

# 适用于小样本的双邻接图判别分析算法

罗璇<sup>1,2</sup> 张莉<sup>1,2</sup> 薛杨涛<sup>1,2</sup> 李凡长<sup>1,2</sup>

- (1. 苏州大学计算机科学与技术学院, 苏州, 215006;
2. 苏州大学机器学习与类脑计算国际合作联合实验室, 苏州, 215006)

**摘要:** 作为一种常用的降维方法, 适用于小样本的监督化拉普拉斯判别分析方法通过使用图嵌入的判别近邻分析得到了很好的降维效果。但该方法在构建近邻图时, 在  $K$  近邻中寻找同类和异类样本点存在数据不平衡问题; 此外, 在优化该方法的目标函数时, 没有全面考虑到类间信息, 从而会在一定程度上降低该方法的性能。针对以上两个问题, 本文提出了适用于小样本的双邻接图判别分析方法。首先该方法分别在同类和异类样本中找出  $K$  个近邻点, 然后使用这  $K$  个类内近邻点和  $K$  个类间近邻点来构造双邻接图, 这样可以确保邻接图中既有同类样本点也有异类样本点, 且数目相同。然后该方法在目标函数的推导结果中加入了类间拉普拉斯散度矩阵, 从而使优化得到的投影矩阵融入更多的类间信息。在 Yale 和 ORL 人脸数据集上进行实验, 并与同类方法相比, 结果表明本文提出的适用于小样本的双邻接图判别分析方法能够得到更好的降维效果。

**关键词:** 人脸识别; 拉普拉斯判别分析; 双邻接图; 降维

**中图分类号:** TP391      **文献标志码:** A

## Double Adjacent Graph-Based Discriminant Analysis for Small Size Sample

Luo Xuan<sup>1,2</sup>, Zhang Li<sup>1,2</sup>, Xue Yangtao<sup>1,2</sup>, Li Fanzhang<sup>1,2</sup>

- (1. School of Computer Science and Technology, Soochow University, Suzhou, 215006, China; 2. Joint International Research Laboratory of Machine Learning and Neuromorphic Computing, Soochow University, Suzhou, 215006, China)

**Abstract:** As a common dimensionality reduction method, the supervised Laplacian discriminant analysis (SLDA) for small size sample achieves a good result of dimensionality reduction via graph embedding discriminant neighborhood analysis. However, when SLDA finds the inter-class and intra-class data points in  $K$  nearest neighbors, there might exist an imbalance problem. Additionally, SLDA does not fully consider the inter-class information, which may decrease the performance of SLDA to a certain extent. To address the two problems mentioned above, we propose a double adjacent graph-based discriminant analysis (DAG-DA) algorithm for small size sample. Firstly, the algorithm tries to find  $K$  nearest neighbors in inter-class and intra-class samples, respectively, and then uses these  $K$  inter-class neighbors and  $K$  intra-class neighbors to construct the double adjacent graph. In this way, we can ensure that the adjacent graph contains both the inter-class and intra-class data points and has the same number. Secondly, the algorithm tries to add the intra-class Laplacian scatter matrix into the objective function of SLDA. Thus,

the projection matrix obtained by optimization takes the information between classes into account fully. We perform experiments on Yale and ORL human face datasets. Experimental results show that the proposed algorithm can get better performance compared with other methods.

**Key words:** face recognition; Laplacian discriminant analysis; double adjacency graph; dimensionality reduction

## 引 言

近年来,随着科学技术的发展,数据的数量和维度呈现出快速增长的趋势。降维不仅可以简化模型、降低模型的训练时间,而且可以提高模型的泛化能力。因为高维数据难以表达且需要较高的计算和存储代价,所以降维方法成为机器学习和模式识别领域中的热门研究方向,并已在人脸识别、手写体识别等系统中取得了良好的效果。

