

广义的监督局部保留投影算法

王晓明 王士同

(江南大学信息工程学院 无锡 214122)

摘要: 针对监督的局部保留投影算法(Supervised Locality Preserving Projection, SLPP)在小样本情况下矩阵的奇异性问题, 该文提出了一种广义的监督局部保留投影算法(Generalized Supervised Locality Preserving Projection, GSLPP)。GSLPP在大样本情况下等价于SLPP, 在小样本情况下却可以等价转换到一个低维空间中求解, 从而有效解决了小样本问题。最后, 实验结果验证了该方法的有效性。

关键词: 特征提取; 局部保留投影; 监督局部保留投影

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2009)08-1840-06

Generalized Supervised Locality Preserving Projection

Wang Xiao-ming Wang Shi-tong

(School of Information, Jiangnan University, Wuxi 214122, China)

Abstract: Supervised Locality Preserving Projection (SLPP) is a generalization of Locality Preserving Projection (LPP) in the case of supervised learning. In this paper the drawback of SLPP in the high-dimensional and small sample size case is pointed out, and a new algorithm called Generalized Supervised Locality Preserving Projection (GSLPP) is proposed. The relationship between SLPP and GSLPP is theoretically analyzed. In the small sample size case GSLPP can be solved equivalently in lower-dimensionality space. Finally, the effectiveness of the proposed algorithm is verified by experimental results.

Key words: Feature extraction; Locality Preserving Projection (LPP); Supervised Locality Preserving Projection (SLPP)

1 引言

模式识别和机器学习中, 经常会遇到数据特征的高维问题, 亦即人们所说的“维数灾难”(Curse of Dimensionality)。为了有效地进行数据分析, 数据特征维数的降低就显得异常重要。特征选择(Feature selection)和特征提取(Feature extraction)就是特征降维的两种主要方法^[1]。

特征提取又可以根据使用训练样本的类标与否来分为监督和无监督的特征提取。最近, 一种叫做局部保留投影(Locality Preserving Projects, LPP)^[2,3]新的特征提取降维方法被提出, 其主要是针对无监督的情况。Shen等人在文献[4]中指出, 由于LPP没有使用类标从而使得不同类别而又靠得较近的数据在投影后还是靠得较近。针对这种情况, 文献[4]中作者提出了监督的局部保留投影算法(Supervised Locality Preserving Projects, SLPP), 在一定程度上解决了这个问题。SLPP形式上类似于

Fisher判别分析(Fisher Discriminant Analysis, FDA)^[5,6], 其优化问题最终也转换为一个广义特征值问题。但是同FDA一样, 其主要的问题是在小样本情况下^[7]会遭遇到矩阵的奇异性问题, 从而不能直接用广义特征值问题来求解。然而, 在实际应用中经常会遇到小样本情况, 如在人脸识别问题中。因此, 针对小样本问题, 进一步研究和探讨怎么来改进SLPP显得十分必要。

针对上述情况, 本文在SLPP的基础上提出了一种叫做广义的监督局部保留投影算法(Generalized Supervised Locality Preserving Projects, GSLPP)。GSLPP在大样本情况下等价于SLPP, 在小样本情况下却可以等价地转换到一个低维空间中求解, 从而有效的解决了小样本问题。这个过程不同于先采用PCA(Principal Components Analysis)^[8,9]来预处理样本数据。后者是把数据预先压缩到一个维数较低的空间, 但是可能丢失数据有用的鉴别信息。然而GSLPP则没有这样的问题, 其解决小样本问题有着坚实的理论基础。对于在小样本情况可能会遇到的计算问题, 文中也介绍一种有效的解决方法。本文用数据的可视

2008-07-24 收到, 2009-03-09 改回

国家863计划项目(2007AA1Z158, 2006AA10Z313)和国家自然科学基金(60704047)资助课题

化来验证了在大样本情况下 GSLPP 和 SLPP 等价性, 并在两个人脸识别问题上测试了其在小样本情况下的性能, 并和其它一些方法进行了比较。

2 监督局部保留投影

这一节将对 SLPP 进行简单介绍, 详细内容可以参考文献[4]。本文假定有一个含有 N 个 M 维样本的数据集 $\{(\mathbf{x}_1, z_1), \dots, (\mathbf{x}_N, z_N)\}$, 该数据集为 c 类。样本数据 $\mathbf{x}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{iM}]^T \in \mathbf{X}$, 类标 $z_i \in \mathbf{Z} = \{1, 2, \dots, c\}$, $i = 1, \dots, N$ 。其中, \mathbf{x}_{im} 为 i 个训练样本的第 m 维特征, z_i 为 i 个的类别标示, \mathbf{T} 表示向量或矩阵转置, $\mathbf{X} \in \mathbf{R}^M$ 。

