

基于贝叶斯优化算法的脸面特征向量量子集选择^{*}

Eigenvector Subset Selection Using Bayesian Optimization Algorithm

郭卫锋 林亚平 罗光平

(湖南大学计算机系 长沙410082)

Abstract Eigenvector subset selection is the key to face recognition. In this paper, we propose ESS-BOA, a new randomized, population-based evolutionary algorithm which deals with the Eigenvector Subset Selection (ESS) problem on face recognition application. In ESS-BOA, the ESS problem, stated as a search problem, uses the Bayesian Optimization Algorithm (BOA) as searching engine and the distance degree as the object function to select eigenvector. Experimental results show that ESS-BOA outperforms the traditional the eigenface selection algorithm.

Keywords Eigenvector subset selection, Face recognition, Distance degree, Bayesian optimization algorithm

1 引言

近年来,人脸识别作为图像分析和理解最成功的应用之一,受到研究人员广泛的关注^[1]。脸部特征抽取选择是人脸识别关键问题之一^[1,2]。现有的脸部特征抽取主要有两类,一类是基于局部几何特征的系统^[3],一类是基于整体模版匹配的系统^[4]。在基于局部几何特征的系统中,通过检测眼睛,鼻子,嘴等面部特征和它们之间的相互关系(距离,面积,角度)来描述脸面。这种方法可以有效降低数据量,但是目前脸面特征的检测和测量技术还不能满足这种方法的要求^[5]。基于整体模版匹配的系统直接用基于外貌的特征表示脸面。与前一种方法相对比,这种方法不需要对脸面特征进行检测和测量,所以更实用可靠^[6]。

基于外貌的特征表达方法通常用 $n \times m$ 维的向量表达 $n \times m$ 个像素的脸面图像。在实际应用中,由于 $n \times m$ 的向量空间的庞大性,不适应人脸识别快速,健壮的要求。目前一般使用降维技术来解决这个问题。最常用的降维方法是主成分分析法(Principle Components Analysis,简称PCA)^[7]。PCA(也称为K-L变换法)根据最小均方差准则选择特征向量。这种选择方法最大化所有训练样本间的离散度,在图像重新组建中表现了良好的性能,但是在每一个图像类有多个训练样本的情况下,不能有效区分类内的变化和类间的变化^[8,9]。因此本文引入距离度量准则^[10]选择特征量。该准则区别考虑类间离散度和类内离散度,使同类相聚,异类相离,能够有效区分类内的变化和类间的变化,识别不同的类。

利用距离度量准则选择特征向量属于空间搜索问题。搜索空间的每一个状态对应识别特征向量的一个可能的子空间。由于计算量的庞大,穷尽所有可能的特征子空间进行评价是不可能的。本文提出一种新的特征向量选择算法即基于贝叶斯优化算法的特征向量量子集选择算法。这种算法以距离度量为目标函数,结合贝叶斯优化算法^[11](Bayesian Optimization Algorithm,简称BOA)作为搜索引擎,优化选择特征向量。实验结果表明,利用该算法进行特征向量的选择和传统的最大特征值选择法相比较,可以获得更高的识别率。

2 PCA 和 BOA

2.1 PCA 方法

考虑 n 维图像空间的 N 个样本脸面图像 $\{x_1, x_2, \dots, x_N\}$, 每个样本属于 c 类 $\{X_1, X_2, \dots, X_c\}$ 中的一类。考虑将 n 维特征空间投影到 m 维特征空间的线性变换。用 $W_{n \times m}$ 表示列向量正交的矩阵。变换后的新的向量 $y_k \in R^m$, 由下式线性变换定义:

$$y_k = W^T x_k, k=1, 2, \dots, N$$

总散度矩阵 S_T :

$$S_T = \sum_{i=1}^{N-1} (x_i - \mu)(x_i - \mu)^T$$

其中 $\mu = \frac{1}{N} \sum_{k=1}^N x_k$ 是所有样本的平均值。经过线性变换得到的新的特征向量 $\{y_1, y_2, \dots, y_N\}$ 的离散度是 $W^T S_T W$ 。PCA 选择最优的投影变换 W_{opt} :

$$W_{opt} = \arg \max_W |W^T S_T W| = [\omega_1, \omega_2, \dots, \omega_m]$$

$\{\omega_i | i=1, 2, \dots, m\}$ 是 S_T 按照降序排列的前 m 个特征值对应的特征向量。由于这些向量和原始的图像有相同的维数,所以这些向量被称为特征脸^[12],也称PCA为特征脸法(eigenface-based method)。这种选择方法在图像重新组建中表现了良好的性能,但只适应于每个类别只有单一样本的情况。PCA的缺点是它最大化的总离散度不仅仅是由类间离散度引起的,同时也是由类内离散度引起的。类间离散度提供有用的分类信息,而类内离散度则会干扰分类。图像和图像之间的变化经常是由于光线,角度而引起的,就是说同一类的图像也会因为光线,角度的变化而变化,这种变化就属于类内离散度问题。这样由于类内离散度的影响, W_{opt} 投影变换保留的主成分信息反映的是光线或角度的变化,而不是分类信息。对应的投影点在特征空间中不能很好地聚类,类与类之间还可能混杂在一起。

