} \end{bmatrix}.$$

偏最小二乘是一种对多组变量进行建模的多元统计数据分析方法, 它既是一种特征提取(数据降

维), 又是一种回归模型 (PLS Regression, PLSR)^[12-13]. PLS 方法研究的焦点是通过抽取潜在变量(latent component)对包含多自变量和多因变量的数据进行建模. 在求解多重多元回归问题中^[20], PLS 方法的优化目标要求潜在变量之间的协方差(covariance)最大化.

给定 n 组样本, 以 X 和 Y 的第 1 个主成分向量 a_1 和 b_1 为例, PLS 的优化目标可以公式化为

$$\max_{a_1, b_1} \langle Xa_1, Yb_1 \rangle \quad \text{s. t.} \quad \|a_1\| = 1, \|b_1\| = 1,$$

其中, 通过引入拉格朗日乘子求出 a_1 和 b_1 . a_1 是对称矩阵 $X^T Y Y^T X$ 的最大特征值对应的特征向量 b_1 是对称矩阵 $Y^T X X^T Y$ 的最大特征值对应的特征向量.

根据线性变换 X 和 Y 的第 1 对潜在变量 u_1 和 v_1 写为 $u_1 = Xa_1$, $v_1 = Yb_1$.

求出第 1 对潜在变量 u_1 和 v_1 后, 可写出 X 和 Y 对潜在变量的回归方程 $X = u_1 p_1^T + E_1$, $Y = v_1 q_1^T + F_1$.

利用 u_1 和 v_1 之间的相关性, Y 可以改写成对 X 的潜在变量 u_1 进行回归建模: $Y = u_1 r_1^T + G_1$, 其中负载向量 $p_1 = X^T u_1 / \|u_1\|^2$, $q_1 = Y^T v_1 / \|v_1\|^2$, $r_1 = Y^T u_1 / \|u_1\|^2$. E_1 、 F_1 和 G_1 分别是对应回归方程的残差矩阵.

将残差矩阵 E_1 作为新的 X , F_1 作为新的 Y , 可以求解出第 2 对潜在变量 u_2 和 v_2 , 得到新的回归方程

$$E_1 = u_2 p_2^T + E_2, \quad F_1 = v_2 q_2^T + F_2.$$

如此重复计算下去, 最终的方法可以提取出 m 对潜在变量, 得

$$Y = u_1 r_1^T + u_2 r_2^T + \cdots + u_m r_m^T + E. \quad (2)$$

将 (2) 式用矩阵形式表示, PLS 模型最终建立 Y 关于 X 的回归方程为 $Y = UR^T + F = XAR^T + F = X\beta + F$, 其中 $\beta = AR^T$ 称为 Y 关于 X 的回归系数 F 为残差矩阵.

在计算得到 PLS 回归模型的回归系数 β 后, 可以对实例 x_i 预测出一个标记分布向量 $\hat{y}_i = x_i \beta$. 预测的 x_i 的年龄标记 $\hat{\alpha}_i$, 即标记分布 \hat{y}_i 中具有最大描述度的年龄 $\hat{\alpha}_i = \arg\max_{\hat{\alpha}_i} \hat{y}_i, \hat{\alpha}_i \in \{1, 2, \dots, n\}$.

2 自适应多重多元回归

由 (1) 式可知高斯分布的标准差 σ 控制着每个年龄上的标记分布的幅度. 传统的标记分布学习为所有的年龄设定统一的标准差, 意味着对于任意一

张人脸图像 相同位置的附近年龄对于其真实年龄的学习贡献度是无差异的. 然而人脸部老化的速度随年龄改变而不同,一方面,在容貌变化越快的年龄点上,越近的年龄信息对于其真实年龄学习的贡献越大,因而其高斯分布的标准差越小,生成更陡的标记分布;另一方面,在容貌变化越慢的年龄上,越远的年龄信息对于其真实年龄的学习也是有一定的贡献程度,因而其高斯分布的标准差越大,生成更平缓的标记分布. 因此,为所有的年龄设定统一的标准差是不合理的. 本文引入标准差的自适应来控制附近年龄信息对于其真实年龄学习的贡献程度.

