doi: 10.12011/1000-6788(2016)02-0528-08 中图分类号: TP181 文献标志码: A

基于 Cholesky 分解的 K2DPCA 人脸识别研究

周水生, 郑 颖, 穆新亮

(西安电子科技大学 数学与统计学院, 西安 710126)

摘 要 K2DPCA (kernel 2D principal component analysis) 是基于非线性特征提取的重要人脸识别方法, 具有成功的应用. 但对大规模训练数据库, 其因核矩阵 $\boldsymbol{K}$ 规模过大、计算代价高而不能有效实现. 采用选主元 Cholesky 分解方法, 仅需计算核矩阵的对角线上元素和部分精选列, 得到迹范数意义下核矩阵 $\boldsymbol{K}$ 的最优 Nyström 型低秩近似 $\boldsymbol{L}\boldsymbol{L}^{\top}$ 来解决该问题. 并只需计算小规模矩阵 $\boldsymbol{L}^{\top}\boldsymbol{L}$ 的特征值和特征向量, 实现大规模 K2DPCA/KPCA (kernel principal component analysis) 的非线性特征提取. 在加噪 ORL 人脸数据库上的实验结果表明, 较 K2DPCA/KPCA 方法, 新方法显著提高了识别率, 并可以很大程度上克服噪声的影响; 在 Extended YaleB 大型人脸数据库上的实验结果表明, 此算法解决了 K2DPCA 核矩阵过大而不能有效实现的缺点.

关键词 人脸识别; KPCA; K2DPCA; Cholesky 分解; Nyström 型低秩近似

K2DPCA methods for face recognition based on Cholesky decomposition

ZHOU Shuisheng, ZHENG Ying, MU Xinliang

(School of Mathematics and Statistics, Xidian University, Xi'an 710126, China)

Abstract K2DPCA (kernel 2D principal component analysis), with successful applications, is an important method for nonlinear face recognition. However, it will meet challenge for the large-scale training problems, because the kernel matrix $\boldsymbol{K}$ is too large to fit the memory and the eigenvalues and eigenvectors of the kernel matrix are not obtained freely. To answer this challenge, we propose a new method, where the kernel matrix $\boldsymbol{K}$ is decomposed as its Nyström-type low-rank approximation $\boldsymbol{L}\boldsymbol{L}^{\top}$ by the pivoted Cholesky decomposition. In the procedure, only the diagonals and some well-chosen columns of the kernel matrix are available, then the optimal Nyström-type approximation is obtained under the trace norm without the full kernel matrix. Owning to this, K2DPCA/KPCA (kernel principal component analysis) is implemented by calculating the eigenvalues and eigenvectors of the small size matrix $\boldsymbol{L}^{\top}\boldsymbol{L}$. The experiments on noise dataset of ORL face database show that the proposed algorithm can significantly improve the recognition rates of K2DPCA/KPCA, and also decrease the effect of noise to some extend. The experimental results on a large scale Extended YaleB face database show that the proposed method overcomes the weakness of K2DPCA on the large-scale training set.

Keywords face recognition; KPCA; K2DPCA; Cholesky decomposition; Nyström-type low-rank approximation

收稿日期: 2014-06-25

作者简介: 周水生 (1972–), 男, 汉, 陕西洛南人, 教授, 博士生导师, 研究方向: 最优化计算方法及其应用, 机器学习及其核学习, 数据挖掘等; 郑颖 (1988–), 女, 汉, 陕西宝鸡人, 硕士, 研究方向: 最优化算法及其应用, 人脸检测等; 穆新亮 (1988–), 男, 汉, 陕西宝鸡人, 硕士, 研究方向: 最优化算法及其应用.

基金项目: 国家自然科学基金 (61179040, 61173089); 陕西省自然科学基金 (2014JM1031); 陕西省科技厅项目 (2013JK0603)
Foundation item: National Natural Science Foundation of China (61179040, 61173089); National Natural Science Foundation of Shaanxi Province (2014JM1031); Foundation of Science and Technology Department of Shaanxi Province (2013JK0603)
中文引用格式: 周水生, 郑颖, 穆新亮. 基于 Cholesky 分解的 K2DPCA 人脸识别研究 [J]. 系统工程理论与实践, 2016, 36(2): 528–535.
英文引用格式: Zhou S S, Zheng Y, Mu X L. K2DPCA methods for face recognition based on Cholesky decomposition[J]. Systems Engineering — Theory & Practice, 2016, 36(2): 528–535.

