

线性判别分析中两种空间信息嵌入方法之比较

牛璐璐¹ 陈松灿^{1,2} 俞 璐³

(南京航空航天大学计算机科学与技术学院 南京 210016)¹

(南京大学计算机软件新技术国家重点实验室 南京 210093)²

(解放军理工大学通信工程学院 南京 210007)³

摘 要 由“没有免费午餐定理(NFL)”可知:只有充分利用了与问题相关的先验知识的学习器才会拥有好的学习性能,而常用的线性判别分析(LDA)在为图像提取判别特征时对图像向量化的处理导致了空间结构的破坏,以致结构信息未被利用,从而限制了学习性能的进一步提升。空间平滑的 LDA(SLDA)通过对 LDA 目标的空间正则化弥补了此不足,同时图像欧氏距离(IMED)则通过空间平滑欧氏距离实现对空间结构的利用,而后将其用于 LDA(IMEDA)。对这两种 LDA 间的内在联系进行了尝试探究:理论上证明了对于中心化样本,SLDA 是 IMEDA 的特例;分析了算法的时间和空间复杂度;经验上通过 Yale、AR 和 FERET 人脸集比较了 SLDA 和 IMEDA 的识别性能和运行时间,同时分析了参数对模型性能的影响。

关键词 线性判别分析,降维,空间结构信息,空间平滑

中图分类号 TP391 **文献标识码** A

Comparison between Two Approaches of Embedding Spatial Information into Linear Discriminant Analysis

NIU Lu-lu¹ CHEN Song-can^{1,2} YU Lu³

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)¹

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)²

(Institute of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China)³

Abstract No Free Lunch Theorem says that only taking full advantage of learning machine of priori knowledge related to the problem under consideration can have a good learning performance. However, the vectorization of the images used in conventional linear discriminant analysis (LDA) damages the spatial structure of initial images, and restricts the improvement of the learning performance of LDA. Spatially smoothing linear discriminant analysis (SLDA) tries to overcome this problem by introducing the spatial regularization to the objective of LDA, whereas IMage Euclidean Distance Discriminant Analysis (IMEDA) substitutes IMage Euclidean Distance (IMED) for the original Euclidean metric in the objective of LDA to utilize the spatially structure information. This paper attempted to explore the intrinsic link between SLDA and IMEDA; theoretically proved that SLDA is the special case of IMEDA when the sample mean of the data set is zero, analyzed the time complexity and the space complexity of the algorithms. The experiments were conducted to compare SLDA with IMEDA on Yale, AR and FERET face datasets, and the influences of the parameters on performance of the algorithms were analyzed.

Keywords Linear discriminant analysis, Dimensionality reduction, Spatial structure information, Spatially smooth

1 引言

特征提取是模式识别研究的关键问题之一,其基本任务是寻求与学习任务(如分类等)尽可能相关的有效特征。特征提取大致分成两大类,即无监督和监督提取。在目前的无监督提取中,主成分分析(PCA)^[1]最为典型,其目的是寻找在最小均方意义下最能代表原数据的投影方法,侧重于信号重建。监督提取中,线性判别分析(LDA)^[1,2]具有代表性,其目标是寻找有效分类的投影方向,使得数据映射到低维空间后,类内

尽可能“紧凑”,类间尽可能“分开”,并已在人脸识别等得到广泛应用^[3]。然而,PCA 和 LDA 面向的操作对象是向量,当要提取图像特征时,向量化会破坏图像原有的空间结构,使部分先验知识未被利用,从而影响其学习性能的进一步提升。“没有免费午餐定理(NFL)”告诉我们:如果能利用更多的与问题“匹配”或相关的先验知识,则能获得性能更好的学习器。因而将空间结构信息融入到(线性)降维过程成为了一个重要研究内容。

目前存在针对该问题的两种主要方法:一是 2007 年由

到稿日期:2013-05-20 返修日期:2013-08-05 本文受 NSFC(61170151,61101202),江苏 NSFC(BK2011728)资助。

牛璐璐(1988—),女,硕士生,主要研究方向为模式识别,E-mail:jndxyilu@126.com;陈松灿(1962—),男,教授,博士生导师,CCF 高级会员,主要研究方向为模式识别、机器学习、神经计算;俞璐(1973—),女,博士,讲师,主要研究方向为图像处理、模式识别。

