

概率二维主分量分析

卿湘运¹ 王行愚¹

摘要 二维主分量分析是一种直接面向图像矩阵表达方式的特征抽取与降维方法. 提出了一个基于二维主分量分析的概率模型. 首先, 通过对此产生式概率模型参数的最大似然估计得到主分量(矢量); 然后, 考虑到缺失数据问题, 利用期望最大化算法迭代估计模型参数和主分量. 混合概率二维主分量分析模型在人脸聚类问题上的应用表明概率二维主分量分析模型能作为图像矩阵的密度估计工具. 含有缺失值的人脸图像重构实验阐述了此模型及迭代算法的有效性.

关键词 主分量分析, 二维主分量分析, 期望最大化算法, 缺失数据
中图分类号 TP391.4

Probabilistic Two-dimensional Principal Component Analysis

QING Xiang-Yun¹ WANG Xing-Yu¹

Abstract Two-dimensional principal component analysis (2DPCA) is an approach to feature extraction and dimensionality reduction for an image represented straightforward as a matrix. In this paper, a probabilistic model for 2DPCA, called P2DPCA, is proposed. First, the principal components (vectors) are derived through maximum-likelihood estimation of parameters in the generative probabilistic model. Then, due to dealing properly with missing data, we present an expectation-maximization (EM) algorithm for estimating the parameters of the model and principal components. The application to cluster face images using mixtures of P2DPCA models shows that P2DPCA model can be a tool for density-estimation of image matrix. Experimental results on face image reconstruction with missing data illustrate the effectiveness of the model and the EM iterative algorithm.

Key words Principal component analysis (PCA), two-dimensional principal component analysis (2DPCA), expectation-maximization (EM) algorithm, missing data

主分量分析 (Principal component analysis, PCA) 作为一种经典的特征抽取与维数降低方法在人脸识别、探索性数据分析等领域得到广泛应用, 其主要目标是使得高维数据经过线性变换在低维空间的方差最大. 特别地, 在人脸识别领域中基于主分量分析的人脸表达方法称作“特征脸”(Eigenface)方法^[1], 该方法首先将二维人脸图像矩阵重新排列为二维图像矢量, 再对图像矢量空间的协方差矩阵进行奇异值分解, 取前几个最大的奇异值对应的特征矢量为特征脸, 然后将图像矢量在特征脸上的投影系数作为特征值进行分类. 由于图像矢量往往是高维的, 因此计算协方差矩阵通常很困难. 虽然一些改进方法可以解决高维小样本主分量分析存在的问题, 但有可能导致无法得到精确的特征矢量. 杨健等提出一种更直观的图像投影技术——二维主分量分

析 (Two-dimensional principal component analysis, 2DPCA)^[2], 该技术直接对图像矩阵进行协方差计算, 而无需将二维图像矩阵转换为一维矢量, 因此更容易精确计算协方差矩阵. 应用改进的 2DPCA 也使得在线掌纹识别计算复杂度降低, 系统的实用性进一步提高^[3]. 类似的思想也被推广到线性判别分析 (Linear discriminant analysis, LDA) 等线性变换特征抽取方法中^[4]. 王立威等证明了 2DPCA 等同于基于块单元的主分量分析的特例^[5] (文献 [6] 也独立地给出了证明), 即图像的每一行为一块, 人脸图像的所有行图像为一维矢量样本, 再应用 PCA 即可得到 2DPCA, 并给出了一个反例^[7], 进而对 2DPCA 的鲁棒性提出了质疑. 但正如文献 [2] 所指出的那样, 2DPCA 的优点主要体现在计算方便, 在几个人脸数据集上进行识别实验的分类性能差异在统计上没有意义.

2DPCA 解决了高维样本数据带来的问题, 但是不能处理含有缺失数据的样本. 例如, 人脸受到眼镜、围巾等其他物体的遮挡, 网络视频传输由于丢包导致解码的视频图像出现马赛克, 或本身图像质量退化引起的图像不完整等都可视为含有缺失数据的二维图像. 常规的方法是将缺失的部分填充均值, 或

收稿日期 2007-05-30 收修改稿日期 2007-09-10
Received May 30, 2007; in revised form September 10, 2007
国家自然科学基金 (60674089), 教育部博士点基金 (20040251010) 资助

Supported by National Natural Science Foundation of China (60674089), Research Foundation for the Doctoral Program of Higher Education of China (20040251010)

1. 华东理工大学信息科学与工程学院 上海 200237
1. College of Information Science and Technology, East China University of Science and Technology, Shanghai 200237
DOI: 10.3724/SP.J.1004.2008.00353