降维方法可分为线性和非线性两种。线性降维方法,如主成分分析法(Principal component analysis, PCA)<sup>[1]</sup>、线性判别分析(Linear discriminant analysis, LDA)<sup>[2]</sup>等,几何表达简单、计算成本较低,是分析高维数据的基础。PCA由Karl Pearson于1901年提出,该方法使用正交线性变换将一组观测的可能相关变量转换为一组线性不相关变量,这些不相关变量叫做主成分。拥有最大方差,即尽可能不相关的数据称为第一主成分,并以此类推。PCA算法的核心是对数据协方差矩阵进行特征分解。LDA算法与方差分析和回归分析密切相关,它试图找到一组特征的线性组合,能将多个对象尽量分开,即找到最佳投影方向。

如果所需的特征在一个低维的流形子空间中,那么传统的线性降维算法并不能很好地提取出该非线性结构,因此很多非线性的流形学习算法被提出,包括等距映射(Isomap)<sup>[3]</sup>、拉普拉斯特征映射(Laplacian eigenmaps, LE)<sup>[4]</sup>及局部线性嵌入(Locally-linear embedding, LLE)<sup>[5]</sup>等。

但是以上流形学习算法只考虑了训练样本,处理过程中未能得到一个显式映射,因此不能处理新样本,即所谓的“out of sample”问题。针对这个问题,He等人提出了保局算法(Locality preserving projection, LPP)<sup>[6]</sup>和局部保持嵌入算法(Neighborhood preserving embedding, NPE)<sup>[7]</sup>:(1) LPP算法首先整合邻近点信息建立一个图,该图可看做一个连续流形结构的线性离散近似<sup>[6]</sup>,然后使用拉普拉斯算子计算出转换矩阵将数据点映射到低维子空间;LPP算法使原始高维空间中距离相近点在降维后依旧保持近距离。(2) NPE算法的目的在于保持局部流形结构。在一定范围内给定一些数据点,首先建立权重矩阵来描述数据点之间的关系,并将每个点用邻近点线性表示;然后找到一个最佳嵌入使得该邻近结构在低维空间中得以保持。以上两个利用局部信息进行降维的算法虽然没有“out of sample”问题,但非局部信息在此过程中被忽略,即算法没有考虑相隔较远的点之间的关系。2007年,Yang等人提出了非监督判别投影(Unsupervised discriminant projection, UDP)算法<sup>[8]</sup>。该方法考虑了样本点在低维表示后的局部和非局部性,最小化局部散度的同时最大化非局部散度,即使非局部散度与局部散度的比值尽可能大。但UDP是非监督的学习方法,也就是不能利用样本的类别信息,不利于分类;而且在处理过程中也出现小样本问题<sup>[9]</sup>;当训练样本个数远小于样本维数时,就会出现矩阵奇异问题。2012年,楼宋江等人提出的小样本有监督拉普拉斯判别分析算法(Supervised Laplacian discriminant analysis, SLDA)<sup>[9]</sup>,通过将类内拉普拉斯散度矩阵投影在去除零空间的总体拉普拉斯散度矩阵上,避免了小样本问题。但是由于该方法中没有明确最大化类间散度,因而导致在投影空间的可分性不够好。2015年,丁春涛等人提出的基于双邻接图的判别近邻嵌入算法(Double adjacency graphs-based discriminant neighborhood embedding, DAG-DNE)<sup>[10]</sup>,融入了构建双邻接图的想法,解决了判别近邻嵌入(Discriminant neighborhood embedding, DNE)图嵌入降维方法在构造权重矩阵时数据不平衡的问题。

受到 DAG-DNE 算法的启发,本文提出了一种新的基于拉普拉斯判别分析的有监督降维方法,称为适用于小样本的双邻接图判别分析方法(Double adjacent graph-based discriminant analysis for small sample size, DAG-DA)。DAG-DA 在构建邻接图时考虑了样本的不平衡问题,每个样本点都至少建立两个关联,并将局部拉普拉斯散度矩阵与非局部拉普拉斯散度矩阵的差投影到总体拉普拉斯散度矩阵的秩空间中,然后在该空间中进行特征问题的求解,求得投影矩阵。通过在简单人工数据集、Yale 人脸库和 ORE 人脸库的实验证明了 DAG-DA 算法的优越性。