原始的 LPP 主要针对的无监督的情况。Shen 等人在文献[4]中把 LPP 进行了推广, 提出了一种叫做监督的局部保留投影算法(SLPP)。在 SLPP 中, 首先是建立类间邻接图和类内邻接图[4], 然后分别定义了样本类内矩阵 \mathbf{B} 和类间的相似性矩阵 \mathbf{W} :

$$B_{ij} = \begin{cases} S_{ij}, & \text{当 } \mathbf{x}_i \text{ 和 } \mathbf{x}_j \text{ 相连} \\ 0, & \text{其他} \end{cases}, \quad W_{ij} = \begin{cases} S_{ij}, & \text{当 } \mathbf{x}_i \text{ 和 } \mathbf{x}_j \text{ 相连} \\ 0, & \text{其他} \end{cases} \quad (1)$$

为了减少构建邻接图的过程, 我们可以采用下式定义样本类内矩阵 \mathbf{B} 和类间矩阵 \mathbf{W} :

$$B_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t), & \text{当 } z_i \neq z_j \\ 0, & \text{其他} \end{cases}, \quad W_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t), & \text{当 } z_i = z_j \\ 0, & \text{其他} \end{cases} \quad (2)$$

这里, $\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t)$ 被叫作热核 (heat kernel)[2,3], 参数为 t ; $\|\mathbf{x}\|$ 为向量的 L_2 范数; B_{ij} 为类间的样本的相似性, W_{ij} 为类内的样本的相似性。SLPP 的优化问题定义为

$$\begin{aligned} \max_{\mathbf{w}} J_{\text{SLPP}}(\mathbf{w}) &= \max_{\mathbf{w}} \frac{\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{y}_i - \mathbf{y}_j)^2 B_{ij}}{\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{y}_i - \mathbf{y}_j)^2 W_{ij}} \\ &= \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{L}^B \mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} \mathbf{L}^W \mathbf{X}^T \mathbf{w}} \end{aligned} \quad (3)$$

其中 $\mathbf{L}^B = \mathbf{D}^B - \mathbf{B}$, $\mathbf{L}^W = \mathbf{D}^W - \mathbf{W}$ 。这里, \mathbf{D}^B 和 \mathbf{D}^W 都是对角矩阵, 它们的元素分别为 \mathbf{B} 和 \mathbf{W} 对应的行或者列上的元素之和(注意, \mathbf{B} 和 \mathbf{W} 都是对称矩阵)。可以看出, SLPP 的主要思想是把样本间的相似性分为类间和类内两部分, 投影结果是保留类内局部结构的同时最大化类间分离度。SLPP 的优化问题也可以转化为了一个广义特征值问题:

$$\mathbf{X} \mathbf{L}^B \mathbf{X}^T \mathbf{w} = \lambda \mathbf{X} \mathbf{L}^W \mathbf{X}^T \mathbf{w} \quad (4)$$

这样, SLPP 寻找的最优投影方向就对应于这个广义特征值问题的最大特征值所对应的特征向量。应该注意到, $\mathbf{X} \mathbf{L}^W \mathbf{X}^T$ 是 $M \times M$ 矩阵, 在小样本情况下^[7] ($M \gg N$) $\mathbf{X} \mathbf{L}^W \mathbf{X}^T$ 可能会出现奇异性, 从而不能直接采用广义特征值问题来解决 SLPP 的优化问题。

3 广义的监督局部保留投影

在这一节, 将首先介绍 GSLPP, 然后分析它和 SLPP 的联系, 并讨论其在小样本情况下的求解过程。

3.1 广义的监督局部保留投影(GSLPP)

为了后面讨论的方便, 设定:

$$\begin{aligned} \mathbf{Z}_B &= \mathbf{X} \mathbf{L}^B \mathbf{X}^T, \quad \mathbf{Z}_W = \mathbf{X} \mathbf{L}^W \mathbf{X}^T, \\ \mathbf{Z}_T &= \mathbf{X} (\mathbf{L}^B + \mathbf{L}^W) \mathbf{X}^T = \mathbf{Z}_B + \mathbf{Z}_W \end{aligned} \quad (5)$$