基于人脸的老化速度是随年龄改变而不同的思想,提出自适应多重多元回归分析模型为不同的年龄设置不同的标准差,再采用偏最小二乘对该回归模型进行求解. 假设 $Y = X\beta$ 是一个回归模型,其中 β 为回归系数,训练集人脸图像的标记分布由(1)式定义. 给定训练集 S ,PLS-ALDL 算法的目标是: (1) 式中离散高斯分布的标准差 σ 要自适应于每一幅训练图像的真实年龄,即为每一个年龄分配一个不同的标准差 $\sigma_{\alpha_i}, \forall \alpha_i \in w$. 为了实现这一目标,PLS-ALDL 算法先根据初始化标准差生成初始标记分布,建立回归模型并得到预测的标记分布;然后用预测的标记分布拟合各年龄的标准差,再根据标准差更新标记分布,建立新的回归模型;最后交替迭代,互相更新,直至算法达到收敛. 具体而言,PLS-ALDL 算法的求解步骤如下:

(i) 生成年龄预测模型. 采用 PLS-ALDL 算法中第 $k-1$ 次迭代中为不同的年龄 α_i 计算得到的标准差 $\sigma_{\alpha_i}^{k-1}, \forall \alpha_i \in w$,为每张人脸图像 x_i 生成其对应的标记分布 $y_i^{k-1} = N(\alpha_i, (\sigma_{\alpha_i}^{k-1})^2)$. 标准差在第 1 次迭代时,统一初始化为一个固定值 $\sigma_{\alpha_i}^0 = \sigma, \forall \alpha_i \in w$,如 $\sigma = 20$.

然后建立 PLS 回归模型,得到因变量矩阵 Y 关于自变量矩阵 X 的回归方程,得到回归系数 β ,使得模型产生的预测标记分布与(1)式产生的标记分布 y_i^{k-1} 最相似.

回归模型 $Y = X\beta$ 求解完毕后,对于任意一张训练集中的人脸图像 x_i ,可以通过该回归模型得到对应的预测标记分布 $\hat{y}_i = x_i\beta$. x_i 的模型预测年龄为 $\hat{\alpha}_i = \arg\max \hat{y}_i$. 模型预测年龄的绝对误差为 $e_i = |\alpha_i - \hat{\alpha}_i|$,则整个训练集上的平均绝对误差(Mean

Absolute Error, MAE) $M_{AE} = \sum_{i=1}^n e_i/n$.

(ii) 更新标准差. 在所有预测标记分布的样本集中,挑选出绝对误差低于 M_{AE} 的样本作为拟合标准差 σ 的训练集. 因为这些训练样本具有更强的学习能力,适合用作求解标准差自适应问题的训练样本.

按照对应人脸图像的真实年龄把“学习能力强”的样本分为 c 组,意味着同组的训练样本具有相同的真实年龄,而相同的年龄则设定生成相同标准差的离散高斯分布. 设真实年龄属于年龄 α 的组中的所有图像的下标集合为 $I_\alpha = \{i \mid |\alpha_i - \hat{\alpha}_i| < M_{AE} \text{ 且 } \alpha_i = \alpha\}, i \in \{1, 2, \dots, n\}$. I_α 中的所有人脸图像的标记分布均由相同的标准差 σ_α 生成. I_α 中的每张人脸图像 $x_g, g \in I_\alpha$ 则与一个年龄标记分布 $N(\alpha_g, \sigma_\alpha^2)$ 相关联,即

$$y_{g,j} = \frac{1}{\sigma_\alpha \sqrt{2\pi Z}} \exp(-(w_j - \alpha_g)^2 / (2\sigma_\alpha^2)), \quad (3)$$

其中 $g \in I_\alpha, j \in \{1, 2, \dots, c\}$.

采用标准差 σ_α 为年龄 α_g 产生的离散高斯分布 $N(\alpha_g, \sigma_\alpha^2)$,应该与年龄回归模型预测的标记分布 \hat{y}_g 尽可能相似. 选择 KL 散度作为度量 2 个分布相似性的标准,则第 k 次迭代年龄 α 的最优标准差 σ_α 可以由如下优化问题进行求解得到:

$$\sigma_\alpha^k = \arg\min_{\sigma} \sum_{g \in I_\alpha} KL(\hat{y}_g \parallel N(\alpha_g, \sigma^2)). \quad (4)$$

将(3)式代入(4)式得到一个非线性优化问题,可以用对数屏障内点法进行求解^[21].

