

# 一种面向数据学习的快速 PCA 算法<sup>\*</sup>

余 映<sup>1</sup>      王 斌<sup>1,2</sup>      张立明<sup>1</sup>

<sup>1</sup>(复旦大学 电子工程系 上海 200433)

<sup>2</sup>(复旦大学 信息科学与工程学院 波散射和遥感信息教育部重点实验室 上海 200433)

**摘 要** 高维数据的主成分分析较难处理, 因为计算时间和空间复杂度随着数据维数的增加而急剧增加. 文中提出一种直接面向数据学习的 PCA 算法, 即在迭代中新的权向量等于所有样本向量的加权和, 因而不需要计算数据协方差矩阵. 在解决给定样本向量或平稳随机过程的 PCA 问题时, 该算法能够弥补目前批量算法和增量算法存在的不足. 此外, 在理论上证明该算法的收敛性. 实验结果表明, 该算法能在很少迭代次数内迅速收敛到精确解.

**关键词** 主成分分析 (PCA), 无监督学习, 快速算法, 子空间, 协方差矩阵  
**中图法分类号** TP 391.41

## A Fast Data-Oriented Algorithm for Principal Component Analysis

YU Ying, WANG Bin<sup>2</sup>, ZHANG Liming

<sup>1</sup>(Department of Electronics Engineering, Fudan University, Shanghai 200433)

<sup>2</sup>(The Key Laboratory of Wave Scattering and Remote Sensing Information (Ministry of Education),  
School of Information Science and Engineering, Fudan University, Shanghai 200433)

### ABSTRACT

Principal components analysis (PCA) for high-dimensional data is a difficult problem because the computational time and the space complexity rapidly increase as the data dimensions increase. A data-oriented and covariance-free PCA algorithm is proposed inspired by the idea that the updated eigenvector in iteration is the weighted average of all samples. In a stationary environment or the condition that all training samples are available, the proposed algorithm is capable of overcoming the shortage of the conventional batch or incremental approaches. Furthermore, the convergence of the proposed algorithm is proved mathematically. Experimental results show that the most accurate solution is converged in a few iterations by the proposed algorithm.

**Key Words** Principal Component Analysis (PCA), Unsupervised Learning, Fast Algorithm, Subspace, Covariance Matrix

<sup>\*</sup> 国家自然科学基金项目 (No. 60672116, 60571052), 上海市重点学科建设项目 (No. B112) 资助  
收稿日期: 2008-05-15, 修回日期: 2008-12-24  
作者简介: 余映, 男, 1977 年生, 博士研究生, 主要研究方向为人工神经网络、计算机视觉. Email: yuyingmai@163.com; 王斌, 男, 1964 年生, 教授, 主要研究方向为信号和图像处理. 张立明, 女, 1943 年生, 教授, 主要研究方向为人工神经网络、图像处理 and 识别.  
©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

# 1 引言

在模式识别中,特征提取是一个重要过程,它将样本从高维数据空间变换到低维特征空间,并保留绝大部分信息.主成分分析(Principal Component Analysis, PCA)就是一种被广泛应用的信号特征提取和数据降维方法<sup>[1]</sup>.虽然在不同应用领域其被赋予不同名称,如KL变换(Karhunen-Loeve Transform)、霍特林变换(Hotelling Transform)、子空间法(Subspace Approach)和特征结构法(Eigen-structure Approach)等,但它们本质相同.

PCA算法可分为两类:批量计算法和增量计算方法.批量算法需要计算数据协方差矩阵,通过特征值分解(Eigenvalue Decomposition, EVD)来求解特征向量<sup>[2]</sup>.然而,当向量空间维数较大时,数据协方差矩阵和矩阵特征值分解的计算复杂度和数据存储量都将急剧增加以至于使系统无法承担.例如,在分析自然图像的主成分时<sup>[3]</sup>,每张大小为256行和256列的自然图像都是一个维数为 $d = 65536$ 的向量,协方差矩阵的元素总数为 $d^2 = 4294967296$ .如果样本数目较小,可用奇异值分解(Singular Value Decomposition, SVD)来求解特征向量<sup>[2]</sup>.但在实际应用中,样本数据量往往非常大.在这种情况下,基于协方差矩阵计算的批量计算方法就变得束手无策.因此,不需要计算协方差矩阵(Covariance-Free)的增量学习PCA算法可用来求解高维数据的主成分.增量算法大多是基于Hebbian学习规则的神经网络<sup>[4]</sup>,具有神经生物学意义.目前,增量学习算法主要有Oja提出的随机梯度上升(Stochastic Gradient Ascent, SGA)算法<sup>[5]</sup>,Sanger提出的广义Hebbian算法(Generalized Hebbian Algorithm, GHA)<sup>[6]</sup>和Weng提出的CCIPCA(Candid Covariance-Free Incremental PCA)算法<sup>[7]</sup>.理论上增量算法经过无穷多次迭代后能够求出特征向量,但是在实际应用中误差总会存在.当用格莱姆施密特正交化(Gram-Schmidt Orthogonalization)方法并行计算各个特征向量时,这种误差便会累积并向前传播<sup>[6-7]</sup>.