Cai 等^[4]提出的空间平滑的子空间学习(SSSL)(或特征提取)方法,具体是将拉普拉斯平滑正则化惩罚嵌入至子空间学习的目标函数中,实现对图像空间结构信息的利用。实验表明:嵌入了图像空间结构信息后所学到的子空间投影(或所提取的特征)能产生比未利用该信息的对应原算法(如 LDA、局部保持投影 LPP^[5]、邻域保持嵌入 NPE^[6]、临界费舍尔分析 MFA^[7]),及 Eigenface^[3,8]和 Fisherface^[8]等更优的分类性能。结果促发了其后许多的跟进工作:如 2009 年,Zuo 等^[9]补充了 SSSL 中的惩罚函数,加入了人脸图像的多尺度特性(multi-scale characteristic),提出了 LOG 和 DOG 惩罚函数。实验表明此方法优于原 SSSL;Hou 等^[10]在 SSSL 的基础上,通过限制投影方向间的正交性,提出了正交的空间平滑算法(OSSL),获得了比 SSSL 更好的性能;李勇周等^[11]通过将空间正则化、LDA 和局部敏感判别投影相结合,设计了一种空间光滑的子空间学习算法,它在性能上优于 SSSL。2010 年,Lu 等^[12]发现 SSSL 虽考虑了图像像素间的空间关系,但仍将每幅人脸看作一个整体,忽略了人脸不同部分可能对人脸表示和识别具有不同辨别力的弊端,从而定义了类间样本权矩阵和类内样本权矩阵的概念,并提出了基于表情的双加权子空间学习方法;2011 年,Wen 等^[13]采用数据图结构(graph of data)定义了类间图矩阵(between-class graph matrix)和类内图矩阵(within-class graph matrix),而后采用空间平滑。此法一方面在数据图结构中嵌入了人脸图像的流形(manifold)特性,另一方面采用空间平滑考虑了像素间的空间关系,因此获得了比 SSSL 更好的效果。此类利用空间信息的方法拥有的共同点是将空间信息以正则化的方式嵌入对应的优化目标,通过优化获得所期望的投影。

另一个利用空间信息的途径是对一般的欧氏距离度量进行空间平滑,使其适合度量图像间的相似性。较早提出距离平滑的工作是 2005 年由 Wang^[14]等提出的图像欧氏距离 IMED。不同于传统欧氏距离^[15,16,17],IMED 通过引入空间平滑,实现了欧氏距离中空间结构的嵌入,保持了对像素小扰动下的相对不敏感性,且计算简单,适合进行图像的高效识别。由于 IMED 的这些优点,促动了其后许多相关研究,如 2006 年,Chen 等^[18]考虑到含噪数据常导致 Isomap 算法不够稳定,提出了基于 IMED 的 Isomap,实验表明它相比欧氏距离的 Isomap 在性能上有较大提升;2008 年,Li 等^[19]不仅利用像素间的空间关系而且也利用了灰度级的先验知识,提出了自适应的图像欧氏距离(AIMED);黄晓华等^[20]利用自适应图像欧氏距离构建近邻矩阵,更好地利用了图像的空间关系和类信息,提出一种针对图像识别的鉴别型局部线性嵌入算法,并应用于人脸识别;2009 年,Sun 等^[21]又从平移不变(SI)的角度给出了 IMED 的解释,证明了 IMED 实际上属于 SI 变换,且每一个 SI 变换可等价地被视为另一个 SI 变换与欧氏距离之和;2010 年,Gu 等^[22]提出了更广的拉普拉斯平滑变换(LST),其通过拉普拉斯矩阵刻画像素间的空间结构,由此形成了一种广义的图像距离度量,推广了 IMED。LST 既可对图像降维,也可在降维后进一步结合至任何学习算法,包括 LDA 和 SVM^[23]。

上述两种方法从不同途径一致地提高了原有算法的分类性能^[4,20],但就我们所知,对它们两者间的内在联系以及各自的优劣,还未有深入探究。本文的目的是基于流行的 LDA,探究它们间的内在联系。为方便,我们以第一类中的 SSSL

和第二类中的 LST 为例,分别将它们结合到 LDA 形成对应的模型 SLDA 和 IMEDA,进而围绕两者开展对比研究。

本文的贡献总结如下:

(1)理论上证明了对于中心化样本,SLDA 是 IMEDA 的特例;分析了算法的时间和空间复杂度。

(2)经验上比较了 SLDA 和 IMEDA 的分类性能,发现结果相当;并在实验中给出了算法的运行时间,发现 IMEDA 较快。

(3)探究了参数对模型性能的影响。

本文第 2 节回顾了与 SLDA 和 IMEDA 相关的算法;第 3 节在理论上给出了 SLDA 和 IMEDA 间的内在联系,并分析了算法的时间和空间复杂度,在经验上比较了 SLDA 和 IMEDA 的分类性能和运行时间,并探究了参数对模型性能的影响;最后总结全文,并给出了下一步研究工作。

2 相关工作回顾

本节主要回顾 SSSL、LST 和由它们形成的比较模型 SLDA 和 IMEDA,由于 SSSL 和 LST 均采用了拉普拉斯惩罚实现平滑,故先简要介绍拉普拉斯平滑(Laplacian Smoothing)。

现假设有 n 张 c 类 $M \times N$ 人脸图像可用,将它们用向量表示为 $X = [x_1, x_2, \dots, x_n]$ 。

2.1 拉普拉斯平滑

设 f 是定义在区域 $\Omega \in R^d$ 的一个实函数, Ω 上的拉普拉斯算子 \mathcal{L} 和拉普拉斯代价函数 $\mathcal{J}(f)$ 分别为:

$$\mathcal{L}f(t) = \sum_{j=1}^d \frac{\partial^2 f}{\partial t_j^2} \quad (1)$$

$$\mathcal{J}(f) = \int_{\Omega} [\mathcal{L}f]^2 dt \quad (2)$$

显然, $\mathcal{J}(f)$ 反映了函数 f 在区域 Ω 上的平滑性。由于本文的研究对象为图像,因此仅讨论 $d=2$ 的情况,且采用的是离散的拉普拉斯平滑,考虑到 SSSL 和 LST 不同的离散化,具体的离散过程分别见各自的小节。

