矩阵计算

李宇峰
liyf@nju.edu.cn
人工智能学院

3.3 特征分析的应用—基于特征脸的人脸识别

1. 问题的背景

进入互联网时代，我们每天都在和各种账号密码打交道。如果为所有账号设置统一的密码，一旦一个账号密码泄密，其他的账号密码都面临泄露的威胁；如果账号密码很多，就会带来很大的记忆负担。生物特征(如指纹、掌纹、人脸、虹膜)设别技术恰巧能解决这样的问题。

3.3 特征分析的应用—基于特征脸的人脸识别

由于生物特征具有特异性和不变性，因此能够作为个人身份凭证；相比于数字账号密码，生物特征不易丢失和伪造，因此具有更高的安全性。

常见的生物特征识别技术包括人脸识别、指纹识别、虹膜识别等。相对于指纹、虹膜等生物特征，人脸的获取采集要简单和方便的多。人脸识别技术通过计算机自动分析人脸图像，进而辨认该人脸的身份信息。目前，人脸识别已经开始广泛应用在海关边检、监控安防、电商金融、门禁系统以及个人电子设备（如手机、数码相机和个人电脑等）。

人脸识别技术的难点在于不同表情、不同角度、不同光照下的人脸差别很大，甚至在一些场合还会有口罩墨镜等遮挡物品。这些因素给人脸识别带来严峻的挑战。因此人脸识别问题成为模式识别领域的重要难题。

2. 问题的提出

人脸识别的算法很多，如基于特征脸的人脸识别算法，上次讲过的稀疏表示算法，以及基于神经网络的人脸识别算法等。特征脸是一种比较简单的人脸的识别算法，也是第一种有效的人脸识别算法。特征脸算法本质上是在一组人脸图像上进行主

成分分析(PCA)，所谓特征脸实际上就是这组人脸图像对应协方差矩阵的特征向量。根据主成分分析原理，对于任意的一张人脸，都可以采用特征脸的线性组合来近似拟合。

前面已经讲过，主成分分析是一种有效的特征降维手段，其目标是将数据投影到低维子空间中，同时保证在子空间中数据的方差尽可能大，每一个特征向量对应一个一维子空间的法方向。通过这种线性变换，能够有效去除数据中的冗余信息。在人脸识别领域，将子空间对应的特征向量形象地称为特征脸。

平均脸

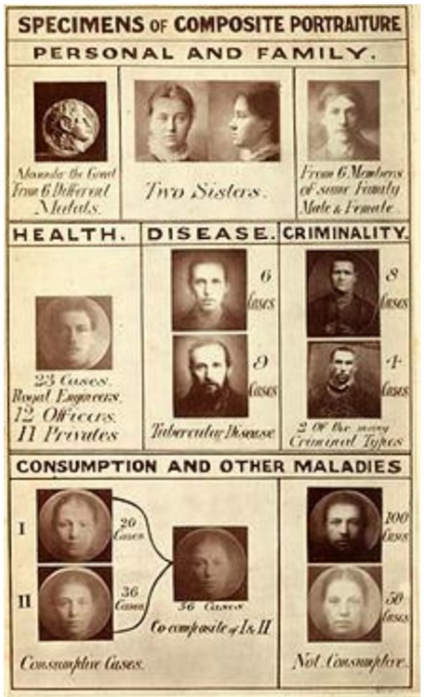


Fig. 1 Composite portraiture, Francis Galton, 1883.