得到所有年龄的标准差 $\sigma_\alpha^k, \alpha \in w$ 后,送入算法的(i)步,算法进入第 $k+1$ 次外层迭代,更新整个训练集人脸图像对应的离散高斯分布,即标记分布 y_i^k ,再建立新的回归模型,得到新的预测标记分布. 重复上述过程,直至 2 次外层迭代对应的 M_{AE} 的差值小于给定的阈值 ε ,即 $|M_{AE}^{k-1} - M_{AE}^k| < \varepsilon$,PLS-ALDL 算法达到收敛. 本文提出的 PLS-ALDL 算法总结如下:

Algorithm 1 自适应多因变量 PLS 标记分布学习算法(PLS-ALDL)

Input: age set $w = \{w_1, w_2, \dots, w_c\}$, training set $S = \{(x_1, \alpha_1), (x_2, \alpha_2), \dots, (x_n, \alpha_n)\} (x_i \in X^{n \times d}, \alpha_i \in w)$, converge threshold ε , numbers of latent components m , maximum iterations L .

Output: the age predication model based on PLS regression.

1: $M_{AE}^0 = 100$;

```

2:  $\sigma_{w_j} = 20 \quad j \in \{1, 2, \dots, c\}$ ;
   //initialize standard deviation each age
3: for  $k = 1$  to  $l$ 
4:  $y_i = N(\alpha_i, \sigma_{\alpha_i}^2) \quad i \in \{1, 2, \dots, n\}$ ;
   //generate label distribution for each image
5: build PLS regression model on  $X$  and  $Y = \{y_1, y_2, \dots, y_n\}$ ;
6:  $\hat{y}_i = x_i \beta \quad i \in \{1, 2, \dots, n\}$ ;
   //predict label distribution by PLS regression model
7:  $\hat{\alpha}_i = \arg\max y_i \quad i \in \{1, 2, \dots, n\}$ ;
   //estimate the age of  $x_i$ 
8:  $M_{AE}^k = \frac{1}{n} \sum_{i=1}^n |\alpha_i - \hat{\alpha}_i|$ ;
9: if  $|M_{AE}^{k-1} - M_{AE}^k| < \varepsilon$  or  $k = l$ 
   //convergence condition
10: break;
11: end if
12: for  $h = 1$  to  $c$ 
   //update standard deviation for each age
13:  $I_{w_h} = \{i \mid |\alpha_i - \hat{\alpha}_i| < M_{AE}^l \ \&\& \ \alpha_i = w_h\}$ ,
    $i \in \{1, 2, \dots, n\}$ ;
14:  $\sigma_{w_h} = \arg\min_{\sigma} \sum_{g \in I_{w_h}} KL(\hat{y}_g \parallel N(\alpha_g, \sigma^2))$ ;
15: end for
16: end for
17: output the age predication model based on
   PLS regression.

```

3 实验结果及分析

3.1 实验及参数设置

本文采用 MORPH^[16] 数据库进行实验. MORPH 数据库是公开的人脸年龄数据库,一共有来自超过 13 000 人的 55 132 张人脸图像,年龄范围为 16 ~ 77 岁,平均每个人有 4 张人脸图像. MORPH 数据库是一个多种族数据库,其中 77% 为非洲人脸图像, 19% 为欧洲人脸图像,剩余 4% 的人脸图像来自西班牙、亚洲、印度和其他种族. 本文实验采用 BIF (Biologically Inspired Features, BIF)^[22] 特征对 MORPH 数据库中的数据进行特征提取,得到 4 096 维的

人脸特征,并对得到的特征采用 MFA (Marginal Fisher Analysis, MFA)^[23] 降维至 200 维.