2.2 SSSL

SSSL 用拉普拉斯惩罚作为正则化项嵌入至子空间学习的降维目标函数中,实现对图像空间结构信息的利用。

2.2.1 SSSL 中拉普拉斯的离散化

对一幅图像, Ω 为一个二维格子阵。为方便,约定其在矩形区域 $[0, 1]^2$ 上。对区域 Ω 作如下网格剖分:令 $\mathbf{h} = (h_1, h_2)$,其中 $h_1 = 1/M, t_{i1} = (i_1 - 0.5)h_1, 1 \leq i_1 \leq M; h_2 = 1/N, t_{i2} = (i_2 - 0.5)h_2, 1 \leq i_2 \leq N$,即可得到一个 $M \times N$ 规格的网格(lattice),并记为 $\Omega_k = \{(t_{i1}, t_{i2}) | 1 \leq i_1 \leq M, 1 \leq i_2 \leq N\}$ 。仿照文献[4],令 D_j 为一个能对 $\partial^2 / \partial t_j^2$ 作离散近似的 $p \times p$ 阶矩阵, p 维向量 $\mathbf{u} = (u(t_1), \dots, u(t_p))$ 为函数 $u(t)$ 的离散化形式,而 D_j 具有如下性质,即:

$$[D_j \mathbf{u}]_i \approx \frac{\partial^2 u(t_i)}{\partial t^2}$$

式中, $i=1, 2, \dots, p$ 。采用修正的 Neuman 离散化方法^[24],取

$$D_j = \frac{1}{h_j^2} \begin{bmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \cdot & \cdot & \cdot \\ & & & 1 & -2 & 1 \\ & & & & & 1 & -1 \end{bmatrix}$$

即可将二维拉普拉斯算子 \mathcal{L} 的离散化近似为式(3)的 $(M \times N) \times (M \times N)$ 矩阵。

$$\Delta = D_1 \otimes I_2 + I_1 \otimes D_2 \quad (3)$$

式中, I_1 和 D_1 分别为 $M \times M$ 的单位阵和方阵, I_2 和 D_2 分别为 $N \times N$ 的单位阵和方阵, \otimes 为 Kronecker 积。

容易证明, 对于一个 $M \times N$ 维向量 b , $\|\Delta \cdot b\|^2$ 正比于相邻网格节点二阶差分的平方和, 它体现了 b 在 $M \times N$ 个网格的平滑度。

2.2.2 SSSL 建模及 SLDA

SSSL 可从如下定义的最大化问题(4)中导出,

$$\max_{b, \alpha} \frac{b^T X W X^T b}{(1-\alpha)b^T X D X^T b + \alpha \mathcal{J}(b)} \quad (4)$$

式中, $\alpha (0 \leq \alpha \leq 1)$ 用于控制平滑度, $\mathcal{J}(b) = \|\Delta \cdot b\|^2 = b^T \Delta^T \Delta b$ 为离散化拉普拉斯平滑函数, W 是一个 $(M \times N) \times (M \times N)$ 对称阵, D 为一个对角阵, 且 $D_{ii} = \sum_j W_{ij}$ 。选取不同的 W , 可得到不同的线性降维算法^[9], 如空间平滑的线性判别分析(SLDA)、空间平滑的局部保持投影(SLPP)、空间平滑的邻域保持嵌入(SNPE)、空间平滑的临界费舍尔分析(SMFA)和空间平滑的局部判别嵌入(SLDE)。

对于 SLDA 算法, 假设 n_i 表示第 $i (1 \leq i \leq c)$ 类样本的个数, 且 $n_1 + n_2 + \dots + n_c = n$, W 对应于:

$$W_{ij} = \begin{cases} 1/n_i, & \text{如果 } x_i \text{ 与 } x_j \text{ 均属于第 } i \text{ 类} \\ 0, & \text{否则} \end{cases}$$

此时, 可验证 $D = I$ 。

显然式(4)最大化问题可通过求解的广义特征值方程即式(5)得到所期望的子空间或投影矩阵。

$$X W X^T b = \lambda((1-\alpha) X D X^T + \alpha \Delta^T \Delta) b \quad (5)$$

2.3 LST

拉普拉斯平滑变换(LST)^[22]作为一种图像的预处理方法, 将二维空间域的图像映射成频域上的一维向量, 直接滤除高频信息, 因而平滑了图像, 同时降低了计算复杂性, 且独立于数据而仅与 2D 空间域相关。LST 不同于拉普拉斯特征映射(Laplacian Eigenmaps)^[25]和 LPP^[5], 它关注像素。

2.3.1 LST 中拉普拉斯的离散化

将 $M \times N$ 维的列向量 v 表示为 $v(i_1 * N + i_2)$, 其中 $i_1 = 0, 1, \dots, M-1, i_2 = 0, 1, \dots, N-1$, 且记 $v(\lceil i_1, i_2 \rceil) = v(i_1 * N + i_2)$, 其中运算符 $\lceil i_1, i_2 \rceil = i_1 * N + i_2$ 。因此一幅图像 $f(i_1, i_2)$ 表示为一个列向量 $f(\lceil i_1, i_2 \rceil)$ 。若 $(M \times N) \times (M \times N)$ 维矩阵 L_{MN} 表示为 $L(i_1 * N + i_2, i_1' * N + i_2')$, 其中 $i_1, i_1' = 0, 1, \dots, M-1, i_2, i_2' = 0, 1, \dots, N-1$, 则 $L(\lceil i_1, i_2 \rceil, \lceil i_1', i_2' \rceil) = L(i_1 * N + i_2, i_1' * N + i_2')$ 。