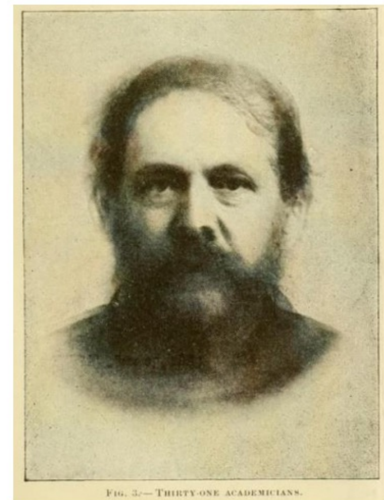


Fig. 2 Eighteen naturalists and thirteen mathematicians, Pumpelly 1885

平均脸

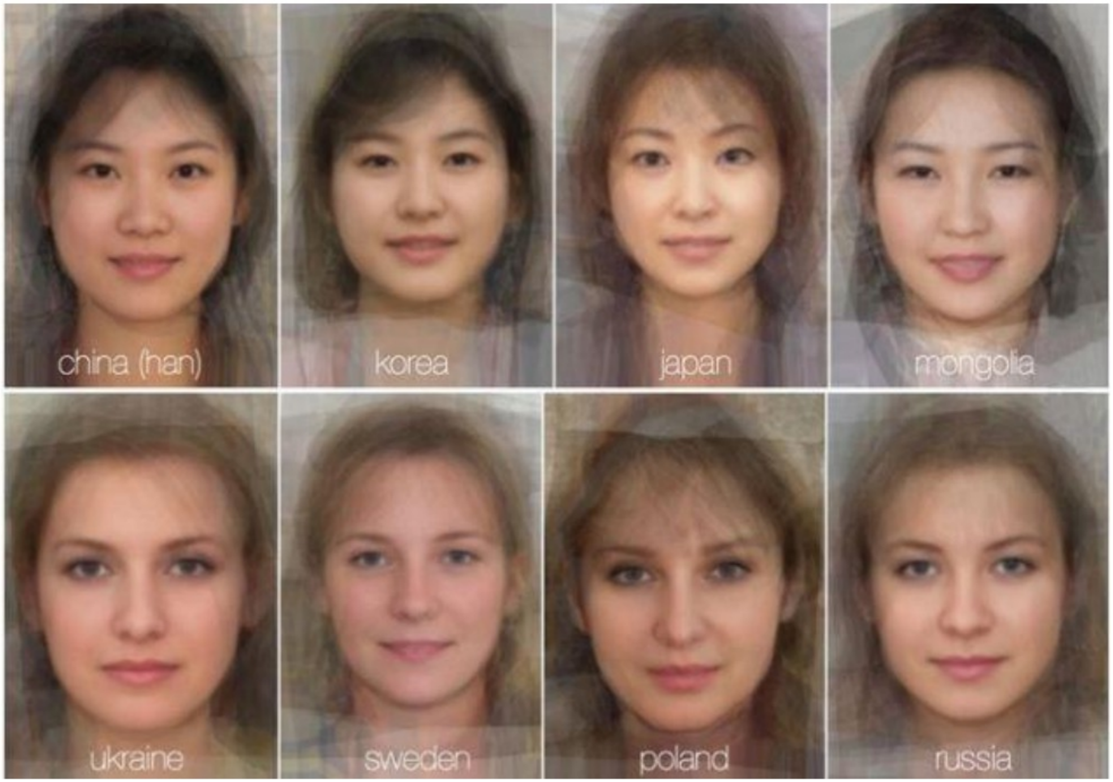


Fig. 3 “Mrs Averages”, University of Glasgow 2013

3. 特征脸数学模型

假设有一组M个人脸图像(灰度图像，大小都为$m \times n$), 每一张人脸图像都可以表示为一个由$m \times n$个元素组成的长向量，即

$$\mathbf{x}_i = (p_{i1}, p_{i2}, \cdots, p_{iN})^T, \qquad i = 1,2,\cdots, M$$

其中p表示图像像素，$N = m \times n$表示向量长度，M表示人脸图像的数量。进行主成分分析之前需要对所有的人脸图像进行零均值化，即将上述的每一个人脸向量都减去均值向量，其中 均值向量为

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}_i$$

零均值化后的人脸向量为$\mathbf{z}_i = \mathbf{x}_i - \boldsymbol{\mu}$, 现在希望找到一组正交投影方向将数据投影到低维特征子空间中，这可以通过计算特征协方差矩阵的特征向量来实现。协方差矩阵定义为

$$\mathbf{C} = \mathbf{Z}\mathbf{Z}^T$$

其中$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_M]$为数据矩阵，设协方差矩阵对应的特征值$\lambda_i$和特征向量$\mathbf{u}_i$满足

$$\mathbf{Z}\mathbf{Z}^T\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

在实际应用中，由于人脸图像的维数$N = m \times n$往往比较大。

例如当人脸图像大小为100 $\times$ 100时，协方差矩阵就是一个 10000×10000的矩阵，而计算矩阵特征值的时间复杂度是$O(N^3)$，因此直接计算协方差矩阵的特征值是不现实的。