本文提出的PCA算法基于一种直接面向数据学习的思想,即在迭代中新的权向量是所有数据向量的加权和,称为直接面向数据学习的PCA算法(Data Oriented PCA, DOPCA).在解决给定数据或平稳随机过程的PCA问题时,DOPCA算法能够弥补目前批量算法和增量算法存在的不足.首先,它不需要计算数据协方差矩阵,运算复杂度低,可解决高维数据问题.其次,DOPCA算法具有单边渐近的收敛特性,能够较

快地求出最优特征向量的精确解.此外,本文证明PCA子空间的收敛一致性,并提出一种快速算法,使计算效率更高,收敛速度加快.

## 2 直接面向数据学习的快速PCA算法

### 2.1 求解第1个特征向量

PCA问题可定义如下:设 $x$ 为 $n$ 维向量空间中的零均值随机向量,即 $E\{x\} = 0$ . $C_x$ 为其协方差矩阵,即 $C_x = E\{xx^T\}$ ,其特征向量 $\{w_0, w_1, \dots, w_m\}$ 对应按大小顺序排列的特征值 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ .内积 $x^T w_0$ 即为 $x$ 的1阶主成分.由特征向量 $w_0, w_1, \dots, w_m$  ( $m < n$ )张成的线性子空间称为 $m$ 维PCA子空间.正交矩阵 $W = [w_0, w_1, \dots, w_m]$ 称为PCA变换矩阵.PCA算法的目标是估计协方差矩阵的特征向量.

设 $w(0)$ 为随机初始迭代权向量,求解第1特征向量 $w_0$ 的DOPCA算法迭代式为

$$w(t+1) = E\{w(t)^T x x\}. \quad (1)$$

由于 $w(t)^T x$ 为标量,所以式(1)看作所有样本向量的加权平均.这种用所有样本向量加权和作为新的权向量的迭代称为直接面向数据的学习.下面2个定理将证明式(1)能收敛到样本空间的第1最大特征向量.

**定理1** 设 $w(t)$ 是式(1)在迭代时刻 $t$ 的权向量, $\theta_1(t)$ 是 $w(t)$ 与最大特征向量 $w_0$ 之间的夹角.如果 $0 < \theta_1(0) < \pi$ ,那么当 $t \rightarrow \infty$ 时, $\theta_1(t) \rightarrow 0$ .

**证明** 假设 $w(0)$ 可表示为所有特征向量的线性组合,

$$w(0) = \sum_{i=1}^n \alpha_i w_i$$

因为式(1)可写为 $w(t+1) = E\{xx^T\}w(t)$ ,所以

$$\begin{aligned} w(t) &= (E\{xx^T\})^t w(0) = (E\{xx^T\})^t \sum_{i=1}^n \alpha_i w_i \\ &= \sum_{i=1}^n \alpha_i (E\{xx^T\})^t w_i = \sum_{i=1}^n \lambda_i^t \alpha_i w_i \end{aligned}$$

于是

$$\begin{aligned} \theta_1(t) &= \arccos \frac{w_0^T w(t)}{\|w(t)\|} \\ &= \arccos \frac{\lambda_1^t \alpha_1}{\sqrt{(\lambda_1^t \alpha_1)^2 + (\lambda_2^t \alpha_2)^2 + \dots + (\lambda_n^t \alpha_n)^2}} \\ &= \arccos \frac{\alpha_1}{\sqrt{\alpha_1^2 + \left(\frac{\lambda_2}{\lambda_1}\right)^2 \alpha_2^2 + \dots + \left(\frac{\lambda_n}{\lambda_1}\right)^2 \alpha_n^2}} \end{aligned}$$

因为  $0 < \theta_1(0) < 0.5\pi$ , 所以  $w_q^T w(0) = \alpha_1 > 0$ . 因为  $\lambda_1$  是  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  中的最大特征值, 所以当  $t \rightarrow \infty, \theta_1(t) \rightarrow 0$  证毕.

