

基于大间距准则的不相关 保局投影分析

龚 劬¹ 唐萍峰¹

摘要 局部保持投影 (Locality preserving projections, LPP) 算法只保持了目标在投影后的邻域局部信息, 为了更好地刻画数据的流形结构, 引入了类内和类间局部散度矩阵, 给出了一种基于有效且稳定的大间距准则 (Maximum margin criterion, MMC) 的不相关保局投影分析方法. 该方法在最大化散度矩阵迹差时, 引入尺度因子 α , 对类内和类间局部散度矩阵进行加权, 以便找到更适合分类的子空间并且可避免小样本问题; 更重要的是, 大间距准则下提取的判别特征集一般情况下是统计相关的, 造成了特征信息的冗余, 因此, 通过增加一个不相关约束条件, 利用推导出的公式提取不相关判别特征集, 这样做, 对正确识别更为有利. 在 Yale 人脸库、PIE 人脸库和 MNIST 手写数字库上的测试结果表明, 本文方法有效且稳定, 与 LPP、LDA (Linear discriminant analysis) 和 LPMIP (Locality-preserved maximum information projection) 方法等相比, 具有更高的正确识别率.

关键词 特征提取, 大间距准则, 保局投影, 不相关判别分析, 人脸识别

引用格式 龚劬, 唐萍峰. 基于大间距准则的不相关保局投影分析. 自动化学报, 2013, 39(9): 1575–1580

DOI 10.3724/SP.J.1004.2013.01575

Uncorrelated Locality Preserving Projections Analysis Based on Maximum Margin Criterion

GONG Qu¹ TANG Ping-Feng¹

Abstract Locality preserving projections (LPP) algorithm can only preserve nearest local quantity, so within-class and between-class local scatter matrices are introduced for characterizing the manifold structure better, and a method called uncorrelated locality preserving projection analysis based on effective and stable maximum margin criterion (MMC) is proposed. When maximizing the trace difference of scatter matrix, weight the within-class and between-class local scatter matrix through a regularized parameter so as to find the better classification subspace and avoid small sample problem. More importantly, the discriminant feature set based on the MMC is generally statistical correlated, which makes the feature information be redundant, so an uncorrelated constraint was added in the paper and the uncorrelated discriminant feature set is extracted by the derived formulas, which are more favorable for the correct recognition. Ultimately experiments on Yale, PIE face database and MNIST handwritten digit database show that the method in this paper is effective and stable and has a higher correct recognition rate compared with the LPP, LDA (Linear discriminant analysis) and LPMIP (Locality-preserved maximum information projection).

Key words Feature extraction, maximum margin criterion (MMC), locality preserving projections (LPP), uncorrelated discriminant analysis, face recognition

Citation Gong Qu, Tang Ping-Feng. Uncorrelated locality preserving projections analysis based on maximum margin criterion. *Acta Automatica Sinica*, 2013, 39(9): 1575–1580

在机器学习、人工智能和计算机视觉等领域, 如何对目标数据进行有效地特征提取是一个十分关键的问题, 研究人员对此已经提出了很多的算法^[1-4]. 其中, 最为经典的两种算法是主成分分析 (Principal component analysis, PCA)^[5] 和线性判别分析 (Linear discriminant analysis, LDA)^[6] 方法. 然而当训练样本的数量较小时, 类内散度矩阵奇异, 便出现了所谓的小样本问题^[7], 已经有不少解决此问题的算法提出, 如大间距准则 (Maximum margin criterion, MMC) 分析^[8]、正则化局部保持投影^[9] 和张量局部 Fisher 判别分析^[10] 等. MMC 用类间与类内散度矩阵的差作为分类的目标函数, 避免了类内散度矩阵求逆, 从而有效地解决了小样本问题, 是一种非常有效且稳定的算法.

局部保持投影 (Locality preserving projections, LPP)^[11] 只考虑了邻域局部信息, 而没有考虑非邻域局部信息, 因此, LPP 只能保证邻近的两个目标在投影后尽可能接近. 边缘判别分析 (Marginal Fisher analysis, MFA)^[12] 通过引入类边缘分离度量信息, 不仅能达到 LPP 的目的, 而且还能使不同类的目标在投影后尽可能分得更开, 但是却存在着小样本问题. 所以本文引入一种新的类内和类间局部散度矩阵用于刻画数据间的局部信息, 在 MMC 基础上, 使类内局部散度矩阵最小, 同时使类间局部散度矩阵最大, 实现特征提取后同类更近不同类相隔更远的目的. 文献 [13] 指出, 统计不相关的判别特征向量分类性能优于正交特征分类. 通过分析发现, 在大间距准则下提取的判别特征集各分量间一定是统计相关的, 因此, 本文提出了一种基于大间距准则的不相关保局投影分析 (Uncorrelated locality preserving projections analysis based on maximum margin criterion, ULPA-MMC) 算法, 充分抓住类间和类内局部信息, 采用 MMC 准则克服小样本问题, 采用奇异值分解降维, 最后, 通过增加一个不相关约束条件, 利用推导出的公式提取不相关判别特征向量用于分类识别.

