

基于双曲余弦矩阵鉴别分析的小样本问题研究*

冉瑞生^{1a}, 张守贵^{1b}, 任银山², 房斌^{3†}

(1. 重庆师范大学 a. 计算机与信息科学学院; b. 数学科学学院, 重庆 401331; 2. 哈工大机器人集团重庆云安科技有限公司, 重庆 400039; 3. 重庆大学 计算机学院, 重庆 400044)

摘要: 线性判别分析(LDA)是模式识别领域的一个经典方法,但是LDA难以克服小样本问题。针对LDA的小样本问题,提出一种双曲余弦矩阵鉴别分析方法(HCDA)。该方法首先给出了双曲余弦矩阵函数的定义及其特征系统,再利用双曲余弦矩阵函数特征系统的特点,将其引入Fisher准则中进行特征提取。HCDA有两方面的优势:a)避免了小样本问题,可以提取更多的鉴别信息;b)HCDA方法隐含了一个非线性映射。该映射具有扩大样本间距离的作用,并且对不同类别样本间距离的扩大尺度要大于同类样本间距离的扩大尺度,从而更有利于模式分类。在手写数字库、手写字母图像库和Georgia Tech人脸图像库上的实验结果表明,相对于具有代表性的解决LDA小样本问题的方法,HCDA具有更好的识别性能。

关键词: 双曲余弦函数; 矩阵函数; 线性判别分析; 小样本问题

中图分类号: TP391.4

文献标志码: A

文章编号: 1001-3695(2020)08-059-05

doi: 10.19734/j.issn.1001-3695.2019.01.0099

Research of small-sample-size based on hyperbolic cosine matrix discriminant analysis

Ran Ruisheng^{1a}, Zhang Shougui^{1b}, Ren Yinshan², Fang Bin^{3†}

(1. a. College of Computer & Information Science, b. College of Mathematical Science, Chongqing Normal University, Chongqing 401331, China; 2. Harbin Institute of Technology Robot Group Chongqing Yun'an Technology Co., Ltd., Chongqing 400039, China; 3. College of Computer Science, Chongqing University, Chongqing 400044, China)

Abstract: Linear discriminant analysis(LDA) was a classical approach in pattern recognition, but LDA has the so-called small-sample-size(SSS) problem. To address this problem, this paper presented a hyperbolic cosine matrix discriminant analysis named as HCDA. On the basis of hyperbolic cosine function, this paper first gave the definition and eigen-system of hyperbolic cosine matrix function, and then based on the characteristics of hyperbolic cosine matrix function eigen-system, it was introduced into LDA. HCDA had two superiorities. First, HCDA had no SSS problem and then it could extract more discriminant information. Second, there was an implicit mapping of samples. The mapping had a diffusion effect on the distance between samples, and the diffusion scale to the between-class distance was larger than that to the within-class distance. The experiment results on the handwritten digit, handwritten letter and Georgia Tech face databases show that, HCDA gets more advantageous recognition performance compared to the representative methods to solve the small-sample-size problem of LDA.

Key words: hyperbolic cosine function; matrix function; linear discriminant analysis; small-sample-size problem

0 引言

线性判别分析(linear discriminant analysis, LDA)算法是一种基于线性子空间的特征提取方法,是模式识别领域的经典算法。线性判别分析基于Fisher判别准则函数,因此也称做Fisher线性判别(Fisher linear discriminant, FLD)。FLD最早由Fisher^[1]于1936年提出,Duda等人^[2,3]基于Fisher思想进一步解决了多类样本的分类问题。Belhumeur等人^[4]首次将线性判别分析引入模式识别。LDA的基本思想是将高维的样本投影到最佳鉴别矢量空间,以达到抽取分类信息和压缩特征空间维数的效果。LDA方法使投影后样本的类间离散度最大且类内离散度最小。就是说,它能够保证投影后的模式样本在新的空间中有最大的类间距离和最小的类内距离,即模式在投影空间中有最佳的可分离性。因此,它是一种有效的特征抽取方法。

LDA是模式识别领域的一个研究热点。近年来,人们从

不同的方面对LDA方法进行了大量研究。例如,将线性判别分析的思想引入泛化的流形学习算法中,提出了局部鉴别保持投影方法^[5]、半监督最大边际分析方法^[6]、流形划分判别分析方法(MPDA)^[7]、判别稀疏邻域保持嵌入(DSNPE)方法等^[8];基于 l_1 范数做距离度量的LDA方法^[9-11];将稀疏限制条件加到LDA方法上的研究,如稀疏无关联线性判别分析SULDA^[12]、稀疏多类最大散度差方法(SMMSD)^[13]、稀疏判别分析的应用研究^[14]等。关于线性判别分析方法的研究还有很多。虽然LDA方法是一个经典且应用广泛的方法,但是LDA方法不得不面临所谓的小样本问题。LDA方法是将图像矩阵拉伸为一个高维的图像向量来处理。在大多数情形,图像向量的维数往往远大于样本数,导致样本类内散度矩阵 S_w 奇异,从而使得该方法的求解变得困难,此即所谓的小样本问题。文献[4]采用PCA与LDA相结合的方法解决了小样本问题。该方法首先利用PCA技术对人脸图像进行降维,使得类内散度矩