因此, 一幅图像 $f(i_1, i_2)$ 的拉普拉斯平滑函数 $\mathcal{J}(f)$ 可表示为:

$$\mathcal{J}(f) = \sum_{i_1=0}^{M-2} \sum_{i_2=0}^{N-1} [f(\lceil i_1+1, i_2 \rceil) - f(\lceil i_1, i_2 \rceil)]^2 + \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{N-2} [f(\lceil i_1, i_2+1 \rceil) - f(\lceil i_1, i_2 \rceil)]^2 \quad (6)$$

经过简单的几何变换, 可得到:

$$\mathcal{J}(f) = f^T L_{MN} f \quad (7)$$

式中, $L_{MN} = D - W$, 记为拉普拉斯矩阵, W 定义如下:

$$W(\lceil i_1, i_2 \rceil, \lceil i_1', i_2' \rceil) = \begin{cases} 1, & \text{if } |i_1 - i_1'| + |i_2 - i_2'| = 1 \\ 0, & \text{if } |i_1 - i_1'| + |i_2 - i_2'| \neq 1 \end{cases}$$

D 为对角阵, 且相应对角元素为 W 所在行(或列)的行和(或列和), 即

$$D(\lceil i_1, i_2 \rceil, \lceil i_1, i_2 \rceil) = \sum_{\lceil i_1', i_2' \rceil} W(\lceil i_1, i_2 \rceil, \lceil i_1', i_2' \rceil)$$

由上可知, 拉普拉斯矩阵 L_{MN} 有如下性质^[22]:

(1) L_{MN} 是稀疏矩阵, 且含有非零元素的个数为 $5MN - 2M - 2N$;

(2) L_{MN} 是半正定对称矩阵, 最小特征值为 0, 秩为 $MN - 1$;

(3) 特征值 0 对应的特征向量 e_0 的所有元素相等。

2.3.2 LST 的计算及 IMEDA

图像的 LST 通过以下 3 步完成:

(1) 根据图像大小, 构造拉普拉斯矩阵 L_{MN} ;

(2) 计算 L_{MN} 的前 $k (0 \leq k \leq MN)$ 个最小特征值和相应的特征向量 $\{e_i\}_{i=0}^{k-1}$;

(3) 令 $\Gamma = (e_0, e_1, \dots, e_{k-1})$, 那么 $g = \Gamma^T f$ 即为图像 f 在低频系数 k 下变换后的向量, 则样本集 $X = [x_1, x_2, \dots, x_n]$, 经变换后记为 $\hat{X} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n] = \Gamma^T X \in R^{k \times n}$ 。

IMEDA 是将经 LST 降维后的图像用于 LDA, 由此形成了我们所谓的 IMEDA。其优化目标为:

$$\max_{\omega} \frac{\omega^T \hat{S}_B \omega}{\omega^T \hat{S}_W \omega} \quad (8)$$

式中, \hat{S}_B 和 \hat{S}_W 分别是变换空间后的类间散布矩阵和类内散布矩阵, 即 $\hat{S}_B = \sum_{i=1}^c n_i (u_i - u)(u_i - u)^T$, $\hat{S}_W = \sum_{i=1}^c \sum_{j=1}^{n_i} (\hat{x}_{ij} - u_i)(\hat{x}_{ij} - u_i)^T$, \hat{x}_{ij} 表示第 i 类的第 j 个样本, $u_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{x}_{ij}$, $u = \frac{1}{n} \sum_{i=1}^n \hat{x}_i$ 。

在处理人脸图像时, 常会遇到小样本问题, 使得经 LST 后的类内散布矩阵仍可能是奇异, 故采用正则化 LDA 来缓与此问题, 其优化目标变为:

$$\max_{\omega, \gamma} \frac{\omega^T \hat{S}_B \omega}{\omega^T \hat{S}_W \omega + \gamma \omega^T \omega} \quad (9)$$

式中, $\gamma > 0$ 为小的正数(本文取为 0.05)。

下节探讨 SLDA 和 IMEDA 的关系。

3 SLDA 与 IMEDA 间的内在联系

3.1 理论探讨

本小节从两者的优化目标出发, 讨论 SLDA 和 IMEDA 的关系。

由文献[26]可知, 对于 LDA, 样本总散布矩阵 $S_T = X(I - \frac{1}{n}ee^T)X^T$, 类内散布矩阵 $S_W = XLX^T$, 类间散布矩阵 $S_B =$