1 ULPA-MMC 方法

1.1 特征提取

假设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 是 N 维欧氏空间 \mathbf{R}^N 中的 n 个训练样本, 共有 c 个类别, 每个类对应的样本数量为 n_1, n_2, \dots, n_c . 一般的线性特征提取方法是找到一个线性映射满足下式:

$$f: \mathbf{x}_i \mapsto \mathbf{y}_i = P^T \mathbf{x}_i \quad (1)$$

其中, $P \in \mathbf{R}^{N \times d}$ 是按照某种优化准则得到的投影矩阵, $\mathbf{y}_i \in \mathbf{R}^{d \times 1}$ 代表 \mathbf{x}_i 的投影特征向量, 下面介绍 LPP 的优化准则.

LPP 算法是为了找到一个能够保持数据局部结构的子空间, 其目标函数如下:

$$\min \text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n (\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T R(i, j) \right) \quad (2)$$

其中, $R(i, j) = \begin{cases} W(i, j), & \text{若 } \mathbf{x}_i \text{ 位于 } \mathbf{x}_j \text{ 的 } k \text{ 邻域} \\ 0, & \text{其他情形} \end{cases}$, $\text{tr}(\cdot)$

Supported by Innovative Talent Training Project, the Third Stage of "211 Project", Chongqing University (S-09110)

本文责任编辑 周杰

Recommended by Associate Editor ZHOU Jie

1. 重庆大学数学与统计学院 重庆 401331 1. College of Mathematics and Statistics, Chongqing University, Chongqing 401331

表示矩阵的迹, W 用来衡量 \mathbf{x}_i 和 \mathbf{x}_j 的相似程度, W 越大表示 \mathbf{x}_i 和 \mathbf{x}_j 越接近, 当 \mathbf{x}_i 和 \mathbf{x}_j 相等时, $W = 1$, W 定义如下:

$$W(i, j) = \exp \left\{ -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma} \right\} \quad (3)$$

从式 (2) 中邻接矩阵 R 的定义可以看出, LPP 是一种无监督的算法, 并且较好地保持了邻域局部信息, 但是却忽略了非邻域局部信息. 值得一提的是, 文献 [14] 给出了一种自适应流形算法, 改进了流形上的邻域刻画方式, 相比于邻接矩阵 R 来说, 能更好地保持数据点原本在流形上的邻域关系, 但是计算比较复杂. 接下来给出 ULPA-MMC 方法的目标函数定义:

$$\max \text{tr} \left(\begin{array}{c} \alpha \sum_{i=1}^n \sum_{j=1}^n (\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T \bar{S}(i, j) - \\ (1 - \alpha) \sum_{i=1}^n \sum_{j=1}^n (\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T S(i, j) \end{array} \right) \quad (4)$$

其中, $\text{tr}(\cdot)$ 表示矩阵的迹, α 是一个尺度因子, $\alpha \in [0, 1]$,

$$S(i, j) = \begin{cases} W(i, j), & \text{若 } \mathbf{x}_i \text{ 和 } \mathbf{x}_j \text{ 是相同的类} \\ 0, & \text{其他} \end{cases} \quad (5)$$

$$\bar{S}(i, j) = \begin{cases} W(i, j), & \text{若 } \mathbf{x}_i \text{ 和 } \mathbf{x}_j \text{ 是不同的类} \\ 0, & \text{其他} \end{cases} \quad (6)$$

由式 (5) 所定义的邻接矩阵 S 与 LPP 算法中的邻接矩阵 R 功能是相似的, 不同的是本文的矩阵充分利用了样本的类别信息. 为了与 LPP 中“邻域局部”的说法相统一, 因此, 本文称式 (7) 和 (8) 所定义的矩阵分别为类间和类内局部散度矩阵. 这样我们的目标函数便同时保持了类间和类内局部信息.

需要特别说明的是, 在特征空间中, 利用式 (4) 得到的最优投影矩阵进行特征提取, 会使得不同类的样本投影后距离间隔越远, 而相同类的样本投影后则会越接近, 这正是我们所期望得到的.

从目标函数式 (4) 中可以得到类间和类内局部散度矩阵 G_b 和 G_w 的定义如下:

$$G_b = \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \bar{S}(i, j) = 2X\bar{L}X^T \quad (7)$$

其中, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, $\bar{L} = \bar{D} - \bar{S}$, \bar{D} 是对角阵, 其第 i 个对角元素为 \bar{S} 的第 i 行元素求和.

$$G_w = \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T S(i, j) = 2XLX^T \quad (8)$$

其中, $L = D - S$, D 是对角阵, 其第 i 个对角元素为 S 的第 i 行元素求和.