收稿日期: 2019-01-27; 修回日期: 2019-04-29 基金项目: 国家自然科学基金资助项目(61876026); 教育部人文社会科学研究项目(20YJAZH084); 重庆市基础研究与前沿探索研究资助项目(cstc2016jcyjA0419, cstc2017jcyjAX0316); 重庆师范大学校级科研项目(16XLB006, 16XZH07)

作者简介: 冉瑞生(1976-),男,教授,博士,主要研究方向为模式识别、机器学习; 张守贵(1974-),男,副教授,博士,主要研究方向为偏微分方程数值计算; 任银山(1994-),男,硕士,主要研究方向为机器学习、模式识别; 房斌(1964-),男(通信作者),教授,博导,主要研究方向为计算机视觉、模式识别(fb@cqu.edu.cn)。

阵 S_w 满秩,再利用 LDA 技术对其进行特征提取,进而进行人脸识别,该方法称为 Fisherfaces 方法或 PCA + LDA 方法。人们也提出了很多解决 LDA 小样本问题的方法。具有代表性的有正则 LDA 方法(regularized LDA, RLDA),该方法给 S_w 加上一个对称非负惩罚矩阵 Δ ,从而规避了 S_w 的奇异性^[15];基于 S_w 零空间的 LDA 方法(null LDA, NLDA)^[16],其思想是在 S_w 零空间中最大化样本类间距离,从而在 S_w 零空间中提取具有最大鉴别能力的信息;广义奇异值分解(GSVD)用于处理矩阵 S_w 的逆,从而提出了基于 GSVD 的 LDA 方法(LDA/GSVD)^[17];文献[18]提出了一个称为 LDA/QR 的两阶段 LDA 方法,该方法首先利用类间离散度矩阵 S_b 的 QR 分解最大化不同类之间的可分性,然后用 LDA 方法进一步对散度矩阵降秩;最大边界准则法(maximum margin criterion, MMC),该方法以最大化 $S_b - S_w$ 为目标函数,避免了计算类内散度矩阵 S_w 的逆,从而克服了小样本问题^[19]。此外,还有直接基于图像矩阵进行特征提取的二维 LDA 方法^[20],以及基于张量分析的多线性 LDA 方法^[21]等改进方法。

近年来,双曲余弦函数逐渐出现在模式识别和信号处理中^[22-23]。文献[22]提出了一种噪声自由长度判别分析方法(NLDA),用于解决噪声自由相关模式选择问题,利用双曲余弦函数构建 NLDA 的目标函数。文献[23]将双曲余弦函数应用到医学图像波束成形的旁瓣抑制中。

受文献[22-23]的启发,基于双曲余弦函数和 Fisher 的思想,本文提出了双曲余弦矩阵鉴别分析方法,用于解决经典 LDA 方法的小样本问题。提出的双曲余弦矩阵鉴别分析方法首先利用双曲余弦函数构建双曲余弦矩阵函数,然后将样本的类间散度矩阵 S_b 和类内散度矩阵 S_w 替换为相应的双曲余弦矩阵函数 $\cosh(S_b)$ 和 $\cosh(S_w)$,之后再应用 LDA 方法进行特征提取。该方法被命名为双曲余弦矩阵鉴别分析(hyperbolic cosine matrix discriminant analysis, HCDA)。由于双曲余弦矩阵函数 $\cosh(S_b)$ 和 $\cosh(S_w)$ 均是非奇异的,从而有效地避免了 LDA 方法的小样本问题,从而可以提取更多的特征信息;同时,这样的处理相当于将类间散度矩阵 S_b 和类内散度矩阵 S_w 映射到另一个空间,并且由于双曲余弦函数的单调性,该映射具有距离扩大的作用,并且对不同类样本距离的扩大作用要大于同类样本距离的扩大作用,从而有利于模式分类。所以,相对于已有的解决 LDA 小样本问题的方法,HCDA 方法具有更好的分类性能。

1 双曲余弦矩阵函数

本文提出的双曲余弦矩阵鉴别分析方法基于双曲余弦函数和矩阵函数之上,所以,首先引入双曲余弦函数和矩阵函数的定义,以及双曲余弦矩阵函数的特征系统问题。

1) 双曲余弦矩阵函数定义

定义 1 对任意自变量 x ,双曲余弦函数定义为

$$\cosh(x) = \frac{\exp(x) + \exp(-x)}{2}$$

对实变量 x ,当 $x > 0$ 时,双曲余弦函数的图像如图 1 所示。易知,当 $x > 0$ 时,双曲余弦函数是严格单调递增的。

定义 2^[24] 设 $f(x)$ 是一个以 x 为自变量的数值函数, A 是一个方阵,在 $f(x)$ 的表达式中以方阵 A 代替自变量 x ,得到的函数 $f(A)$ 称为以方阵 A 为自变量的矩阵函数。

