

基于广义矩阵指数的判别局部保持投影方法^{*}

任银山¹, 冉瑞生¹, 房 斌²

(1. 重庆师范大学 计算机与信息科学学院, 重庆 401331; 2. 重庆大学 计算机学院, 重庆 400044)

摘要:【目的】局部保持投影(LPP)是一种经典的非线性数据降维方法。在 LPP 方法基础上人们提出了判别局部保持投影方法(DLPP),并取得了良好的效果,但 DLPP 方法存在小样本问题,针对该问题提出了广义矩阵指数判别局部保持投影(GEDLPP)算法。【方法】基于矩阵函数的性质,使用广义矩阵指数函数来重构 DLPP,即为 GEDLPP 算法。【结果】提出的算法有两个优点:一是解决了 DLPP 方法的小样本问题;二是 GEDLPP 所隐含的非线性映射拉伸了不同类别样本之间的距离,从而提高了模式分类的能力。【结论】在 COIL-20 数据库, Yale, Extended Yale B 和 CMU-PIE 人脸数据集上的实验结果表明:与最近提出的解决 DLPP 小样本问题的改进方法相比,GEDLPP 的识别率优于其他方法。

关键词:流形学习;数据降维;局部保持投影;小样本问题;矩阵函数

中图分类号:TP181

文献标志码:A

文章编号:1672-6693(2022)04-0114-10

数据降维一直是机器学习和计算机视觉领域的一个研究热点。目前人们提出了很多基于流形学习的非线性降维方法,经典方法有局部线性嵌入(LLE)^[1], ISOMAP^[2], 拉普拉斯特征映射(LE)^[3]和局部切空间排列(LTSA)^[4]。但是,这些经典方法的嵌入或映射仅定义在训练数据集上,而测试样本不得不再次嵌入或映射,因此它们不适用于模式分类,例如人脸识别任务。为了改进这一缺陷,人们提出了线性化方法,对 LLE, ISOMAP, LE 等方法进行线性化处理。其中局部保持投影(LPP)^[5]是 LE 的线性化版本,也是线性化方法的一个典型代表。基于 LPP 的思想,文献[6]提出了“拉普拉斯脸”,并用于人脸识别。

LPP 是一种无监督的方法,该方法训练阶段未采用样本的类别标签信息。当在训练阶段考虑样本的类别标签信息时,在 LPP 基础上人们提出了判别局部保持投影(DLPP)方法^[7-10]。DLPP 方法表现出比 LPP 更有效的性能,例如它在人脸识别任务中实现了更好的分类能力。然而,在机器学习和计算机视觉等领域,样本的维度通常很高,而获取到的样本数量比样本的维度小得多,这造成 DLPP 的广义特征值问题中的矩阵是奇异的,进而很难求解 DLPP 的广义特征方程,这就是 DLPP 方法的“小样本问题”^[11-14]。

解决小样本问题的常用方法是:首先使用主成分分析(PCA)方法^[15]降低初始样本的维数,从而使得 DLPP 的广义特征值问题中的矩阵是非奇异的,然后使用 DLPP 在降维后的空间中提取样本的特征。但是,当使用该方法降维时,初始样本中一些重要信息也会同时丢失。

为了解决小样本问题,Yang 等人^[11]提出了零空间判别局部保持投影(NDLPP)方法,该方法解决了零空间中的特征值问题,从而避免了小样本问题。基于最大边际准则的判别局部保持投影(DLPP/MMC)^[12]也是解决 DLPP 的小样本问题的一种方法,它基于最大边际准则构造 DLPP 的目标函数,即最大化局部类间散度和局部类内散度之间的差,而不是比率。DLPP/MMC 方法在求解特征值问题时不需要计算矩阵的逆,与特征方程中的矩阵奇异与否没有直接关系,从而自然地避免了 DLPP 的小样本问题。

文献[13-14]提出了基于矩阵指数的 DLPP 方法(MEDLPP)来解决 DLPP 的小样本问题。它用矩阵指数替换了 DLPP 中的散度矩阵,因此避免了散度矩阵的奇异性。

近些年来,人们提出很多基于流形学习的数据降维方法,如邻域保持嵌入(NPE)、局部判别嵌入(LDE)和半

* 收稿日期:2020-12-07 修回日期:2022-06-14 网络出版时间:2022-06-30 20:15

资助项目:国家自然科学基金(No. 61876026);教育部人文社会科学基金项目(No. 20YJAZH084);重庆市技术创新与应用发展专项项目(No. cstc2019jscx-mbdxX0061, No. cstc2020jscx-msxmX0190);重庆市自然科学基金(No. cstc2016jcyjA0419);重庆师范大学校级基金项目(No. 16XLB006, No. 16XZH07)

第一作者简介:任银山,男,研究方向为机器学习,E-mail:viprys@163.com;通信作者:冉瑞生,男,教授,博士,E-mail:rshran@cqu.edu.cn

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20220630.1347.022.html

监督判别嵌入(SDE)方法等,但这些方法都存在小样本问题。由于矩阵指数方法在解决小样本问题上的有效性,它的思想也可用于解决上述数据降维方法的小样本问题,人们相继提出了基于矩阵指数的邻域保持嵌入方法(ENPE)^[16]、基于矩阵指数的LPP方法(ELPP)^[17]、基于矩阵指数的局部判别嵌入方法(ELDE)^[18]和基于矩阵指数的半监督判别嵌入方法(ESDE)^[19]。文献[20]提出了一个通用的基于矩阵指数的数据降维框架,并基于该框架提出了矩阵指数边际Fisher分析方法(EMFA)和矩阵指数无监督判别投影方法(EUDP)。这些方法都解决了对应方法的小样本问题,并都表现出更好的性能。

本文基于矩阵函数的性质,利用一般性的矩阵指数重构DLPP方法,提出了一种通用的矩阵指数判别局部保持投影(GEDLPP)方法。本文首先介绍矩阵函数并讨论它的特征系统,接着使用一般的指数函数构造矩阵指数函数,然后基于矩阵指数函数的性质,提出一种广义的矩阵指数判别局部保留投影方法,称为GEDLPP。GEDLPP可以避免DLPP中散度矩阵的奇异性,从而解决了DLPP的小样本问题。更为重要的是,GEDLPP可以扩大不同类别样本之间的距离,与现有方法相比,对模式分类更为有利。因此,与最新方法相比,GEDLPP可以得到更好的分类性能。另一方面,作为一种通用方法,GEDLPP的主要思想可以扩展到上述基于矩阵指数的数据降维方法上,如ELDE,ESDE,EMFA等。