简单地丢弃。此外, 如何给出一个概率框架对 2DPCA 予以明确的概率解释, 也是一个需要解决的问题。受到 Bishop 等提出的概率主分量分析 (Probabilistic principal component analysis, PPCA) 的启发^[8], 本文提出了一种概率二维主分量分析方法 (Probabilistic two-dimensional principal component analysis, P2DPCA)。作为概率主分量分析在二维矩阵样本上的扩充, P2DPCA 一方面保持了 2DPCA 的优点, 计算快速方便; 另一方面根据概率模型的特点, 应用期望最大化 (Expectation-maximization, EM) 算法处理含有缺失数据的二维主分量分析。另外, 由单个概率密度推广至混合密度形式, 也可对具有流形结构的数据作非线性投影。如果对模型参数设置合适的先验分布, 则可应用贝叶斯方法进行模型选择。

本文首先给出 P2DPCA 的概率模型, 然后根据最大似然得到参数估计的闭形式解。再对此模型应用 EM 算法推断模型参数, 在 EM 算法中, 对于含有缺失数据的问题, 在 E 步不仅计算中间隐变量的期望值, 也计算缺失数据的估计值; 在 M 步最大化似然目标函数以推断各参数值。最后给出了 P2DPCA 的两个应用: 混合概率二维主分量分析在人脸聚类上的应用和含有缺失数据的人脸图像重构。

1 概率二维主分量分析模型

设期望均值为 $0_{m \times n}$ 的二维矩阵 X 的产生式层次概率模型如下

$$X = WH + \epsilon \tag{1}$$

$$\epsilon \sim N_{m \times n}(0_{m \times n}, I_m \otimes \sigma^2 I_n) \tag{2}$$

$$W \sim N_{m \times d}(0_{m \times d}, I_m \otimes I_d) \tag{3}$$

其中, $d \times n$ 矩阵 H 为因子装载矩阵, $m \times d$ (一般地, $d \leq n$) 矩阵 W 为 $m \times n$ 观测矩阵 X 的低维投影, 在此模型中也被称作隐变量, $0_{m \times n}$ 表示所有元素为 0 的 $m \times n$ 矩阵, I_m 表示单位 $m \times m$ 矩阵, \otimes 表示矩阵的 Kronecker 积。服从 $N_{m \times n}(M, W \otimes V)$ 分布的 Y 有如下密度函数

$$f(Y) = (2\pi)^{-mn/2} \times |W|^{-n/2} \times |V|^{-m/2} \times \exp\{-\text{tr}[W^{-1}(Y - M)V^{-1}(Y - M)^T]/2\} \tag{4}$$

其中 $\text{tr}(\cdot)$ 表示矩阵的迹运算。此模型将观测矩阵当作服从矩阵正态分布的随机矩阵。噪声 ϵ 也服从矩阵正态分布。容易看出此模型为各向同性高斯噪声模型, 虽然在应用上受到一些限制, 但能给模型参数推断带来很大的便利。

根据式 (1) 和 (2), X 的条件概率为

$$X|W, H, \sigma^2 \sim N_{m \times n}(WH, I_m \otimes \sigma^2 I_n) \tag{5}$$

再结合式 (3), 将 W 进行积分

$$p(X|H, \sigma^2) = \int p(X|W, H, \sigma^2)p(W)dW = \int (2\pi)^{-mn/2} \times (\sigma^2)^{-mn/2} \times \exp\{-\text{tr}[(X - WH)(\sigma^2 I_n)^{-1} \times (X - WH)^T]/2\} \times (2\pi)^{-md/2} \times \exp\{-\text{tr}(WW^T)/2\}dW = (2\pi)^{-mn/2} \times (\sigma^2)^{-m(n-d)/2} \times |\Delta|^{-m/2} \times \exp\{-\frac{1}{2\sigma^2} \text{tr}[X(I_n - H^T \Delta^{-1} H)X^T]\} \tag{6}$$

其中 $\Delta = HH^T + \sigma^2 I_d$ 。设观测到的图像矩阵个数为 L , 其概率 2DPCA 模型可用一有向无环图表示, W_i 为隐变量, 如图 1 所示。

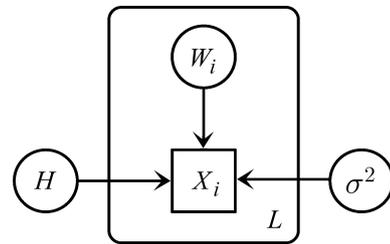


图 1 L 个观测二维矩阵 X 的概率 2DPCA 模型

Fig.1 The probabilistic 2DPCA model for L observations of two-dimensional matrix X

根据此概率模型图, 略去常数项的似然为

$$\mathcal{L}h = -\frac{m(n-d)L}{2} \ln \sigma^2 - \frac{mL}{2} \ln |\Delta| - \frac{L}{2\sigma^2} [\text{tr}(S) - \text{tr}(H^T \Delta^{-1} HS)] \tag{7}$$

其中协方差矩阵 $S = (\sum_{i=1}^L X_i^T X_i)/L$ 。最大化此似然目标函数, 可得到参数 H 与 σ^2 的驻点值。

应用贝叶斯理论, 可得到在已知观测矩阵 X_i 下隐变量矩阵 W_i 的条件分布

$$W_i|X_i \sim N(X_i H^T \Delta^{-1}, I_m \otimes \sigma^2 \Delta^{-1}) \tag{8}$$