为了说明本文提出的自适应多重多元回归方法的有效性,选取了以下 4 种人脸年龄估计方法进行对比实验,并考察潜在成分数量对模型性能的影响. PLS-LDL: 曾雪强等^[10] 采用的多因变量 PLS 标记分布学习的人脸年龄估计方法. IIS-ALDL: Geng Xin 等^[16] 提出的基于最大熵模型的自适应标记分布学习算法. BFGS-ALDL: Geng Xin 等^[16] 提出的基于拟牛顿法的自适应标记分布学习算法. PLS + SVR: 曾雪强等^[17] 提出的采用 PLS 降维和 SVR 进行年龄估计算法.

本文采取十折交叉验证的方法测试所有人脸年龄估计算法的性能,并保证同一个人的人脸图像都划分在同一折中. 在每次交叉验证中,PLS-ALDL 算法将训练集中的每一张人脸图像初始化为均值为其真实年龄、标准差为 20 的离散高斯分布,算法收敛阈值 ε 设为 0.02,提取的 PLS 潜在变量个数为 20. 本文测试了对比算法的多种参数配置,并选择其中的最优结果作为对比算法的实验结果. 具体而言, PLS-LDL 算法的潜在变量个数设置为 21,各个年龄的高斯分布的标准差统一定义为 20; IIS-ALDL 算法和 BFGS-ALDL 算法的初始化标准差设置为 3,收敛阈值为 0.02. PLS + SVR 算法采用 PLS 将特征维度降至 26 维. 本文将十折交叉验证的平均性能(平均绝对误差 M_{AE} ,累积得分 C_s)和各年龄自适应的标准差作为最终的实验结果. 累积得分 (Cumulative Score, C_s) 为 $C_s(j) = N_{e \leq j} / N \times 100\%$,其中 $N_{e \leq j}$ 为算法的估计年龄和对应人脸图像的真实年龄的绝对误差 e 小于等于 j 岁的人脸图像张数, N 为所有的人脸图像张数.

3.2 标记分布自适应迭代次数的影响分析

为了分析标记分布自适应迭代 (LD adaption iteration) 次数对年龄预测模型性能的影响,图 3 给出了 PLS-ALDL 算法、BFGS-ALDL 算法和 IIS-ALDL 算法中 M_{AE} 得分随着 LD 自适应迭代次数增加的变化情况. 图 3 中的 PLS-ALDL 算法 M_{AE} 曲线的起点即为 PLS-LDL 算法的 M_{AE} ,横坐标为标记分布自适应迭代次数. 从图 3 可以看出 PLS-ALDL 算法的 M_{AE} 变化曲线随着 LD 自适应迭代次数的增加,在开始的几次迭代中迅速下降,然后保持相对稳定,最后在第 8 次达到收敛; IIS-ALDL 算法的 M_{AE} 变化曲线在第 5 次达到收敛; BFGS-ALDL 算法的 M_{AE} 变化曲线在第

4次快速达到收敛.图3中 M_{AE} 变化曲线的对比表明,PLS-ALDL算法采用偏最小二乘法建立的自适应多重多元回归模型,随着标记分布自适应迭代次数的增加,在模型的稳定性和精确度方面都是表现最佳的.

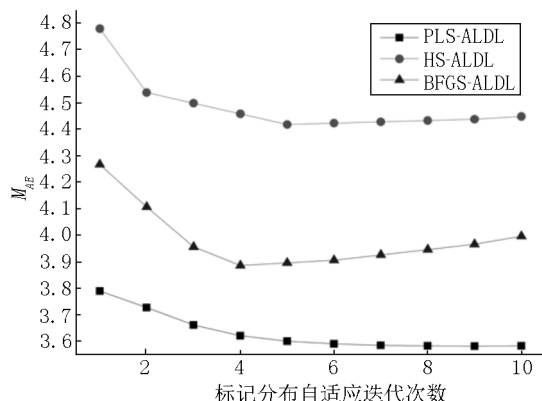


图3 在标记分布自适应迭代次数变化情况下的自适应年龄估计方法的 M_{AE} 得分

3.3 潜在变量个数对模型性能的影响

为了对比 PLS-ALDL 和 PLS-LDL 这 2 种年龄估计方法的性能,以及潜在变量个数 m 对模型性能的影响;将 m 从 1 增加到 25,并分别记录下 PLS-ALDL 和 PLS-LDL 对应的 M_{AE} ,详细实验结果如图 4 所示.