根据定义 1 和 2,双曲余弦矩阵函数可定义为

定义 3 设 $\cosh(x)$ 是一个以 x 为自变量的双曲余弦函数, A 是一个方阵,双曲余弦矩阵函数 $\cosh(A)$ 可定义为

$$\cosh(A) = \frac{\exp(A) + \exp(-A)}{2}$$

2) 双曲余弦矩阵函数的特征系统

定理 1^[24] 设 A 为任意一个 n 阶实对称矩阵,则 A 可以对角化,也即,存在一个 n 阶正交矩阵 T 和矩阵 D ,使得 $D = T^T A T = T^{-1} A T$,且 D 是一个对角矩阵。其中 T^T 为矩阵 T 的转置矩阵,为矩阵 T 的逆矩阵。

上述的对角矩阵 $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$,其中 $\lambda_i (i = 1, 2, \dots, n)$ 是矩阵 A 的特征值;矩阵 T 的第 i 列为矩阵 A 的第 i 个特征值 $\lambda_i (i = 1, 2, \dots, n)$ 所对应的特征向量。

定理 2^[24] 设 A 为任意一个 n 阶矩阵, A 可以对角化,也即,存在一个 n 阶矩阵 X ,使得 $A = X \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) X^{-1}$ 。设 $f(A)$ 是一个矩阵函数,有

$$f(A) = X \text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)) X^{-1}$$

定理 2 表明了一个矩阵函数 $f(A)$ 与矩阵 A 的特征系统之间的紧密关系。根据定理 2,容易得到如下结论:

定理 3 设 A 为任意一个 n 阶实对称矩阵, $f(A)$ 是一个矩阵函数。设 $v_i (i = 1, 2, \dots, n)$ 是矩阵 A 的属于特征值 $\lambda_i (i = 1, 2, \dots, n)$ 的特征向量,即

$$A v_i = \lambda_i v_i \quad i = 1, 2, \dots, n$$

则 $v_i (i = 1, 2, \dots, n)$ 仍然是矩阵函数 $f(A)$ 的属于特征值 $f(\lambda_i) (i = 1, 2, \dots, n)$ 的特征向量,即

$$f(A) v_i = f(\lambda_i) v_i \quad i = 1, 2, \dots, n$$

定理 3 说明,如果 λ 是矩阵 A 的特征值,则 $f(\lambda)$ 是矩阵函数 $f(A)$ 的特征值;如果向量 v 是矩阵 A 属于特征值 λ 的特征向量,则向量 v 仍是矩阵函数 $f(A)$ 属于特征值 $f(\lambda)$ 的特征向量。显然,定理 3 的结论对于双曲余弦矩阵函数 $f(A) = \cosh(A)$ 也是成立的。

2 双曲余弦矩阵鉴别分析

设 S_b 为类间散度矩阵, S_w 为类内散度矩阵。LDA 方法通过最优化下面的准则函数找到一组最优鉴别矢量构成的投影矩阵 W :

$$J(W) = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} \quad (1)$$

若矩阵 S_w 非奇异,最优鉴别矢量,也即矩阵 W 的列向量可以通过求解如下的广义特征值问题而得到:

$$S_b w = \lambda S_w w \quad (2)$$

W 的每一列为前 $C-1$ (C 表示样本的类别数)个较大特征值对应的特征向量。LDA 是将一个图像拉伸为一个向量进行处理。但是,在大多数情形,图像向量的维数往往远大于样本数,使得样本类内散度矩阵 S_w 奇异,导致上述广义特征值问题不可解,这就是所谓的小样本问题。

本文中,为解决小样本问题,可以作如下处理。设 $A_b = \text{diag}(\lambda_{b1}, \lambda_{b2}, \dots, \lambda_{bn})$ 为 S_b 的特征值做成的对角矩阵, $\varphi_b = (\varphi_{b1}, \varphi_{b2}, \dots, \varphi_{bn})$ 为 S_b 的特征向量矩阵,且 λ_{bi} 与 φ_{bi} 一一对应 ($i = 1, 2, \dots, n$); 设 $A_w = \text{diag}(\lambda_{w1}, \lambda_{w2}, \dots, \lambda_{wn})$ 为 S_w 的特征值做成的对角矩阵, $\varphi_w = (\varphi_{w1}, \varphi_{w2}, \dots, \varphi_{wn})$ 为 S_w 的特征向量矩阵,且 λ_{wi} 与 φ_{wi} 一一对应 ($i = 1, 2, \dots, n$)。LDA 的准则函数式(1)可表示为

$$J(W) = \arg\max_W \frac{|W^T \varphi_b^T A_b \varphi_b W|}{|W^T \varphi_w^T A_w \varphi_w W|} \quad (3)$$

事实上,当小样本问题出现时,特征值矩阵 A_w 中含有 0 特征值,导致 S_w 奇异。为避免 S_w 的奇异性,作如下处理。设 $f(x) = \cosh(x) (x \geq 0)$ 是一个双曲余弦函数。在上述 LDA 准则函数式(3)中,将 A_b 中的 λ_{bj} 替换为 $\cosh(\lambda_{bj})$, A_w 中的 λ_{wj} 替换为 $\cosh(\lambda_{wj})$,并且记

$$\cosh(A_b) = \text{diag}(\cosh(\lambda_{b1}), \cosh(\lambda_{b2}), \dots, \cosh(\lambda_{bn}))$$

$$\cosh(\mathbf{A}_w) = \text{diag}(\cosh(\lambda_{w1}), \cosh(\lambda_{w2}), \dots, \cosh(\lambda_{wn}))$$