2 概率二维主分量分析模型的最大似然解

对似然目标函数关于 H 求导, 利用 $\text{tr}(AB) = \text{tr}(BA)$ 及一些矩阵微分公式^[9], 假定 Δ 非奇异, H

$\neq 0$, 解以下方程

$$\frac{\partial}{\partial H} \mathcal{L}h = -2L(m\Delta^{-1}H - \frac{1}{\sigma^2}\Delta^{-1}H + \frac{1}{\sigma^2}\Delta^{-1}HSH^T\Delta^{-1}H) = 0 \quad (9)$$

对 H 作奇异值分解 (d 一般远小于 n)

$$H = U\Omega V^T \quad (10)$$

其中, U 为 $d \times d$ 维正交矩阵, $\Omega = \text{diag}(\omega_1, \dots, \omega_d)$ 为 $d \times d$ 维由奇异值组成的对角矩阵, $V = (v_1, \dots, v_d)$ 为 $n \times d$ 维由正交列矢量组成的矩阵. 将上式代入式 (9) 得到

$$\begin{aligned} U(m\Omega)V^T - U[\frac{1}{\sigma^2}\Omega - \frac{1}{\sigma^2}(\Omega^2 + \sigma^2I_d)^{-1}\Omega^3]V^TS \\ = 0 \\ \Rightarrow UV^TS = U(m(\Omega^2 + \sigma^2I_d))V^T \end{aligned} \quad (11)$$

对 $\omega_j \neq 0$, 上式意味着 $v_j^TS = m(\omega_j^2 + \sigma^2)v_j^T$, V 的每一列为 S 的特征矢量, 相应的特征值 $\varphi = m(\omega_j^2 + \sigma^2)$, 故 $\omega_j = (\varphi_j/m - \sigma^2)^{1/2}$; 如果 $\omega_j = 0$, v_j 可取任意值. H 所有可能的解表示为

$$H = R_d(K_d - \sigma^2)^{1/2}V^T \quad (12)$$

其中 R_d 为任意 $d \times d$ 维正交旋转矩阵, V 为 $n \times d$ 维矩阵, 每一列是 S 的特征矢量, K_d 为 $d \times d$ 维对角矩阵, 其每个对角元素如下

$$K_j = \begin{cases} \varphi_j/m, & \text{对应特征向量 } \mathbf{v}_j \\ \sigma^2, & \text{否则} \end{cases}$$

再将式 (12) 代入似然目标函数 (7), 经过一些线性代数运算, 得到

$$\begin{aligned} \mathcal{L}h = -\frac{L}{2}\{m(n-d')\ln\sigma^2 + \sum_{j=1}^{d'}\varphi_j + \\ \frac{1}{\sigma^2}\sum_{j=d'+1}^n\varphi_j + md'\} \end{aligned} \quad (13)$$

其中 d' 为协方差矩阵 S 保留下来的非零的特征值个数, $\varphi_1, \dots, \varphi_{d'}$ 为相应保留下来的特征值, $\varphi_{d'+1}, \dots, \varphi_n$ 为丢弃的特征值. 最大化式 (13) 得到 σ^2 的估计值

$$\sigma^2 = \frac{1}{m(n-d')} \sum_{j=d'+1}^n \varphi_j \quad (14)$$

至此得到 P2DPCA 模型各参数最大似然的闭形式解. 在实际求解中, 一般先对方差 S 特征值分解,

设定理想的低维数 d , 根据式 (14) 求得 σ^2 , 再根据式 (12) 得到 H , 或者得到 H^T :

$$H^T = V(K_d - \sigma^2)^{-1/2}R_d$$

一般地, R_d 可取为单位矩阵, 故直接得到低维投影 $W_i = X_iH^T\Delta^{-1}$.

3 概率二维主分量分析的 EM 算法

既然能得到 P2DPCA 模型各参数最大似然的闭形式解, 为何还要应用 EM 算法迭代求解模型参数值? 主要原因有以下几点: 1) 当观测图像含有缺失数据, 如人脸图像受到眼镜、围巾等遮拦, 或者由于成像条件受到影响, 致使图像的某一部分数据不能正确反映目标对象, 则不能直接应用上述最大似然闭形式解, 但可以应用 EM 算法迭代求解模型参数和中间变量的估计值; 2) 类似概率主分量分析的一些推广工作^[10], 如果将此概率模型扩展到二维因素分析或混合概率二维主分量分析模型, 则没有闭形式解, 需要利用 EM 算法推断模型参数; 3) 不需要直接计算协方差矩阵 S . 概率主分量分析应用 EM 算法的主要原因在于 3), 因为对于高维样本计算协方差矩阵困难, 甚至难以存储, 同时进行奇异值分解计算量大; 而概率二维主分量分析应用 EM 算法则主要顾虑到原因 1) 和 2), 二维主分量分析的一个重要的优点就是计算协方差矩阵 S 可行, 且计算量少, 所以理由 3) 的优势不是很突出.