**定理 2** 假设协方差矩阵  $C_x$  的  $n$  个特征向量  $w_1, w_2, \dots, w_n$  把整个向量空间划分成  $2^n$  个象限, 则  $w(t+1)$  和  $w(t)$  必在同一象限内.

**证明** 根据特征向量的数学定义, 有  $w_{oi}^T w(t+1) = w_{oi}^T E\{w(t)^T x x^T\} = w_{oi}^T E\{x x^T\} w(t) = \lambda_i w_{oi}^T w(t)$

其中  $i = 1, 2, \dots, n$ . 由于特征值  $\lambda_i > 0$  所以  $w_{oi}^T w(t+1)$  和  $w_{oi}^T w(t)$  同号. 故  $w(t+1)$  和  $w(t)$  必在同一象限内. 证毕.

定理 1 说明  $w(t)$  与最大特征向量之间的夹角在迭代过程中单调减小. 定理 2 说明  $w(t)$  必与初始权向量  $w(0)$  在同一象限. 因此, 如果每次迭代后将权向量归一化, 那么式 (1) 便能收敛到最大特征向量.

DOPCA 算法的直观解释如下. 假设有一组零均值 2 维数据, 用一个椭圆形状分布的数据云来表示, 如图 1(a) 所示, 根据特征向量的几何意义,  $w_1$  对应椭圆长轴,  $w_2$  对应椭圆短轴, 不妨假设  $w$  在第 1 象限. 根据定理 1 和定理 2  $w(t+1)$  一定在  $w(t)$  与  $w_1$  之间. 此外, 如果定义  $\tilde{x}$  的“等效向量”为

$$\tilde{x} = \begin{cases} x & w(t)^T x \geq 0 \\ -x & w(t)^T x < 0 \end{cases}$$

那么式 (1) 可写成  $w(t+1) = E\{w(t)^T x \tilde{x}\}$ , 因为  $w(t)^T x x = w(t)^T x \tilde{x}$  注意到  $w^T x \geq 0$  故对称向量的“作用力”不会相互抵消, 其几何意义如图 1(b) 所示.

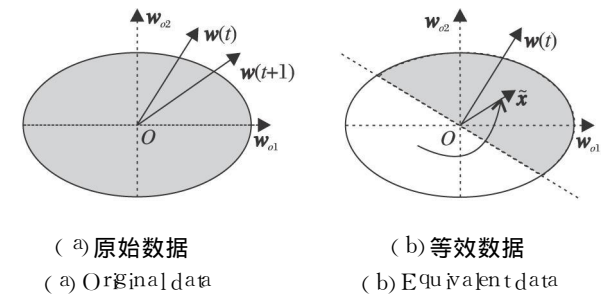


图 1 DOPCA 算法的直观解释

Fig 1 Intuitive explanation of DOPCA algorithm

2.2 求解高阶特征向量

2.2.1 数据紧缩法

迭代式 (1) 只能求解最大特征向量. 因为特征向量相互正交, 所以可利用这种正交性求解其它特征向量. 例如, 假设已找到最大特征向量  $w_1$ , 并将所有样本投影到与  $w_1$  正交的子空间中. 由于紧缩样

本的协方差矩阵有最大特征值  $\lambda_2$  和最大特征向量  $w_2$ , 因此, 利用式 (1)  $w(t)$  将收敛到  $w_2$ . 所以, 求解下一特征向量的方法是在迭代之前将所有样本数据做数据紧缩:

$$x \leftarrow x - \sum_{i=1}^k w_{oi} w_{oi}^T x$$

这种紧缩方法使下一特征向量的迭代在与前面特征向量相正交的子空间内进行.

2.2.2 权向量紧缩法

求解高阶特征向量的另一方法是将更新后的权向量投影到与前面特征向量相正交的子空间内. 定理 3 表明用这种方法式 (1) 将收敛到“紧缩”子空间内的最优特征向量.

**定理 3** 设  $w(t)$  是式 (1) 在迭代时刻的权向量,  $\theta_k(t)$  是  $w(t)$  与第  $k$  个特征向量  $w_{ok}$  之间的夹角. 如果  $w(0) \perp \{w_{o1}, w_{o2}, \dots, w_{ok-1}\}$  且  $0 < \theta_k(0) < 0.5\pi$ , 那么当  $t \rightarrow \infty$  时,  $\theta_k(t) \rightarrow 0$ .