## 1 适用于小样本的 SLDA 算法

针对小样本问题,已有相应的一些解决方案<sup>[9,11-12]</sup>,其中监督化拉普拉斯判别分析算法和本文的工作紧密相关。该算法是一种有监督的流形学习算法,其目标是 minimized 局部散度,同时最大化非局部散度,即非局部散度与局部散度的比值尽可能大<sup>[13]</sup>,表达式为

$$\max J(\mathbf{W}) = \frac{J_N(\mathbf{W})}{J_L(\mathbf{W})} \Leftrightarrow \min J(\mathbf{W}) = \frac{J_L(\mathbf{W})}{J_T(\mathbf{W})} \quad (1)$$

式中: $\mathbf{W}$  为投影矩阵, $J_L$  为局部散度, $J_N$  为非局部散度, $J_T$  为总体散度, $J_T = J_L + J_N$ 。

以上优化问题(1)可以转化为如下目标函数

$$\mathbf{S}_L \mathbf{w} = \lambda \mathbf{S}_T \mathbf{w} \quad (2)$$

式中: $\lambda$  为特征分解得到的特征值, $\mathbf{w}$  为特征向量, $\mathbf{S}_L$  为局部拉普拉斯散度矩阵, $\mathbf{S}_T$  为丢弃零空间的总体拉普拉斯散度矩阵。

## 2 适用于小样本的 DAG-DA 算法

### 2.1 双邻接图判别分析

由于适用于小样本的监督化拉普拉斯判别分析算法在分子上仅考虑将局部拉普拉斯散度矩阵进行投影,忽略了非局部拉普拉斯散度矩阵。为了强调非局部信息,本文考虑在最优化函数里引入非局部散度,表达式为

$$\min J(\mathbf{W}) = \frac{J_L(\mathbf{W}) - J_N(\mathbf{W})}{J_T(\mathbf{W})} = \frac{J_D(\mathbf{W})}{J_T(\mathbf{W})} \quad (3)$$

局部散度为

$$J_L = \frac{1}{2nm} \sum_{i=1}^n \sum_{j=1}^n F_{ij} (y_i - y_j)^2 \quad (4)$$

非局部散度为

$$J_N = \frac{1}{2nm} \sum_{i=1}^n \sum_{j=1}^n B_{ij} (y_i - y_j)^2 \quad (5)$$

式(4,5)中: $\mathbf{F}$  为相似度矩阵; $\mathbf{B}$  为差异度矩阵。与 SLDA 方法从数据点的  $K$  近邻中找到同类和异类样本点来构建邻接图不同,本文在构造邻接图时借用双邻接图的思想,从同类和异类样本点间分别找到  $K$  个近邻点,再构建  $\mathbf{F}$  和  $\mathbf{B}$ 。

设样本集  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ , 其中  $\mathbf{x}_i \in \mathbf{R}^d$ ,  $y_i \in \{1, 2, \dots, c\}$  为  $\mathbf{x}_i$  的类别标签,  $n$  为样本的个数,  $d$  为样本的维数,  $c$  为类别总数。对于这个样本集,利用样本的类别信息,对每一个数据点都构造相似度矩阵  $\mathbf{F}$  和差异度矩阵  $\mathbf{B}$ , 根据样本分布赋予不同的权值,表达式为

$$F_{ij} = \begin{cases} \exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}\right\} & \mathbf{x}_i \in N_k^+(j) \text{ 或 } \mathbf{x}_j \in N_k^+(i) \\ 0 & \text{其他} \end{cases} \quad (6)$$

$$B_{ij} = \begin{cases} -\exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma}\right\} & \mathbf{x}_i \in N_k^-(j) \text{ 或 } \mathbf{x}_j \in N_k^-(i) \\ 0 & \text{其他} \end{cases} \quad (7)$$

式中:  $N_k^+(i)$  和  $N_k^+(j)$  分别表示样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  的同类近邻集合, 且  $\mathbf{x}_i$  和  $\mathbf{x}_j$  类别相同;  $N_k^-(i)$  和  $N_k^-(j)$  分别表示样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  的异类近邻集合, 且  $\mathbf{x}_i$  与  $\mathbf{x}_j$  类别不相同;  $t$  和  $\sigma$  为可变参数。由式(6,7)可以看出, 当同类样本距离越近, 它们的相似度越高, 反之亦然。当异类样本距离越近, 它们的差异度越小, 反之亦然。