至此, 将式 (1) 代入式 (4) 便可得到化简的目标函数如下:

$$\begin{aligned} \max J(P) &= \text{tr}(P^T[\alpha G_b - (1 - \alpha)G_w]P) = \\ &= \text{tr}(P^T X (\alpha \bar{L} - (1 - \alpha)L) X^T P) = \\ &= \text{tr}(P^T X G X^T P) \end{aligned} \quad (9)$$

其中, P 是线性投影矩阵, $P \in \mathbf{R}^{N \times d}$, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, $G = \alpha \bar{L} - (1 - \alpha)L$, 而 $\alpha \in [0, 1]$, 当 α 取值越大时, 目标函

数更注重类间局部信息对分类的影响, 当 α 取值越小时, 目标函数更注重类内局部信息对分类的影响, 通过调整 α 的值, 能更好地反映出高维空间中的流形结构.

在式 (9) 基础上增加一个正交约束 $P^T P = I$, 与文献 [8] 的求解方法类似, 于是可以得到目标函数 (9) 在正交约束条件下的最大值为 $J_{\max} = \sum_{i=1}^d \lambda_i$, $\lambda_1, \dots, \lambda_d$ 是 XGX^T 的前 d 个最大特征值, 最优解为 $P = [\eta_1, \dots, \eta_d]$, η_1, \dots, η_d 是 $\lambda_1, \dots, \lambda_d$ 所对应的特征向量.

但是矩阵 XGX^T 维数较高, 直接求其特征值, 计算复杂度高, 所以要先对它进行降维处理. 设 X 的秩为 r , $X = U_1 \sum_r V_1^T$ 是 X 的奇异值分解, 其中, $U_1 \in \mathbf{R}^{N \times r}$, $V_1 \in \mathbf{R}^{r \times n}$. 令 $M = U_1^T XGX^T U_1$, $\phi \in \mathbf{R}^{r \times d}$, 可以得到如下优化问题:

$$\begin{aligned} \max J(\phi) &= \text{tr}(\phi^T M \phi) \\ \text{s.t. } &\phi^T \phi = I \end{aligned} \quad (10)$$

定理 1. 若问题 (10) 的最优解为 ϕ , 则优化问题 (9) 的最优解为 $P = U_1 \phi$.

证明. 设 λ 和 η 分别是 XGX^T 的特征值和特征向量, μ 和 φ 分别是 M 的特征值和特征向量.

令 $XGX^T \eta = \lambda \eta$, $M \varphi = \mu \varphi$, 由于 $U_1^T U_1 = I$, 则有 $M U_1^T U_1 \varphi = \mu \varphi$ 成立. 两边同时左乘 U_1 可得

$$U_1 M U_1^T U_1 \varphi = \mu U_1 \varphi$$

对上式左边进一步化简可得

$$\begin{aligned} U_1 M U_1^T U_1 \varphi &= U_1 U_1^T XGX^T U_1 U_1^T U_1 \varphi = \\ &U_1 U_1^T U_1 \sum_r V_1^T G(U_1 \sum_r V_1^T)^T U_1 U_1^T U_1 \varphi = \\ &U_1 \sum_r V_1^T G(U_1 \sum_r V_1^T)^T U_1 \varphi = \\ &XGX^T U_1 \varphi \end{aligned}$$

于是 $XGX^T U_1 \varphi = \mu U_1 \varphi$ 成立, 所以 XGX^T 和 M 有相同的特征值即 $\lambda = \mu$, 并且二者的特征向量满足 $\eta = U_1 \varphi$.

又因为问题 (10) 的最优值 $J_{\max} = \sum_{i=1}^d \mu_i$, 其中 μ_1, \dots, μ_d 是 M 的前 d 个最大的特征值, 最优解 $\phi = [\varphi_1, \dots, \varphi_d]$, $\varphi_1, \dots, \varphi_d$ 是 μ_1, \dots, μ_d 所对应的特征向量, 所以 $P = U_1 \phi$ 成立. \square

直接计算 XGX^T 的特征值计算复杂度为 $O(N^3)$, 对 X 进行奇异值分解的计算复杂度为 $O(Nn^2)$, 对 M 进行特征分解的计算复杂度为 $O(r^3)$. 因此, 定理 1 给出了求解投影矩阵 P 的更高效的方法.

1.2 不相关分析

结合定理 1 的结论和文献 [15] 定义的不相关变换, 可以得到本文的不相关变换如下:

$$Y = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_d \end{pmatrix} = \begin{pmatrix} \varphi_1^T \\ \varphi_2^T \\ \vdots \\ \varphi_d^T \end{pmatrix} U_1^T X \quad (11)$$

其中, $\zeta_i \in \mathbf{R}^{1 \times n}$.