**证明** 因为  $w(0) \perp \{w_{o1}, w_{o2}, \dots, w_{ok-1}\}$ , 所以  $w(0)$  可表示为  $w(0) = \sum_{i=k}^n \alpha_i w_{oi}$  于是

$$\begin{aligned} w(t) &= (E\{x x^T\})^t w(0) = (E\{x x^T\})^t \sum_{i=k}^n \alpha_i w_{oi} \\ &= \sum_{i=k}^n \alpha_i (E\{x x^T\})^t w_{oi} = \sum_{i=k}^n \lambda_i^t \alpha_i w_{oi} \end{aligned}$$

从而

$$\begin{aligned} \theta_k(t) &= \arccos\left(\frac{w_{ok}^T w(t)}{\|w(t)\|}\right) \\ &= \arccos\left(\frac{\alpha_k}{\sqrt{\alpha_k^2 + \left(\frac{\lambda_{k+1}}{\lambda_k}\right)^2 \alpha_{k+1}^2 + \dots + \left(\frac{\lambda_n}{\lambda_k}\right)^2 \alpha_n^2}}\right) \end{aligned}$$

同定理 1 的证明有, 当  $t \rightarrow \infty$  时,  $\theta_k(t) \rightarrow 0$  证毕.

根据定理 3 只需将迭代向量  $w(t)$  投影到与前面特征向量相正交的子空间内, 便可使用式 (1) 逐一估计出后面的特征向量. 假设已找到前  $k$  个特征向量  $\{w_{o1}, w_{o2}, \dots, w_{ok}\}$ , 为了使式 (1) 能收敛到第  $k+1$  个特征向量  $w_{ok+1}$ , 需将迭代向量进行紧缩, 即

$$w \leftarrow w - \sum_{i=1}^k w_{oi} w_{oi}^T w$$

此方法称为权向量紧缩法. 因为数据紧缩只需进行一次, 迭代过程中不再重复进行, 所以本文实验均采用数据紧缩的办法.

2.3 快速计算

通常在解决正交投影追踪问题时, 总是随机初始化一个权向量进行迭代. 然而这样做不是高效的, 没有充分利用前面特征向量迭代过程中隐含的有用信息. 事实上, 当迭代向量收敛到某一子空间的最优

特征向量时,其在下一紧缩子空间内的投影也在朝向次优特征向量.如果将此投影信息用于下一个特征向量的迭代,那么将会加快迭代的收敛速度.定理4即为此方法的理论依据.

为方便表示,先定义如下符号,设  $S_k$  是与前  $k$  个特征向量  $\{w_1, w_2, \dots, w_{ok}\}$  相正交的  $k$  阶紧缩子空间.  $S_k$  也可看作是由特征向量  $\{w_{\alpha(k+1)}, w_{\alpha(k+2)}, \dots, w_{\alpha n}\}$  张成的线性子空间.

**定理 4** 假定  $w^{(t)}$  是式 (1) 在迭代时刻  $t$  的权向量,其初始值  $w(0)$  在子空间  $S_{k-1}$  内.对于  $m = k+1, \dots, n-2$  设  $w^{(m)}(\eta)$  是  $w^{(t)}$  在子空间  $S_m$  上的投影,  $\theta^{(m)}(\eta)$  是  $w^{(m)}(\eta)$  与第  $m+1$  个特征向量  $w_{\alpha(m+1)}$  之间的夹角.如果  $0 < \theta^{(m)}(0) < 0.5\pi$ , 那么当  $t \rightarrow \infty$  时,  $\theta^{(m)}(\eta) \rightarrow 0$

**证明** 根据定理 1、定理 3 的证明和迭代式 (1), 可得

$$w(0) = \sum_{i=k}^n \alpha_i w_{\alpha_i}, \quad w(\eta) = \sum_{i=k}^n \lambda_i^t \alpha_i w_{\alpha_i}.$$

从而

$$w^{(m)}(0) = w(0) - \sum_{i=k}^m \alpha_i w_{\alpha_i} = \sum_{i=m+1}^n \alpha_i w_{\alpha_i}$$

且

$$w^{(m)}(\eta) = w(\eta) - \sum_{i=k}^m \lambda_i^t \alpha_i w_{\alpha_i} = \sum_{i=m+1}^n \lambda_i^t \alpha_i w_{\alpha_i},$$

$$m = k+1, \dots, n-2$$

于是

$$\begin{aligned} \theta^{(m)}(\eta) &= \arccos\left(\frac{w_{\alpha(m+1)}^T w^{(m)}(\eta)}{\|w^{(m)}(\eta)\|}\right) \\ &= \arccos\left(\frac{\alpha_{m+1}}{\sqrt{\alpha_{m+1}^2 + \left(\frac{\lambda_{m+2}^t}{\lambda_{m+1}^t}\right)^2 \alpha_{m+2}^2 + \dots + \left(\frac{\lambda_n^t}{\lambda_{m+1}^t}\right)^2 \alpha_n^2}}\right), \end{aligned}$$

$$m = k+1, \dots, n-2.$$

因为  $0 < \theta^{(m)}(0) < 0.5\pi$ , 所以

$$w_{\alpha(m+1)}^T w^{(m)}(0) = \alpha_{m+1} > 0 \quad m = k+1, \dots, n-2$$

因为  $\lambda_{m+1}$  是  $\{\lambda_{m+1}, \dots, \lambda_n\}$  中的最大特征值, 所以当  $t \rightarrow \infty$  时,  $\theta^{(m)}(\eta) \rightarrow 0 \quad m = k+1, \dots, n-2$  证毕.