由式(4,6)可以得到

$$J_L = \frac{1}{2nm} \sum_{i=1}^n \sum_{j=1}^n F_{ij} (y_i - y_j)^2 = \frac{1}{nm} \mathbf{W}^T \mathbf{S}_L \mathbf{W} \quad (8)$$

式中  $\mathbf{S}_L$  为局部拉普拉斯散度矩阵

$$\mathbf{S}_L = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{ij} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T = \mathbf{X} \mathbf{F}' \mathbf{X}^T - \mathbf{X} \mathbf{F} \mathbf{X}^T = \mathbf{X} \mathbf{L} \mathbf{X}^T \quad (9)$$

式中  $\mathbf{F}'$  为一个对角矩阵, 对角线上的元素为:  $F'_{ii} = \sum_j F_{ij}$ ,  $\mathbf{L} = \mathbf{F}' - \mathbf{F}$ ,  $\mathbf{L}$  即为拉普拉斯类内矩阵。

同样, 由式(5,7)可以得到

$$J_N = \frac{1}{2nm} \sum_{i=1}^n \sum_{j=1}^n B_{ij} (y_i - y_j)^2 = \frac{1}{nm} \mathbf{W}^T \mathbf{S}_N \mathbf{W} \quad (10)$$

$$\mathbf{S}_N = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n B_{ij} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T = \mathbf{X} \mathbf{B}' \mathbf{X}^T - \mathbf{X} \mathbf{B} \mathbf{X}^T = \mathbf{X} \mathbf{N} \mathbf{X}^T \quad (11)$$

式中:  $\mathbf{B}'$  为一个对角矩阵, 对角线上的元素为  $B'_{ii} = \sum_j B_{ij}$ ;  $\mathbf{N} = \mathbf{B}' - \mathbf{B}$  为拉普拉斯类间矩阵。

由式(3,8,10)可推导出目标函数等价于

$$\min \frac{J_D(\mathbf{W})}{J_T(\mathbf{W})} = \frac{\mathbf{W}^T \mathbf{S}_D \mathbf{W}}{\mathbf{W}^T \mathbf{S}_T \mathbf{W}} = \frac{\boldsymbol{\alpha}^T \mathbf{S}_D \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{S}_T \boldsymbol{\alpha}} \Leftrightarrow \mathbf{S}_D \mathbf{w} = \lambda \mathbf{S}_T \mathbf{w} \quad (12)$$

式中:  $\mathbf{S}_D = \mathbf{S}_L - \mathbf{S}_N$ ,  $\boldsymbol{\alpha} \in \text{range}(\mathbf{S}_T)$ 。

## 2.2 秩空间

首先对  $\mathbf{S}_T$  进行特征分解。由于总体拉普拉斯散度矩阵  $\mathbf{S}_T$  非负定, 所以对它进行特征值分解后的特征值可能为零, 可能为正。取非零的特征值  $\tilde{\boldsymbol{\Lambda}}$  所对应的特征向量  $\mathbf{U}_1$  组成  $\mathbf{S}_T$  的主空间  $\boldsymbol{\alpha}$ 。即  $\mathbf{S}_T$  特征分解后, 特征值为

$$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r, \lambda_{r+1}, \dots, \lambda_n)$$

其中特征值  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r, \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$ 。特征值对应的特征向量矩阵为

$$\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_n\}$$

其中特征值  $\lambda_i$  对应的特征向量为  $\mathbf{u}_i$ 。

令非负特征值矩阵  $\tilde{\boldsymbol{\Lambda}} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$ , 部分特征向量矩阵  $\mathbf{U}_1 = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ ,  $\mathbf{U}_\perp = \{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_n\}$ 。令  $\mathbf{W}_1 = \mathbf{U}_1 \tilde{\boldsymbol{\Lambda}}$ , 将差异拉普拉斯散度矩阵  $\mathbf{S}_D$  投影到  $\mathbf{W}_1$  上为