定理 2. 对于判别特征集 Y 中的任意两个特征分量 ζ_i 和 ζ_j , 若 $\varphi_i^T U_1^T S_i U_1 \varphi_j = 0$, 则 ζ_i 和 ζ_j 是统计不相关的. (注: 总体散度矩阵 $S_t = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \text{EX})(\mathbf{x}_i - \text{EX})^T$)

证明. 由式 (11) 可以得到 $\zeta_i = \varphi_i^T U_1^T X$, 于是有:

$$\begin{aligned} E[(\zeta_i - E\zeta_i)(\zeta_j - E\zeta_j)^T] &= \\ \varphi_i^T U_1^T E[(X - \text{EX})(X - \text{EX})^T] U_1 \varphi_j &= \\ n \cdot \varphi_i^T U_1^T S_i U_1 \varphi_j &= 0 \end{aligned}$$

所以 ζ_i 和 ζ_j 是统计不相关的. \square

值得说明的一点是, 由定理 2 知, 只有当 $\varphi_i^T U_1^T S_i U_1 \varphi_j = 0$ 恒成立时, 才有判别特征集 Y 统计不相关; 否则便不成立. 因为很多经典的算法, 如 PCA、LDA 等, 不用添加不相关约束条件, 通过分析可知它们得到的判别特征集 Y 是统计不相关的. 因此, 接下来有必要先验证问题 (10) 得到的最优判别特征集 Y 是否是统计不相关的.

由第 1.1 节可知 $\varphi_i^T \varphi_j = 0$, 而 $M \varphi_j = \lambda_j \varphi_j$, 所以有 $\varphi_i^T M \varphi_j = 0$. 因为 $U_1^T U_1 = I$, 则 $\varphi_i^T U_1^T U_1 M U_1^T U_1 \varphi_j = 0$ 等价于

$$\varphi_i^T U_1^T X (\alpha \bar{L} - (1 - \alpha)L) X^T U_1 \varphi_j = 0 \quad (12)$$

而不相关判别条件 $\varphi_i^T U_1^T S_i U_1 \varphi_j = 0$ 等价于

$$\varphi_i^T U_1^T X L_t X^T U_1 \varphi_j = 0 \quad (13)$$

其中, $L_t = D_t - W_t$, $W_t \in \mathbf{R}^{n \times n}$ 每个元素都为 1, D_t 是对角矩阵且 $D_t(i, i) = \sum_{j=1}^n W_t(i, j)$.

至此可以得出, 一般情况下不能由式 (12) 推出式 (13) 恒成立, 因此, 问题 (10) 得到的最优判别特征集 Y 一般是统计相关的, 并且大量的实验也证明了这一点.

接下来通过在问题 (10) 中增加一个不相关判别约束条件, 重新求解. 首先, 需要构造第一个 φ , 即 $\varphi_1 = \arg \max_{\varphi} \varphi^T M \varphi$, 假设已经求出了 K 个统计不相关向量 $\varphi_1, \dots, \varphi_K$, 于是第 $K+1$ 个统计不相关向量便可通过下式得到:

$$\begin{aligned} \max J(\varphi) &= \varphi^T M \varphi \\ \text{s.t. } &\varphi^T \varphi = 1 \\ &\varphi^T U_1^T S_i U_1 \varphi_i = 0, \quad i = 1, \dots, K \end{aligned} \quad (14)$$

下面的定理 3 给出了求解问题 (14) 的方法.

定理 3. 第 $K+1$ 个统计不相关向量 φ_{K+1} 是下式最大特征值所对应的特征向量.

$$[I - A \phi^T (\phi A A \phi^T)^{-1} \phi A] M \varphi = \lambda \varphi \quad (15)$$

其中, $A = U_1^T S_t U_1$, $\phi = [\varphi_1, \dots, \varphi_K]^T$, I 是单位矩阵.

证明. 对于优化问题 (14), 首先, 令 $A = U_1^T S_t U_1$, 再用拉格朗日乘法可得:

$$f(\varphi) = \varphi^T M \varphi - \lambda(\varphi^T \varphi - 1) - \sum_{i=1}^K \mu_i \varphi^T A \varphi_i \quad (16)$$

对 f 求偏导并令 $\frac{\partial f}{\partial \varphi} = 0$, 有

$$2M \varphi - 2\lambda \varphi - \sum_{i=1}^K \mu_i A \varphi_i = 0 \quad (17)$$