$S_T - S_W = X((I - \frac{1}{n}ee^T) - L)X^T$, 其中 $L = I - W$, W 同 2.2.2

节, $n \times 1$ 维列向量 $e = (1, 1, \dots, 1)^T$ 。

故 SLDA 的优化目标可等价于:

$$\begin{aligned} & \max_{b, \alpha} \frac{b^T X W X^T b}{(1-\alpha)b^T X D X^T b + \alpha b^T \Delta^T \Delta b} \\ & \stackrel{\text{令 } \frac{\alpha}{1-\alpha} = \gamma}{\Leftrightarrow} \max_{b, \gamma} \frac{b^T X W X^T b}{b^T X D X^T b + \gamma b^T \Delta^T \Delta b} \\ & \Leftrightarrow \max_{b, \gamma} \frac{1}{\frac{b^T X D X^T b + \gamma b^T \Delta^T \Delta b}{b^T X W X^T b}} \end{aligned}$$

$$\begin{aligned} & \stackrel{\text{SLDA中 } D=I, \text{ 故 } D=I=L+W}{\Leftrightarrow} \max_{b, \gamma} \frac{1}{\frac{b^T X(L+W)X^T b + \gamma b^T \Delta^T \Delta b}{b^T X W X^T b}} \end{aligned}$$

$$\begin{aligned}
& \Leftrightarrow \max_{b, \gamma} \frac{1}{b^T X L X^T b + \eta b^T \Delta^T \Delta b} \\
& \Leftrightarrow \max_{b, \gamma} \frac{b^T X W X^T b}{b^T X L X^T b + \eta b^T \Delta^T \Delta b} \Leftrightarrow \\
& \begin{cases} \max_{b, \gamma} \frac{b^T (S_B + \frac{1}{n} X e e^T X^T) b}{b^T (S_W + \eta \Delta^T \Delta) b}, & \text{样本未中心化} \\ \max_{b, \gamma} \frac{b^T S_B b}{b^T (S_W + \eta \Delta^T \Delta) b}, & \text{样本中心化} \end{cases}
\end{aligned} \quad (10)$$

式中, $0 \leq \eta < \infty$ 。

对于 IMEDA, 因 $\hat{X} = \Gamma^T X$, 所以 $\hat{S}_B = \Gamma^T S_B \Gamma$, $\hat{S}_W = \Gamma^T S_W \Gamma$ 。故其优化目标式(9)可等价于:

$$\max_{\omega, \gamma} \frac{\omega^T \Gamma^T S_B \Gamma \omega}{\omega^T \Gamma^T S_W \Gamma \omega + \gamma \omega^T \omega} \stackrel{\omega = \Gamma^T b}{\Leftrightarrow} \max_{b, \gamma} \frac{b^T S_B b}{b^T (S_W + \gamma (\Gamma^+)^T (\Gamma^+)) b} \quad (11)$$

从式(10)、式(11)可见, 当样本中心化后, 变换矩阵 $\Gamma = \Delta^+$ 且 $\gamma = \eta$ 时, SLDA 和 IMEDA 优化目标相同, 即 SLDA 是 IMEDA 的特例, IMEDA 更具有一般性。理论上, SLDA 和 IMEDA 算法的时间复杂度均为 $O(d^3)$, 空间复杂度均为 $O(M^2 N^2)$ 。

3.2 实验比较

本实验采用 Yale、AR 和 FERET 人脸集, 相关信息见表 1, 使用正确识别率指标评估算法的性能, 并探究参数对模型的影响。

表 1 实验数据集

数据集	类别数	样本数	像素	Gm/Pn
Yale	15	165	32×32	G8/P3
AR	120	1680	50×40	G10/P4
FERET	200	1400	80×80	G5/P2

3.2.1 人脸数据集实验

在实验中, 我们对每类图像, 随机选择其中 m 张作为训练样本, 剩下 n 张作为测试样本 (标记为 Gm/Pn), 重复执行 20 轮, 并对不同子空间维数均采用最近邻法分类, 以最高平均正确识别率作为最终性能的衡量指标。

实验比较了原空间中未结合空间信息的正则化 LDA (记为 RLDA)、结合了空间关系但未正则化的 2DLDA^[27] (其直接在 2D 图像上执行 LDA, 因此某种程度上通过非正则化的方式利用了空间信息)、SLDA 和 IMEDA 等 4 种方法, 对比结果见表 2。我们观测到:

1) SLDA 和 IMEDA 由于考虑了空间信息, 识别性能和稳定性相当, 差异不足 1%, 但均一致高于 RLDA;

2) 尽管 2DLDA 也利用了空间信息, 但由于在计算行 (或列) 投影矩阵时, 忽略了列 (或行) 方向上的空间信息, 相对 SLDA 和 IMEDA 而言, 空间信息未被更多地利用, 因此其识别率仍低于 SLDA 和 IMEDA。

上述观测再次印证了 NFL 定理, 即先验知识的充分利用能产生更优的学习性能。

表 2 在人脸数据集上的识别率对比 (%)

方法	Yale	AR	FERET
RLDA	83.82±5.70	96.88±0.61	90.80±2.33
2DLDA	70.27±5.93	97.45±0.48	83.75±3.91
SLDA	85.39±4.08	97.58±0.65	91.99±2.46
IMEDA	84.47±4.54	98.50±0.52	92.16±2.50

实验中还比较了 SLDA 和 IMEDA 算法的运行时间 (单位秒)。在不同数据集下, SLDA 主要统计的是求解投影阵时特征值分解的运行时间, IMEDA 主要统计的是拉普拉斯矩阵的特征值分解的运行时间, 对图像降维后, 再求解投影阵的过程 (记为 D_IMEDA), 统计的是降维后求解投影阵时特征值分解的运行时间, 对比结果见表 3。由表看出:

1) SLDA 和 IMEDA 由于涉及特征值分解的矩阵维数相同, 故运行时间相当;

2) D_IMEDA 由于图像被预先降维, 大大降低了被涉及矩阵的维数, 故运行时间远小于前两者。

尽管理论上, SLDA 和 IMEDA 时间复杂度相当, 但由于拉普拉斯矩阵的构造和特征值分解可独立于数据进行, 故 IMEDA 算法整体的运行时间主要由 D_IMEDA 决定, 所以在实际应用上, IMEDA 比 SLDA 有更少的运行时间, 因而具有更高的效率。

表 3 在人脸数据集上的运行时间对比

方法	Yale	AR	FERET
SLDA	1.2794	6.4564	247.2283
IMEDA	2.1317	6.5132	245.1619
D_IMEDA	0.0123	0.3994	0.0182

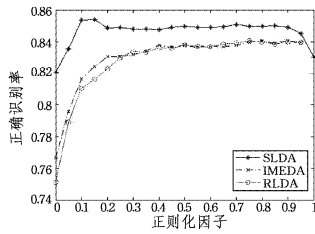
3.2.2 参数选择实验

本小节分别讨论 SLDA、IMEDA 和 RLDA 的正则化因子, IMEDA 的低频系数和 2DLDA 的迭代次数及降维维度对实验结果的影响。从图 1 可见, SLDA、IMEDA 和 RLDA 识别率均是在正则化因子取 0 时最小, 而后随着正则化因子增大而增大, 最后平稳波动; SLDA 的正则化因子用来控制空间平滑度, 在其较大范围内识别率变化均比较平稳, 故 SLDA 对空间正则化因子不太敏感, 但当其过大时, 过度的空间正则化反而不利于提高识别性能; IMEDA 和 RLDA 的正则化因子用来解决奇异性问题, 当其过大时, 并不能持续提高识别率, 只会增加计算量; 在讨论 IMEDA 正则化因子对实验的影响时, Yale、FERET 数据集上低频系数分别取图 2(a) 和 (c) 的局部峰值 80 和 100, 对 AR 数据集, 当图 2(b) 的低频系数大于 600 时, 识别率基本不再提高, 若继续增大低频系数只会增加计算复杂度, 故综合考虑, 我们将低频系数取为 600。对于 IMEDA, 其正则化因子仅仅是为避免奇异性而引入, 因此在固定 $\gamma=0.05$ 下讨论 IMEDA 中低频系数对实验效果的影响。由图 2(a) 和图 2(c) 可观察到, IMEDA 在 Yale 和 FERET 人脸集上的识别率均存在局部峰值, 也验证了 LST 作为一种过滤高频图像信息的预处理方法平滑了图像, 提高了识别率; 同时从图 2(b) 可见, 虽不存在明显的局部峰值, 但当低频系数大于 600 时, 正确识别率较高并逐渐趋于稳定。对于 2DLDA, 当降维维度大于 12 时, 增大降维维度并不会显著提升识别率, 当迭代次数大于 8 时, 识别率已趋于平稳, 故从计算成本和识别性能等综合考虑, 降维维度和迭代次数分别取为 12 和 8。图 3(a) 给出了在降维维度为 12 时, 性能随迭代次数的变化趋势, 可见当其大于 8 时识别率基本趋于稳定; 而图 3(b) 考察了迭代次数为 8 时, 性能曲线随降维维度的变化, 发现识别率开始时随其增大而增大, 当维度大于 12 时基本趋于稳定。