中间隐变量 $\{W_i\}$ 与观测矩阵 $\{X_i\}$ 的联合似然为

$$\mathcal{L}h_c = \sum_{i=1}^L \ln p(X_i, W_i) \quad (15)$$

其中, 联合概率密度 $p(X_i, W_i) = p(X_i|W_i, H, \sigma^2) \times p(W_i)$. 在 E 步, 求当 H 与 σ^2 固定时在分布 $p(X_i|W_i, H, \sigma^2)$ 下 $\mathcal{L}h_c$ 的期望值, 可得到 W_i 的期望值 $\langle W_i \rangle$ 及 $W_i^TW_i$ 的期望值 $\langle W_i^TW_i \rangle$

$$\langle W_i \rangle = X_iH^T\Delta^{-1} \quad (16)$$

$$\langle W_i^TW_i \rangle = m\sigma^2\Delta^{-1} + \langle W_i \rangle^T\langle W_i \rangle \quad (17)$$

在 M 步, 即在已知 $\langle W_i \rangle$ 及 $\langle W_i^TW_i \rangle$ 的情况下, 最大化联合似然 $\mathcal{L}h_c$, 分别得到 H 与 σ^2 的估计值

$$H = [\sum_{i=1}^L \langle W_i^TW_i \rangle]^{-1} [\sum_{i=1}^L \langle W_i \rangle^T X_i] \quad (18)$$

$$\begin{aligned} \sigma^2 = \frac{1}{mnL} \{L\text{tr}(S) + \sum_{i=1}^L [-2\text{tr}(X_iH^T\langle W_i \rangle^T) + \\ \text{tr}(H^T\langle W_i^TW_i \rangle H)]\} \end{aligned} \quad (19)$$

迭代执行式 (16)~(19), 直至收敛条件满足. 经 EM 算法得到的 H 矩阵各行矢量不正交, 因此其重构不是均方误差准则下的最优重构.

由于 d 经常远小于 n , $\langle W_i^T W_i \rangle$ 与 Δ 均为 $d \times d$ 矩阵, 求 $\langle W_i^T W_i \rangle$ 逆的复杂度为 $O((md)^2 + d^3)$, 求 Δ 逆的复杂度为 $O((nd)^2 + d^3)$, 因此对它们求逆的计算量远小于推断 PPCA 模型参数, EM 算法计算类似中间变量求逆的计算量, 从文献 [11](此文献也独立地提出了 PPCA 的 EM 迭代算法) 中可以看到, 对应的一维特征矢量的维数为 mn , 因子装载矩阵 C 的维数为 $(mn) \times d$, 一般 PCA 的主分量 d 大于 2DPCA 的主分量 d . 为避免对 $(mn) \times (mn)$ 矩阵 $(CC^T + \sigma^2 I_{(mn)})$ 求逆, 文献 [11] 应用 Woodbury 公式对 $(I_d + C^T C / \sigma^2)$ 求逆, 再左乘 C , 右乘 C^T , 其复杂度为 $O(2mnd^2 + (mnd)^2 + d^3)$. 两者的求迹运算都可运用相同技巧以减少计算量. 因此应用 EM 算法推断 P2DPCA 模型参数及隐变量期望值仍保持了 2DPCA 计算方便、节约计算量的优点.

当有缺失数据时, 需对每次迭代的 E 步进行适当的修改. 当缺失数据较少时, 可以直接利用上次迭代得到的 $\langle W_i \rangle$ 和 H 来估计缺失数据值 $\langle X_i \rangle = \langle W_i \rangle H$; 而当缺失数据较多时, 设观测到的数据矩阵 X_i 中共 p 个元素是有值的, 将其按行重新排列成矢量 $\mathbf{Y} = [y_1, \dots, y_p] \in \mathbf{R}^p$, 相应地也将待估计的 W_i 按行排列成矢量 $\boldsymbol{\omega}$, 首先计算约束矩阵 $A^* = I_{m \times m} \otimes H^T$, 再将 A^* 中对应缺失数据的约束去除, 得到行数为 p 、列数为 $m \times d$ 的约束矩阵 A , 解最小平方问题 $\|A\boldsymbol{\omega} - \mathbf{Y}\|_2^2$ 则能得到 W_i 的估计值 $\langle W_i \rangle$, 其中 $\|\cdot\|_2$ 表示 l_2 范数, 一般需利用 QR 分解方法求解 $\boldsymbol{\omega}$.

因此 P2DPCA 较之 2DPCA 有能处理含有缺失数据的矩阵的优点, 不仅能得到低维空间投影值, 还能推测在最大联合似然准则下缺失数据的估计值.