在式 (17) 两边同时乘以 φ^T 可得 $\lambda = \varphi^T M \varphi$, 于是要使目标函数值最大, 即 λ 最大即可. 在式 (17) 两边同时乘以 $\varphi_j^T A$ 可得:

$$2\varphi_j^T A M \varphi - \sum_{i=1}^K \mu_i \varphi_j^T A A \varphi_i = 0 \quad (18)$$

将 $j = 1, \dots, K$ 带入到式 (18) 可得:

$$2 \begin{pmatrix} \varphi_1^T \\ \vdots \\ \varphi_K^T \end{pmatrix} AM\varphi - \begin{pmatrix} \varphi_1^T \\ \vdots \\ \varphi_K^T \end{pmatrix} AA(\varphi_1, \dots, \varphi_K) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix} = 0 \quad (19)$$

令 $\phi = [\varphi_1, \dots, \varphi_K]^T$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_K \end{pmatrix}$, 代入式 (19) 可得:

$$\mu = 2(\phi AA\phi^T)^{-1} \phi AM\varphi \quad (20)$$

式 (17) 可以写成下面的形式:

$$2M\varphi - 2\lambda\varphi - A\phi^T\mu = 0 \quad (21)$$

将式 (20) 带入式 (21), 化简即得:

$$[I - A\phi^T(\phi AA\phi^T)^{-1}\phi A] M\varphi = \lambda\varphi \quad (22)$$

□

假设已经求出了前 d 个不相关向量, 当 φ_{d+1} 对应的最大特征值小于 ε (ε 取 0.01) 时, φ_{d+1} 对目标函数贡献极小, 即终止求解, 得到不相关投影矩阵 $\phi = [\varphi_1, \dots, \varphi_d]^T$.

总结本文 ULPA-MMC 算法的具体实现步骤如下:

- 1) 给定训练样本集 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, 通过式 (5) 和 (6) 计算邻接矩阵, 由式 (7) 和 (8) 计算 \bar{L} 和 L , 得到矩阵 G ;
- 2) 对 X 进行奇异值分解, 得到其左奇异值矩阵 U_1 , 从而得到矩阵 $M = U_1^T X G X^T U_1$;
- 3) 取 M 的最大特征值对应的特征向量, 作为第一个不相关向量 φ_1 , 由式 (15) 计算所有不相关相关向量, 令 $\phi = [\varphi_1, \dots, \varphi_d]^T$, 从而得到本文的投影矩阵 $P = U_1\phi$;
- 4) 给定一个中心化了的训练样本 \mathbf{x}_0 , 得到其不相关判别特征 $\mathbf{y}_0 = P^T \mathbf{x}_0$, 最后采用最近邻分类得到 \mathbf{x}_0 的类别。

1.3 讨论

本文的工作是在文献 [16] 提出的 LPMIP/QR (Locality-preserved maximum information projection/QR) 算法基础上进行的. 但是需要指出的是, 本文与 LPMIP/QR 方法有着本质的区别, 首先, LPMIP 算法采用的是基于不完全 Cholesky 分解的 QR 分解方法, 而本文采用的是用 X 的左奇异矩阵进行降维, 计算复杂度低且不会破坏原始数据的结构; 其次, 用 LPMIP 算法得到的最优判别特征集 Y 一般都是统计相关的, 而本文方法能得到最优不相关判别特征集 Y , 具有更好的分类能力; 还有 LPMIP 算法采用的是无监督的邻接矩阵, 需要调整 k 邻域参数, 而本文的有监督邻接矩阵则不需要, 而且更简单有效. 而本文与文献 [17-18] 相比, 虽然出发点比较一致, 但是所使用的分类判别准则完全不一样, 求解的方法也完全不同, 再者本文更兼有 LPMIP 算法的优势.

2 实验结果与分析

2.1 实验准备知识

本文有两个重要的参数 σ 和 α 需要设定, 与文献 [16] 的参数设定方法类似, 我们令 $\sigma = 2^l \sigma_0$, 其

中, $l = \{-5, -4, \dots, 0, 1, \dots, 10\}$, σ_0 是训练样本矩阵 X 的二范数. 尺度因子 $\alpha = 2^{a/4.5} \lambda_{\max}(U_1^T G_w U_1) / \lambda_{\max}(U_1^T G_b U_1)$, 其中, $\lambda_{\max}(\cdot)$ 表示矩阵的最大特征值, $a = \{-20, \dots, -1, 0, 1, \dots, 20\}$, 这里 α 的设置与文献 [19] 中的参数 M 设置方法相似, 因为 α 是一个尺度因子, 主要用于平衡矩阵 G_w 和 G_b , 通过定理 1 可知亦可用于平衡矩阵 $U_1^T G_w U_1$ 和 $U_1^T G_b U_1$, 因此, 我们用这两个矩阵的最大特征值的比, 乘以一个非负常数来近似 α 值. 在以后的所有实验中, l 和 a 取值为最大识别率时所对应的值, 参数 a 与识别率的关系如图 1.