1 LPP和DLPP方法回顾

1.1 LPP

假设 $\mathbf{X}=(x_1, x_2, \dots, x_m)$ 是 \mathbf{R}^n 空间中的原始 n 维数据集。LPP首先构造一个近邻图, W_{ij} 表示近邻图上点对 x_i 和 x_j 的连接权重,通常选择热核函数来构造权重 W_{ij} 。LPP方法的思想在于寻找一个线性投影:

$$\mathbf{y}_i = \mathbf{U}^T \mathbf{x}_i, \quad (1)$$

将原始样本 x_i 映射到 d 维($d \ll n$)空间中的一个点 y_i ,并使得降维后的样本点尽可能地保持原样本之间的邻接关系。LPP定义如下目标函数:

$$\sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 W_{ij} = \sum_{i,j} \|\mathbf{U}^T \mathbf{x}_i - \mathbf{U}^T \mathbf{x}_j\|^2 W_{ij} = \text{tr}(\mathbf{U}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{U}), \quad (2)$$

并施加约束条件 $\mathbf{U}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{U} = \mathbf{I}$,从而获得投影矩阵 \mathbf{U} 。(2)式中矩阵 \mathbf{L} 是拉普拉斯矩阵, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, \mathbf{W} 是由 W_{ij} 组成的权重矩阵, \mathbf{D} 是矩阵 \mathbf{W} 的行和(或列和)矩阵,即 $D_{ii} = \sum_j W_{ij}$, tr 表示矩阵的迹。基于(2)式和约束条件,LPP的目标函数可以表述为:

$$\arg \min_{\mathbf{U}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{U} = \mathbf{I}} \text{tr}(\mathbf{U}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{U}). \quad (3)$$

1.2 DLPP

LPP是一种无监督方法,在训练阶段没有采用样本的类别信息。与LPP不同的是,DLPP在LPP方法基础上,在训练阶段采用样本的类别信息进行学习,是一种有监督的方法。DLPP为第 c 个类内样本构造一个邻接矩阵 $\mathbf{W}^c = (W_{ij}^c)$,并构造一个总的类内邻接矩阵 $\mathbf{W} = \text{diag}(\mathbf{W}^1, \mathbf{W}^2, \dots, \mathbf{W}^C)$,其中 C 是样本类别数;对于类间样本,DLPP在每个类别中找到一个均值点,然后对于 C 个均值点,构造一个邻接矩阵 $\mathbf{B} = (B_{ij})$ 。DLPP寻求线性投影 $\mathbf{y}_i = \mathbf{U}^T \mathbf{x}_i$,以将原始样本 x_i 映射到低维空间中的一个点 y_i ,其中 \mathbf{U} 是所需的投影矩阵。

DLPP的目标函数为:

$$\max \frac{\sum_{i,j=1}^C (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_j)^2 B_{ij}}{\sum_{c=1}^C \sum_{i,j=1}^{n_c} (\mathbf{y}_i^c - \mathbf{y}_j^c)^2 W_{ij}^c}, \quad (4)$$

其中: n_c 代表第 c 类中的样本数, \mathbf{y}_i^c 是第 c 类中的第 i 个投影向量, $\bar{\mathbf{y}}_i = \sum_{k=1}^{n_i} \mathbf{y}_k^i$ 。(4)式可以简化为:

$$J(\mathbf{U}) = \max_{\mathbf{U}} \frac{\text{tr}(\mathbf{U}^T \bar{\mathbf{X}} \mathbf{H} \bar{\mathbf{X}}^T \mathbf{U})}{\text{tr}(\mathbf{U}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{U})} = \max_{\mathbf{U}} \frac{\text{tr}(\mathbf{U}^T \mathbf{S}_H \mathbf{U})}{\text{tr}(\mathbf{U}^T \mathbf{S}_L \mathbf{U})} \quad (5)$$

其中: $\bar{\mathbf{X}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_C]$, $\mathbf{H} = \mathbf{E} - \mathbf{B}$, \mathbf{E} 是对角阵,且 E_{ii} 是矩阵 \mathbf{B} 的行和矩阵,即 $E_{ii} = \sum_j B_{ij}$, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, \mathbf{D} 是对角阵,且 D_{ii} 矩阵 \mathbf{W} 的行之和,即 $D_{ii} = \sum_j W_{ij}$, $\mathbf{S}_H = \bar{\mathbf{X}} \mathbf{H} \bar{\mathbf{X}}^T$, $\mathbf{S}_L = \mathbf{X} \mathbf{L} \mathbf{X}^T$ 。(5)式可以简化为以下广义特征

向量问题:

$$S_H u = \lambda S_L u. \quad (6)$$

上述方程中,矩阵 S_L 的秩有如下不等式:

$$\text{rank}(S_L) = \text{rank}(XLX^T) \leq \text{rank}(X) \leq \min(m, n),$$

其中: $\text{rank}(\cdot)$ 代表矩阵的秩, m 是训练样本的数量, n 是样本的维数。在大多数情形,训练样本的数量远小于样本的维数,即 $m \ll n$, 因此有 $\text{rank}(S_L) \leq \min(m, n) \ll n$ 。注意到矩阵 $S_L = XLX^T$ 是 $n \times n$ 的矩阵 S_L , 因此矩阵是奇异的, 导致方程(6)不易求解。这就是 DLPP 方法的“小样本问题”。该问题备受人们关注, 文献[11-14]即是解决 DLPP 的“小样本问题”所提出的算法。

2 GEDLPP

2.1 矩阵函数及本征系统

矩阵函数是提出的 GEDLPP 的理论基础, 因此这里首先介绍矩阵函数的定义。

定义 1^[21] 令幂级数 $\sum_{m=0}^{\infty} c_m x^m$ 对于 $|x| < R$ 收敛, 当 x 在收敛域中时, 函数 $f(x) = \sum_{m=0}^{\infty} c_m x^m$ 收敛。设 A 为

方阵, 它的谱半径 $\rho(A) < R$, 定义 $f(A) = \sum_{m=0}^{\infty} c_m A^m$, 则函数 $f(A)$ 是矩阵 A 的矩阵函数。