则式(3)变为

$$J(\mathbf{W}) = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \boldsymbol{\varphi}_b^T \cosh(\mathbf{A}_b) \boldsymbol{\varphi}_b \mathbf{W}|}{|\mathbf{W}^T \boldsymbol{\varphi}_w^T \cosh(\mathbf{A}_w) \boldsymbol{\varphi}_w \mathbf{W}|} \quad (4)$$

由于散度矩阵 S_b 和 S_w 均为对称半正定矩阵^[19], 根据定理 3, 式(4)可表示为

$$J(\mathbf{W}) = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \cosh(\mathbf{S}_b) \mathbf{W}|}{|\mathbf{W}^T \cosh(\mathbf{S}_w) \mathbf{W}|} \quad (5)$$

通过 Lagrange 乘数法, 最优鉴别矢量, 也即矩阵 \mathbf{W} 的列向量可以通过求解如下的广义特征值问题而得到:

$$\cosh(\mathbf{S}_b) \mathbf{w} = \lambda \cosh(\mathbf{S}_w) \mathbf{w} \quad (6)$$

对式(6)所求得特征值按从大到小排序, 取前面最大的 d 个特征值所对应的特征向量为列做成的矩阵即为所求 \mathbf{W} 。

3 理论分析

3.1 解决小样本问题

线性判别分析方法有小样本问题。通常, 为了克服小样本问题, 经典的 PCA + LDA 方法首先采用 PCA 方法对原始样本进行降维, 使得类间散度矩阵 S_w 非奇异, 然后再用 LDA 方法提取最优鉴别向量。然而, 在提取特征时, 由于采用了降维后的样本, 而不是采用初始样本, 势必导致矩阵 S_b 和 S_w 的零特征值所对应的特征信息被抛弃。但是, 零特征值所对应的特征信息可能含有有用的鉴别信息。进一步, 这些被抛弃的重要信息将不会体现在 LDA 所提取的图像特征信息中, 这对于模式分类是不利的^[25]。

根据第 2 章的讨论, HCDA 方法通过求解式(6), 也即下面的广义特征值问题而得到最优投影轴:

$$\cosh(\mathbf{S}_b) \mathbf{w} = \lambda \cosh(\mathbf{S}_w) \mathbf{w}$$

注意到类间散度矩阵 S_b 和类内散度矩阵 S_w 均为对称半正定矩阵^[19], 其所有特征值均非负, 即 $\lambda_{bi} \geq 0, \lambda_{wi} \geq 0 (i = 1, 2, \dots, n)$ 。根据定理 3, 矩阵 $\cosh(\mathbf{S}_b)$ 的特征值为 $\cosh(\lambda_{bi})$, 且根据定义 1 可知 $\cosh(\lambda_{bi}) \geq 1$; 同理, 矩阵 $\cosh(\mathbf{S}_w)$ 的特征值为 $\cosh(\lambda_{wi})$, 且 $\cosh(\lambda_{wi}) \geq 1$ 。所以 $\cosh(\mathbf{S}_b)$ 和 $\cosh(\mathbf{S}_w)$ 是满秩矩阵。即使小样本问题出现时, 矩阵 S_b 和 S_w 为其中的所有信息, 包括零特征值所对应的特征信息, 都能被 HCDA 方法提取到。所以, 相对于 LDA 方法, 可以相信 HCDA 方法将有更强的鉴别能力, 第 4 章的实验结果也证实了这一点。

3.2 距离扩大映射

在模式识别领域, 核方法通常用于非线性数据的处理。核方法的基本思想是通过一个非线性映射 Φ 将非线性的原始样本映射到一个新的特征空间:

$$\Phi: \mathbf{R}^n \rightarrow F, \quad x \mapsto \Phi(x) \quad (7)$$

然后新的特征空间中提取特征。基于核方法的思想, 核 Fisher 判别方法 (kernel Fisher's discriminant, KFD) 于 1999 年被提出^[26]。

事实上, HCDA 方法含有一个隐式的非线性映射:

$$\Theta: \mathbf{R}^{n \times n} \rightarrow \mathbf{R}^{n \times n} \quad (8)$$

$$S_b \rightarrow \Theta(S_b) = \cosh(S_b) \quad (9)$$

$$S_w \rightarrow \Theta(S_w) = \cosh(S_w) \quad (10)$$

所以, HCDA 与 KFD 方法具有相似性, 两者均采用非线性映射作变换, 然后新的特征空间中应用 Fisher 准则。两者的区别在于核方法是对原始样本进行映射, 而 HCDA 方法是对散度矩阵进行映射。需注意到, 散度矩阵是由原始样本构成的, 所以可以认为 HCDA 方法隐含着对原始样本的非线性映射, 也具有核方法的属性。KFD 方法是基于 LDA 方法处理非线性问