$$\tilde{\mathbf{S}}_D = \mathbf{W}_1^T \mathbf{S}_D \mathbf{W}_1 \quad (13)$$

因为差异拉普拉斯散度矩阵  $\mathbf{S}_D$  也是非负定的, 所以对它进行特征值分解后的特征值可能为零, 也可能为正。设分解后  $d$  个最小的特征值为  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_d$ , 其对应向量分别为  $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_d$ , 令  $\mathbf{W}_2 = \{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_d\}$ 。最终的投影矩阵  $\mathbf{W} = \mathbf{W}_1 \mathbf{W}_2$ 。DAG-DA 算法的步骤如下:

输入: 训练样本集合  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ , 令样本矩阵为  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$ , 降维子空间维数  $r$ 。

输出: 投影矩阵  $\mathbf{W} \in \mathbf{R}^{d \times r}$ 。

(1) 构建相似度矩阵  $S$  和差异性矩阵  $B$ ;

(2) 计算局部拉普拉斯散度矩阵  $S_L$ , 非局部拉普拉斯散度矩阵  $S_N$ , 以及它们的和  $S_T$  与差  $S_D$ ;  $S_L = XLX^T$ , 其中  $L = S' - S$ ,  $S'$  为对角矩阵,  $S'_i = \sum_j S_{ij}$ ;  $S_N = XNX^T$ , 其中  $N = B' - B$ ,  $B'$  为对角矩阵,  $B'_i = \sum_j B_{ij}$ ;  $S_T = S_L + S_N$ , 且  $S_D = S_L - S_N$ ;

(3) 对矩阵  $S_T$  进行特征分解, 取其非零特征值  $\lambda_i$  和非零特征值对应的向量  $u_i$ , 分别组成矩阵  $\Lambda$  和  $U$ , 其中  $\Lambda$  为对角矩阵且  $\Lambda_i = \lambda_i$ ,  $U$  的第  $i$  列为  $u_i$ , 令  $W_1 = UA$ ;

(4) 将  $S_D$  在  $W_1$  上投影, 得到  $\tilde{S}_D = W_1^T S_D W_1$ , 对  $\tilde{S}_D$  进行特征分解, 取最小的  $r$  个特征值对应的特征向量, 组成一个矩阵  $W_2$ ;

(5) 得到最终投影矩阵:  $W = W_1 W_2$ 。

### 2.3 DAG-DA 算法和 SLDA 算法比较

(1) 在用邻接图来保持样本的局部结构信息时, SLDA 算法找到近邻后, 利用类别信息再构建类内和类间邻接图, 这样可能会出现只有类内近邻或只有类间近邻的数据不平衡问题。而在计算权值时, 同等距离下类间近邻比类内近邻的权值更大, 这显然不利于保持数据的结构。DAG-DA 算法分别在类内和类间找近邻点, 从热核函数可以看出: 若是类内近邻点, 则赋予较大的权值; 若是类间近邻点, 则权值较小。这样可以扩大类内距离缩小类间距离的影响, 同时也解决了数据不平衡的问题, 更好地保持了数据的流形结构。

(2) 在优化问题选取目标函数时, SLDA 算法的目标函数(式(1))分子上仅考虑了将局部拉普拉斯散度矩阵进行投影, 而忽略了非局部拉普拉斯散度矩阵; 而 DAG-DA 算法在最优化函数(式(3))里引入了非局部散度, 强调了类间散度, 使得降维后的数据可分性更好。

(3) 在对总体拉普拉斯散度矩阵进行特征分解后, SLDA 算法选取的投影空间为特征向量乘以一个小于 1 的数; 而 DAG-DA 算法直接将非零向量作为投影空间, 相当于少乘一个缩小因子, 使数据更加准确。

## 3 实验结果与分析

为验证 DAG-DA 算法的有效性, 在各类数据上对 DAG-DA 算法进行实验, 并与 SLDA 算法、LPP 算法和 DNE 算法进行比较。这 4 种算法均可以对近邻个数进行调节, 其中 SLDA 算法和 DAG-DA 算法中的  $t$  和  $\sigma$  为可变参数。本文实验均采用最近邻分类器对降维后的数据进行分类。