定理 1 假设 $f(x)$ 是任一函数, A 是 n 阶实对称方阵, 矩阵 A 对应的矩阵函数为 $f(A)$ 。设 A 的特征值为 λ_i , 对应的特征向量为 v_i , 即 $Av_i = \lambda_i v_i (i=1, 2, \dots, n)$, 则有:

$$f(A)v_i = f(\lambda_i)v_i, i=1, 2, \dots, n, \quad (7)$$

即 $f(\lambda_i)$ 是 $f(A)$ 的特征值, v_i 是 $f(A)$ 的特征向量。

证明 由于矩阵 A 是实对称方阵, 因此 A 可以对角化。因此, 存在一个使 A 对角化的 n 阶正交矩阵 $V = [v_1, v_2, \dots, v_n]$, 使得 $A = V \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) V^{-1}$, 其中 λ_i 是 A 的特征值, v_i 是 A 的对应特征向量。那么, 根据矩阵函数的性质, $f(A)$ 也可以对角化^[21], 它的形式为: $f(A) = V \text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)) V^{-1}$, 这个对角化形式也可以改写为: $f(A)v_i = f(\lambda_i)v_i, i=1, 2, \dots, n$ 。证毕

2.2 GEDLPP

如 1.2 节所述, DLPP 方法是寻找一个最优投影矩阵 U 使目标函数(5)式最大化。为方便起见, 将(5)式表示为:

$$J(U) = \max_U \frac{\text{tr}(U^T S_H U)}{\text{tr}(U^T S_L U)}. \quad (8)$$

令 λ_{bi} 为矩阵 S_H 的特征值, v_{bi} 为属于特征值 $\lambda_{bi} (i=1, 2, \dots, n)$ 的特征向量, 从而 $S_H = V_H^T \Lambda_H V_H$, 其中 $\Lambda_H = \text{diag}(\lambda_{b1}, \lambda_{b2}, \dots, \lambda_{bn})$, $V_H = (v_{b1}, v_{b2}, \dots, v_{bn})$ 。类似地, 令 λ_{wi} 为矩阵 S_L 的特征值, v_{wi} 为特征值 $\lambda_{wi} (i=1, 2, \dots, n)$ 的特征向量, 从而 $S_L = V_L^T \Lambda_L V_L$, 其中 $\Lambda_L = \text{diag}(\lambda_{w1}, \lambda_{w2}, \dots, \lambda_{wn})$, $V_L = (v_{w1}, v_{w2}, \dots, v_{wn})$ 。因此, (8)式可变换为:

$$J(U) = \max_U \frac{\text{tr}(U^T V_H^T \Lambda_H V_H U)}{\text{tr}(U^T V_L^T \Lambda_L V_L U)}. \quad (9)$$

设 $f(x) = a^x$ 为一般指数函数, $f(A) = a(A)$ 为对应的矩阵函数。本文使用指数函数 $f(x)$ 映射矩阵 S_H 和 S_L 的特征值, 然后将任一特征值映射为相应的指数形式, 即:

$$\lambda_{bi} \rightarrow a^{\lambda_{bi}}, \lambda_{wi} \rightarrow a^{\lambda_{wi}}. \quad (10)$$

做这种映射的目的来自两个方面。首先, 矩阵 S_H 和 S_L 是奇异的, 因此特征值必为 0。通过映射(10)将这些零特征值映射为 1, 则可以避免矩阵 S_H 和 S_L 的奇异性; 其次, 映射(10)具有距离扩散效应, 这将在下一节中详细说明。

根据映射(10), 令 $a(\Lambda_H) = \text{diag}(a^{\lambda_{b1}}, a^{\lambda_{b2}}, \dots, a^{\lambda_{bn}})$ 和 $a(\Lambda_L) = \text{diag}(a^{\lambda_{w1}}, a^{\lambda_{w2}}, \dots, a^{\lambda_{wn}})$, 这样(9)式可变为:

$$J(U) = \max_U \frac{\text{tr}(U^T V_H^T a(\Lambda_H) V_H U)}{\text{tr}(U^T V_L^T a(\Lambda_L) V_L U)}. \quad (11)$$

注意到矩阵 S_H 和 S_L 是半正定的,这样定理 1 成立。也就是说, $a^{\lambda_{bi}}$ 是矩阵函数 $a(S_H)$ 的特征值, v_{bi} 仍然是 $a(S_H)$ 的特征向量, $a^{\lambda_{wi}}$ 是矩阵函数 $a(S_L)$ 的特征值, v_{wi} 仍然是 $a(S_L)$ 的特征向量,也即有: $a(S_H) = V_H^T a(\Lambda_H) V_H$, $a(S_L) = V_L^T a(\Lambda_L) V_L$ 。

因此, (11) 式可表示为:

$$J(U) = \max_U \frac{\text{tr}(U^T a(S_H) U)}{\text{tr}(U^T a(S_L) U)} \quad (12)$$

(12) 式可通过代数变换转化为以下广义特征向量问题:

$$a(S_H)u = \lambda a(S_L)u \quad (13)$$

求解(13)式,并取前面最大的 d 个特征值(按降序排列)所对应的特征向量构成矩阵 U ,也即是所求的最优投影矩阵。

为使降维之后的向量保持原始样本之间的内部结构,进一步对最优投影矩阵 U 做了正交化处理。正交化的处理过程参见文献[22]。设 u_1, u_2, \dots, u_d 是由(13)式求得的 d 个投影向量。设 $z_1 = u_1$, 假设已经对 u_1, u_2, \dots, u_d 中的前 k 个做了正交化,即由 u_1, u_2, \dots, u_k 得到 k 个正交化向量 $z_1, z_2, \dots, z_k (1 \leq k \leq d-1)$, 则第 $k+1$ 个正

$$\text{交向量 } z_{k+1} = u_{k+1} - \sum_{i=1}^k \frac{z_i^T u_{k+1}}{z_i^T z_i} z_i。$$

由此过程确定的向量 $z_1, z_2, \dots, z_k, z_{k+1}$ 是正交的。因此得到经过正交处理之后的投影矩阵 $Z = [z_1, z_2, \dots, z_d]$ 是一个正交矩阵,也即 $ZZ^T = I$ 。

下面从样本之间的欧氏距离来说明正交化的好处。对投影向量正交化后, GEDLPP 的线性投影变为 $y_i = Z^T x_i$, 初始样本点对 x_i 和 x_j 经过降维后得到点对 y_i 和 y_j 。此时, y_i 和 y_j 之间的欧式距离为:

$$d(y_i, y_j) = \|y_i - y_j\| = \|Z^T x_i - Z^T x_j\| = \sqrt{(x_i - x_j)^T Z Z^T (x_i - x_j)} = \|x_i - x_j\|。$$