从图 4 可以看出,除了潜在变量为 1 和 2 时,采用自适应多因变量 PLS 标记分布学习的 PLS-ALDL 方法的性能优于传统标记分布学习的 PLS-LDL 方法.这一实验结果再次证明生成和人脸老化趋势一致的标记分布能更有效地提高人脸年龄估计的精度.对于最优的潜在变量个数,PLS-ALDL 在 $m = 20$ 时 M_{AE} 取得最小值,PLS-LDL 在 $m = 21$ 时 M_{AE} 取得最小值.这一结果显示采用偏最小二乘法建立的回归模型最佳的潜在变量个数在 20 附近.

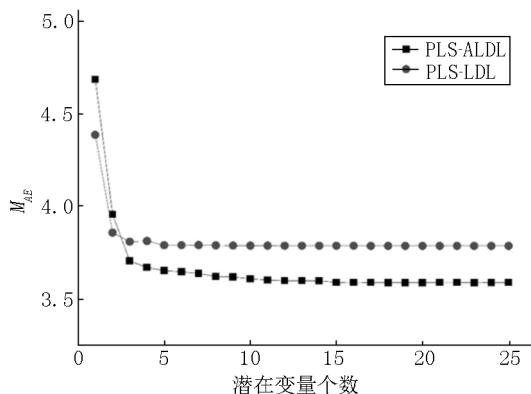


图4 在潜在成分数量变化情况下的 PLS-ALDL 和 PLS-LDL 的 M_{AE} 得分

3.4 各年龄标准差自适应调整的结果

图 5 展示了每个年龄 $\alpha \in w$ 的最终自适应标

准差 σ_{α} .图 5 中的每一个方块点表示一个年龄对应的标准差.标准差越大,人脸外观的变化越慢;标准差越小,人脸外观的变化越快.从图 5 可以看出,PLS-ALDL 算法的标准差曲线在 20 岁之前呈上升趋势,20~45 岁之间变化相对比较平缓,而 45 岁之后呈急速下降的趋势,这表明老年时期的容貌的改变比在中年时期更加明显.特别地,PLS-ALDL 算法的非自适应版本,即 PLS-LDL 算法,将各个年龄的高斯分布的标准差统一定义为 20,这一假定是无法准确地描述人脸外观随年龄改变而变化的过程.注意,由于大于 59 岁的训练样本很少,实验结果不能够具有说服力,所以图 5 中没有展示 59 岁之后的标准差情况.

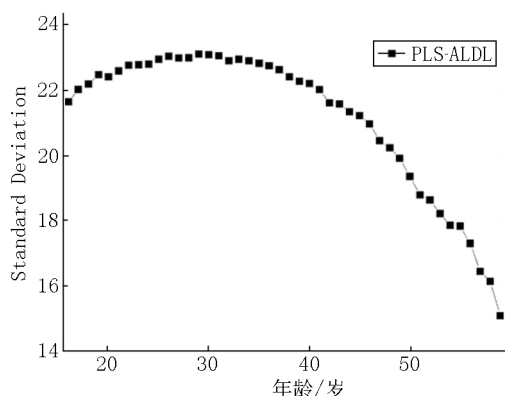


图5 在 MORPH 数据集上 PLS-ALDL 为各年龄自适应计算得到的高斯分布标准差

3.5 多种人脸年龄估计算法的预测性能对比

表 1 所示为各年龄估计算法在十折交叉验证实验后的 M_{AE} 以及训练时间(单位: s)(均值 \pm 标准差).从表 1 可以看出,PLS-ALDL 算法的 M_{AE} 为 3.58,低于其余 4 种方法,这个实验结果表明本文提出的 PLS-ALDL 算法是有效的.特别地,PLS-ALDL 算法的 M_{AE} 低于它的非自适应版本,即 PLS-LDL 算法显示了标记分布自适应迭代的有效性,表明标记分布的自适应过程可以根据人脸老化速度随年龄改变而不同的特点来学习对应年龄的高斯分布的标准差,说明生成和人脸老化趋势一致的标记分布是可以更有效地提高人脸年龄估计的精度.同样 PLS-ALDL 算法的 M_{AE} 显著低于 PLS + SVR 算法,表明 PLS-ALDL 建立的回归模型性能优于 PLS + SVR.然而就计算复杂度而言,标记分布自适应迭代过程会增加算法的计算复杂度. PLS-ALDL 算法的计算复杂度低于 IIS-ALDL 算法,但高于 BFGS-ALDL 算法,这说明 PLS-ALDL 算法在明显提高人脸年龄估计算法精度的同时,计算复杂度并没有大幅度增加.