4 实验结果

当利用 P2DPCA 最大似然闭形式解进行人脸识别实验时, 由于本方法的解与 2DPCA 的解相似, 其不同点是本算法对低维投影损失的方差有明确的解释, 如式 (14) 所示, σ^2 完全由所丢弃的特征分量对应的特征值决定, 且 σ^2 值通常较小, 如忽略其值, 对比 2DPCA 方法只是 H 矩阵都除以 m , 而相应的投影系数都乘上 m , 所以应用本方法在对 ORL、AR 和 Yale 数据集进行人脸识别的分类准确率与利用 2DPCA 的分类准确率相似, 故本文没有给出人脸识别分类的结果. 关于 2DPCA 与 PCA 人脸识别效果的比较, 文献 [2, 5, 7] 等已有深入的讨论, 而且

分类效果受到所选分类器的影响^[12-13], 因此这里不再赘述. 特别指出的是以下所有实验图像的像素灰度值都已归一化到 $[0, 1]$ 区间.

4.1 混合概率二维主分量分析模型

P2DPCA 只定义了单个线性投影, 对于具有流形结构的数据 (如人脸图像经常只受到光照、角度变换等几个自由度的作用), 可利用多个 P2DPCA 模型的混合获得非线性投影, 同时可得到其聚类结果. K 个概率二维主分量分析模型混合的对数似然目标函数为

$$\mathcal{L}h_{mix} = \sum_{i=1}^L \ln\{p(X_i)\} = \sum_{i=1}^L \ln\left\{\sum_{c=1}^K \pi_c p(X_i|c)\right\} \quad (20)$$

其中 π_c 为第 c 个混合分量的权重, 满足 $\pi_c \geq 0$, $\sum_{c=1}^K \pi_c = 1$. 对于此模型, 每个混合分量有相应的均值 M_c (不一定是 0), 只需对前述公式作相应的改变即可: $X_i = X_i - M_c$. 同样应用 EM 算法迭代可得到各分量的参数. 设观测数据 X_i 属于类 c 的后验概率 $\phi_{ic} = p(c|X_i)$, 其 EM 算法如下:

E 步:

$$\langle W_{ic} \rangle = (X_i - M_c) H_c^T \Delta_c^{-1} \quad (21)$$

$$\langle W_i^T W_i \rangle = m \sigma_c^2 \Delta_c^{-1} + \langle W_{ic} \rangle^T \langle W_{ic} \rangle \quad (22)$$

其中 $\Delta_c = \sigma_c^2 I_d + H_c H_c^T$.

M 步:

$$H_c = \left[\sum_{i=1}^L \phi_{ic} \langle W_{ic}^T W_{ic} \rangle \right]^{-1} \left[\sum_{i=1}^L \phi_{ic} \langle W_{ic}^T \rangle (X_i - M_c) \right] \quad (23)$$

$$\sigma_c^2 = \frac{1}{mn \sum_{i=1}^L \phi_{ic}} \left\{ \sum_{i=1}^L \phi_{ic} \text{tr}[(X_i - M_c)(X_i - M_c)^T] - \right.$$

$$2 \sum_{i=1}^L \phi_{ic} \text{tr}[(X_i - M_c) H_c^T \langle W_{ic} \rangle^T] + \left. \sum_{i=1}^L \phi_{ic} \text{tr}(H_c^T \langle W_{ic}^T W_{ic} \rangle H_c) \right\} \quad (24)$$

$$M_c = \left[\sum_{i=1}^L \phi_{ic} (X_i - W_{ic} H_c) \right] / \sum_{i=1}^L \phi_{ic} \quad (25)$$

$$\pi_c = \left(\sum_{i=1}^L \phi_{ic} \right) / L \quad (26)$$

$$\phi_{ic} \propto p(X_i|c, \langle W_{ic} \rangle, H_c, M_c, \sigma_c^2) \pi_c \quad (27)$$

且 $\sum_{c=1}^K \phi_{ic} = 1$.

以下应用上述混合 P2DPCA 模型解决一个人脸聚类的问题. 数据来源于 Stirling 人脸数据库 (<http://pics.psych.stir.ac.uk/cgi-bin/PICS/New/pics.cgi>), 选择所有的 68 个前视 (类 1) 和 105 个侧视 (类 2) 灰度人脸图像, 部分图像如图 2 所示, $K = 2$, $L = 173$. 每张人脸图像大小为 $128 \text{ 行} \times 128 \text{ 列}$.