值得注意的是, 其中 \mathbf{Z}_B , \mathbf{Z}_W 和 \mathbf{Z}_T 都是对称的, 而且都是半正定的。为了有效地处理小样本问题, 通过修改 SLPP 的目标函数, 定义广义的监督局部保留投影算法(GSLPP)的优化问题定义为

$$\begin{aligned} \max_{\mathbf{w}} J_{\text{GSLPP}}(\mathbf{w}) &= \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{L}^B \mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} (\mathbf{L}^W + \mathbf{L}^B) \mathbf{X}^T \mathbf{w}} \\ &= \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{Z}_B \mathbf{w}}{\mathbf{w}^T \mathbf{Z}_T \mathbf{w}} \end{aligned} \quad (6)$$

其求解方式也可以转化为一个广义特征值问题:

$$\mathbf{Z}_B \mathbf{w} = \lambda \mathbf{Z}_T \mathbf{w} \quad (7)$$

这样, GSLPP 寻找的最优投影方向就对应于这个广义特征值问题的最大特征值所对应的特征向量。GSLPP 和 SLPP 看上去有着非常相似的形式和求解过程, 甚至在小样本情况下同样会遇到矩阵 \mathbf{Z}_T (与 \mathbf{Z}_W 一样是 $M \times M$ 矩阵) 的奇异问题。但是通过后面的分析我们将发现, GSLPP 在大样本情况下等价于 SLPP, 而在小样本情况的求解过程中, 有着 SLPP 所没有的而又至关重要的优点。它可以等价转换到一个低维空间中求解, 而在这个低维空间中则不再有矩阵的奇异问题。

3.2 GSLPP 与 SLPP 的联系

这节将介绍在 $\mathbf{Z}_W = \mathbf{X} \mathbf{L}^W \mathbf{X}^T$ 非奇异情况下 GSLPP 与 SLPP 的等价性。首先来介绍几个定理。

定理 1 当 \mathbf{Z}_W 非奇异时, 广义特征方程 $\mathbf{Z}_B \mathbf{w} = \lambda \mathbf{Z}_T \mathbf{w}$ 的任意特征值 λ 满足 $0 \leq \lambda < 1$ 。

证明 当 \mathbf{Z}_W 非奇异时, \mathbf{Z}_W 为正定矩阵, 所以有 $\mathbf{w}^T \mathbf{Z}_W \mathbf{w} > 0$ 。由于

$$\mathbf{w}^T \mathbf{Z}_T \mathbf{w} = \mathbf{w}^T (\mathbf{Z}_B + \mathbf{Z}_W) \mathbf{w} > \mathbf{w}^T \mathbf{Z}_B \mathbf{w} \quad (8)$$

从而, $\lambda = \frac{\mathbf{w}^T \mathbf{Z}_B \mathbf{w}}{\mathbf{w}^T \mathbf{Z}_T \mathbf{w}} < 1$ 。

又由于 Z_B 半正定, 即 $w^T Z_B w \geq 0$, 所以有 $\lambda = \frac{w^T Z_B w}{w^T Z_T w} \geq 0$. 证毕

定理 2 当 Z_W 非奇异时, 假定 β 是广义特征方程 $Z_B w = \lambda Z_T w$ 属于 λ 特征值所对应的特征向量, 则 β 是广义特征方程 $Z_B w = \lambda Z_W w$ 属于特征值 $\lambda/(1-\lambda)$ 所对应的特征向量。

证明 如果 β 是广义特征方程 $Z_B w = \lambda Z_T w$ 属于 λ 特征值所对应的特征向量, 则可以得出 $Z_B \beta = \lambda Z_T \beta$ 。

由于 $Z_T = Z_B + Z_W$, 则 $Z_B \beta = \lambda Z_T \beta = \lambda(Z_B + Z_W)\beta$, 即 $(1-\lambda)Z_B \beta = \lambda Z_W \beta$ 。由于定理 1 易得 $1-\lambda > 0$, 所以有 $Z_B \beta = [\lambda/(1-\lambda)] Z_W \beta$ 。命题得证。