题的不足而提出的; 而 HCDA 方法是为了解决 LDA 的小样本问题而提出, 其优势在于克服了小样本问题, 并具有样本距离扩大作用。

LDA 方法主要在于找到一组最优鉴别矢量, 使得投影后的样本同时具有最大的类间距离和最小的类内距离。这两个距离可以用散度矩阵的迹来表示:

$$d_b = \text{trace}(S_b) = \lambda_{b1} + \lambda_{b2} + \dots + \lambda_{bn} \quad (11)$$

$$d_w = \text{trace}(S_w) = \lambda_{w1} + \lambda_{w2} + \dots + \lambda_{wn} \quad (12)$$

由于 S_b 和 S_w 均为对称半正定矩阵, 其所有特征值均非负。从统计意义上看, 式(11)中的特征值 λ_{bi} 反映了不同类别样本的分离度, 式(12)中的特征值 λ_{wi} 反映了同类别样本的紧密度; 其比率 $\lambda_{bi}/\lambda_{wi}$ 度量了模式分类性, 比率越大则更有利于模式分类。

通过映射(8)~(10)和定理 3, 式(11)和(12)中的距离 d_b 和 d_w 被替换为距离 d_b^h 和 d_w^h :

$$d_b^h = \text{trace}(\cosh(S_b)) = \cosh(\lambda_{b1}) + \cosh(\lambda_{b2}) + \dots + \cosh(\lambda_{bn}) \quad (13)$$

$$d_w^h = \text{trace}(\cosh(S_w)) = \cosh(\lambda_{w1}) + \cosh(\lambda_{w2}) + \dots + \cosh(\lambda_{wn}) \quad (14)$$

类似于 LDA 的分析, 从统计意义上看, 式(13)中的特征值 $\cosh(\lambda_{bi})$ 反映了经映射后在新的特征空间里不同类别样本的分离度, 式(14)中的特征值 $\cosh(\lambda_{wi})$ 反映了在新的特征空间里同类别样本的紧密度; 其比率 $(\cosh(\lambda_{bi}) / (\cosh(\lambda_{wi})))$ 度量了映射后的模式分类性, 比率越大则更有利于模式分类。

现在讨论 HCDA 的距离扩大作用。本文提出的 HCDA 方法的非线性映射 Θ 具有距离扩大作用, 并且对于不同类别样本距离的扩大尺度要大于对同类别样本距离的扩大尺度。可以用非线性映射对特征值作用前、后的几何意义, 也即 LDA 和 HCDA 方法散度矩阵的特征值的几何意义来进行说明, 如图 2 所示。图中, 直线为函数 $y = x$, 表达映射前的特征值; 曲线为函数 $y = \cosh(x)$, 表达映射后的特征值。

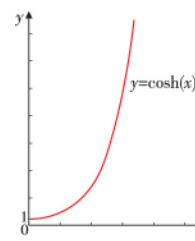


图 1 双曲余弦函数的曲线 ($x > 0$)

Fig. 1 The curve of hyperbolic cosine function ($x > 0$)

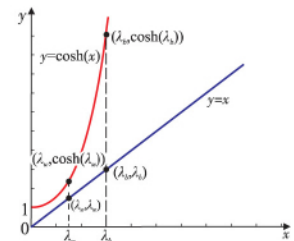


图 2 非线性映射距离扩大作用的分析

Fig. 2 The analysis of distance diffusion of the nonlinear mapping

通常, 不同类别样本间的距离要大于同类别样本间的距离, 也即 $d_b > d_w$, 所以, 对于式(11)和(12)中的大多数特征值, 有 $\lambda_{bi} > \lambda_{wi} > 0$ 。根据 λ_{wi} 的取值, 作如下讨论:

a) 当 $\lambda_{wi} > 0$ 时, 根据图 2, 有不等式:

$$\cosh(\lambda_{bi}) > \lambda_{bi}, \cosh(\lambda_{wi}) > \lambda_{wi} \quad (15)$$

也即 HCDA 的非线性映射 Θ 对不同类别样本和同类别样本的距离均具有扩大作用。进一步, 因为 $\lambda_{bi} > \lambda_{wi}$, 根据双曲余弦函数的单调性, 有不等式:

$$\cosh(\lambda_{bi}) > \cosh(\lambda_{wi}) \quad (16)$$

也即非线性映射 Θ 对于不同类别样本距离的扩大尺度要大于对同类别样本距离的扩大尺度。

另外, 根据图 2, 有不等式:

$$\frac{\cosh(\lambda_{bi})}{\cosh(\lambda_{wi})} > \frac{\lambda_{bi}}{\lambda_{wi}} \quad (17)$$

根据前述比率 $\cosh(\lambda_{bi}) / \cosh(\lambda_{wi})$ 和 $\lambda_{bi} / \lambda_{wi}$ 所表达的意