1 引言

特征提取通过投影实现高维数据的降维, 是人脸识别的关键环节. 主成分分析 (PCA) 是最早提出且应用广泛的一种线性特征抽取方法 (参阅专著 [1]), 最近也有很多应用研究 [2−4]. PAC 通过计算训练样本协方差矩阵的特征向量, 求投影矩阵, 来产生低维特征数据. 在处理人脸识别问题时, PCA 首先需将各个图像矩阵重排成高维向量作为训练样本, 计算所有训练样本协方差矩阵的特征向量. 这对高分辨率图像, 重排导致训练样本维数过高、协方差矩阵规模大, 进而特征向量求解困难. 幸运的是, 可以采用奇异秩分解 (SVD) 技巧, 通过计算训练数据 Gram 矩阵的特征值和特征向量来降低计算量. 但由于将图像矩阵重排成向量破坏了原图像二维数据结构或其邻域信息, 不利于后期的识别检测、重构等处理. 为了尽量保留原始图像的邻域信息, Yang 等 [5] 提出了基于图像矩阵的 2DPCA(线性) 特征抽取方法. 此方法不需要将图像矩阵重排成向量, 而是将 m 个图像矩阵 $\boldsymbol{A}_i$ 直接并排为 $\boldsymbol{X}=[\boldsymbol{A}_1\ \boldsymbol{A}_2\ \cdots\ \boldsymbol{A}_m]$, 利用 PCA 直接处理. 其实质是将图像的每一行/列作为一个样本. 较原始 PCA 方法大大降低了样本维数, 缩小了协方差矩阵的规模, 这可有效求得投影矩阵. 并且在 2DPCA 中, 由于图像矩阵的行数或列数是固定不变的, 不受样本个数的影响, 因而对于大规模样本集, 2DPCA 算法仍能有效抽取线性特征.

虽然 PCA 和 2DPCA 都能有效降低数据维数, 但它们都是线性特征抽取方法. 而人脸图像信息中往往含有非线性特征, 因此有必要研究非线性特征提取方法. 作为 PCA 的非线性推广, KPCA[6−8] 的主要思想是通过非线性映射, 将样本数据从输入空间映射到高维甚至无穷维的 Hilbert 特征空间, 再在此特征空间中进行线性 PCA 降维. 利用在支持向量机等领域应用广泛的核技巧 [9−11], KPCA 仅通过对核矩阵进行特征值分解就可以有效抽取数据的非线性特征. 文 [12–13] 研究了 KPCA 在股票分类和监控诊断等领域的应用.

文 [14] 提出了 K2DPCA: 将 m 个图像矩阵 $\boldsymbol{A}_i$ 并排为 $\boldsymbol{X}=[\boldsymbol{A}_1\ \boldsymbol{A}_2\ \cdots\ \boldsymbol{A}_m]$, 把 $\boldsymbol{X}$ 的每一列作为一个样本进行非线性映射, 然后利用 KPCA 技巧进行非线性特征提取, 取得了较好的实验效果. 由于 KPCA 算法核矩阵的规模是训练样本的个数, 故 K2DPCA 算法核矩阵的规模是图像的个数与图像的行数或列数的乘积, 其存储复杂度和计算复杂度都急剧增加.

为了降低复杂度, Sun 等 [15] 和 Eftekhari 等 [16] 提出对核矩阵分块, 将大的核矩阵特征向量的计算, 转换成若干小核矩阵特征向量的计算, 最后用所得的多个小矩阵的特征向量合并计算所需的投影矩阵, 进行抽取特征. 整个过程较复杂, 且没能完全解决存储复杂度大的问题; 且当训练样本数目较多时, 小核矩阵数量增多, 特征向量计算也很耗时. Wang 等 [17] 研究了图像矩阵的向量、矩阵混合表示方法, 来减小核矩阵规模以降低计算复杂度.

这几种方法虽可以缩小核矩阵的规模, 如 160 张分辨率为 $28*23$ 的图像的 K2DPCA 训练, [15–16] 是将 $4480*4480$ 的大核矩阵分成 $28*28$ 块维数为 $160*160$ 的小核矩阵, 再对小矩阵计算特征值分解然后合并各自特征向量得到投影矩阵. 而实际仅考虑原核矩阵的主对角分块矩阵, 需计算 $160*160$ 的核矩阵 28 次. [17] 是先将每个图像矩阵相邻的若干行 (或列) 形成的块拉成一个行 (列) 向量执行 KPCA, 即对图像分块, 将每一块作为一个单元做非线性映射, 构造核矩阵, 进而使核矩阵规模缩小, 但这在某种程度上也破坏了图像的 2D 结构. 当训练样本多时, 核矩阵规模主要由训练样本个数决定, 这些方法不能有效减小核矩阵规模.