证毕

这里我们假定 λ_1 和 λ_2 是广义特征方程 $Z_B w = \lambda Z_T w$ 的两个不同特征值, 并且 $\lambda_1 > \lambda_2$, 其对应的特征向量分别为 β_1 和 β_2 。由定理 2 易得, $\lambda_1/(1-\lambda_1)$ 和 $\lambda_2/(1-\lambda_2)$ 即是 $Z_B w = \lambda Z_W w$ 的特征值, 其对应的特征向量也是 β_1 和 β_2 , 而且有 $\lambda_1/(1-\lambda_1) > \lambda_2/(1-\lambda_2)$ 。所以, 这两个特征方程的最大特征值分别所对应的特征向量是相同的。因而在 Z_W 非奇异的情况下, SLPP 和 GSLPP 的优化问题式(3)和式(6)转化为的广义特征值问题式(4)和式(7)有着相同的最大特征值对应的特征向量, 从而式(3)和式(6)在相同的点分别取得了目标函数的最优值。所以当 Z_W 非奇异时, GSLPP 等价于 SLPP。

3.3 小样本情况下的 GSLPP

在上一节中, 我们指出了在 Z_W 非奇异的情况下, SLPP 和 GSLPP 是等价的。然而在小样本情况下, Z_W 则往往奇异, 甚至 Z_T 也是奇异的。此时它们不能直接转化为广义特征值问题来求解。一个可行的办法就是把问题转换到 R^M 的一个低维子空间中进行求解。后面将会看到, GSLPP 在子空间进行求解不会损失任何有用的鉴别信息, 也就是说这个转换是等价的, 然而 SLPP 则不是。这也就是 GSLPP 的最大优势所在。下面将具体分析和探讨 GSLPP 在小样本情况下的求解问题。

设 $\beta_1, \beta_2, \dots, \beta_M$ 表示 Z_T 的标准正交的特征矢量系, 则空间 R^M 可以用这个标准正交的特征矢量系张成, 表示为 $R^M = \text{span}\{\beta_1, \dots, \beta_M\}$ 。

定义 1 定义 R^M 的子空 Ω 由 Z_T 的非零特征值所对应的特征向量所张成, 即 $\Omega = \text{span}\{\beta_1, \dots, \beta_r\}$ 。其正交补空间为 $\Psi = \text{span}\{\beta_{r+1}, \dots, \beta_M\}$, 即由 Z_T 的零特征值所对应的特征向量所张成。

可以看出, $r = \text{rank}(Z_T)$ 为 Z_T 的秩, $R^M =$

$\Omega + \Psi$ 。下面介绍几个定理。

定理 3 假定矩阵 A 为半正定矩阵, 则 $w^T A w = 0$ 当且仅当 $A w = 0$ 。这里 $0 = [0, \dots, 0]^T \in R^M$ 为元素全零向量。

证明 由于 A 为半正定矩阵, 其特征值均为非负实数, 对其进行特征分解得到:

$$A = U \Lambda U^T = U \Lambda^{1/2} U^T U \Lambda^{1/2} U^T = P^T P \quad (9)$$

其中 Λ 为对角阵, 对角线元素为 A 的特征值; U 为标准正交阵, 列矢量为 Λ 中特征值对应的特征向量; $P = U \Lambda^{1/2} U^T$ 。因此, 有:

$$w^T A w = w^T P^T P w = \|P w\|^2 = 0 \Rightarrow$$

$$P w = 0 \Rightarrow P^T P w = 0 \Rightarrow A w = 0 \quad (10)$$

由 $A w = 0$, 容易得到 $w^T A w = 0$ 。所以, $w^T A w = 0$ 当且仅当 $A w = 0$ 。证毕

定理 4 假如 Z_T 奇异, $w^T Z_T w = 0$ 当且仅当 $w^T Z_B w = 0$ 和 $w^T Z_W w = 0$ 。

证明 因为 Z_B 和 Z_W 半正定的, 所以 $w^T Z_B w \geq 0$, $w^T Z_W w \geq 0$ 。又由于 $w^T Z_T w = w^T (Z_B + Z_W) w = w^T Z_B w + w^T Z_W w$, 所以有 $w^T Z_T w = 0$ 与 $w^T Z_B w = 0$ 和 $w^T Z_W w = 0$ 等价。证毕

利用上面几个定理, 可以得到下面一个重要的定理。

定理 5 假定 $\beta \in \Omega$, 则有 $J_{\text{GSLPP}}(w) = \frac{w^T Z_B w}{w^T Z_T w} = \frac{\beta^T Z_B \beta}{\beta^T Z_T \beta} = J_{\text{GSLPP}}(\beta)$ 。