**定理 4** 说明式 (1) 的迭代在所有紧缩子空间内具有收敛同步性.例如,在估计  $w_q$  的迭代过程中,权向量  $w(\eta)$  将逐渐朝向  $w_q$ ;与此同时,其在  $\{w_{\alpha(m+1)}, w_{\alpha(m+2)}, \dots, w_{\alpha n}\}$  张成的线性子空间内的投影也在逐渐靠近  $w_{\alpha(m+1)}$ ,  $m = 1, 2, \dots, n-2$

以上分析说明,求解一个特征向量的迭代过程蕴含着对求解其它特征向量非常有用的信息.假定  $w^{(t)}$  是式 (1) 在估计第  $k$  个特征向量  $w_{ok}$  时的迭代

向量,  $w^{(k)}(\eta)$  是它在  $S_k$  上的投影.根据定理 4 当  $t \rightarrow \infty$  时,  $w^{(k)}(\eta)$  将指向  $w_{\alpha(k+1)}$ .如果利用  $w^{(t)}$  在  $S_k$  内的投影作为下一个特征向量  $w_{\alpha(k+1)}$  的初始值,则可加速迭代,以节约计算成本.

## 2.4 具体算法步骤

**算法 1** 前  $P$  个特征向量的 DOPCA 基本算法

step 1 选择需要计算的特征向量数  $P$  并使  $k = 1$ .

step 2 随机初始化权向量  $w_k$ .

step 3 更新权向量:  $w_k^+ = E\{w_k^T x x\}$ .

step 4 归一化权向量:  $w_k = w_k^+ / \|w_k^+\|$ .

step 5 若  $w_k$  未收敛,则返回到 step 3.否则,转到 step 6.

step 6 紧缩样本数据:  $X = X - w_k (w_k^T X)$ .

step 7  $k = k+1$  回到 step 2 直到  $k = P$

在 step 6 紧缩所有样本向量,数据矩阵  $X$  的每一列对应一个样本向量.当第  $k$  个特征向量的当前值  $w_k$  和更新以前的值  $w_k^*$  满足  $w_k^T w_k^* = 1$  时迭代收敛.通常地,可允许有很小的收敛容差,即判断迭代收敛的标准为

$$abs(w_k^T w_k^* - 1) < \epsilon$$

其中,  $\epsilon$  是收敛容差,  $abs(\cdot)$  是求绝对值运算.

**算法 2**  $P$  个特征向量的 (DOPCA) 快速算法 (Fast DOPCA)

step 1 使  $k = 1$  随机初始化权向量  $w_k$  并使  $w_k^* = w_k$ .

step 2 更新权向量:  $w_k^+ = E\{w_k^T x x\}$ .

step 3 归一化权向量:  $w_k = w_k^+ / \|w_k^+\|$ .

step 4 若  $w_k$  未收敛,则  $w_k^* = w_k$  返回到 step 2 否则,转到 step 5

step 5 紧缩样本向量:  $X = X - w_k (w_k^T X)$  并计算差值:  $\xi = w_k^* - w_k (w_k^T w_k^*)$ .

step 6  $k = k+1$ ,  $w_k = \xi / \|\xi\|$ ,  $w_k^* = w_k$  回到 step 2 直到  $k = P$

在 step 5 中,  $\xi$  是迭代向量  $w_k$  在  $\{w_{\alpha(k+1)}, w_{\alpha(k+2)}, \dots, w_{\alpha n}\}$  张成的  $k$  阶紧缩子空间内的投影.根据定理 4 当  $w_k$  收敛时,  $\xi$  与  $w_{\alpha(k+1)}$  之间的夹角将会很小.在 step 6 中,将其进行归一化,作为第  $k+1$  个特征向量的迭代初始值.