可以看出 y_i 和 y_j 之间保持了原样本对 x_i 和 x_j 的欧式距离,因此,保持了原样本之间的本质内部结构。

对于 GEDLPP, 基于特征值即(10)式的映射, 矩阵 S_H 和 S_L 分别映射为矩阵指数 $a(S_H)$ 和 $a(S_L)$ 。注意到矩阵 S_H 和 S_L 是由原始样本构成的, 可以认为 GEDLPP 方法隐含了一个从原始样本空间到另一个空间的非线性映射:

$$\begin{aligned} \Theta: \mathbf{R}^{n \times n} &\rightarrow \mathbf{R}^{n \times n}, \\ S_H &\rightarrow \Theta(S_H) = a(S_H), S_L \rightarrow \Theta(S_L) = a(S_L)。 \end{aligned} \quad (14)$$

然后, 基于目标函数(12)式寻找最优投影矩阵。

2.3 GEDLPP 的理论分析

2.3.1 避免小样本问题 从 2.2 节可知, GEDLPP 方法基于新的目标函数(12)式, 并简化为广义特征向量问题。

注意到, 矩阵指数 $a(S_H)$ 和 $a(S_L)$ 的特征值是 $a^{\lambda_{bi}}$ 和 $a^{\lambda_{wi}}$ 的表达形式, 其中 λ_{bi} 和 λ_{wi} 分别是矩阵 S_H 和 S_L 的特征值。显然, $a^{\lambda_{bi}} > 0$ 和 $a^{\lambda_{wi}} > 0$, 所以矩阵指数 $a(S_H)$ 和 $a(S_L)$ 是非奇异矩阵, 则广义特征值问题(13)总是可解的, 从而避免了 DLPP 的小样本问题。

2.3.2 距离扩散效应 对于 DLPP 方法, W_{ij}^c 和 B_{ij} 是权重系数, W_{ij}^c 代表同类样本 x_i 和 x_j 之间的相似性, B_{ij} 是第 i 类和第 j 类均值矢量之间的相似性。样本空间中的类间距离 d_b 和类内距离 d_w 可以表示为:

$$d_b = \frac{1}{2} \sum \|\bar{x}_i - \bar{x}_j\|^2 B_{ij} = \text{tr}(\bar{X} H \bar{X}^T) = \text{tr}(S_H), \quad (15)$$

$$d_w = \frac{1}{2} \sum \|x_i - x_j\|^2 W_{ij}^c = \text{tr}(X L X^T) = \text{tr}(S_L)。 \quad (16)$$

由于矩阵的迹等于矩阵的所有特征值之和, 所以以上两个距离可以写为:

$$d_b = \text{tr}(S_H) = \lambda_{b1} + \lambda_{b2} + \dots + \lambda_{bn}, \quad (17)$$

$$d_w = \text{tr}(S_L) = \lambda_{w1} + \lambda_{w2} + \dots + \lambda_{wn}。 \quad (18)$$

对于 GEDLPP 方法, 通过映射(14), 矩阵 S_H 和 S_L 被映射为 $a(S_H)$ 和 $a(S_L)$, 根据定理 1, (17)和(18)式中的距离 d_b 和 d_w 变为 d_b^a 和 d_w^a :

$$d_b^a = \text{tr}(a(S_H)) = a^{\lambda_{b1}} + a^{\lambda_{b2}} + \dots + a^{\lambda_{bn}}, \quad (19)$$

$$d_w^a = \text{tr}(a(S_L)) = a^{\lambda_{w1}} + a^{\lambda_{w2}} + \cdots + a^{\lambda_{wn}}. \quad (20)$$

通常不同类样本的类间距离大于同类样本的类内距离,即 $d_b > d_w$, 因此对于(17)、(18)式中的大多数 λ_{bi} , λ_{wi} , 应有 $\lambda_{bi} > \lambda_{wi}$ 。注意到如果 $a > 0$, 那么指数函数 $f(x) = a^x$ 是严格的单调递增函数, 则有:

$$\frac{a^{\lambda_{bi}}}{a^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \quad (21)$$

然后, 根据(19)、(20)和(21)式, 有:

$$\frac{d_b^a}{d_w^a} > \frac{d_b}{d_w}. \quad (22)$$

作为一种特殊情况, 如果指数函数的底数 $a = e$, 则不等式(21)仍然成立, 并且有:

$$\frac{a^{\lambda_{bi}}}{a^{\lambda_{wi}}} > \frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \quad (23)$$

从而, 有以下不等式:

$$\frac{d_b^a}{d_w^a} > \frac{d_b^e}{d_w^e} > \frac{d_b}{d_w}. \quad (24)$$

上述不等式(21)和(23)可以通过图 1 来解释, 它表示了特征值在映射前后的几何含义。图 1 中, 点划线表示函数 $y = x$, 它表示映射前的特征值, 虚线表示函数 $y = e^x$, 表示通过函数 $f(\lambda) = e^\lambda$ 映射后的特征值, 实线表示函数 $y = a^x$, 表示通过函数 $f(\lambda) = a^\lambda$ 映射后的特征值。

通过不等式(22), 利用非线性映射函数 Θ , 类间距离 d_b 和类内距离 d_w 分别被扩散到 d_b^a 和 d_w^a , 同时类间距离的扩大尺度比类内距离的扩大程度更大, 这对模式分类是有利的; 另一方面, MEDLPP 也具有这种能力, 但根据不等式(23)、(24), MEDLPP 的扩散效应远小于 GEDLPP。这样, 与 MEDLPP 相比, GEDLPP 将获得更好的模式分类能力。