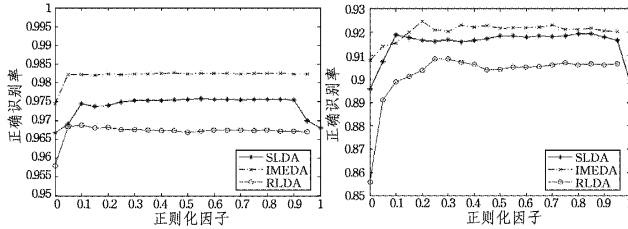
述分析方法,二是将其结合应用于多视图数据分析^[28]。

参 考 文 献

- [1] Duda R O, Hart P E, Stork D G. Pattern Classification(2nd edition)[M]. Wiley-Interscience, Hoboken, NJ, 2000
- [2] Fisher R A. The Use of Multiple Measurement in Taxonomic Problems[J]. Annals of Eugenics, 1937, 20(1): 139-152
- [3] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86
- [4] Cai Deng, He Xiao-fei, Hu Yu-xiao, et al. Learning a spatially smooth subspace for face recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, 2007: 1-7
- [5] He Xiao-fei, Niyogi P. Locality Preserving Projections[C]// Proceedings of Neural Information Processing Systems. Vancouver, Canada, 2003: 153-160
- [6] He Xiao-fei, Cai Deng, Yan Shui-cheng, et al. Neighborhood preserving embedding[C]// Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 2005: 1208-1213
- [7] Yan Shui-cheng, Xu Dong, Zhang Ben-yu, et al. Graph embedding and extension: A general framework for dimensionality reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51
- [8] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720
- [9] Zuo Wang-meng, Liu Lei, Wang Kuan-quan, et al. Spatially smooth subspace face recognition using LOG and DOG penalties [C]// Lecture Notes in Computer Science. Wuhan, China, 2009: 439-448
- [10] Hou Chen-ping, Nie Fei-ping, Zhang Chang-shui, et al. Learning an orthogonal and smooth subspace for image classification[J]. IEEE Signal Processing Letters, 2009, 16(4): 303-306
- [11] 李勇周, 罗大庸, 刘少强. 空间光滑且完整的子空间学习算法 [J]. 模式识别与人工智能, 2009, 22(3): 400-405
- [12] Lu Ji-wen, Tan Yap-peng. A doubly weighted approach for appearance-based subspace learning methods [J]. IEEE Transactions on Information Forensics and Security, 2010, 5(1): 71-81
- [13] Wen Hao, Wen You-kui. Face recognition using spatially smooth and maximum minimum value of manifold preserving[C]// International Conference on Applied Informatics and Communication. Xi'an, China, 2011: 194-204
- [14] Wang Li-wei, Zhang Yan, Feng Ju-fu. On the Euclidean Distance of Images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 1334-1339
- [15] Li Jun-li, Chen Gang, Chi Zhe-ru. A Fuzzy Image Metric with Application to Fractal Coding[J]. IEEE Transactions on Image Processing, 2002, 11(6): 636-643
- [16] Simard P, Cun Yan-le, Dender J S. Efficient Pattern Recognition Using a New Transformation Distance[C]// Advances in Neural Information Processing Systems 5. Denver, CO, USA, 1993: 50-58
- [17] Huttenlocher DP, Klanderman G A, Rucklidge WJ. Comparing Images Using the Hausdorff Distance[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(9): 850-863



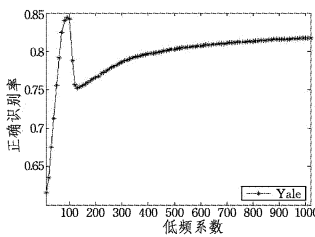
(a) Yale 人脸数据集



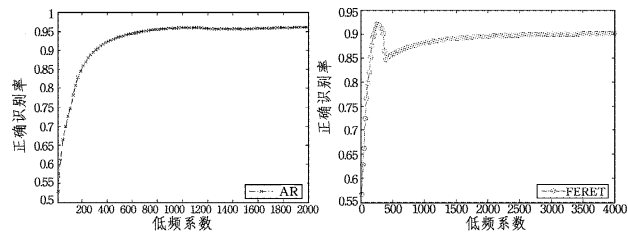
(b) AR 人脸数据集

(c) FERET 人脸数据集

图 1 正则化因子对实验的影响



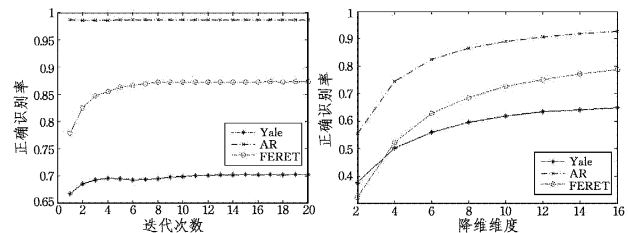
(a) Yale 人脸数据集



(b) AR 人脸数据集

(c) FERET 人脸数据集

图 2 IMEDA 中不同低频系数下的识别率



(a) 迭代次数

(b) 降维维度

图 3 2DLDA 算法中迭代次数和降维维度对实验的影响

结束语 本文尝试探究了空间信息在线性判别分析(LDA)中不同利用方式的性能比较,侧重讨论了空间信息在优化目标和距离度量两种嵌入方式下的内在联系。理论上证明了对中心化样本,SLDA 是 IMEDA 的特例;分析了算法的时间和空间复杂度;经验上表明了 SLDA 和 IMEDA 的识别性能和稳定性相当;并比较了其运行时间,结果 IMEDA 较快。像 SLDA 这类将空间信息嵌入优化目标中的研究方法,嵌入方法因被嵌的优化目标而异;而像 IMEDA 这类将空间信息嵌入距离度量的研究方法,嵌入方法独立于被嵌算法,操作简单,效率高,有较强的推广性。下一步工作,一是将结合 Li 等^[19]提出的自适应图像欧氏距离(AIMED)进一步完善上

- [18] Chen Jie, Wang Rui-ping, Shan Shi-guang, et al. Isomap based on the image Euclidean distance[C]// IEEE International Conference on Pattern Recognition. Hong Kong, China, 2006; 1110-1113
- [19] Li Jing, Lu Bao-liang. An adaptive image Euclidean distance[J]. Pattern Recognition, 2009, 42(3): 349-357
- [20] 黄晓华, 梁超, 郑文明. 图像空间中的鉴别型局部线性嵌入方法[J]. 中国图象图形学报, 2010, 15(12): 1776-1782
- [21] Sun Bing, Feng Ju-fu, Wang Li-wei. Learning IMED via Shift-Invariant transformation[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA, 2009; 1398-1405
- [22] Gu Sui-cheng, Tan Ying, He Xin-gui. Laplacian smoothing transform for face recognition[J]. Science China Information Sciences, 2010, 53(12): 2415-2428
- [23] Vapnik V. The nature of statistical learning theory [M]. New York: Springer Verlag, 1995; 1-50
- [24] O'Sullivan Finbarr. Discretized Laplacian Smoothing by Fourier Method [J]. Journal of the American Statistical Association, 1991, 86(415): 634-642
- [25] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]// Proceedings of Advances in MIT Press, Neural Information Processing Systems. MA, USA; 2001; 585-591
- [26] He Xiao-fei, Yan Shui-cheng, Hu Yu-xiao, et al. Face recognition using laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340
- [27] Ye Jie-ping, Janardan R, Qi Li. Two-dimensional linear discriminant analysis[C]// Proceedings of Advances in Neural Information Processing Systems. Vancouver, British, 2004; 1569-1576
- [28] Chen Xiao-hong, Chen Song-can, Xue Hui, et al. A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data[J]. Pattern Recognition, 2012, 45(5): 2005-2018