## 2.5 运算复杂度

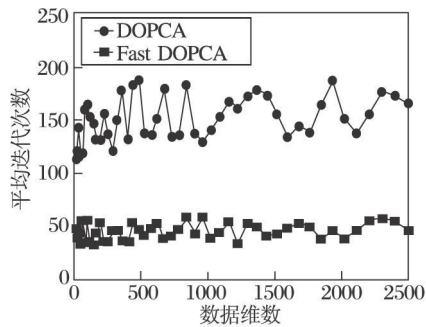
假定样本向量数目为  $n$  向量空间维数为  $d$  于是求协方差矩阵的运算复杂度为  $O(d^2 n)$ .大多批量学习 PCA 算法运算复杂度约为  $O(d^2 + d n)$ ,至少与向量空间维数的平方成正比<sup>[2]</sup>. DOPCA 计算前  $P$  个

特征向量的运算复杂度为  $O(dnlp)$ , 其中, 为求解一个特征向量需要的平均迭代次数, 与收敛容差  $\epsilon$  有关. 因为本文算法的运算复杂度仅与向量空间维数成正比, 所以在解决高维 PCA 问题时具有较显著优势.

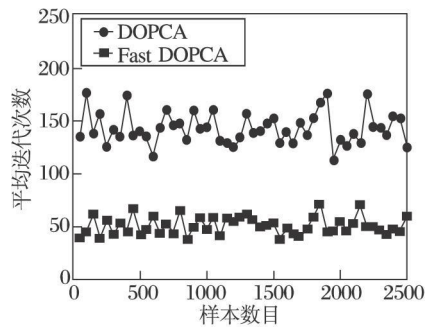
3 实验结果和评价

本文实验均在 Matlab 7.0 环境下运行, 计算机配置 Intel i. 50C 处理器, 1GB 内存, 32 位操作系统.

其中,  $P$  为需要计算的特征向量数目,  $l_k$  为计算第  $k$  个特征向量的迭代次数. 本实验计算前 10 个特征向量. 最终平均迭代次数从 5 次独立实验取平均得到, 其中初始权向量随机选取. 图 2 (a) 给出平均迭代次数曲线, 可看出平均迭代次数具有统计平稳性, 而且快速算法收敛更快. 再来观察 与样本向量数目之间的关系. 图 2 (b) 中给出不同样本向量数目下的平均迭代次数曲线. 从相同图像集中随机采样大小为  $30 \times 30$  ( $d=900$ ) 的  $n=100\ 200\ 300\ \cdots, 5\ 000$  个图像子块进行计算. 实验独立重复 5 次, 得到最终平均迭代次数. 可以看出, 迭代次数也是随机平稳的. 为保证计算精度, 收敛容差  $\epsilon = 1.0 \times 10^{-15}$ .



(a)  $n=1000$



(b)  $d=900$

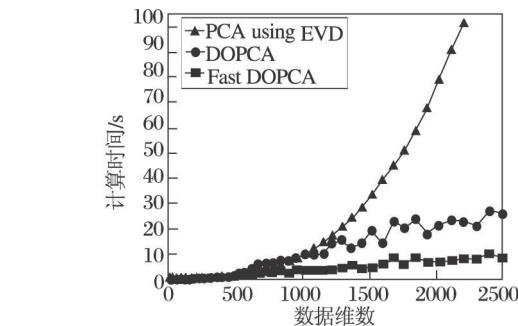
图 2 不同条件下 DOPCA 迭代次数的平稳性

Fig 2 Statistical stationarity of iteration number of DOPCA in different conditions

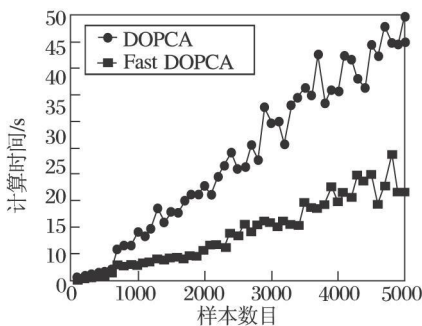
3.1 实验 1

如前所述, 平均迭代次数 是 DOPCA 运算复杂度的一个关键因素. 首先来观察 与样本向量维数之间的关系. 随机地从自然图像数据集 (数据集见 <http://www.cis.hut.fi/projects/ica/data/image>) 中选取大小为  $s \times s$  ( $s=4\ 5\ \cdots, 50$ , 相应向量维数  $d=16\ 25\ \cdots, 2\ 500$ ) 的 1 000 个图像子块, 并计算平均迭代次数:

$$l = \frac{1}{P} \sum_{k=1}^P l_k$$



(a)  $n=100$



(b)  $d=100$

图 3 不同条件下 DOPCA 求解前 10 个特征向量的计算时间  
Fig 3 Computation time of DOPCA for first 10 eigenvectors in different conditions

3.2 实验 2

在相同计算精度前提下, 算法迭代时间为计算所有特征向量的时间总和:

$$t = \sum_{k=1}^P t_k$$

其中,  $P$  为需要计算的特征向量数目,  $t_k$  为第  $k$  个特征向量的计算时间. 图 3 (a) 给出迭代时间随向量空间维数变化的曲线图. 本实验求解 100 个样本数据

的前 10 个特征向量,可以看出,EVD算法计算时间随向量空间维数的增加而急剧增加,而本文算法的计算时间是线性增长的.图 3(b)为在不同数据量时求解大小为  $10 \times 10$  像素子图像前 10 个特征向量的时间.计算时间随着向量空间维数或样本数目的增加总体呈现线性增长趋势.此外,DOPCA快速算法明显好于普通的 DOPCA算法.