图 2 Stirling 人脸数据库人脸图像

Fig. 2 Example images from the Stirling faces database

设置 H_c 每个元素的初值服从标准正态分布, $\phi_{ic} \approx 1/K$, 根据 ϕ_{ic} 得到初始类别, 求得每类的均值作为 M_c 的初始值, $\sigma_1^2 = 0.01$, $\sigma_2^2 = 0.009$, $d = 1$. 应用上述 EM 算法, 约迭代执行 30 次就收敛, 根据不同的初始值运行 20 次 (以下简称此方法为 MixP2DPCA). 与之比较的算法有:

1) 首先利用 2DPCA 降维, 再利用 K 均值聚类算法, $d = 1$ (以下简称 2DPCA + K-means);

2) 首先利用 P2DPCA 最大似然闭形式解降维, 再利用 K 均值聚类算法, $d = 1$ (以下简称 P2DPCA + K-means);

3) 如文献 [14] 所述由 Roth 等提出的聚类方法, 首先运行 PCA 得到低维投影系数, 再同时进行特征选择和聚类 (以下简称 Roth);

4) 利用流形学习思想的局部保持聚类算法 (Locality preserving clustering, LPC)^[15]. 在此方法中, 设定最近邻数为 10, 通过映射降至低维数 3, 热核参数 σ 的选取方法为 10 倍数据最近邻距离的均值.

在以上算法中, K 均值聚类算法采用欧氏距离作为距离度量, 聚类数设定为真实聚类个数 2, 随机设定类中心点, 最大迭代次数为 300, 根据不同初始值运行 20 次, 在此数据集上 K 均值聚类算法的结果均很稳定.

所有 5 个算法的实验结果如表 1 所示. 2DPCA + K-means 与 P2DPCA + K-means 方法没有考虑此数据集本质低维和非线性映射特性, 映射关系是全局线性的, 同时由于降维后的维数较高 ($m \times d$),

也使得 K 均值聚类算法的性能下降, 所以两方法聚类结果较差, 其中 P2DPCA 算法的 $\sigma^2 = 0.023$; Roth 算法则极富技巧性, 因而得到较好的聚类结果; LPC 方法在低维空间保持高维空间的局部特性, 对于此人脸数据库实质上只有人脸的转动, 所以 LPC 特别适合对此类本质低维、高维嵌入的图像数据进行聚类, 故聚类结果最好; 应用混合概率二维主分量分析模型的聚类算法也得到较好的聚类结果, 因为该算法也能通过局部线性子模型的混合对全局非线性图像数据建模, 且不需对图像作任何变换, 直接建立概率模型, 同时对每张图像的聚类不确定性可以用后验概率 ϕ_{ic} 来度量. 虽然在此实验中 MixP2DPCA 比 LPC 的聚类效果稍差, 但是 LPC 的降维本质上还是一种线性映射, 因此对于非常复杂、高度非线性的数据, LPC 的降维仍摆脱不了线性映射的局限. 文献 [16] 对当前的非参数流形学习方法和混合概率参数模型在非线性映射应用方面各自的优缺点均有详细阐述, 同时本概率模型能很自然地推广到文献 [16] 的流形对齐模型中.

表 1 人脸聚类实验结果

Table 1 Results on face clustering

| 聚类算法 | 误分人脸数 |
|------------------|-------|
| DPCA + K-means | 32 |
| P2DPCA + K-means | 32 |
| Roth | 7 |
| LPC | 3 |
| MixP2DPCA | 8 |

4.2 缺失值问题

本实验素材取自 ORL 人脸数据库 (<http://www.cam-orl.co.uk>), 共 40 人, 每张人脸图像大小为 $112 \text{ 行} \times 92 \text{ 列}$, 每人 10 张人脸图像, 共 400 张人脸. 每人保留一张完整的人脸图像, 对其余图像随机设定 $40 \text{ 行} \times 30 \text{ 列}$ 大小的区域为缺失块. 文献 [11] 的方法能处理 2^{12} 维数据矢量的 PCA, 但是就此问题如不对图像进行下抽样, 以目前普通计算机的存储容量来说, 文献 [11] 的方法还是无法处理 2^{14} 维数据矢量的 PCA. 对于 2DPCA 重构方法, 某位置缺失数据预先以所有图像此位置的均值 (只计算不含缺失值图像的均值) 代替, 再利用 2DPCA 获得其重构图像 (在此实验中简称此方法为 2DPCA). 根据本文提出的模型应用 EM 算法估计重构图像及原图像的缺失部分 (以下简称此方法为 P2DPCA), 其中初始值 $\sigma^2 = 0.01$, 设定 H 每个元素的初值服从标准正态分布, 为简化计算, 设定收敛标准为最大迭

代执行 500 次或 $(\|H^{(i)} - H^{(i-1)}\|_2 / (n \times d)) < 10^{-5}$, 其中 $H^{(i)}$ 表示第 i 次迭代后经 EM 算法 M 步得到的 H . 由于缺失数据较少, 所以在 E 步不采取 QR 分解求 W_i 的估计值 $\langle W_i \rangle$ (当观测数据较多时, 进行 QR 分解的计算量太大).