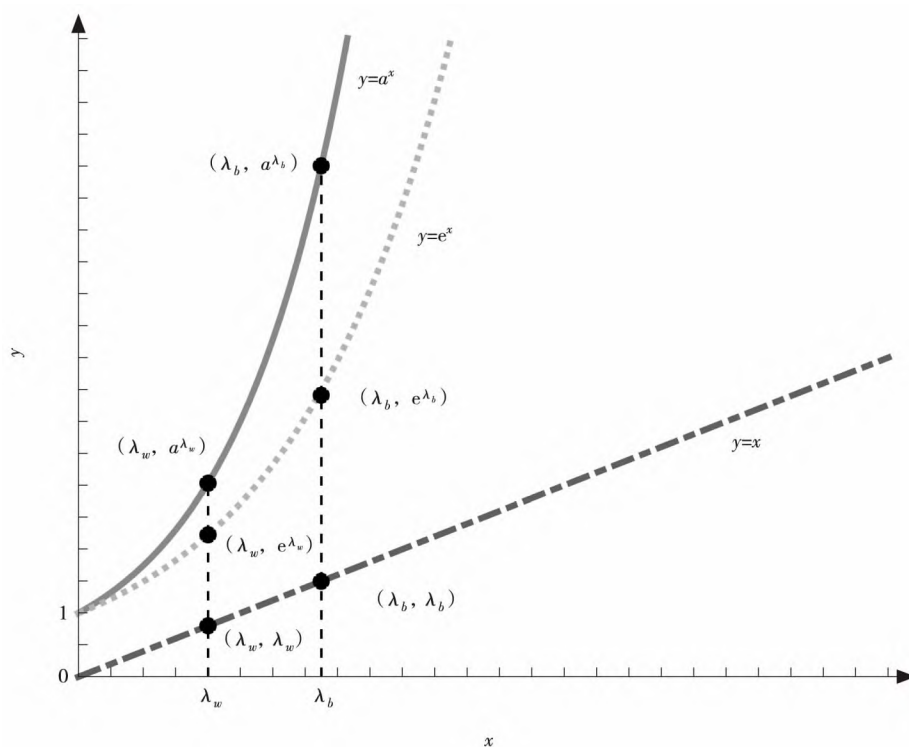
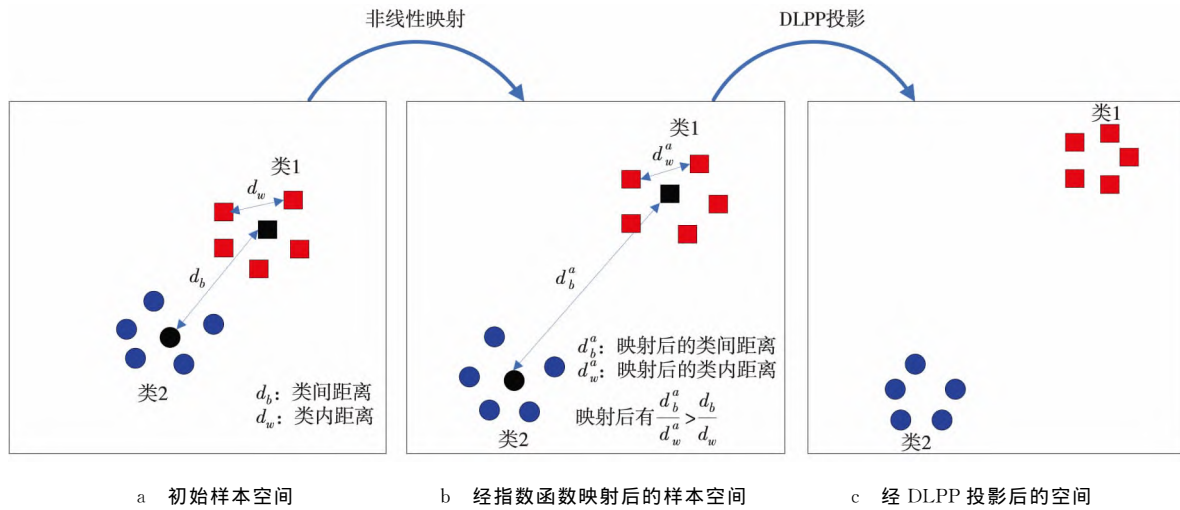


图 1 非线性映射的几何含义

Fig. 1 The geometric meanings of the non-linear mapping

为了直观地解释 GEDLPP 方法的主要思想, 以及非线性映射的距离扩散效应, 图 2 表示了 GEDLPP 的几何解释。为了方便起见, 使用两类样本进行说明。如图 2 所示, 通过非线性映射, GEDLPP 方法将原始样本映射到

另一个空间,在这样一个新的空间上,类间距离和类内距离都被扩大了,但类间距离的扩大尺度比类内距离的扩大程度更大,这样有利于模式分类。最后,在新的空间上,用 DLPP 方法进行投影和特征提取,这样投影后类间距离被拉大,类内距离被缩小。



注:红色正方形表示类别 1,黑色正方形表示类 1 的中心;蓝色圆圈表示类 2,黑色圆圈表示类 2 的中心

图 2 GEDLPP 方法的几何解释

Fig. 2 The geometrical interpretation of GEDLPP using an illustrative example

2.4 指数函数底数 a 的讨论

在 2.2 节中,在一般的指数函数 $f(x)=a^x$,以及相应的矩阵函数 $f(A)=a(A)$ 基础上,提出了 GEDLPP 方法。事实上,指数函数的底数 a 是一个可变的参数。本节讨论底数 a 在取不同值的时候 GEDLPP 方法所代表的情况。

1) 由 GEDLPP 算法的目标函数(12)可知,当 $a=1$ 时,目标函数(12)为: $J(U)=\max_U \frac{\text{tr}(U^T S_H U)}{\text{tr}(U^T S_L U)}$, 此函数即为 DLPP 算法的目标函数,即(5)式。

2) $a=e$ 时,(12)式为: $J(U)=\max_U \frac{\text{tr}(U^T \exp(S_H)U)}{\text{tr}(U^T \exp(S_L)U)}$, 此函数正是 MEDLPP 算法的目标函数^[13-14]。

由此可见,指数函数底数 a 的取值可确定不同的算法,DLPP 和 MEDLPP 方法可视为 GEDLPP 方法的特殊形式,因此 GEDLPP 更具有通用性和代表性。

3 实验结果

为了验证 GEDLPP 方法的有效性,在 4 个公开的图像库上进行了实验。DLPP 以及它的改进算法,如 NDLPP,DLPP/MMC 和 MEDLPP 方法,均采用人脸图像库做实验。为了便于 GEDLPP 方法和上述方法做对比,本文采用了 Yale,Extended Yale B 和 CMU-PIE 共 3 个人脸数据库做实验。此外,也采用一个非人脸图像库,即 COIL-20 物体图像库做实验。