证明 根据前面所述, 易得 $w = \beta + \gamma$, 这里 $w \in R^M$, $\beta \in \Omega$, $\gamma \in \Psi$ 。

由于 $\gamma \in \Psi$, 所以有 $\gamma^T Z_T \gamma = 0$, 由定理 4 进一步可得 $\gamma^T Z_B \gamma = 0$ 。再通过定理 3, 进而有 $Z_T \gamma = 0$, $Z_B \gamma = 0$ 。又由于 Z_T 和 Z_B 都是对称矩阵, 所以有 $\gamma^T Z_T \beta = (\beta^T Z_T \gamma)^T = 0$, $\gamma^T Z_B \beta = (\beta^T Z_B \gamma)^T = 0$ 。所以

$$\left. \begin{aligned} w^T Z_T w &= (\beta + \gamma)^T Z_T (\beta + \gamma) = \beta^T Z_T \beta \\ &\quad + \gamma^T Z_T \beta + \beta^T Z_T \gamma + \gamma^T Z_T \gamma = \beta^T Z_T \beta \\ w^T Z_B w &= (\beta + \gamma)^T Z_B (\beta + \gamma) = \beta^T Z_B \beta \\ &\quad + \gamma^T Z_B \beta + \beta^T Z_B \gamma + \gamma^T Z_B \gamma = \beta^T Z_B \beta \end{aligned} \right\} (11)$$

因此, $J_{\text{GSLPP}}(w) = \frac{w^T Z_B w}{w^T Z_T w} = \frac{\beta^T Z_B \beta}{\beta^T Z_T \beta} = J_{\text{GSLPP}}(\beta)$ 。

证毕

定理 5 提示我们, 对于 GSLPP 目标函数的优化问题, 可以等价在由 Z_T 的非零特征值所对应的特征向量所张成的 R^M 的子空 Ω 中求解。由于空间 Ω 为 r 维, 根据线性代数理论, 其必然同构于 R^r 空

间。相应同构映射可以定义为 $\beta = P v$, $\beta \in \Omega$, $v \in R^r$, $P = [\beta_1, \dots, \beta_r]$ 。这样, $J_{GSLPP}(\beta)$ 又可以改写为

$$J_{GSLPP}(\beta) = \frac{\beta^T Z_B \beta}{\beta^T Z_T \beta} = \frac{(P v)^T Z_B (P v)}{(P v)^T Z_T (P v)} = \frac{v^T P^T Z_B P v}{v^T P^T Z_T P v} = \frac{v^T \tilde{Z}_B v}{v^T \tilde{Z}_T v} = \tilde{J}_{GSLPP}(v) \quad (12)$$

其中 $\tilde{Z}_B = P^T Z_B P$, $\tilde{Z}_T = P^T Z_T P$, $\tilde{Z}_T = P^T Z_T P = \Lambda$ 。这里 Λ 为对角阵, 其对角线上的元素为 Z_T 的非零特征值。这里易得 Λ 为正定矩阵, 将不再存在其奇异性问题。这样, 原来 GSLPP 的准则函数就可以转换到一个低维的空间来求解, 而且没有损失任何鉴别信息, 也就是这个过程是等价的, 而又成功解决了矩阵的奇异性问题。但是, 如果在 $J_{SLPP}(w)$ 中对于 Z_W 为采用同样的方法, 则不能保证 Z_W 零特征值所对应的特征向量 γ , 使得 $\gamma^T Z_B \gamma = 0$ 成立, 从而不能保证在 Z_W 非零特征值所对应的特征向量张成的空间中求解的目标函数是等价于原来的目标函数。这就是 GSLPP 算法优势所在。这样, 广义特征值问题式(7)就转化为为了:

$$\tilde{Z}_B v = \lambda \tilde{Z}_T v \quad (13)$$

下面简要介绍 GSLPP 算法的过程。

步骤 1 分别构造类内和类间邻接图, 并计算 B 和 W (详细参考文献[6]);

步骤 2 根据式(5)计算 Z_T , 并求得其标准正交的特征矢量系。令 $P = [\beta_1, \dots, \beta_r]$, 这里 β_1, \dots, β_r 是 Z_T 的非零特征值所对应的特征向量;