选取两个量化指标, 一个为重构图像误差 E_{rec}

$$E_{rec} = \sum_l \|X_l - X_l^*\|_2^2$$

其中, X_l 表示第 l 张人脸图像 (假定知道真正的缺失数据值), X_l^* 表示含有缺失数据的图像 l 经低维投影重构得到的图像. E_{rec} 越小, 反映 l_2 范数意义下的逼近性能越好, 但 E_{rec} 忽视了图像内容对人眼的影响, 不能完全反映出人眼对图像的感知质量, 因此引进一个基于结构失真的图像质量评价标准 SSIM-index^[17], 其值在 $[0, 1]$ 之间, 越接近 1, 说明两图像越相似. 本实验取所有重构图像与真实图像的 SSIM-index 均值作为评价标准, 标识为 ssim-index.

当 d 取不同值时, 应用 2DPCA 与 P2DPCA 的重构图像误差和 ssim-index 结果分别如图 3 和图 4 所示. 对于相同的 d 值, P2DPCA 的 ssim-index 总是比 2DPCA 的 ssim-index 高, d 越大两者的 ssim-index 越接近; 当 d 较小时, P2DPCA 的 E_{rec} 比 2DPCA 的 E_{rec} 小, 但当 d 较大时, 反而更大. 两种方法在量化指标上的差异不显著, 主要原因有以下几点:

1) 缺失数据只占整个图像数据的 10.48%, 因此对缺失数据估计所带来的量化指标的改善也相应地只占很少一部分;

2) 毕竟我们并不知道真正的缺失数据值, P2DPCA 重构也只是在最大似然框架下对缺失数据的一个推测与估计, 因此重构图像误差 E_{rec} 在实际的实验中是没办法计算的;

3) 迄今为止, 图像质量评价指标还不是很完善.

图 5 (见下页) 是 $d = 8$ 时根据两种方法得到的部分图像的重构结果. 其中补缺图像的意思是将原图像缺失部分填补由各算法得到的估计值. 从图 5 可观察到, 应用 P2DPCA 得到的补缺图像能够大致得到相应部位的轮廓图像, 如最上一张图像的眼睛和最下一张的嘴巴, 而应用 2DPCA 得到的补缺图像出现较严重的马赛克现象. 在重构图像中, 应用本方法得到的重构图也能较清楚地恢复相应部位的轮廓, 而由 2DPCA 得到的重构图像虽马赛克效应较补缺图像弱, 但缺失部分像素值较一致, 视觉效果较模糊.

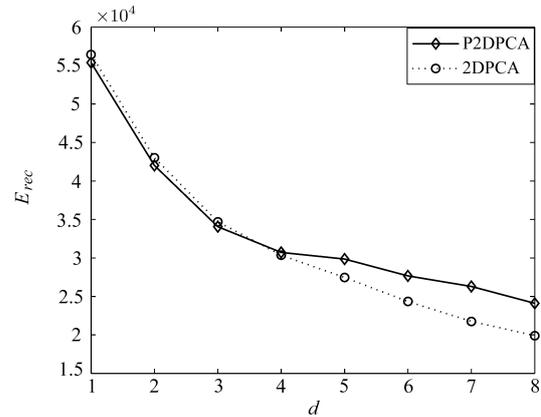


图 3 重构图像误差

Fig. 3 Reconstruction errors

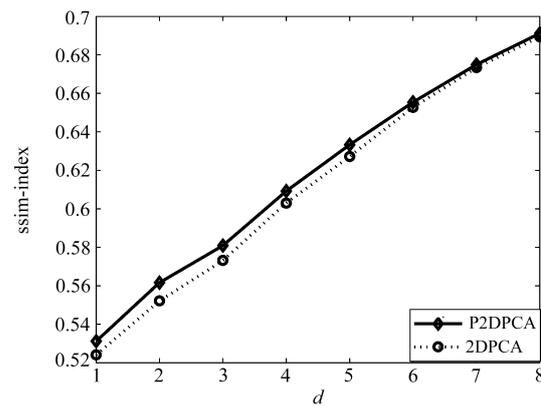


图 4 图像质量评价 ssim-index

Fig. 4 Image quality assessment using ssim-index

5 结论

对 2DPCA 建立一个概率模型具有其他概率模型也具备的优点^[8,18], 如: 通过似然函数能与其他基于概率密度的方法进行比较; 结合合适的先验能得到贝叶斯推断方法和进行模型选择; 当观测数据有缺失值时概率模型也能获得主分量投影; 组合单个模型可得到混合模型, 得到数据的非线性主分量分析, 以探索、发现数据的流形结构; 当混合模型应用于聚类时, 可以度量聚类结果的不确定性等. 此外, P2DPCA 还具有 2DPCA 的主要优点: 计算简单方便, 节省计算量.

以下是作者下一步研究的方向: 1) 对模型参数赋予合适的先验, 特别是若能将类别信息考虑进去, 则能得到具有判别信息的主分量; 2) 利用变分贝叶斯方法推断混合概率二维主分量分析贝叶斯模型的参数; 3) 如文献 [19] 所述, 对双向二维主分量分析也建立类似的概率模型.