本文实验所采用的数据库都是模式识别领域广泛采用的, 主要有 MNIST 手写数字库, Yale 人脸库和卡耐基梅隆大学的 PIE 人脸库, 两个人脸库按眼部位置对齐的方式, 分别被裁剪为 64 像素 \times 64 像素和 32 像素 \times 32 像素大小.

MNIST 手写数字库包含 60 000 张训练图像和 10 000 测试图像, 本文只选取其中的 2 000 张图像 (每类有 200 张) 作为训练集, 再从剩下的测试图像中选取 2 000 张作为测试集; Yale 人脸库¹含有在不同光照和表情下的 165 张人脸图像 (15 人, 每人 11 张), 随机选取每个人的 n ($n = 2, 3, 4, 5, 6$) 张图像作为训练集, 余下的作为测试集, 对每个 n 值, 作 10 次随机划分实验; PIE 人脸库共有 68 人的 41 368 张人脸图像, 它包含了 13 种不同的姿势、43 种不同的光照条件、4 种不同的表情, 本文从中随机选取 15 个人的 5 种接近正面 (C05、C07、C09、C27、C29) 的人脸图像 (每人 170 张) 作为实验数据, 然后, 随机选取每个人的 n ($n = 5, 10, 20, 30$) 张图像作为训练集, 余下的作为测试集, 对每一个 n 值, 作 10 次随机划分实验. 对于以上的每个 10 次随机划分实验, a 保持不变时, α 会随着训练样本集的变化而变化.

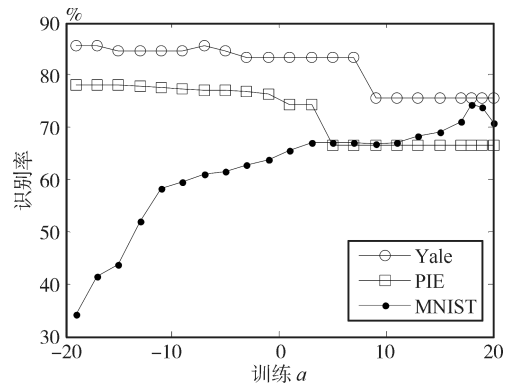


图 1 本文方法参数 a 与识别率的关系

Fig. 1 Relationship of recognition accuracy and a

从图 1 可以看出: 对人脸库参数 a 取小于 -5 时, 识别率达到最大且比较稳定; 对手写数字库, 则取 18 或 19 时, 识别率最大.

2.2 不相关性验证

本文将直接求解问题 (10) 得到投影矩阵的方法称之为正交方法, 而求解问题 (14) 得到投影矩阵的方法称之为不相关方法. 二者的各项参数设置都一致, 采用基于欧氏距离的最近邻分类, 令 l 为 5 或 6, a 取 -13 或者 -12 . 做 10 次随机划分实验, 每次都能获得一个最高识别率, 表 1 给出了最高识别率的均值和标准差.

¹ <http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html>

表1 人脸库中不相关与正交方法的识别能力比较 (%)
Table 1 Comparison of classification performance (%)

人脸库	训练集	正交方法	不相关方法
Yale ($l = 5, a = -12$)	2	66.33 ± 4.50	68.78 ± 4.19
	3	77.00 ± 1.23	77.96 ± 2.33
	4	84.24 ± 2.38	85.29 ± 1.97
PIE ($l = 6, a = -13$)	5	76.01 ± 2.29	78.54 ± 2.90
	10	91.13 ± 1.31	92.40 ± 1.22

从表1可以看出, 不相关方法在各种条件下都能表现出一定的优势, 说明了不相关判别特征是有效可行的.

2.3 实验结果分析

本文用于实验对比的算法有 PCA+LDA^[7]、LPP^[11]、ULDA^[15]、MMC^[8] 和 LPMIP^[16], 同样采用基于欧氏距离的最近邻分类器进行分类. 对于人脸库, 本文方法令 l 为 5 或 6, a 取 -12 或 -13 , 而对于 MNIST 库, 令 l 为 -1 , a 为 18. 表 2~表 4 分别比较了不同算法在三种数据库不同训练集下的 10 次随机划分实验的最高识别率的均值和标准差.