步骤 3 求解广义特征值问题式(13), 令其前 d 个最大特征值所对应的特征向量分别为: v_1, \dots, v_d , 并令 $A = [v_1, \dots, v_d]$;

步骤 4 对于原始空间的向量 x , 通过如下的方式嵌入到 d 维空间中: $y = A^T P^T x$ 。这里, $x \in R^M$, $y \in R^d$ 。

4 计算问题

在小样本问题中, 尤其是在人脸数据中, 样本维数 M 往往非常高, 所以 Z_T (是 $M \times M$ 矩阵) 将是一个很大的矩阵, 其标准正交的特征矢量系和特征值的求解将会耗费很长时间。为了解决这个问题, 本文采用类似于文献[10]中的技巧, 对其进行转换求解。

假定有一特征值问题 $(L^B + L^W) X^T X u = \lambda u$, 则有 $X(L^B + L^W) X^T X u = \lambda X u$, 即 $Z_T(X u) = \lambda(X u)$ 。所以可以先求特征方程 $(L^B + L^W) X^T X u = \lambda u$, 然后再做变换 $\beta = X u$ 得到原始的特征向量。这里应该注意到, $(L^B + L^W) X^T X$ 是 $N \times N$ 矩阵,

在小样本情况中其大大小于 $M \times M$ 矩阵, 因而通过这样的方式提高了 GSLPP 在小样本情况下的运算效率。而 $\tilde{J}_{GSLPP}(v)$ 已经在一个维数较低的 R^M 的子空间中了, 所以其不必再进行转换了。

5 实验

在这一节中, 我们将进行实验研究。实验分为两部分, 第 1 部分为数据的二维可视化; 第 2 部分是小样本的人脸识别问题。实验中, 本文把 GSLPP 和 FDA, LFDA^[11], SLPP 进行了比较。其中, LFDA 是一种结合 LPP 和 FDA 的监督算法, 也面临着小样本的问题。文献[4]指出, 对于热核参数 t 一般情况下可以采用所有样本间距离的平均值。

5.1 数据可视化

在这一部分的实验中, 我们将进行数据的二维可视化测试。实验数据采用了 Pipeline Flow(<http://www.ncrg.aston.ac.uk/GTM/3PhaseData.html>) 数据。Pipeline Flow 数据包含 oil, water 和 gas 三类 12 维样本数据, 训练和测试样本数据各共 1000 个。我们取出其中的训练样本来进行可视化测试。

图 1 给出了 Pipeline Flow 数据的二维可视化效果。图 1(a)和图 1(b)是采用了式(1)定义相似性的结果, 图 1(c)和图 1(d)是采用了式(2)定义相似性的结果。从二维可视化结果我们可以看出, GSLPP 和 SLPP 所得到的结果一样, 这也验证了前面的理论分析, 即在大样本情况下它们是等价的。也说明了采用式(2)定义相似性是可行的。采用式(2)定义相似性最大好处就是不用构建邻接图, 节省构建邻接图的时间。

5.2 人脸识别

针对小样本情况, 本文将在在两个人脸图像数据上对 GSLPP 进行实验测试。第 1 个数据是 ORL

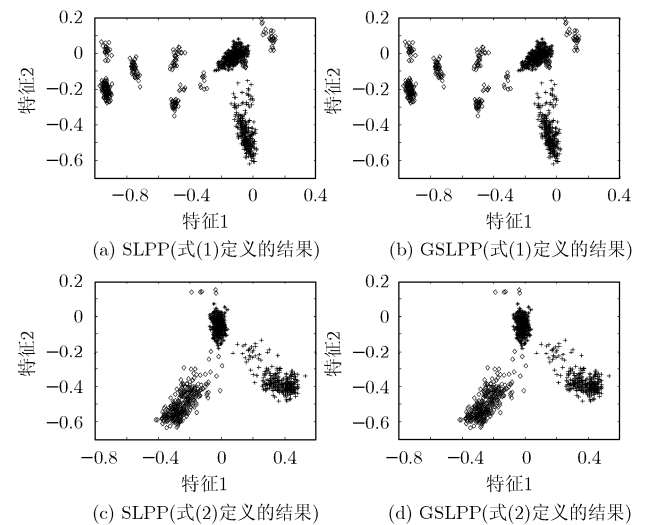


图 1 Pipeline Flow 数据的可视化结果

(<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>) 人脸数据。ORL 标准人脸库由 40 个人、每人 10 幅分辨率为 92×112 的人脸图像组成。第 2 个是 Yale (<http://www.uk.research.att.com/facedatabase.htm>) 人脸数据库。Yale 数据库包含 15 个人共 165 个人脸，每个人都有 11 个样本。在