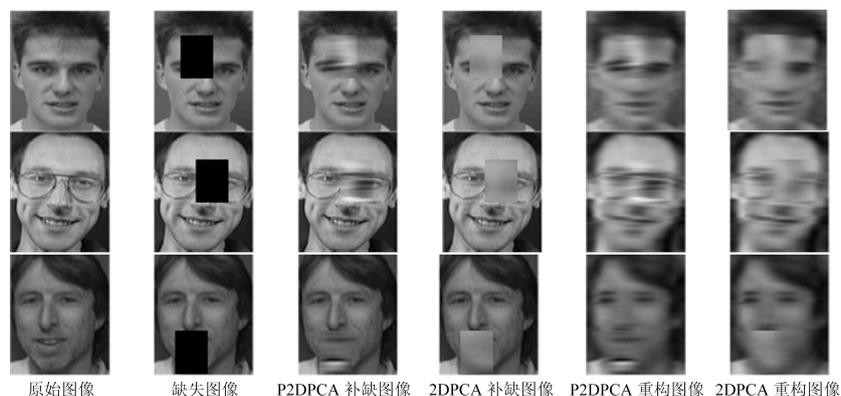


图 5 含有缺失数据的人脸图像重构

Fig. 5 Face image reconstruction with missing data

References

- Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71–86
- Yang J, Zhang D, Frangi A F, Yang J Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(1): 131–137
- Li Qiang, Qiu Zheng-Ding, Sun Dong-Mei, Liu Lu-Lu. Online palmprint identification based on improved 2D PCA. *Acta Electronica Sinica*, 2005, **33**(10): 160–163
(李强, 裘正定, 孙冬梅, 刘陆陆. 基于改进二维主成分分析的在线掌纹识别. *电子学报*, 2005, **33**(10): 160–163)
- Jing X Y, Wong H S, Zhang D. Face recognition based on 2D Fisherface approach. *Pattern Recognition*, 2006, **39**(4): 707–710
- Wang L W, Wang X, Zhang X R, Feng J F. The equivalence of two-dimensional PCA to line-based PCA. *Pattern Recognition Letters*, 2005, **26**(1): 57–60
- Kong H, Wang L, Teoh E K, Li X C, Wang J G, Venkateswarlu R. Generalized 2D principal component analysis for face image representation and recognition. *Neural Networks*, 2005, **18**(5-6): 585–594
- Wang Li-Wei, Wang Xiao, Chang Ming, Feng Ju-Fu. Is two-dimensional PCA a new technique? *Acta Automatic Sinica*, 2005, **31**(5): 782–787
(王立威, 王潇, 常明, 封举富. 关于二维主成分分析方法的研究. *自动化学报*, 2005, **31**(5): 782–787)
- Tipping M E, Bishop C M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 1999, **61**(3): 611–622
- Zhang Xian-Da. *Matrix Analysis and Applications*. Beijing: Tsinghua University Press, 2004. 271–283
(张贤达. *矩阵分析与应用*. 北京: 清华大学出版社, 2004. 271–283)
- Tipping M E, Bishop C M. Mixtures of probabilistic principal component analysers. *Neural Computation*, 1999, **11**(2): 443–482
- Roweis S. EM algorithms for PCA and SPCA. In: Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10. Denver, USA: MIT Press, 1997. 626–632
- Zuo W M, Zhang D, Wang K Q. An assembled matrix distance metric for 2DPCA-based image recognition. *Pattern Recognition Letters*, 2006, **27**(3): 210–216
- Meng J C, Zhang W B. Volume measure in 2DPCA-based face recognition. *Pattern Recognition Letters*, 2007, **28**(10): 1203–1208
- Roth V, Lange T. Feature selection in clustering problems. In: Proceedings of the 2003 Conference on Advances in Neural Information Processing Systems 16. Cambridge, UK: MIT Press, 2003. 1237–1244
- Zheng X, Cai D, He X F, Ma W Y, Lin X Y. Locality preserving clustering for image database. In: Proceedings of the 12th Annual ACM International Conference on Multimedia. New York, USA: ACM Press, 2004. 885–891
- Verbeek J. Learning nonlinear image manifolds by global alignment local linear models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(8): 1236–1250
- Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600–612
- Bishop C M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. 570–582
- Zhang D Q, Zhou Z H. (2D)²PCA: two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing*, 2005, **69**(1-3): 224–231



卿湘运 华东理工大学博士研究生. 主要研究方向为统计机器学习与模式识别. E-mail: xytsing@yahoo.com.cn
(QING Xiang-Yun Ph.D. candidate at East China University of Science and Technology. His research interest covers statistical machine learning and pattern recognition.)



王行愚 华东理工大学教授. 主要研究方向为智能控制和模式识别. 本文通信作者. E-mail: xywang@ecust.edu.cn
(WANG Xing-Yu Professor at East China University of Science and Technology. His research interest covers intelligent control theory and pattern recognition. Corresponding author of this paper.)