义可知,HCDA 的模式分类性更好,从而更有利于分类。

b) 当 $\lambda_{wi} = 0$ 时,显然有 $\cosh(\lambda_{wi}) = 1$ 。因为 $\lambda_{bi} > \lambda_{wi} = 0$, 所以在这种情形总有不等式:

$$\cosh(\lambda_{bi}) > \lambda_{bi}, \cosh(\lambda_{wi}) > \lambda_{wi} \quad (18)$$

$$\cosh(\lambda_{bi}) > \cosh(\lambda_{wi}) \quad (19)$$

进而有不等式:

$$\frac{\cosh(\lambda_{bi})}{\cosh(\lambda_{wi})} > \lambda_{bi} \quad (20)$$

综上所述,HCDA 的非线性映射 Θ 具有距离扩大的作用。虽然映射后不同类别样本间的距离与同类样本间的距离均被拉大,但根据式(16)和(19),该映射对不同类别样本距离的扩大尺度要大于对同类样本距离的扩大尺度。由不等式(17)和(20)可知,HCDA 方法比 LDA 方法的分类能力更强,将表现出更好的分类性能。这就是 HCDA 方法的优势。

4 实验结果

本章对提出的 HCDA 方法进行了手写数字识别、字母识别和人脸识别实验。

手写数字库和字母库为纽约大学公开的手写数字和手写字母图像库^[27]。手写数字库由手写的数字 0~9 构成,每个数字有 39 个样本;手写字母由字母 A~Z 构成,每个字母也有 39 个样本。图 3 给出了手写数字库中的数字 0~9 的示例图像,图 4 给出了手写字母库中部分字母的示例图像。



图 3 手写数字 0~9 示例图像

Fig. 3 The sample image of handwritten digit 0~9



图 4 手写字母库中部分字母的示例图像

Fig. 4 Some sample images from handwritten letter database

人脸图像库为公开的 Georgia Tech 人脸库,由乔治亚理工学院信号与图像处理中心构建。该人脸库由 50 个人,每个人 15 幅彩色 JPEG 样本图像构成。样本图像有正面和倾斜的,并有表情、光照变化,图像大小为 150×150 像素。实验中每幅图像被校正了人脸的位置,并被转换为灰度图像,大小设定为 32×32 像素。图 5 给出了该人脸库中某人的部分原始样本图像。



图 5 Georgia Tech 人脸库中某人的原始样本图像

Fig. 5 Some sample images of one individual from Georgia Tech face database

在实验中,为了体现 HCDA 的模式分类效果,将 HCDA 的实验结果与经典的主成分分析方法 PCA^[28],以及解决线性判别分析小样本问题具有代表性的 PCA+LDA、LDA/GSVD、LDA/QR、RLDA、NLDA、MMC 方法的实验结果进行了比较。

实验中,PCA 方法提取主成分的阈值设置为 98%;PCA+LDA 方法中,用 PCA 降维时的阈值也设置为 98%;RLDA 方法的原理在于给类内散度矩阵 S_w 加上一个对称非负惩罚矩阵来规避 S_w 的奇异性,实验中设置 $\tilde{S}_w = S_w + rI$,这里 $r = 0.1$, I 是单位矩阵。

实验中,对每个样本随机选取 p 幅人脸图像作为训练样本集,剩余的样本作为测试样本集。注意到 LDA 方法最多获取

$C-1$ 维的子空间维数^[2,3],其中 C 表示样本的类别数。为了实验结果比较的对等性,在实验中对所有方法都提取 $C-1$ 维的子空间维数。为了得到稳定的实验结果,对每一个给定的训练样本个数 p 进行 30 次实验。然后将 30 次实验的识别结果的平均值作为相应方法的识别率。

在手写数字库、手写字母库和 Georgia Tech 人脸库上的实验结果分别如表 1~3 所示。

表 1 八种方法在手写数字库上的平均识别率

Tab. 1 Average recognition rate of the eight methods on handwritten digit database

方法	3 个训练样本	5 个训练样本	7 个训练样本	9 个训练样本
PCA	63.22	70.145	74.22	78.77
PCA+LDA	69.61	71.38	78.35	80.87
LDA/GSVD	67.19	70.39	77.35	78.99
LDA/QR	70.18	72.56	80.23	80.98
RLDA	69.77	71.34	79.83	78.98
NLDA	70.65	79.56	80.02	81.23
MMC	72.39	79.79	81.16	82.47
HCDA	72.63	80.00	82.50	83.73

表 2 八种方法在手写字母库上的平均识别率

Tab. 2 Average recognition rate of the eight methods on handwritten letter database

方法	7 个训练样本	9 个训练样本	11 个训练样本	13 个训练样本
PCA	62.71	65.36	66.01	67.74
PCA+LDA	60.40	63.50	64.34	67.03
LDA/GSVD	60.17	63.33	65.35	64.78
LDA/QR	61.38	64.16	66.26	65.19
RLDA	59.95	60.75	61.13	62.74
NLDA	62.65	65.56	66.62	65.53
MMC	58.60	57.27	52.29	47.94
HCDA	63.82	66.15	66.97	67.99

表 3 八种方法在 Georgia Tech 人脸库上的平均识别率

Tab. 3 Average recognition rate of the eight methods on Georgia Tech face database