### 3.1 人工数据

随机生成两类符合均匀分布的人工数据, 一类是 0~1 之间的随机数, 另一类是 0.7~1.7 之间的随机数, 分别用 SLDA 算法和 DAG-DA 算法学习得到投影矩阵。图 1 显示了 SLDA 算法和 DAG-DA 算法学习投影矩阵后投影后的效果。从图 1 中看到, 图 1(b) 中的类间点距离更大, 类内点距离更小, 也就是 DAG-DA 算法相对比 SLDA 而言, 降维效果有了明显的提升, 也更适合于分类。在对优化函数进行等价替代后, SLDA 算法的目标函数里并没有考虑类间距离, 仅将局部拉普拉斯散度矩阵投影到总空间。DAG-DA 算法对其进行改进, 将局部拉普拉斯散度矩阵与非局部拉普拉斯散度矩阵的差进行投影, 是一个更理想的目标函数。

除了图 1 能直观地看出两种算法的分类效果外, 表 1 还给出了两种算法在人工数据测试后的具体结果。从表 1 中可以看出, 用 DAG-DA 算法降维比 SLDA 算法得到的类内距离和更小, 类间距离和更大, 说明降维后分类效果更佳。

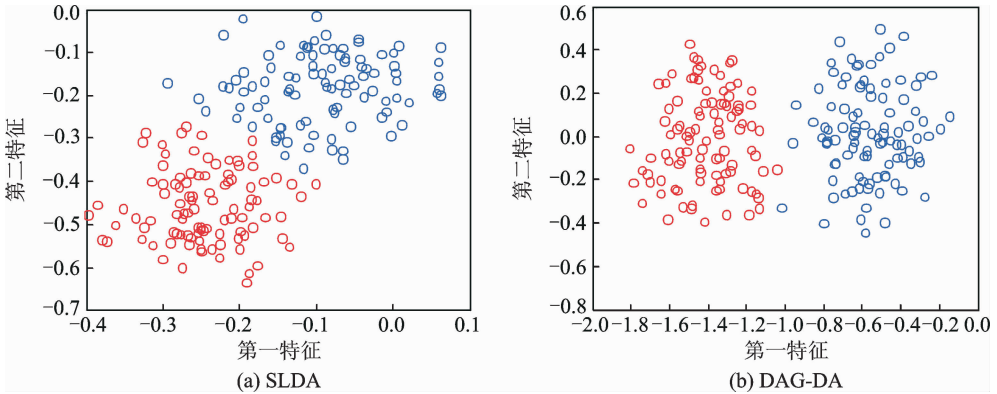


图1 SLDA与DAG-DA在简单人工数据集上的比较

Fig. 1 Comparison of SLDA and DAG-DA on artificial dataset

表1 SLDA与DAG-DA在人工数据上测试结果比较

Tab. 1 Test results of SLDA and DAG-DA on artificial dataset

测试项	类间距离和	类内距离和	类内距离和/类间距离和
原始数据	$1.7364 \times 10^4$	$3.5626 \times 10^4$	0.4869
降维数据(SLDA)	$2.3782 \times 10^3$	$9.9643 \times 10^3$	0.4190
降维数据(DAG-DA)	$3.2325 \times 10^4$	$9.8289 \times 10^3$	0.3041

### 3.2 Yale人脸数据库

Yale人脸数据库<sup>[14]</sup>共有165张人脸灰度图像,每张图像数字为 $32 \times 32$ 的像素矩阵,来自15个不同的人,每人11张。每个人的各张图像有不同的光照条件(中央光照、左侧光照、右侧光照)和不同的表情(眼镜/无眼镜、高兴、常态、悲伤、睡眠、惊讶和眨眼)。Yale人脸数中不同条件下的一些人脸图像样本如图2所示。该实验中,随机选择60%样本作为训练样本,其余的40%作为测试样本。