然而由于我们知道矩阵$\mathbf{ZZ}^T$是一个和$\mathbf{Z}^T\mathbf{Z}$具有相同的非零特征值，而$\mathbf{Z}^T\mathbf{Z}$是一个$M \times M$的实对称矩阵，在特征脸问题中人脸图像的个数M往往远小于人脸图像的维数N。因此我们可以通过求解低阶矩阵$\mathbf{Z}^T\mathbf{Z}$的特征值间接得到高阶协方差矩阵$\mathbf{ZZ}^T$的非零特征值和相应的特征向量。

问题：高阶矩阵$\mathbf{ZZ}^T$和M阶矩阵$\mathbf{Z}^T\mathbf{Z}$的特征向量之间有什么关系？

设矩阵$\mathbf{Z}^T\mathbf{Z}$的特征值和特征向量分别是$\tilde{\lambda}_i$和$\tilde{\mathbf{u}}_i$, 则有

$$\mathbf{Z}^T\mathbf{Z}\tilde{\mathbf{u}}_i = \tilde{\lambda}_i\tilde{\mathbf{u}}_i$$

两边左乘数据矩阵$\mathbf{Z}$, 即得

$$\mathbf{Z}\mathbf{Z}^T(\mathbf{Z}\tilde{\mathbf{u}}_i) = \tilde{\lambda}_i(\mathbf{Z}\tilde{\mathbf{u}}_i)$$

这说明，若$\tilde{\lambda}_i$是矩阵$\mathbf{Z}^T\mathbf{Z}$的一个特征值，$\tilde{\mathbf{u}}_i$是对应的特征向量，那么$\tilde{\lambda}_i$也是矩阵$\mathbf{Z}\mathbf{Z}^T$的特征值，对应的特征向量是$\mathbf{Z}\tilde{\mathbf{u}}_i$.

也就是说:

$$\lambda_i = \tilde{\lambda}_i，\ \mathbf{u}_i = \mathbf{Z}\tilde{\mathbf{u}}_i$$

对于人脸图像数据，每一个特征向量$\mathbf{u}_i$在显示为图像时都很像人脸，因此也就称为特征脸。

由于我们仅需要计算投影方向向量，由此一般需要得到的特征向量$\mathbf{u}_i$归一化为单位向量，记为$\mathbf{e}_i$。

特征值的大小在一定程度上反映了投影到对应的子空间后的数据方差的大小，或者说原始数据的信息量。

通常我们将特征值按照从大到小的次序排列，在实际应用中，往往前10%左右的特征值就包含了原始数据90%左右的信息量。因此可以选择合适的正整数d，将数据从N维降到d维，降维后的数据向量为

$$\mathbf{y}_i = [\mathbf{e}_1^T\mathbf{z}_i, \mathbf{e}_2^T\mathbf{z}_i, \cdots, \mathbf{e}_d^T\mathbf{z}_i]^T, \quad i = 1,2,\cdots,M$$

利用降维后的数据，可以重建或者重构原始图像，也就是

$$\hat{\mathbf{y}}_i = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_d]\mathbf{y}_i + \boldsymbol{\mu}$$

利用降维后的数据，可以判别人脸所属类型。

在实际应用中我们可以用$d_1 = \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2$表示降维重建后的图像$\hat{\mathbf{x}}_i$与原始图像$\mathbf{x}_i$之间的距离。如果$d_1$大于一定的阈值，则认为该图像$\mathbf{x}_i$不是人脸图像。

令$r_i = \min\|\mathbf{y}_u - \mathbf{y}_i\|_2^2$表示降维后数据$\mathbf{y}_u$与训练集样本降维数据$\mathbf{y}_i$之间的最小欧氏距离。若$r_i$小于一定的阈值，则认为该图像的类别为其最近邻所属类别，否则认为该人脸图像类别未知。

4. 实验设计与实现

在一个公开的人脸数据库*ORL*上测试基于特征脸的人脸识别算法。

这个实验包含来自40个不同志愿者的共400张人脸灰度图像，每个志愿者采集的10张人脸图像中包含了不同表情、姿态的变化，图像大小以及归一化为112×92，图1展示了 该数据集中的部分的人脸图像。