2.2 BOA

BOA 是一种基于贝叶斯概率模型的进化算法。贝叶斯网络是一无环有向图,它既可以对所给的数据进行描述,又可以产生与所给定数据性质相同的数据,因此常用于对离散或

^{*} 本文得到国家教育部科学技术重点研究项目(99092)资助。郭卫锋 硕士研究生,研究方向为人脸识别。林亚平 教授,博士生导师,主要研究方向为网络计算和机器学习。

硕士研究生,研究方向为人脸识别。林亚平 教授,博士生导师,主要研究

连续变量的多项式数据建模。利用贝叶斯网络求解问题的关键在于学习网络,学习网络指的是找到一个网络,使之在约定的评价标准下,最好地匹配现有训练数据集。网络的评价标准说明网络所分析的数据的好坏程度,也就是对网络质量的评价;搜索过程则是要寻找具有最高评价标准值的网络。BOA 算法利用 Bayesian Dirichlet 评价标准和贪婪算法建立和搜索网络。研究表明,BOA 算法求解许多测试函数,包含高阶构造快问题都得到了很好的优化结果。本文采用 BOA 算法作为优化算法。有关 BOA 算法的细节参见文[11]。

3 基于贝叶斯优化算法的特征向量子集选择

一个胜任的脸面识别系统应该有较强的适应性,性能应该不受脸面角度、光线及表情的显著变化而下降^[14]。现有的识别体系都采用每类多个训练样本^[15]的方法解决这个问题。如前所述特征脸法适应于每个类别只有单一样本的情况,而且脸面识别信息不仅存在于特征值大的特征向量,同样也存在于特征值小的特征向量^[16]。也有学者提出样本按照角度分类,分别建立不同的子特征空间,这种方法不能适应光线、表情的变化^[17]。因此本文引入距离度量准则优化选择特征向量。距离度量准则定义如下:

$$J = \text{tr}(S_w + S_B)$$

其中 S_B 是类间离散度矩阵, S_w 是类内离散度矩阵。

$$S_B = \sum_{i=1}^c |\lambda_i| (\mu_i - \mu)(\mu_i - \mu)^T$$

$$S_w = \sum_{i=1}^c \sum_{x_k \in \lambda_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

λ_i 是对应每一类的先验概率, μ_i 为类别 i 的样本平均值。评价

准则同时考虑了类内、类间离散度。利用该准则选择 m 个特征向量作为正交变换矩阵 $W_{n \times m}$ 的列向量。变换后的新的向量 $z_k \in R^m$, 由下式线性变换定义:

$$z_k = W^T x_k, k=1, 2, \dots, N$$

变换后的样本点在对应的特征空间具有同类相聚, 异类相离的特点。

根据距离度量 J 准则, 进行特征向量的选择是一个搜索问题。本文提出了基于贝叶斯优化算法的特征向量子集选择法, 具体算法描述如下:

- (1) 用 PCA 变换产生训练集的特征向量;
- (2) 将训练集按照特征向量投影;
- (3) $t=0$, 随机生成初始群体 $P(0)$;
- (4) 从 $P(t)$ 选择较好的子集 $S(t)$;
- (5) 按照选定的搜索方法和评价图的准则组建贝叶斯图 B ;
- (6) 按照 B 编码的联合分布概率产生新的字符串集 $O(t)$;
- (7) 按照适应度函数就是距离度量 J 用 $O(t)$ 替代部分 $P(t)$ 生成新的群体 $P(t+1)$;
- (8) 如果没达到终止条件, 转到(4)。

4 实验

本文采用权威的 FERET(The Facial Recognition Technology) 数据库^[18]进行实验。2001年 FERET 发布的 CD 包含 10451个8位的灰度脸面图。用 FERET 提供的预处理程序对图像进行标准预处理, 处理后的图像为 150×130 个像素的原始标准图像, 如图1所示。



图1 标准图

实验中建立的训练集为 training460, 库集 gallery46, 测试集 probe46。Training460是从 FERET 中选择的460个脸面图, 这460个脸面图是46个人在不同的光线、角度、时间或表情下拍摄的图像, 每人10张图像。从 training460中随机选择, 每个人选一张脸面图分别构成 Gallery46, Probe46。本文主要作了以下实验:

1) 特征脸选择法。设特征向量选择数和训练集脸面数的百分比为 cutoff, 分别做 cutoff=10, cutoff=40两个实验。

2) 基于贝叶斯优化算法的特征向量子集选择法。同样做两个实验 cutoff=10, cutoff=40。

文[12]中指出训练集的脸面图像数为115, 特征向量数为40就足够表达脸面特征了。也就是 cutoff=40。FERET 1996/97的研究中也是用的这种选择尺度。为了便于比较, 本文分别做了 cutoff=40, cutoff=10的实验。实验结果是脸面的识别率。具体实验数据见表1。

实验采用的分类算法是 K 临近算法^[19], 表1中的 Rank 是指采用 K 临近分类方法的 K 值。从表1可以得到以下结论:

1) 基于贝叶斯优化算法的特征向量子集选择法的性能优于特征脸法。

2) 对于训练集的脸面图像数 $N=460$, 特征向量数 $m=46$ 也能够很好地表达特征。也就是说对于大型脸面训练库, 同样

不需要更多的特征数表达脸面。

表1 特征脸和 ESS-BOA 算法比较

识别率 %	特征脸 cutoff=10	特征脸 cutoff=40	ESS-BOA cutoff=10	ESS-BOA cutoff=40
Rank1	63.04	58.70	65.22	60.87
Rank2	71.71	65.22	71.74	67.39
Rank3	71.74	65.22	71.74	69.57
Rank4	73.91	69.57	76.09	71.74
Rank5	76.09	71.74	76.09	73.91
Rank6	82.61	73.91	82.61	78.26
Rank7	82.61	84.78	82.61	80.43
Rank8	84.78	86.69	84.78	89.13
Rank9	89.13	86.69	91.30	91.30
Rank10	91.30	89.13	93.48	91.30

结论 本文针对 PCA 变化的特征向量选择, 以新的评判准则提出新的选择算法即基于贝叶斯优化算法的特征向量子集选择。该算法以距离度量准则为适应度函数, 将贝叶斯优化算法引入进行优化特征向量选择。实验表明该算法优于以往的特征脸算法。由于特征脸法同时是其他脸面识别方法的基础, 所以本文的算法一样既可以直接用于脸面识别, 也可以作

(下转第194页)

例2 设 $G=PSL(2,17)$, 由上述算法求得 $\pi_r(G)=\{1,2,3,4,8,9,17\}$,

$x_1=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ 对应1阶元 Dx_1 , $x_2=\begin{pmatrix} 0 & 1 \\ 16 & 0 \end{pmatrix}$ 对应2阶元 Dx_2 , $x_3=\begin{pmatrix} 0 & 1 \\ 16 & 1 \end{pmatrix}$ 对应3阶元 Dx_3 , $x_4=\begin{pmatrix} 0 & 1 \\ 16 & 6 \end{pmatrix}$ 对应4阶元 Dx_4 , $x_5=\begin{pmatrix} 0 & 1 \\ 16 & 5 \end{pmatrix}$ 对应8阶元 Dx_5 , $x_6=\begin{pmatrix} 0 & 1 \\ 16 & 3 \end{pmatrix}$ 对应9阶元 Dx_6 , $x_7=\begin{pmatrix} 0 & 1 \\ 16 & 2 \end{pmatrix}$ 对应17阶元 Dx_7 . 其中 $D=\{E_1, E_9\}$, 而 $E_1=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ 及 $E_9=\begin{pmatrix} 16 & 0 \\ 0 & 16 \end{pmatrix}$.

各阶元个数如下:

1阶元1个; 2阶元153个; 3阶元272个; 4阶元306个; 8阶元612个; 9阶元816个; 17阶元288个; 群的阶为2448.

4 复杂度分析

本算法的输入尺寸由 q 决定, 算法执行的运算次数取决于 q 的大小. 下面只对在最坏情况下的时间复杂度作粗略的估计.

在过程一中, 求一个二级矩阵的行列式要2次乘法, 主程序中由于二级矩阵的4个元分别从0到 $q-1$ 循环, 共调用过程一 q^4 次, 共计算了 $2q^4$ 次乘法.

特殊线性群 $SL(2, q)$ 的阶为 $q(q^2-1)$, 说明其中行列式为1的矩阵 x 有 $q(q^2-1)$ 个, 每个这样的矩阵都要调用过程二求 x 的阶. 在求阶过程中每个 r 阶元 x 要进行矩阵乘法 $r-1$ 次, 每次矩阵乘法要进行8次数的乘法, 在最坏情况下 r 取 q

(q^2-1) (而在平均情况下 r 取 q), 因此最坏要进行 $8q(q^2-1)$ 次乘法. 此外每个矩阵 x 还需进行 $2r$ 次判断与 E_1 及 E_q 是否相等, 每判断一次要进行4次数的比较, 最坏时要进行 $8q^2(q^2-1)^2$ 次数的比较.