3.3 实验 3

本实验采用 ORL 人脸图像数据库进行特征脸提取实验.数据库中共有 400 张  $112 \times 92$  像素的人脸图像,向量空间维数  $d = 10304$ ,样本维数比为  $400/10304 \approx 0.039$ ,所以是一个奇异 (Singularity) 问题.用 SVD 矩阵分解法计算出前 10 个特征脸,将其作为对照基准<sup>[2]</sup>.然后,用估计的特征向量  $w$  和基准特征向量  $v$  之间的相关度来衡量迭代收敛的准确性:

$$c = ||w^T v||.$$

由于  $w$  和  $v$  都是单位向量,所以  $0 \leq c \leq 1$ .

本实验用 CCIPCA 算法来做实验比较.把 CCIPCA 的 400 次迭代 (对应 400 个样本) 看作 1 个

迭代周期.因此,CCIPCA 的 1 个迭代周期对应 DOPCA 的 1 次迭代.实验记录 CCIPCA 每个迭代周期后的前 10 个特征向量的收敛准确度并绘制收敛曲线,如图 4(a) 所示.不难看出,CCIPCA 估计的后 5 个特征向量存在明显的偏差.图 4(b)、(c) 分别给出 DOPCA 算法和 Fast DOPCA 算法的收敛曲线.可以看出,本文的两种算法不仅单调收敛而且收敛迅速,经过 20 次迭代后收敛准确性就已接近 1.其中 Fast DOPCA 收敛更快,因为其迭代向量初始值的准确性已经很高.

DOPCA 以逐一方式计算出每个特征向量,而 CCIPCA 采用并行计算方式.表 1 给出本实验估计前 10 个特征脸的平均计算时间和收敛准确度 ( $C_1 - C_{10}$ ).可以看出,CCIPCA 算法 (迭代 100 个周期) 用时较多,收敛准确度低.本文算法能够得到特征脸的精确解,而且 Fast DOPCA 算法用时更少.

表 1 3 种算法估计前 10 个特征脸的实验结果对比  
Table 1 Result comparison of estimating first 10 eigenvectors among 3 algorithms

	CCIPCA	DOPCA	FastDOPCA
平均计算时间 / s	425.94	106.98	43.26
收敛准确度	$C_1$ 1.00000	1.00000	1.00000
	$C_2$ 0.99979	1.00000	1.00000
	$C_3$ 0.99972	1.00000	1.00000
	$C_4$ 0.99255	1.00000	1.00000
	$C_5$ 0.99021	1.00000	1.00000
	$C_6$ 0.91221	1.00000	1.00000
	$C_7$ 0.90946	1.00000	1.00000
	$C_8$ 0.64084	1.00000	1.00000
	$C_9$ 0.56899	1.00000	1.00000
	$C_{10}$ 0.97047	1.00000	1.00000

4 讨 论

现在来考虑当存在相等特征值时的情况.假定按降序排列后在  $\lambda_1$  与  $\lambda_m$  之间的特征值都相等,即  $\lambda_{+1} \geq \lambda_1 = \lambda_{+1} = \dots = \lambda_m > \lambda_{m+1}$ .根据定理 1,迭代向量首先收敛到最大特征值对应的特征向量.因此,前  $1 - 1$  个特征向量的迭代不受影响.而特征向量  $\{w_{\alpha_1}, \dots, w_{\alpha_m}\}$  的迭代将收敛到由它们张成线性子空间内的一组正交基,收敛方位取决于初始值的选择.

根据定理 1 和定理 3,如果  $w(0)$  恰好垂直于最优特征向量,那么式 (1) 迭代就不能收敛到最优解.但在实际应用中,随机设置的初始向量恰好垂直于最优特征向量的概率很小,而且在计算过程中引入的舍入误差也会让本文算法总能收敛到最优解.