图1 ORL数据集部分人脸图像

实验中是把人脸图像分成3部分，将前30个类别中每个类别的前6张人脸图像作为训练集，用于计算特征脸，也就是总共有180个训练样本；剩下的人脸图像作为测试集，其中前30个类别中每个类别的后4张人脸图像为正样本类(带有标记)，共120个；后10个类别的所有人脸图像作为负样本类(不带有标记)，共100个。

实验中计算180个训练样本的协方差矩阵并计算其特征值和特征向量，保留最大的40个特征值对应的特征向量作为特征脸。对于任意一张人脸图像，都可以用这40个特征脸的线性组合近似表示。

因此经过*PCA*变换后，每张人脸图像都被映射为一个40维的向量。以下图2给出了前20个特征脸。

图2 前20个特征脸

图3展示了部分人脸图像经*PCA*降维后的重建结果。可以看到大部分人脸图像的重建结果和原始图像比较接近，说明采用经过*PCA*降维的数据保留了原始数据中的绝大部分信息。

图3 人脸图像重建结果

5. 结果分析

实验中，我们保留了40个特征脸。如果保留的特征脸很少，重建图像和原始图像差别很大，不同类型的人脸难以区分；如果保留的特征脸很多，重建图像精度更高，但也意味着 要更多计算量，可以通过特征值的累计贡献来确定合适的特征脸个数。

3.4 广义特征值分析

特征分析的基础是对于矩阵$A \in C^{n\times n}$要确定数λ，以及非零向量$u \in C^n$, 使得$Au = \lambda u$。

但在信息科学的一些领域（如流形学习）中还会遇到这样的问题：对于矩阵$A, B \in C^{n\times n}$, 要确定数λ，以及非零向量$u \in C^n$, 使得

$$Au = \lambda Bu$$

这就是广义特征值问题，矩阵$A, B \in C^{n\times n}$组成一对矩阵束(*matrix pair*), 记成(A, B), 常数λ和非零向量u分别称为矩阵束的广义特征值(*value generalized eigenvalue*)和广义特征向量(*generalized eigenvector*)。

一个广义特征值和与之对应的广义特征向量组成一广义特征对，记作(λ, u)。$Au = \lambda Bu$也称为广义特征方程。由此可以看出，特征值问题是当矩阵束(A, I)时广义特征值问题的一个特例。

虽然广义特征值和广义特征向量总是成对出现，但是广义特征值可以单独求出。这一情况与通常特征值可以单独求出相类似。由特征方程可以得到

$$(A - \lambda B)u = 0$$

由于要求广义特征向量$u \neq 0$，所以一定要有系数矩阵$A - \lambda B$为奇异矩阵, 也就是

$$A - \lambda B\text{奇异} \quad \Leftrightarrow \quad |A - \lambda B| = 0.$$

$|A-\lambda B|=0$称为广义特征多项式，因此矩阵束(A,B)又往往表示成$A-\lambda B.$

因此，矩阵束(A,B)的广义特征值λ是满足广义特征多项式$|A-zB|=0$的所有z（包括零值在内）。显然若B为单位矩阵时，广义特征多项式就退化为$|A-\lambda I|=0$, 从这一角度讲，广义特征多项式是一般特征多项式的推广。

若将矩阵束的广义特征值记作$\lambda(A,B)$, 则广义特征值定义为

$$\lambda(A,B)=\{z\in C\colon\ |A-zB|=0\}$$

下面是关于广义特征值问题$Ax = \lambda Bx$的一些性质：

（1）若矩阵A和B互换，且$\lambda \neq 0$，则广义特征值将变为其倒数，但广义特征向量保持不变，即有

$$Ax = \lambda Bx \implies Bx = \frac{1}{\lambda} Ax$$

（2）若B非奇异，则广义特征值问题简化为通常的特征值问题

$$Ax = \lambda Bx \implies (B^{-1}A)x = \lambda x$$

（3）若A和B均为实对称的正定矩阵，则广义特征值一定是正的。

(4) 如果A为奇异矩阵，则$\lambda=0$必定是一个广义特征值。

(5) 若A和B均为正定的$Hermite$矩阵，则广义特征值必定是实的，并且不同的广义特征值(λ_i,λ_j)对应的广义特征向量$(x_i$ $x_j)$相对于正定矩阵A和B分别正交，即有