表1 多种人脸年龄估计算法的 M_{AE} 对比结果

方法	M_{AE}	训练时间/s
PLS-ALDL	3.58 ± 0.03	$4\,878.11 \pm 743.60$
PLS-LDL	3.79 ± 0.03	3.89 ± 0.25
IIS-ALDL	4.42 ± 0.05	$7\,366.41 \pm 885.10$
BFGS-ALDL	3.89 ± 0.11	$2\,235.58 \pm 300.83$
PLS + SVR	4.39 ± 0.07	$1\,480.20 \pm 150.03$

图6为各年龄估计算法在0~10岁错误级上的CS曲线.从图6可看出,PLS-ALDL算法的CS曲线在所有错误级上均明显高于BFGS-ALDL算法;类似地,PLS-ALDL算法和IIS-ALDL算法的CS曲线之间的差距也很明显.实验结果表明PLS-ALDL算法在所有对比的人脸年龄估计算法中性能是最优秀的.

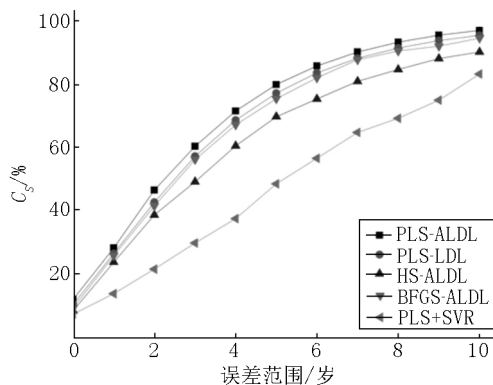


图6 多种人脸年龄估计算法的CS曲线对比结果

4 结束语

针对基于多重多元回归的人脸年龄估计方法不能够生成和脸部老化趋势一致的标记分布的问题,提出一种自适应多重多元回归的方法来实现人脸年龄估计.相对于已有的多重多元回归模型,该方法基于人脸老化速度随年龄的改变而不同的思想,为不同的年龄生成具有适当标准差的离散高斯分布,使模型可以更有效地利用邻近年龄信息进行年龄估计.在MORPH数据库上的人脸年龄估计算法的对比实验表明,本文提出的PLS-ALDL算法表现出最佳性能.解决标准差的自适应问题有助于提高算法的精度.下一步,希望考虑到性别和肤色识别等其它人脸图像信息对PLS-ALDL算法的影响,有望取得更好的人脸年龄预测性能.

5 参考文献

[1] Bekhouche S E, Ouafi A, Taleb-Ahmed A, et al. Facial age estimation using BSIF and LBP [EB/OL]. [2018-01-06]. https://www.researchgate.net/publication/270276061_Facial_age_estimation_using_BSIF_and_LBP.

Facial_age_estimation_using_BSIF_and_LBP.

- [2] Lanitis A, Taylor C J, Cootes T F. Toward automatic simulation of aging effects on face images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002, 24(4): 442-455.
- [3] Fu Yun, Huang T S. Human age estimation with regression on discriminative aging manifold [J]. IEEE Transactions on Multimedia 2008, 10(4): 578-584.
- [4] Guo Guodong, MuGuowang. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, Colorado, USA. IEEE 2011: 657-664.
- [5] Chao Weilun, Liu Junzuo, Ding Jianjun. Facial age estimation based on label-sensitive learning and age-oriented regression [J]. Pattern Recognition, 2013, 46(3): 628-641.
- [6] Rothe R, Timofte R, Van Gool L. Deep expectation of real and apparent age from a single image without facial landmarks [J]. International Journal of Computer Vision, 2018, 126(2/4): 144-157.
- [7] Liu Hao, Lu Jiwen, Feng Jianjiang, et al. Label-sensitive deep metric learning for facial age estimation [J]. IEEE Transactions on Information Forensics and Security 2018, 13(2): 292-305.
- [8] Sahoo T K, Banka H. Multi-feature-based facial age estimation using an incomplete facial aging database [J]. Arabian Journal for Science and Engineering, 2018: 43(12): 1-22.
- [9] Geng Xin, Yin Chao, Zhou Zhihua. Facial age estimation by learning from label distributions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013, 35(10): 2401-2412.
- [10] Zeng Xueqiang, Run Xiang, Zou Huaxing. Partial least squares regression based facial age estimation [C]//IEEE International Conference on Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), Guangzhou, China. IEEE, 2017, 1: 416-421.
- [11] Singer M, Krivobokova T, Munk A, et al. Partial least squares for dependent data [J]. Biometrika, 2016, 103(2): 351-362.
- [12] Stott A E, Kanna S, Mandic D P. Widely linear complex partial least squares for latent subspace regression [J]. Signal Processing 2018, 152: 350-362.
- [13] Ramanathan N, Chellappa R. Modeling age progression in young faces [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA. IEEE 2006: 387-394.