方法	5 个训练样本	6 个训练样本	7 个训练样本	8 个训练样本
PCA	71.09	72.49	76.08	76.57
PCA+LDA	74.14	75.13	79.28	80.63
LDA/GSVD	72.23	72.45	76.36	77.90
LDA/QR	75.99	77.23	80.15	81.22
RLDA	74.12	74.36	78.10	79.67
NLDA	75.35	76.32	78.13	79.96
MMC	77.48	79.57	79.7	81.77
HCDA	79.65	81.87	83.98	84.69

通常,在子空间方法中,人脸识别的精度与所提取的子空间的维度有关。为了体现 HCDA 方法关于子空间维度的稳定性和识别性能的优越性,在手写数字库、手写字母库和 Georgia Tech 人脸库上也进行关于子空间维度的实验。注意到,由于 LDA 方法最多提取 $C-1$ 维的子空间^[2,3],其中 C 为样本的类别数;NLDA 方法受到类内散度矩阵 S_w 奇异性的限制^[13];LDA/QR 方法依赖于类间散度矩阵 S_b 奇异性^[15],所以 PCA+LDA、LDA/QR、NLDA 方法所提取的子空间特征维度也受到限制。而由于 HCDA 方法避免了小样本问题,矩阵 $\cosh(S_b)$ 和 $\cosh(S_w)$ 是满秩矩阵,所以避免了子空间维度的限制。其他的 RLDA、LDA/GSVD 以及 MMC 方法所提取的子空间维度均不受限制^[15,17,19]。于是,选择了 LDA/GSVD、RLDA、MMC 以及本文的 HCDA 方法进行子空间维度的实验。

实验中,随机选取 p 幅图像作为训练样本集,剩余的样本作为测试样本。对于给定的 p 个训练样本,让子空间维数从 10 开始按步长 10 逐次递增到 100,在每一个维度下,用 LDA/GSVD、RLDA、MMC 和本文提出的 HCDA 方法分别提取图像特征,用于训练和识别。为了实验结果的稳定性和合理性,在每一个维度下进行 30 次实验,并将 30 次实验结果的平均识别率作为 p 个训练样本以及该维度下的识别率。

实验结果如图6~8所示。图6是在手写数字库上的实验,训练样本的个数为9;图7是在手写字母库上的实验,训练样本的个数为9;图8是在Georgia Tech人脸图像库上的实验,训练样本的个数为6。

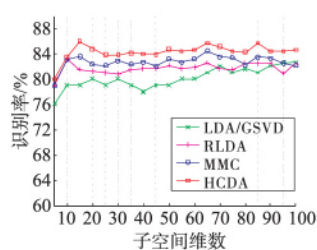


图6 四种方法在 handwritten 数字库上关于子空间维度的识别率

Fig. 6 Recognition accuracy of the four methods versus the subspace dimensions on the handwritten digit database

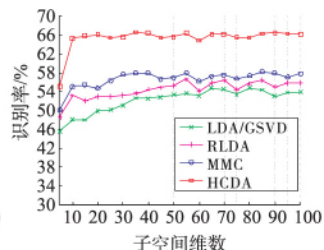


图7 四种方法在 handwritten 字母库上关于子空间维度的识别率

Fig. 7 Recognition accuracy of the four methods versus the subspace dimensions on the handwritten letter database

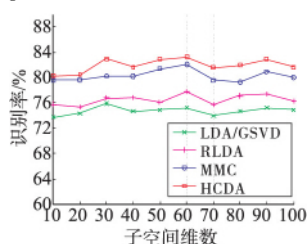


图8 四种方法在 Georgia Tech 人脸库上关于子空间维度的识别率

Fig. 8 Recognition accuracy of the four methods versus the subspace dimensions on the Georgia Tech face database

5 结束语

本文提出了一种双曲余弦矩阵鉴别分析方法,用于解决经典线性判别分析方法的小样本问题。通过双曲余弦矩阵函数,HCDA方法将样本的散度矩阵映射到一个新的空间,然后在新空间上应用LDA方法。通过这样的变换,HCDA方法有效地避免了LDA方法的小样本问题,可以提取更多的样本鉴别信息;其次,这种变换具有距离扩大作用,并且对于不同类别样本间距离的扩大尺度大于对同类样本间距离的扩大尺度,这有利于模式分类。在 handwritten 数字库、handwritten 字母库和 Georgia Tech 人脸库上的实验表明,相对于经典的PCA方法和解决LDA小样本问题具有代表性的方法,HCDA具有更好的识别性能。

参考文献

- [1] Fisher R A. The use of multiple measurements in taxonomic problems [J]. *Annals of Eugenics*, 1936(7): 179-188.
- [2] Rao C R. The utilization of multiple measurements in problems of biological classification [J]. *Journal of the Royal Statistical Society B*, 1948(10): 159-203.
- [3] Duda R, Hart P, Stork D. *Pattern classification* [M]. 2nd ed. Piscataway, NJ: Wiley, 2000.
- [4] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720.
- [5] Huang Shucheng, Zhuang Lu. Exponential discriminant locality preserving projection for face recognition [J]. *Neurocomputing*, 2016(208): 373-377.
- [6] Huang Hong, Liu Jiamin, Pan Yinsong. Semi-supervised marginal Fisher analysis for hyperspectral image classification [C]//Proc of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2012: 377-382.
- [7] Zhou Yang, Sun Shiliang. Manifold partition discriminant analysis [J]. *IEEE Trans on Cybernetics*, 2017, 47(4): 1-11.
- [8] 邹雪城, 刘尹, 邹连英, 等. 改进的鉴别稀疏保持投影人脸识别算

法[J]. *华中科技大学学报: 自然科学版*, 2018, 46(1): 53-57. (Zou Xuecheng, Liu Yin, Zou Lianying, et al. Improved discriminant sparsity preserving projecting face recognition algorithm [J]. *Journal of Huazhong University of Science and Technology: Natural Science Edition* 2018, 46(1): 53-57.)