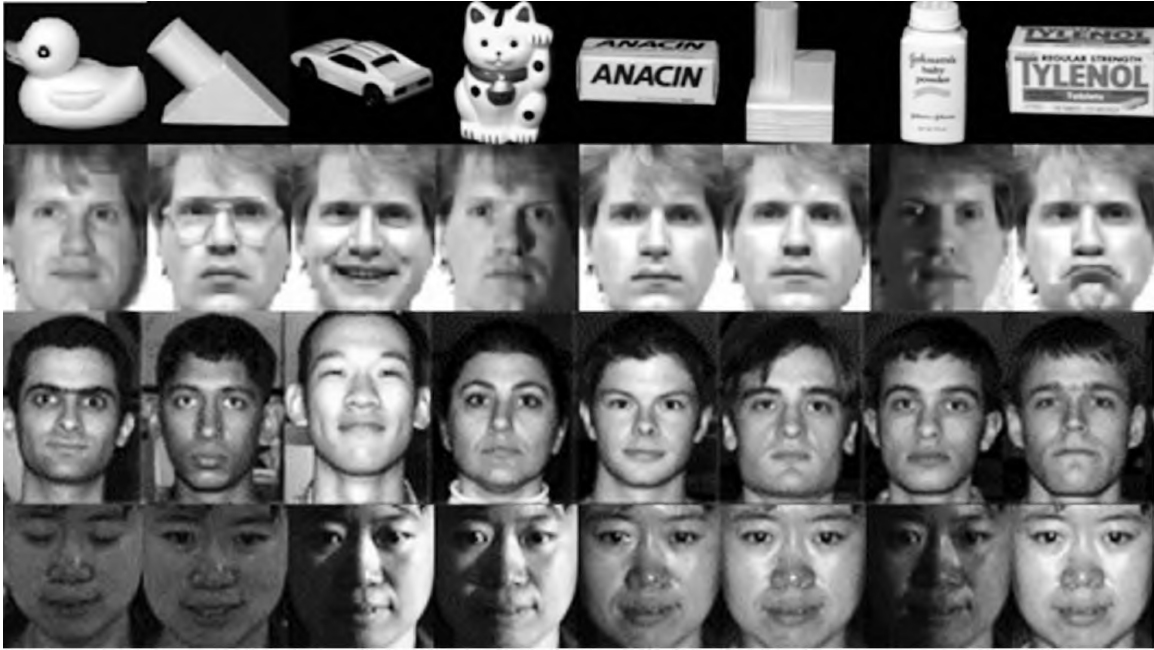
3.1 实验数据集

COIL-20 数据库由哥伦比亚大学提供,该库包含 20 个物体,每个物体每 5° 拍摄一张照片,每个物体共 72 张图像构成。在实验中,图像大小调整为 28×28 像素。

Yale 人脸数据库由 11 个人,每个人 15 张人脸图像构成。所采集的人脸图像有光照、面部表情和是否戴眼镜等变化,在实验中,图像的大小调整为 24×24 像素。

Extended Yale B 人脸库由 38 个人,每个人有 64 张不同的光照条件下的图像组成。图 3 给出了 Extended Yale B 数据库的一些示例图像。在实验中,人脸图像大小调整为 32×32 像素。

在 CMU-PIE 人脸库中,有 68 个人,每个人有 24 个样本。在实验中,图像大小为 28×28 像素。



注:第 1 行是 COIL-20 数据库的样本图像,第 2 行是 Yale 人脸数据库的样本图像,第 3 行是 Extended Yale B 人脸库的样本图像,第 4 行是 CMU-PIE 人脸库的样本图像

图 3 4 个数据集中的部分样本图像

Fig. 3 Some sample images from the four datasets

3. 2 实验设置和结果

在实验中,本文把 GEDLPP 与两种全局降维方法进行了比较:PCA 和 LDA,并与 LPP 和 DLPP 的改进算法做了比较,包括:LPP,DLPP,NDLPP,DLPP/MMC 和 MEDLPP 方法,其中 NDLPP,DLPP/MMC 和 MEDLPP 是解决 DLPP 小样本问题的方法。对于 DLPP 方法,首先使用 PCA 降低原始样本的维数,从而避免 DLPP 的小样本问题。

在实验中,对每个类别设定 p 个训练样本,其余的样本用作测试样本;将降维后的子空间维度按照步长 10 从 10 取到 100,即有 10 个子空间维度。对每一个方法均按照上述的实验设置进行实验。这样,对 p 个训练样本,在每一个子空间维度下,均可得到一个识别率。为了获得稳定的实验结果,在每一个子空间维度下将实验重复 20 次(每次实验中,每个类别随机设定 p 个训练样本)。这样,对于 20 次实验,在每一个维度下就有 20 个识别率,然后将它们的平均值作为当前子空间维度下的识别率。

基于上述实验结果,本文首先评估了这些方法的最佳识别率,用于分析模式分类的性能。对于 p 个训练样本,将 10 个子空间维度下的最大值作为最佳识别率。这样,对每一个方法,在 p 个训练样本下都有一个最佳识别率和相应的最佳子空间维度,以及 20 次实验的标准差。在表 1~4 中列出了在 4 个图像库上这些方法的识别结果。

表 1 COIL-20 图像库实验结果
Tab. 1 The results on COIL-20 database

	% %				
p	8	9	10	11	12
LDA	89. 29±2. 63 (10)	89. 83±2. 43 (20)	90. 32±2. 24 (20)	90. 24±2. 17 (20)	92. 08±2. 62 (10)
DLPP	89. 96±2. 31 (20)	90. 56±1. 82 (30)	91. 38±2. 57 (10)	91. 28±1. 35 (10)	93. 71±2. 13 (20)
NDLPP	90. 03±2. 52 (30)	90. 73±1. 91 (80)	91. 76±1. 78 (90)	91. 37±0. 69 (50)	93. 97±1. 78 (20)
DLPP/MMC	90. 89±2. 26 (80)	91. 81±2. 62 (80)	92. 30±2. 23 (10)	92. 67±0. 79 (80)	94. 19±1. 53 (10)
MEDLPP	91. 96±2. 14 (80)	92. 72±2. 48 (90)	93. 19±2. 42 (10)	93. 78±0. 96 (80)	94. 56±1. 39 (10)
GEDLPP	93. 45±1. 65 (80)	94. 56±1. 29 (90)	94. 39±2. 23 (10)	94. 76±0. 98 (80)	95. 90±1. 31 (10)

注:表中数据为“识别准确度±标准偏差”,括号中数据为最佳子空间维度,数据加粗表示同等条件下最优结果,下同

表2 Yale图像库实验结果
Tab.2 The results on Yale database

p	5	6	7	8	9	%
LDA	75.61±2.13 (10)	78.81±1.68 (10)	81.76±2.65 (20)	83.92±3.46 (10)	87.36±2.85 (20)	
DLPP	78.36±2.84 (70)	79.65±3.29 (80)	85.54±4.61 (10)	88.53±5.12 (20)	89.43±8.10 (30)	
NDLPP	79.62±1.67 (80)	80.56±1.46 (50)	86.81±1.43 (70)	88.82±1.54 (60)	89.98±1.76 (50)	
DLPP/MMC	80.91±4.69 (30)	81.32±3.57 (10)	86.97±6.25 (50)	89.36±7.84 (40)	90.32±5.36 (40)	
MEDLPP	81.69±2.19 (70)	81.65±3.31 (80)	87.31±4.92 (10)	89.78±3.71 (20)	90.97±3.15 (70)	
GEDLPP	84.92±2.43 (70)	83.89±3.10 (80)	88.65±3.16 (10)	90.59±2.39 (20)	92.56±2.34 (70)	