在过程三中进行统计时, 还要多次对数组 $paieg()$ 进行搜索, 由搜索算法^[5]知, 在最坏的情况下每搜索一次的运算量为 N , 其中 N 为数组 $paieg()$ 的大小, 从而总的搜索次数 $W(N) < q^2(q^2-1)^2 < q^6$.

综上所述, 整个计算过程中的工作量在最坏的情况下也不超过

$$2q^4 + q(q^2-1)[8q(q^2-1) + 8q^2(q^2-1)^2 + q(q^2-1)] < 10q^9.$$

可见, 其工作量是与输入尺寸 q 成多项式增长的.

参考文献

- 1 施武杰. 元的阶给定的有限群[J]. 科学通报, 1997, (11): 1702~1706
- 2 施武杰. A5的一个特征性质[J]. 西南师范学院学报(自然科学版), 1986, (3): 11~14
- 3 Lipschutz S, Shi WujiE. FINITE GROUPS WHOSE ELEMENT ORDERS DO NOT EXCEED TWENTY [J]. PROGRESS IN NATURAL SCIENCE. 2000
- 4 Shi wujiE. A CHARACTERISTIC PROPERTY OF $PSL_2(7)$ [J]. J. AusTral. Math. (SeriEs A0 36 (1984). 354~356
- 5 徐士良. 计算机常用算法[M]. 北京: 清华大学出版社, 1989. 43~47
- 6 Proc. European Conf. on Computer Vision, April 1996
- 7 Swets D, Weng J. Hierarchical Discriminant Analysis for Image Retrieval. IEEE PAMI, 1999, 21(5)
- 8 Bian Zhao-Qi, Zhang Xue-Gong, et al. Pattern Recognition. Beijing: Publishing House of QingHua University, 2000 (in Chinese)
- 9 Pelikan M, Goleberg D E, Cauntr-Paz E. BOA: The Bayesian optimization algorithm. GECCO-99. In: Proc. of the Genetic and Evolutionary Computation Conf. CA, Morgan Kaufmann, 1999. 525~532
- 10 Turk M A, Pentland A P. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991, 3(1): 71~86
- 11 Moses Y, Adimi Y, Ullman S. Face recognition: The problem of compensating for changes in illumination direction. In: Proc. of the European Conf. on Computer Vision, 1994, A. 286~296
- 12 Sandy A, Pentland, Choudhury T. Face Recognition for Smart Environments. IEEE Computer, 2000, 33(2): 34~36
- 13 Li S Z, Lu Juwei. Face Recognition Using the Nearest Feature Line Method. IEEE Transactions on Neural Networks, 1999
- 14 O'Toole A, Abdi H, Deffenbacher K A, Valentin D. Low-dimensional representation of faces in higher dimensions of the face space. Journal of The Optical society of America, 1993, A(10): 405~411
- 15 Pentland A, Moghaddam B, Starner T. View-based and modular eigenspaces for face recognition. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition CVPR 94, (Seattle, Washington), 1994. 84~91
- 16 Phillips P J, Wechsler H, Huang J, Rauss P. The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing, 1998, 16(5): 295~306
- 17 Cover T M, Hart P E. Nearest Neighbor Pattern Classification. IEEE Transactions on Information Theory, 1967, IT-13(1): 21~27
- 18 Zhao W, Chellappa P, Rosenfeld A. Face recognition: A Literature Survey. ACM Journal of Computing Surveys, 2000
- 19 Zhao W, Krishnaswamy A, Chellappa R, Swets D, Weng J. Discriminant analysis of principal components for face recognition. In: Wechsler, Philips, Bruce, Fogelman-Soulie, Huang, eds. Face Recognition: From Theory to Applications, Springer-Verlag, 1998. 73~85
- 20 Brunelli R, Poggio T. Face recognition: Features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15: 1042~1052
- 21 Chellappa R, Wilson C L, Sirohey S. Human and Machine Recognition of Faces, A Survey. Proc. of the IEEE, 1995, 83: 705~740
- 22 Cox I J, Ghosn J, Yianilos P N. Feature-based face recognition using mixture-distance. In: Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 1996. 209~216
- 23 Chellappa R, Wilson C L, Sirohey S. Human and machine recognition of faces: A survey. PIEEE, 1995, 83: 705~740
- 24 Kirby M, Sirovich L. Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. Pattern Analysis and Machine Intelligence, 1990, 12(1)
- 25 Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisher-faces: Recognition using class specific linear projection. In:

(上接第163页)

为其他脸面识别算法的基础. 同时本文实验还发现对于大型脸面训练库, 同样不需要更多的特征数表达脸面.

致谢 作者感谢 J. Ross Beveridge 对本文实验的帮助, 感谢 Defense Advanced Research Products Agency (DARPA) 为本文的研究提供的 FERET 脸面数据库及标准化程序.

参考文献