通过对表 2~表 4 和图 2 识别结果进行比较和分析, 可以得到如下的结论:

1) 在三种不同的数据库, 各种不同的训练集下, 本文算法都能获得一致较高的正确识别率, 而且在人脸库中识别率的标准差也比较小, 说明了算法的稳定性较好;

2) 由于本文算法和 LPMIP 算法引入了局部信息, 并且对其进行了加权处理, 更好地刻画了数据的流形结构, 因此, 识别率高于其他算法, 特别地, 当训练样本数较少时, 这种提升效果更加的明显;

3) 与 LPMIP 算法相比, 本文算法有不相关判别特性的优势, 因此, 识别结果更优, 当训练样本数少时, 提升效果会更明显;

4) 从 2) 和 3) 可知, 当训练样本数较少时, 本文算法识

别结果提升会更好, 所以本文方法更适合于人脸识别, 这是因为人脸库的训练样本数远小于其空间维数, 而手写数字库的训练样本数远大于其空间维数, 人脸库的平均识别率相对于 LPMIP 算法提高大于 1%, 而手写库的提高则较小.

3 结论

本文首先引入类内和类间局部散度矩阵, 在 MMC 准则下保持判别特征向量集的不相关性质, 于是提出了一种有效且稳定的 ULPA-MMC 算法. 一方面利用奇异值分解对包含类内和类间局部信息的散度矩阵进行特征提取, 能够得到更有效的判别特征向量集; 另一方面不相关判别特征能够消除信息的冗余性, 因此具有更好的分类能力, 而在 Yale、PIE 和 MNIST 数据库上的实验有力地证明了这一点. 但是, 与 LPP 等流形学习算法类似, ULPA-MMC 算法也需要手工调整参数, 接下来的一项工作便是如何设计一种自适应算法, 根据不同的数据库快速自动地调整参数, 得到最好的识别效果; 而且可以借鉴文献 [19-20] 的核思想将本文算法推广到核方法版本.

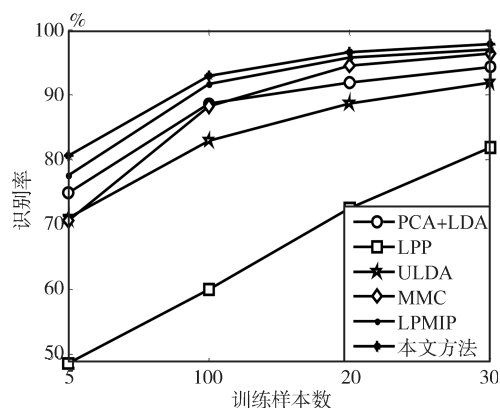


图2 PIE 人脸库 6 种算法的最高识别率
Fig. 2 Top recognition accuracy of six methods

表2 Yale 人脸库最高识别率的均值和标准差比较 ($l = 5, a = -12$) (%)

Table 2 Comparison of top recognition accuracy of different methods in Yale face database ($l = 5, a = -12$) (%)

训练集	PCA+LDA	LPP	ULDA	MMC	LPMIP	本文方法
2	57.85 ± 4.06	58.67 ± 3.49	63.04 ± 4.17	63.04 ± 3.55	66.44 ± 4.72	68.59 ± 4.09
3	76.67 ± 2.72	63.92 ± 2.99	71.25 ± 2.87	72.92 ± 2.33	77.67 ± 2.57	78.83 ± 2.43
4	84.57 ± 3.14	67.71 ± 3.09	76.0 ± 3.7	78.48 ± 1.75	85.05 ± 2.46	85.62 ± 2.56
5	88.44 ± 2.73	72.33 ± 4.43	80.22 ± 3.43	83.22 ± 3.87	88.56 ± 3.78	88.89 ± 3.27
6	88.53 ± 2.45	73.47 ± 4.72	82.0 ± 4.5	84.53 ± 4.23	89.6 ± 3.07	90.67 ± 3.27

表3 PIE 人脸库最高识别率的均值和标准差比较 ($l = 6, a = -13$) (%)

Table 3 Comparison of top recognition accuracy of different methods in PIE face database ($l = 6, a = -13$) (%)

训练集	PCA+LDA	LPP	ULDA	MMC	LPMIP	本文方法
5	71.94 ± 3.09	48.65 ± 0	67.65 ± 3.45	66.8 ± 3.8	75.13 ± 2.5	78.25 ± 2.49
10	87.15 ± 1.4	57.82 ± 2.18	80.6 ± 2.28	86.48 ± 1.83	90.27 ± 1.37	91.53 ± 1.45
20	88.61 ± 3.26	69.77 ± 2.73	87.2 ± 1.5	93.23 ± 1.25	95.07 ± 0.79	95.85 ± 0.73
30	92.52 ± 1.91	79.69 ± 2.17	91.05 ± 0.79	95.9 ± 0.56	96.93 ± 0.19	97.68 ± 0.17

表 4 MNIST 手写数字库的正确识别率和特征维数比较 ($l = -1, a = 18$)Table 4 Comparison of recognition accuracy of different methods in MNIST handwritten digit database ($l = -1, a = 18$)

	PCA+LDA	LPP	ULDA	MMC	LPMIP	本文方法
正确识别率 (%)	43.5	73.25	69.6	72.9	73.86	74.38
特征维数	(625,13)	29	9	624	225	368