表3 Extended Yale B图像库实验结果
Tab.3 The results on Extended Yale B database

p	18	24	30	36	42	%
LDA	85.98±3.96 (20)	88.56±2.23 (10)	88.91±3.56 (10)	90.27±1.63 (10)	91.53±2.27 (20)	
DLPP	86.81±3.73 (40)	88.76±4.51 (30)	89.73±4.89 (30)	90.73±3.69 (30)	91.72±2.19 (30)	
NDLPP	87.71±2.43 (60)	89.53±2.45 (40)	90.67±2.21 (70)	91.36±2.17 (50)	92.93±2.65 (50)	
DLPP/MMC	87.92±2.34 (100)	89.87±2.97 (40)	91.82±2.01 (90)	91.48±2.04 (90)	93.78±2.72 (80)	
MEDLPP	88.54±3.41 (40)	91.53±3.19 (60)	93.81±4.85 (30)	92.72±3.69 (60)	95.92±3.32 (50)	
GEDLPP	95.73±2.93 (40)	95.74±2.41 (60)	98.13±1.26 (30)	98.21±1.51 (20)	98.67±0.02 (50)	

表4 CMU-PIE图像库实验结果
Tab.4 The results on CMU-PIE database

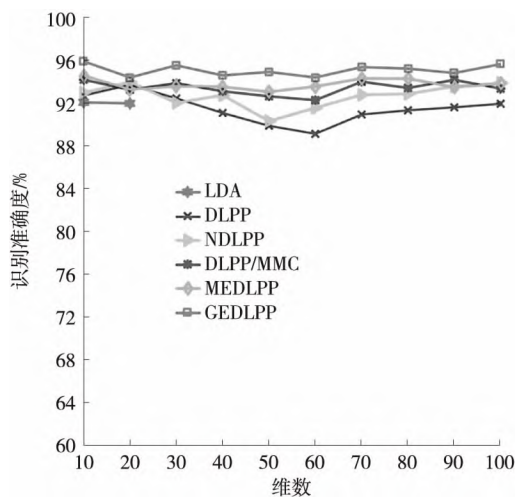
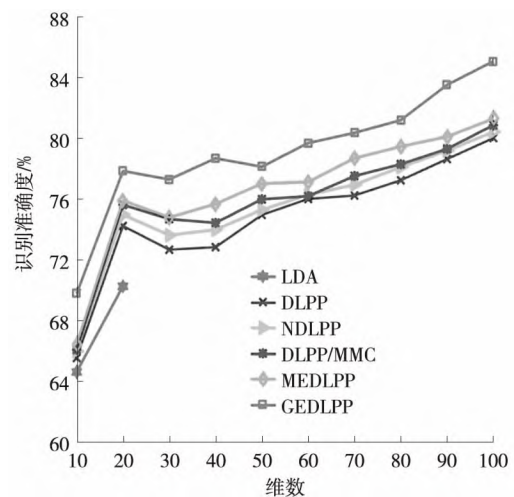
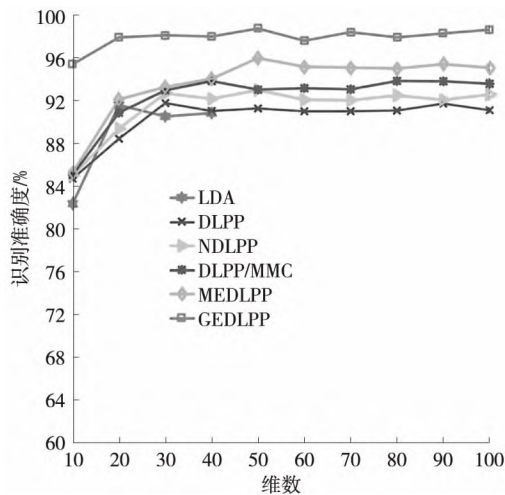
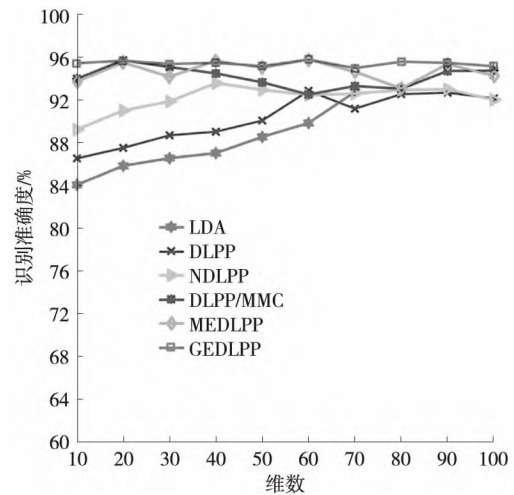
p	7	9	11	13	15	%
LDA	91.73±3.31 (70)	92.91±2.40 (60)	92.71±1.12 (90)	92.91±1.16 (80)	91.69±2.71 (90)	
DLPP	93.49±3.78 (60)	94.51±4.41 (60)	92.82±2.13 (60)	93.81±3.82 (70)	92.76±3.84 (60)	
NDLPP	93.64±2.30 (30)	95.11±2.43 (50)	93.54±2.16 (40)	94.71±2.75 (30)	93.94±2.84 (70)	
DLPP/MMC	94.39±2.83 (60)	95.23±2.91 (70)	95.67±3.24 (20)	96.60±3.26 (60)	97.54±1.98 (80)	
MEDLPP	94.56±4.76 (20)	95.52±3.02 (80)	95.72±2.02 (60)	96.81±1.13 (50)	97.97±1.92 (60)	
GEDLPP	95.45±2.64 (20)	96.38±2.68 (80)	95.74±1.56 (60)	97.45±1.21 (70)	98.28±1.07 (60)	

本文还评估了这些方法在降低到不同子空间维度下的识别性能。根据上述实验结果,对 p 个训练样本,在每一个子空间维度下,均可得到一个平均识别率。为了直观地比较所有方法在不同子空间维度下的识别能力,在4个图像库上,将这些方法在子空间维度从10取到100下的识别率绘制成曲线进行了对比,如图4~7所示。