$$x_i^H A x_j = x_i^H B x_j = 0,\ i \neq j$$

在很多应用中，往往只要用到广义特征值，这时等价矩阵束就是一个非常有用的概念。

定义 所有广义特征值相同的两个矩阵束就称为等价矩阵束。

由广义特征值的定义 $|A-\lambda B|=0$ 和行列式的性质容易知道，如果X和Y是两个非奇异矩阵，则

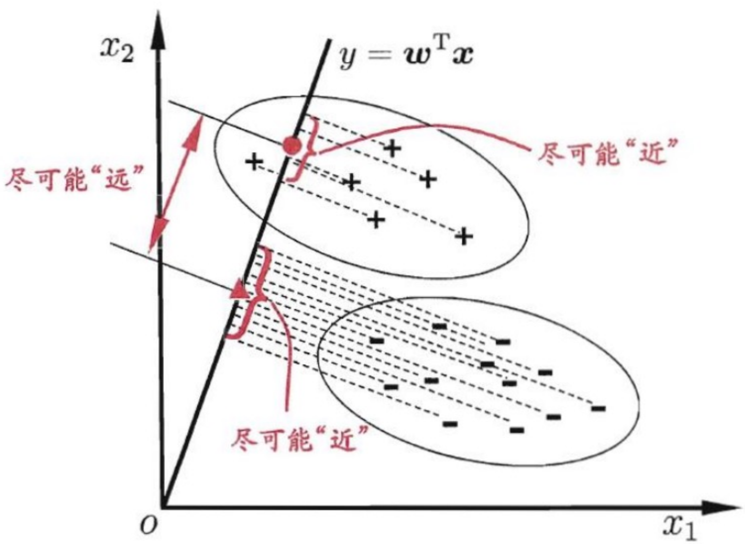
$$|XAY-\lambda XBY|=0 \quad \Leftrightarrow \quad |A-\lambda B|=0$$

因此，矩阵束左乘一个任意非奇异矩阵与（或）右乘任意一个非奇异矩阵，都不会改变矩阵束的广义特征值。这一结果可以总结为下面的命题：

命题 若X和Y是两个非奇异矩阵，则(XAY, ABY)和(A, B)是两个等价矩阵束。

等价矩阵束的概念给我们计算广义特征值问题提供了思路，可以通过选择合适的非奇异矩阵X, Y, 求得等价矩阵束(XAY, ABY)的广义特征值来求得矩阵束(A, B)的广义特征值。

应用：线性判别分析 (LDA)



- LDA: Linear Discriminant Analysis
- 两组数据投影到一维直线上
- 投影后：
  - 本组样本投影间距离尽量近
  - 不同组样本投影间距离尽量远
- 确定直线方向$\boldsymbol{w}$：

应用：线性判别分析 (LDA)

令$\boldsymbol{\mu}_i, \Sigma_i, (i = 0,1)$, 分别表示两组数据集合的均值向量和协方差矩阵。将这两组数据投影到直线$\mathbf{w}$上，则两类数据的中心在直线上的投影分别是$\mathbf{w}^T\boldsymbol{\mu}_0$和$\mathbf{w}^T\boldsymbol{\mu}_1$；两个样本投影的方差分别是$\mathbf{w}^T\Sigma_{\mathbf{0}}\mathbf{w}$和$\mathbf{w}^T\Sigma_{\mathbf{1}}\mathbf{w}$.

找出$\mathbf{w}$使同组数据投影点尽可能近：同组数据投影方差尽可能小；不同组数据投影点尽可能远：不同组数据中心投影尽可能远。

$$\max: \quad J = \frac{\|\mathbf{w}^T\boldsymbol{\mu}_0 - \mathbf{w}^T\boldsymbol{\mu}_{\mathbf{1}}\|^2}{\mathbf{w}^T\Sigma_{\mathbf{0}}\mathbf{w} + \mathbf{w}^T\Sigma_1\mathbf{w}}$$

应用：线性判别分析 (LDA)

- LDA的应用
  - 银行破产预测（Edward Altman, 1968）
  - 脸部识别（费舍尔脸 Fisher Faces）
  - 地球科学（区分蚀变带）

矩阵的相似化简与特征分析

- Hermitian矩阵的特征值分解
- 特征值和特征向量的性质
- 矩阵的相似变换与对角化
- 矩阵的可对角化条件
- 广义特征值分析
- 应用
  - 主成分分析（PCA）
  - 特征脸
  - 线性判别分析（LDA）