文献[9]中，所有的人脸图像都处理为 32×32 的图像，并进行了诸如两个眼睛被处理成同一位置等的其它一些处理。这两个人脸数据处理后的 matlab 数据可以从 <http://www.cs.uiuc.edu/homes/dengcai2/Data/data.html> 处获得。图 2 分别为它们经过预处理后的部分人脸图像。

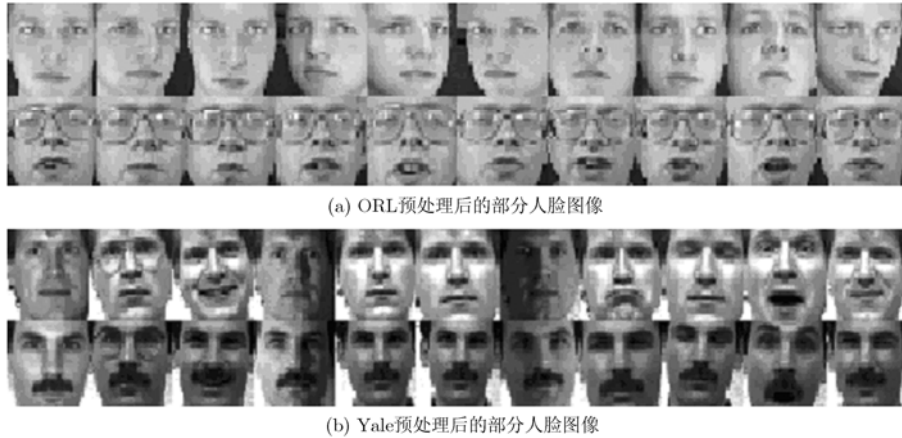


图 2 Yale 预处理后的部分人脸图像

实验中，我们采用了 KNN(K Nearest Neighbor)^[9, 12]来作为人脸图像数据进行投影后的测试分类器，其中采用了 $k = 5$ 。我们从每一类数据中随机抽取出 8 个样本作为训练样本，剩下的作为测试样本。重复同样的过程 20 次，最后把这 20 次的平均识别率作为最后的识别率。由于 FDA, LFDA, SLPP 都面临着矩阵的奇异性问题，所以我们先采用了 PCA 来对原始数据进行维数压缩，然后再在这个低维的压缩后的数据上采用这几种算法进行实验。对于 Yale 数据，我们先用 PCA 把其压缩到 50 维，而 ORL 是 100 维。而对于 GSLPP，我们则是直接采用前面描述的算法过程，没有预先对数据进行维数的压缩。

图 3 给出了这几种算法把 Yale 和 ORL 人脸图像数据投影到不同维数后的识别率。注意，FDA 最多能把数据投影到 $c - 1$ 维空间中，这里 c 为数据的类别数。在实验中我们把其第 $c - 1$ 维的实验结果作为后续的维数虚拟的实验结果。这样是为了其它方法能和其进行对比。从实验结果可以看出，在把数据投影到维数较低的空间时，相对于其它几种算法 GSLPP 表现并不理想，然而伴随着投影维数的增加，GSLPP 的性能迅速上升，在多数情况下都表现最佳。其中原因可能是其它几种方法都预先采用了 PCA 来对数据进行了维数的压缩。由于 PCA 具有一定抑制噪音的作用，但是进行维数压缩也可能损