根据以上结果可以看到:1)与现有方法相比,提出的GEDLPP方法在所有实验数据集上均获得最好的识别率;2)当降维到不同维度时,GEDLPP在所有子空间维度下均取得最好的识别精度,从而表明本文提出的GEDLPP是一种有效的降维方法;3)根据2.4节的讨论,根据指数函数底数 a 的取值情况,DLPP和MEDLPP方法可视为GEDLPP方法的特殊形式。从DLPP、MEDLPP和GEDLPP方法在4个图像库的实验结果可以看到,GEDLPP算法的识别效果最好,MEDLPP算法识别效果其次,DLPP算法识别效果靠后。由此表明,GEDLPP除了更为通用,也更为有效。

4 结束语

本文提出了一种新的降维方法,即GEDLPP。基于矩阵指数函数,GEDLPP用一般的矩阵指数重构DLPP。GEDLPP解决了DLPP的小样本问题,可以扩大类间样本的距离,因此具有更好的模式分类能力。在COIL-20数据库和3个人脸数据库上的实验结果验证了GEDLPP的有效性,并表明GEDLPP相对于现有方法具有更佳的性能。另外,GEDLPP可以作为一种通用方法扩展到其他基于流形的降维方法上,例如NPE,LDE和SDE方法等。

图 4 COIL-20 图像库中算法的性能比较($p=12$)Fig. 4 The comparison of performances versus subspace dimension on COIL-20 database ($p=12$)图 5 Yale 图像库中算法的性能比较($p=3$)Fig. 5 The comparison of performances versus subspace dimension on the Yale face database ($p=3$)图 6 Extended Yale-B 图像库中算法的性能比较($p=42$)Fig. 6 The comparison of performances versus subspace dimension on the Extended Yale B database ($p=42$)图 7 CMU-PIE 图像库中算法的性能比较($p=11$)Fig. 7 The comparison of performances versus subspace dimension on the CMU-PIE database ($p=11$)

参考文献:

- [1] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [2] TENENBAUM J B. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.
- [3] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.
- [4] ZHANG Z, ZHA H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment[J]. Journal of Shanghai University, 2002, 8(4): 406-424.
- [5] HE X F, NIYOGI P. Locality preserving projections[J]. Advances in Neural Information Processing Systems, 2002, 16(1): 186-197.
- [6] HE X F, YAN S, HU Y, et al. Face recognition using laplacian faces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340.
- [7] YU W, TENG X, LIU C. Face recognition using discriminant locality preserving projections[J]. Image and Vision Computing,

- 2006,24(3):239-248.
- [8] ZHU L,ZHU S. Face recognition based on orthogonal discriminant locality preserving projections[J]. Neurocomputing,2007,70(7/8/9):1543-1546.
- [9] JIE G,CHAO W,LING Z. Locality preserving discriminant projections[C]//International Conference on Intelligent Computing Emerging Intelligent Computing Technology and Applications. Berlin:Springer,2009:566-572.
- [10] ZHAO X,TIAN X. Locality preserving fisher discriminant analysis for face recognition[C]//International Conference on Intelligent Computing. Berlin:Springer,2009:261-269.
- [11] YANG L,GONG W,GU X,et al. Null space discriminant locality preserving projections for face recognition[J]. Neurocomputing,2008,71(16/17/18):3644-3649.
- [12] LU G F,LIN Z,JIN Z. Face recognition using discriminant locality preserving projections based on maximum margin criterion [J]. Pattern Recognition,2010,43(10):3572-3579.
- [13] HUANG S C,ZHUANG L. Exponential discriminant locality preserving projection for face recognition[J]. Neurocomputing, 2016,208:373-377.
- [14] LU G F,WANG Y,ZOU J,et al. Matrix exponential based discriminant locality preserving projections for feature extraction [J]. Neural Networks,2018,97:127-136.
- [15] TURK M,PENTLAND A P. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience,1991,3(1):71-86.
- [16] RAN R S,FANG B,WU X G. Exponential neighborhood preserving embedding for face recognition[J]. IEICE Transactions on Information and Systems,2018,E101.D(5):1410-1420.
- [17] WANG S J,CHEN H L,PENG X J,et al. Exponential locality preserving projections for small sample size problem[J]. Neurocomputing,2011,74(17):3654-3662.
- [18] DORNAIKA F,BOSAGHZADEH A. Exponential local discriminant embedding and its application to face recognition[J]. IEEE Transactions on Cybernetics,2013,43(3):921-934.
- [19] DORNAIKA F,TRABOULSI Y E. Matrix exponential based semi-supervised discriminant embedding for image classification [J]. Pattern Recognition,2017,61:92-103.
- [20] WANG S J,YAN S,YANG J,et al. A general exponential framework for dimensionality reduction[J]. IEEE Transactions on Image Processing,2014,23(2):920-930.
- [21] GENE H,CHARLES F. Matrix computations [M]. 4th edition. Baltimore:The Johns Hopkins University Press,2013:530-535
- [22] SONG F,LIU S,YANG J. Orthogonalized Fisher discriminant[J]. Pattern Recognition,2005,38(2):311-313.

Discriminant Locality Preserving Projections Method Based on the Generalized Matrix Exponential

REN Yinshan¹, RAN Ruisheng¹, FANG Bin²

(1. College of Computer and Information Science, Chongqing Normal University, Chongqing 401331;

2. College of Computer Science, Chongqing University, Chongqing 400044, China)

Abstract: [Purposes] Locality Preserving Projections (LPP) is a classic non-linear dimensionality reduction method, which has paid much attention by researchers. Based on the LPP method, people have proposed the Discriminant Locality Preserving Projections method (DLPP), and achieved good results. However, the DLPP method has the small-sample-size (SSS) problem. [Methods] Based on the nature of matrix exponential function, the generalized matrix exponential function is used to reconstruct the DLPP method, and a new dimensionality reduction method is proposed, called General Exponential Discriminant Locality Preserving Projections (GEDLPP). [Findings] The GEDLPP method has two advantages. First, it addresses the SSS problem of the DLPP. Second, the non-linear mapping implied by GEDLPP enlarges the distance between samples of different categories, thereby improving the performance of pattern classification. [Conclusions] The experimental results on COIL-20 database, Yale, extended Yale B and CMU-PIE face datasets show that, compared with the recently proposed improved method to solve the SSS problem of the DLPP, the recognition rate of GEDLPP is superior to state-of-art methods.

Keywords: manifold learning; dimensionality reduction; locality preserving projections; the small-sample-size problem; matrix function

(责任编辑 许 甲)