在 Fast DOPCA 算法中,用前一个特征向量的最

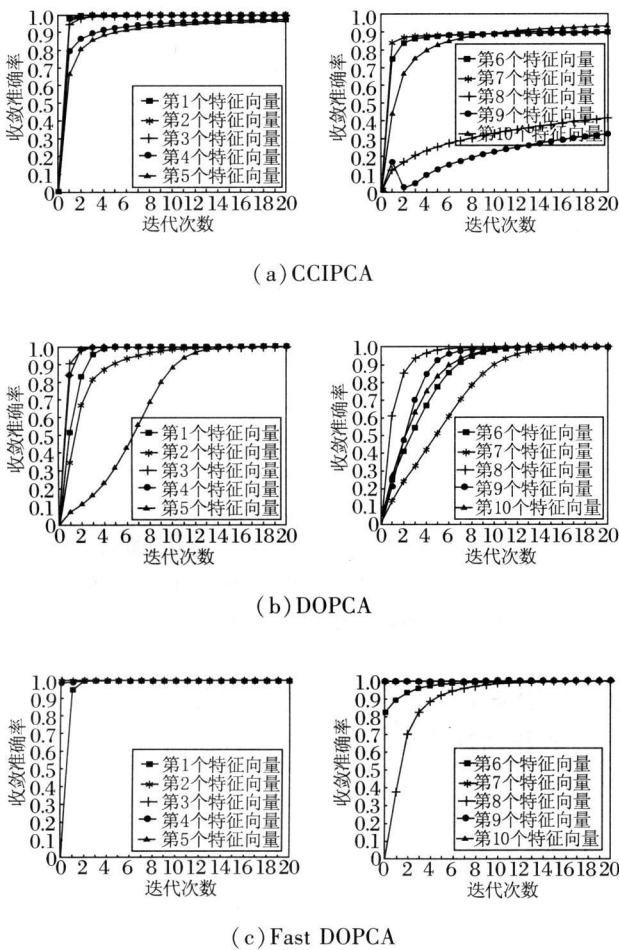


图 4 3 种算法收敛准确性曲线图

Fig. 4 Curves for convergence accuracy by 3 algorithms  
©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

后一次迭代差值作为后一个特征向量的迭代初始值,这样做是非常有效的.通常情况下,特征向量迭代的初始值与特征向量就已有了很高的相关度,所以大大节约迭代次数、节约计算时间、提高估计效率.但是,当收敛容差  $\epsilon$  取值非常小时,最后一步迭代的权向量改变非常小,于是导致差值向量的模很小.由于计算机的计算精度有限,所以可能会造成差值向量方向的偏差,这是本文特征脸实验中第 6 和第 8 个特征脸初始值与基准特征向量相关度不高的原因.

由于 CCIPCA 比 SGA 和 GHA 有更好的收敛特性,所以仅用 CCIPCA 作为增量学习算法的代表进行比较.在特征脸实验中,由于数据库中只有 400 张图像,所以可用矩阵计算方法来求特征向量.但是,当数据量较大时,矩阵计算方法就很难处理了,而用 DOPCA 就可求出特征向量的精确解.

## 5 结 束 语

本文旨在研究不需计算协方差矩阵而求解高维数据特征向量精确解的问题.基于面向数据学习的思想,本文提出一种简单的 DOPCA 算法.经过分析,本文算法可精确收敛到数据协方差矩阵的特征向量.利用迭代向量在高阶线性子空间投影, Fast DOPCA 可

进一步加快收敛速度.实验结果表明, DOPCA 算法收敛速度快,能求出最优特征向量的精确解,是一种解决大数据量高维 FCA 问题的有效方法.

致 谢 特别感谢复旦大学数学系蔡志杰老师在定理证明上给予有益指导!

## 参 考 文 献

- [ 1 ] Haykin S. Neural Networks: A Comprehensive Foundation, 2nd Edition. Upper Saddle River, USA: Prentice Hall, 2001
- [ 2 ] Golub G H, van Loan C F. Matrix Computation, 3rd Edition. Baltimore, USA: John Hopkins University Press, 1996
- [ 3 ] Torralba A, Oliva A. Statistics of Natural Image Categories. Netw. Comput. in Neural Systems, 2003, 14(3): 391—412
- [ 4 ] Oja E. A Simplified Neuron Model as a Principal Component Analyzer. Journal of Mathematical Biology, 1982, 15(3): 267—273
- [ 5 ] Oja E, Karhunen J. On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix. Journal of Mathematical Analysis and Applications, 1985, 106: 69—84
- [ 6 ] Sanger T D. Optimal Unsupervised Learning in a Single Layer Linear Feedforward Neural Network. IEEE Trans on Neural Networks, 1989, 2(6): 459—473
- [ 7 ] Weng Juyang, Zhang Yili, Hwang W S. Candid Covariance Free Incremental Principal Component Analysis. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(8): 1034—1040