References

- Jain A K, Duin R P W, Mao J C. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(1): 4–37
- Liu Qing-Shan, Lu Han-Qing, Ma Song-De. A survey: subspace analysis for face recognition. *Acta Automatica Sinica*, 2003, **29**(6): 900–911
(刘青山, 卢汉清, 马颂德. 综述人脸识别中的子空间方法. *自动化学报*, 2003, **29**(6): 900–911)
- Gao Jun, Sun Chang-Yin, Wang Shi-Tong. (2D) 2UFFCA: two-directional two-dimensional unsupervised feature extraction method with fuzzy clustering ability. *Acta Automatica Sinica*, 2012, **38**(4): 549–562
(皋军, 孙长银, 王士同. 具有模糊聚类功能的双向二维无监督特征提取方法. *自动化学报*, 2012, **38**(4): 549–562)
- Wang R P, Shan S G, Chen X L, Dai Q H, Gao W. Manifold-manifold distance and its application to face recognition with image sets. *IEEE Transactions on Image Processing*, 2012, **21**(10): 4466–4479
- Jolliffe I T. *Principal Component Analysis*. Berlin: Springer-Verlag, 2002
- Fisher R A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, **7**(2): 179–188
- Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711–720
- Li H F, Jiang T, Zhang K S. Efficient and robust feature extraction by maximum margin criterion. *IEEE Transactions on Neural Networks*, 2006, **17**(1): 157–165
- Lu J W, Tan Y P. Regularized locality preserving projections and its extensions for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010, **40**(3): 958–963
- Zheng Jian-Wei, Wang Wan-Liang, Yao Xiao-Ming, Shi Hai-Yan. Face recognition using tensor local Fisher discriminant analysis. *Acta Automatica Sinica*, 2012, **38**(9): 1485–1495
(郑建伟, 王万良, 姚晓敏, 石海燕. 张量局部 Fisher 判别分析的人脸识别. *自动化学报*, 2012, **38**(9): 1485–1495)
- He X F, Niyogi P. Locality preserving projections. In: Proceedings of the 16th Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2003. 153–160
- Yan S C, Xu D, Zhang B Y, Zhang H J. Graph embedding: a general framework for dimensionality reduction. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 830–837
- Yang J, Yang J Y, Zhang D. What's wrong with Fisher criterion? *Pattern Recognition*, 2002, **35**(11): 2665–2668
- Zhang Z Y, Wang J, Zha H Y. Adaptive manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(2): 253–265
- Jin Z, Yang J Y, Hu Z S, Lou Z. Face recognition based on the uncorrelated discriminant transformation. *Pattern Recognition*, 2001, **34**(7): 1405–1416
- Wang H X, Chen S B, Hu Z L, Zheng W M. Locality-preserved maximum information projection. *IEEE Transactions on Neural Networks*, 2008, **19**(4): 571–585
- Lin Yu-E, Gu Guo-Chang, Liu Hai-Bo, Shen Jing. A recognition method using uncorrelated discriminant locality preserving projections. *Journal of Harbin Engineering University*. 2010, **31**(1): 98–101, 114
(林玉娥, 顾国昌, 刘海波, 沈晶. 不相关局部保持鉴别分析算法. *哈尔滨工程大学学报*, 2010, **31**(1): 98–101, 114)
- Lin Yu-E, Gu Guo-Chang, Liu Hai-Bo, Shen Jing, Zhao Jing. An orthogonal feature extraction method based on the within-class preserving for small sample size problem. *Acta Automatica Sinica*, 2010, **36**(5): 644–649
(林玉娥, 顾国昌, 刘海波, 沈晶, 赵靖. 适用于小样本问题的具有类内保持的正交特征提取算法. *自动化学报*, 2010, **36**(5): 644–649)
- Liu Q S, Tang X O, Lu H Q, Ma S D. Face recognition using kernel scatter-difference-based discriminant analysis. *IEEE Transactions on Neural Networks*, 2006, **17**(4): 1081–1085
- Li W, Prasad S, Fowler J E, Bruce L M. Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 2011, **8**(5): 894–898

龚 劬 重庆大学数学与统计学院教授. 主要研究方向为数字图像处理和小波分析. E-mail: gong-qu@163.com
(GONG Qu Professor at the College of Mathematics and Statistics, Chongqing University. Her research interest covers digital image processing and wavelet analysis.)

唐萍峰 重庆大学数学与统计学院硕士研究生. 主要研究方向为数字图像处理 and 模式识别. 本文通信作者. E-mail: tpf@cqu.edu.cn
(TANG Ping-Feng Master student at the College of Mathematics and Statistics, Chongqing University. His research interest covers digital image processing and pattern recognition. Corresponding author of this paper.)