(上接第 35 页)

从图 2 可以观察到项目近邻数量 K 对推荐算法平均绝对误差 MAE 有较大的影响。随着项目近邻数量 K 的增加, 两种推荐算法的 MAE 值首先不断降低, 推荐质量随 K 值增加而不断改进; $K > 40$ 后, MAE 趋于平缓, 推荐算法的推荐质量改进不大。在项目近邻数量 K 取任何值的情况下, 基于耦合对象相似的推荐算法 (COSRec) 的 MAE 都低于基于简单匹配相似度的推荐算法 (SMSRec) 的 MAE 值, 说明在相同条件下, 基于耦合对象相似的推荐算法 (COSRec) 的推荐质量高于基于简单匹配相似度的推荐算法 (SMSRec)。

然而, 从两种推荐算法的 MAE 值变化曲线可以看出, 虽然基于耦合对象相似度的推荐算法 (COSRec) 推荐质量在不同 K 值条件都高于基于简单匹配相似度的推荐算法 (SMSRec), 但是两者的差距不大, 保持在 0.01~0.02 之间。这主要是因为从 MovieLens100K 数据集中只能抽取电影的 genre 特征, 而且其它如 director, actor 等重要特征的缺失在很大程度上限制了基于耦合对象相似度推荐算法的性能。我们认为在能抽取更多项目特征的情形下, 基于耦合对象相似度的推荐算法的推荐质量比传统基于内容的推荐算法有较大的改进。

结束语 推荐系统在为用户解决信息过载问题方面发挥着越来越重要的作用, 它可以为其推荐符合其偏好的商品、新闻、电影和音乐等, 甚至在社交网络系统中可以推荐与用户有共同爱好的好友。特别是在电子商务系统中, 推荐系统根据用户的信息资料和推荐项目的文本信息, 为用户推荐个性化的产品和服务信息, 在提高用户体验的同时, 也增加了电子商务企业的效益。针对传统基于内容的推荐算法的项目之间缺少合适度量手段计算项目之间的相似度的问题, 本文提出了一种基于耦合对象相似度的项目推荐算法。该推荐算法由项目特征抽取、项目耦合相似度模型构建和评分预测等 3 个主要成分组成, 其中的核心在于利用耦合对象相似度 (COS) 度量推荐项目之间的相似度, 通过耦合对象相似度可以有效捕获特征值的频率分布情况和特征依赖聚合度。在现实数据集上的实验验证了在缺失项目特征的情况下, 基于耦合对象相似度的项目推荐算法比传统的基于内容的推荐算法在推荐精

度上有一定程度的改进。

由于目前公开的推荐算法评测数据集普遍包含用户的评分数据, 且项目本身的特征信息较少, 如 MovieLens100K 数据集中仅电影的风格特征可以用来计算电影的相似度, 而且其他如导演、演员等特征信息缺失, 本文下一步的研究将抽取更多的项目特征信息来改进基于耦合对象相似度的推荐算法。另外, 冷启动问题是推荐系统的研究热点, 基于内容的推荐算法在解决冷启动问题上较协同过滤的推荐算法有本质上的优势, 本文下一步将研究基于耦合对象相似度推荐算法在解决推荐系统冷启动问题上的性能表现。

参 考 文 献

- [1] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems; A survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749
- [2] 赵亮, 胡乃静, 张守志. 个性化推荐算法设计[J]. 计算机研究与发展, 2002, 39(8): 986-991
- [3] Balabanovic M, Shoham Y. Fab: content-based, collaborative recommendation[J]. Communications of the ACM, 1997, 40(3): 66-72
- [4] Melville P, Mooney R, Nagarajan R. Content-boosted collaborative filtering for improved recommendations[C]// Proceedings of the National Conference on Artificial Intelligence. AAAI Press, MIT Press, 1999(2002): 187-192
- [5] Wang C, Cao L, Wang M, et al. Coupled nominal similarity in unsupervised learning[C]// CIKM, ACM, 2011; 973-978
- [6] Yu Y, Wang C, Gao Y, et al. A Coupled Clustering Approach for Items Recommendation[C]// The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Gold Coast, Australia, 2013
- [7] Boriah S, Chandola V, Kumar V. Similarity measures for categorical data; a comparative evaluation[C]// SDM 2008, 2008; 243-254
- [8] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]// WWW. ACM, 2001; 285-295
- [9] <http://www.grouplens.org/node/73>