- [14] Albert A M ,Ricanek Jr K ,Patterson E. A review of the literature on the aging adult skull and face: implications for forensic science research and applications [J]. Forensic Science International 2007 ,172(1) : 1-9.
- [15] Panis G ,Lanitis A ,Tsapatsoulis N ,et al. Overview of research on facial ageing using the FG-NET ageing database [J]. Iet Biometrics 2016 ,5(2) : 37-46.
- [16] Geng Xin ,Wang Qin ,Xia Yu. Facial age estimation by adaptive label distribution learning [C] // International Conference on Pattern Recognition (ICPR) ,Stockholm ,Sweden IEEE 2014: 4465-4470.
- [17] 曾雪强 赵丙娟 向润 等. 基于偏最小二乘的人脸年龄估计 [J]. 南昌大学学报: 工科版 2017 ,39(4) : 380-385.
- [18] Ricanek K ,Tesafaye T. Morph: a longitudinal image database of normal adult age-progression [C] // International Conference on Automatic Face and Gesture Recognition (FGR) ,Southampton ,UK IEEE 2006: 341-345.
- [19] Geng Xin ,Ji R. Label distribution learning [J]. IEEE Transactions on Knowledge and Data Engineering ,2016 ,28(7) : 1734-1748.
- [20] Hartung J ,Knapp G. Multivariate multiple regression [M]. Wiley StatsRef: Statistics Reference Online 2014.
- [21] Waltz R A ,Morales J L ,Nocedal J ,et al. An interior algorithm for nonlinear optimization that combines line search and trust region steps [J]. Mathematical Programming ,2006 ,107(3) : 391-408.
- [22] Guo Guodong ,Mu Guowang ,Fu Yun ,et al. Human age estimation using bio-inspired features [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ,Miami ,Florida ,USA IEEE 2009: 112-119.
- [23] Lan YuanDong ,Deng Huifang ,Chen Tao. Dimensionality reduction based on neighborhood preserving and marginal discriminant embedding [J]. Procedia Engineering 2012 ,29(4) : 494-498.

The Facial Age Estimation Based on Adaptive Multivariate Multiple Regression

ZENG Xueqiang^{1 2} ,LUO Mingzhu¹ ,CHEN Sufen³ ,WU Shuixiu² ,WAN Zhongying²

(1. Information Engineering School ,Nanchang University ,Nanchang Jiangxi 330031 ,China;

2. School of Computer and Information Engineering ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China;

3. School of Information Engineering ,Nanchang Institute of Technology ,Nanchang Jiangxi 330099 ,China)

Abstract: In order to address the problem that traditional multivariate multiple regression based label distribution learning methods cannot generate the label distribution according to the tendency of facial aging ,a facial age estimation method based on adaptive multivariate multiple regression has been proposed. The proposed method generates the discrete Gaussian distributions with different standard deviations adapted to different ages ,and using partial least square model to effectively utilize adjacent facial ageing information to predict facial age. Our experimental results on the MORPH face database show that the facial age estimation model in the paper has better performance than existing correlation models.

Key words: facial age estimation; adaptive multivariate multiple regression; label distribution learning; partial least square

(责任编辑: 冉小晓)