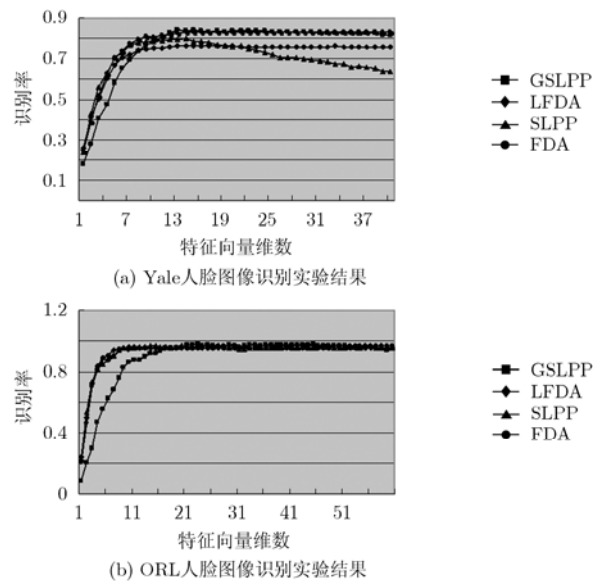


图 3 8 个训练样本识别率

失数据的鉴别信息。然而，对于 SLPP 和 GSLPP 算法而言，从 Yale 数据的实验结果可以看出，GSLPP 明显表现更好，而且是比较稳定；在 ORL 数据的结果中 GSLPP 也表现出了略好或相当的结果。这说明了 GSLPP 对于处理小样本的有效性和优势。

6 结束语

LPP 是一种近来备受关注的降维方法。SLPP

是对其监督的推广, 而本文则在 SLPP 的基础上进一步进行了推广, 提出了 GSLPP。本文从理论上分析了 GSLPP 在大样本情况下等价于 SLPP, 而且能有效地解决 SLPP 在小样本情况下的问题。实验结果验证了 GSLPP 的可行性。然而, 相对于其它几种监督算法, 对于小样本在提取低维特征的时候表现并不理想。尽管文中我们指出, 这可能与其它算法采用了 PCA 来对数据进行了维数的压缩有关, 但是进一步研究其根本原因, 进而改进将是需要进一步研究的方向之一。

参考文献

- [1] 宋枫溪, 高秀梅, 刘树海等. 统计模式识别中的维数削减与低损降维[J]. 计算机学报, 2005, 28(11): 1915-1922.
Song Feng-xi, Gao Xiu-mei, and Liu Shu-hai, *et al.* Dimensionality reduction in statistical pattern recognition and low loss dimensionality reduction [J]. *Chinese Journal of Computers*, 2005, 28(11): 1915-1922.
 - [2] He X and Niyogi P. Locality preserving projections [C]. Proc. Conf. Advances in Neural Information Processing Systems, Vancouver, Canada, 2003: 585-591.
 - [3] Kokiopoulou E and Saad Y. Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(12): 2143-2156.
 - [4] 申中华, 潘永惠, 王士同. 有监督的局部保留投影降维算法[J]. 模式识别与人工智能, 2008, 21(2): 233-239.
Shen Zhong-hua, Pan Yong-hui, and Wang Shi-tong. A supervised locality preserving projection algorithm for dimensionality reduction [J]. *Pattern Recognition and Artificial Intelligence*, 2008, 21(2): 233-239.
 - [5] Tao Q, Wu G W, and Wang J. The theoretical analysis of FDA and applications [J]. *Pattern Recognition*, 2006, 39(6): 1199-1204.
 - [6] 杨键, 杨静宇, 叶晖等. Fisher线性鉴别分析的理论研究及其应用[J]. 自动化学报, 2003, 29(4): 482-493.
Yang Jian, Yang Jing-yu, Ye Hui, *et al.* Theory of fisher linear discriminant analysis and its application [J]. *Acta Automatic Sinica*, 2003, 29(4): 482-493.
 - [7] Liu J, Chen S C, and Tan X Y. A study on three linear discriminant analysis based methods in small sample size problem. *Pattern Recognition*, 2008, 41(1): 102-116.
 - [8] Zhuang X S and Dai D Q. Improved discriminant analysis for high-dimensional data and its application to face recognition [J]. *Pattern Recognition*, 2007, 40(5): 1570-1578.
 - [9] Cai D, He X, and Han J, *et al.* Orthogonal laplacianfaces for face recognition [J]. *IEEE Transactions on Image Processing*, 2006, 15(11): 3608-3614.
 - [10] Turk M and Pentland A. Eigenfaces for recognition [J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.
 - [11] Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis [J]. *Journal of Machine Learning Research*, 2007, 8(5): 1027-1061.
 - [12] Jiang L X, Cai Z H, and Wang D H, *et al.* Survey of improving k-nearest-neighbor for classification [C]. Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Haikou, China, Aug 24-27, 2007: 679-683.
- 王晓明: 男, 1977 年生, 博士生, 从事人工智能、模式识别的研究。
王士同: 男, 1964 年生, 教授, 博士生导师, 从事人工智能、机器学习等研究。