- [9] Wang Haixian, Lu Xuesong, Hu Zilan, et al. Fisher discriminant analysis with l_1 -norm [J]. *IEEE Trans on Cybernetics*, 2014, 44(6): 828-842.
- [10] Li Mengyuan, Wang Jing, Wang Qianqian, et al. Trace ratio 2DLDA with l_1 -norm optimization [J]. *Neurocomputing*, 2017, 266(29): 216-225.
- [11] 余景丽, 胡恩良, 张涛. 一种新的 l_1 度量 Fisher 线性判别分析研究 [J]. *计算机工程与应用*, 2018, 54(4): 128-134. (Yu Jingli, Hu Enliang, Zhang Tao. Study of Fisher linear discriminant analysis based on l_1 -norm [J]. *Computer Engineering and Applications* 2018, 54(4): 128-134.)
- [12] Zhang Xiaowei, Chu Delin, Roger T C E. Sparse uncorrelated linear discriminant analysis for undersampled problems [J]. *IEEE Trans on Neural Networks and Learning Systems*, 2017, 27(7): 1469-1485.
- [13] Yu Lei, Yang Dan, Wang Huiqi. Sparse multiple maximum scatter difference for dimensionality reduction [J]. *Digital Signal Processing*, 2017(62): 91-100.
- [14] 周亚文, 董广军, 薛志祥, 等. 基于稀疏判别分析的高光谱影像特征提取 [J]. *测绘科学技术学报*, 2017, 34(4): 370-375. (Zhou Yawen, Dong Guangjun, Xue Zhixiang, et al. Hyperspectral imagery feature extraction based on sparse discriminant analysis [J]. *Journal of Geomatics Science and Technology* 2017, 34(4): 370-375.)
- [15] Hastie T, Buja A, Tibshirani R. Penalized discriminant analysis [J]. *Annals of Statistics*, 1995, 23(1): 73-102.
- [16] Chen Lifan, Mark Liao H Y, Ko M T, et al. A new LDA-based face recognition system which can solve the small sample size problem [J]. *Pattern Recognition*, 2000, 33(10): 1713-1726.
- [17] Howland P, Park H. Generalizing discriminant analysis using the generalized singular value decomposition [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2004, 26(8): 995-1006.
- [18] Ye Jieping, Li Qi. A two-stage linear discriminant analysis via QR-decomposition [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2005, 27(6): 929-941.
- [19] Li Haifeng, Jiang Tao, Zhang Keshu. Efficient and robust feature extraction by maximum margin criterion [J]. *IEEE Trans on Neural Networks*, 2006, 17(1): 157-165.
- [20] Nousath S, Kumar G H, Shivakumara P. (2D) 2LDA: an efficient approach for face recognition [J]. *Pattern Recognition*, 2006, 39(7): 1396-1400.
- [21] Yuan Sen, Mao Xia, Chen Lijiang. Multilinear spatial discriminant analysis for dimensionality reduction [J]. *IEEE Trans on Image Processing*, 2017, 26(6): 2669-2681.
- [22] Murthy K R, Ghosh A. Noisy-free length discriminant analysis with co-sine hyperbolic framework for dimensionality reduction [J]. *Expert Systems with Applications*, 2017(81): 88-107.
- [23] Aghdam S A, Bagby J, Pla R J. Malleable hyperbolic cosine weighting function for side lobe suppression in medical image beamforming [C]//Proc of IEEE International Symposium on Phased Array Systems & Technology. Piscataway, NJ: IEEE Press, 2017: 1-6.
- [24] Golub G H, Van Loan C F. *矩阵计算* [M]. 袁亚湘, 等译. 北京: 科学出版社, 1983: 644-647. (Golub G H, Van Loan C F. *Matrix computations* [M]. Yuan Yaxiang, Trans. Beijing: Science Press, 1983: 644-647.)
- [25] Dai Daoqing, Yuen P C. Face recognition by regularized discriminant analysis [J]. *IEEE Trans on Systems Man & Cybernetics: Part B*, 2007, 37(4): 1080-1085.
- [26] Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels [C]//Proc of IEEE Neural Networks for Signal Processing. Piscataway, NJ: IEEE Press, 1999: 41-48.
- [27] Roweis S. Data for MATLAB hackers [EB/OL]. [2015-03-24]. <https://cs.nyu.edu/~roweis/data.html>.
- [28] Turk M A, Pentland A P. Eigenfaces for recognition [J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.