图2 Yale人脸图像示例

Fig. 2 Face examples of Yale dataset

在实验时,为了减少大量的计算,先采用PCA方法对测试数据进行初始降维,初始降维的投影矩阵为 $A$ 。再采用DAG-DA,SLDA,LPP,DNE四种算法进行再次降维,二次降维矩阵为 $W$ ,最终投影矩阵为 $P=WA$ 。利用 $P$ 把训练样本和训练样本投影到低维空间,再用近邻分类器进行分类,得到各个方法降维后的识别率。在构造邻接图时,选取 $k=3$ ,通过循环找到最佳的 $t$ 和 $\sigma$ 。图3显示了DAG-DA算法,SLDA算法,LPP算法和DNE算法在不同维数下的识别率。表2显示了不同算法在Yale人脸库上的最高识别率和特征维度。

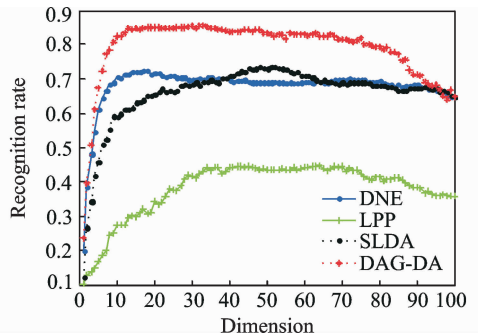


图3 在Yale人脸库上不同维数的识别率

Fig. 3 Recognition rate of different algorithms on Yale dataset

表 2 Yale 人脸库各算法最高识别率和特征维数

Tab. 2 Recognition rate and dimension of different algorithms on Yale dataset

算法	识别率	最优维数
LPP	44.83% ± 8.1%	64
DNE	72.67% ± 4.0%	18
SLDA	73.83% ± 3.9%	48
DAG-DA	85.50% ± 5.6%	32

### 3.3 ORL 人脸数据库

ORL 人脸数据库<sup>[15]</sup>共有 400 张人脸灰度图像, 每张图像为  $32 \times 32$  的像素矩阵, 来自 40 个不同的人, 每人 10 张。每个人的各图像有不同的大小、人脸旋转程度和不同的表情(睁眼/闭眼、高兴/悲伤、戴眼镜/不戴眼镜等)。ORL 人脸数据库中不同条件下的一些人脸图像样本如图 4 所示。该实验中, 随机选择 60% 样本作为训练样本, 其余的 40% 作为测试样本。



图 4 ORL 人脸图像示例

Fig. 4 Face examples of ORL dataset

同样, 为了减少大量的计算, 先采用 PCA 方法对测试数据进行初始降维, 再采用 DAG-DA, SLDA, LPP 和 DNE 四种算法进行再次降维, 最后用近邻分类器进行分类, 得到各个方法降维后的识别率。本文选取  $k=3$ , 通过循环找到最佳的  $t$  和  $\sigma$ 。图 5 显示了 DAG-DA 算法, SLDA 算法, LPP 算法和 DNE 算法在不同维数下的识别率。表 3 显示了不同算法在 ORL 人脸库上的最高识别率和特征维度。

从实验结果可以看出, DAG-DA 算法达到的最佳识别率最高。SLDA 算法的识别率高于 LPP 算法, 因为 SLDA 算法考虑了非局部信息和样本的类别信息。DNE 算法也是监督化学习方法, 它构建了类内与类间的邻接图, 让同类数据在低维空间紧凑, 不同类数据尽可能分开, 识别率高于 LPP 算法。DAG-DA 算法不仅考虑了非局部信息和样本类别信息, 而且考虑了数据不平衡问题, 并在目标函数内加入了非局部拉普拉斯散度矩阵, 所以最佳识别率不仅优于 LPP 算法, 也比 SLDA 和 DNE 算法更胜一筹。

### 4 结束语

本文为了处理监督化拉普拉斯判别分析方法中的邻接图数据不平衡问题, 以及目标函数中类间信息缺失问题, 提出了适用于小样本的双邻接图判别分析算法 DAG-DA, 并在 Yale 和 ORL 人脸数据集上验证了该方法的有效性。

但是, 本文算法在实现过程中需要寻找热核函数的最佳参数值, 同时去掉了总体拉普拉斯散度矩阵

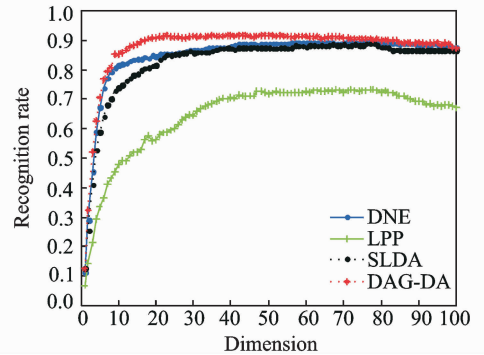


图 5 在 ORL 人脸库上不同维数的识别率  
Fig. 5 Recognition rate of different algorithms on ORL dataset

表 3 ORL 人脸库各算法最高识别率和特征维数

Tab. 3 Recognition rate and dimension of different algorithms on ORL dataset

方法	识别率	最优维数
LPP	73.29% ± 2.1%	77
DNE	89.64% ± 0.5%	71
SLDA	89.14% ± 0.9%	76
DAG-DA	91.79% ± 0.5%	23

的零空间,这两步使得该算法的时间复杂度比 LPP 和 DNE 算法高。下一步工作将尝试寻找时间复杂度较低且降维效果好的算法。

### 参考文献:

- [1] Jolliffe I T. Principal component analysis [M]. Berlin:Springer-Verlag, 2002.
- [2] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: A general framework for dimensionality reduction [J]. IEEE Trans Pattern Anal Mach Intell, 2006, 29 (1):40.
- [3] Tenenbaum J B, De S V, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290 (5500):2319-2323.
- [4] Belkin M, Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6):1373-1396.
- [5] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2009, 290(5500):2323-2326.
- [6] Niyogi X. Locality preserving projections[J]. Advances in Neural Information Processing Systems, 2004, 16:153.
- [7] He X, Cai D, Yan S, et al. Neighborhood preserving embedding[C]// Tenth IEEE International Conference on Computer Vision. [S. l. ]:IEEE, 2005:1208-1213.
- [8] Yang J, Zhang D, Yang J, et al. Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(4):650-664.
- [9] 楼宋江, 张国印, 潘海为, 等. 人脸识别中适合于小样本情况下的监督化拉普拉斯判别分析[J]. 计算机研究与发展, 2012, 49(8):1730-1737.  
Lou Songjiang, Zhang Guoyin, Pan Haiwei, et al. Supervised Laplacian discriminant analysis for small sample size problem with its application to face recognition[J]. Journal of Computer Research and Development, 2012, 49(8):1730-1737.
- [10] Ding C T, Zhang L. Double adjacency graphs-based discriminant neighborhood embedding [J]. Pattern Recognition, 2015, 48(5):1734-1742.
- [11] Huang R, Liu Q, Lu H, et al. Solving the small sample size problem of LDA[C]// 16th International Conference on Pattern Recognition. [S. l. ]:IEEE Computer Society, 2002:30029.
- [12] Jain A K, Duin R P W, Mao J. Statistical pattern recognition: A review [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(1):34-37.
- [13] Tao Y J, Yang J. Quotient vs. difference: Comparison between the two discriminant criteria [J]. Neurocomputing, 2010, 73 (10/11/12):1808-1817.
- [14] University of California, San Diego. The Yale face database[EB/OL]. <http://vision.ucsd.edu/content/yale-face-database>, 2013-12-21.
- [15] AT&T Laboratories Cambridge. The ORL face database[EB/OL]. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>, 2013-12-21.

### 作者简介:



罗璇 (1993-), 女, 硕士研究生, 研究方向: 机器学习与模式识别。



张莉 (1975-), 女, 博士, 教授, 博士生导师, 研究方向: 机器学习与模式识别, E-mail: zhangliml@suda.edu.cn。



薛杨涛 (1992-), 女, 硕士研究生, 研究方向: 机器学习与模式识别。



李凡长 (1964-), 男, 教授, 博士生导师, 研究方向: 李群机器学习、数据挖掘和动态模糊。