本文研究 K2DPCA 问题. 注意到核矩阵本身是训练样本的一种相似性度量方式的体现, 故相似样本所对应的核矩阵的列也是相似或线性相关的. 通常由于训练数据中有大量相似样本, 故核矩阵一般是低秩的或者是可以被低秩近似的. K2DPCA 的核矩阵更是如此. 本文利用选主元的 Cholesky 分解方法, 逐步得到大规模核矩阵 $\boldsymbol{K}$ 的低秩近似 $\boldsymbol{L}\boldsymbol{L}^{\top}$, 整个过程中只需计算核矩阵的对角线元素和部分精选列, 且仅需存储低秩矩阵 $\boldsymbol{L}$. 进一步通过计算小规模矩阵 $\boldsymbol{L}^{\top}\boldsymbol{L}$ 的特征值分解, 求得到训练数据的非线性特征. 这不但大大降低了计算复杂度, 而且有效节省了存储复杂度. 实验结果表明, 所提方法可以有效求解大规模训练问题, 克服 K2DPCA 由于超出内存空间而不能有效实现的缺点; 且对小规模含噪训练问题, 该方法可以起到降噪效果.

2 相关算法介绍

本节简要介绍相关的主成分分析算法, 包括 PCA/KPCA/2DPCA/K2DPCA, 并讨论了各种方法的计算复杂度、存储复杂度, 详细参阅专著 [1, 7] 等以及文献 [2, 6, 14–17] 等.

2.1 线性主成份分析 (PCA)

设 $\boldsymbol{x}_i\in\mathbb{R}^n\ (i=1,2,\cdots,m)$ 是训练集中的 m 个 n 维样本数据, 线性 PCA 是先将原样本数据中心化

后 (仍记为 $\boldsymbol{x}_i$), 得到矩阵 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m]$, 再由其协方差矩阵 $\boldsymbol{S}=\frac{1}{m}\boldsymbol{X}\boldsymbol{X}^\top$ 的前 p 个最大特征值对应的正交特征向量, 计算出投影矩阵 $\boldsymbol{W}\in\mathbb{R}^{n\times p}(p\le n)$. 对任一训练/测试样本 $\boldsymbol{x}$, 将其投影到该投影矩阵 $\boldsymbol{W}$ 上, 得到保持了其主要特征的降维样本 $\boldsymbol{y}=\boldsymbol{W}^\top\boldsymbol{x}\in\mathbb{R}^p$. 对于高维数据 ($m<n$), 为了避免求高维矩阵的特征值、特征向量, 通过计算训练样本的 Gram 矩阵 $\boldsymbol{X}^\top\boldsymbol{X}$ 的前 p 最大特征值分解 $\sum_{i=1}^{p}\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^\top$, 可得到 $\boldsymbol{S}$ 的前 p 个最大的特征值为 $\frac{1}{m}\lambda_i$ $(i=1,2,\cdots,p)$, 而相应的特征向量为 $\boldsymbol{X}\boldsymbol{u}_i$ $(i=1,2,\cdots,p)$, 再进行标准化可得投影矩阵 $\boldsymbol{W}$, 进一步提取训练/测试样本的降维特征. PCA 的计算复杂度为 $\mathcal{O}(\min\{m^3,n^3\})$.

2.2 非线性主成份分析 (KPCA)

KPCA[6−8] 是 PCA 算法的非线性推广. 思想是通过非线性映射 $\boldsymbol{\psi}:\mathbb{R}^n\to\mathbb{F}$, 将原样本向量映射到高维、甚至无穷维的 Hilbert 特征空间 $\mathbb{F}$ 中, 即 $\boldsymbol{x}_i\to\boldsymbol{\psi}(\boldsymbol{x}_i)(i=1,2,\cdots,m)$, 再在特征空间 $\mathbb{F}$ 中进行类似 PCA 的处理. 为了避免在特征空间的高维计算, 利用 SVM 等领域应用广泛的核技巧 [9−11], 选择合适的满足 Meccer 条件的核函数 $k(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{\psi}(\boldsymbol{x})^\top\boldsymbol{\psi}(\boldsymbol{y})$, 计算核矩阵 $\boldsymbol{K}\in\mathbb{R}^{m\times m}$ 满足 $\boldsymbol{K}_{ij}=k(\boldsymbol{x}_i,\boldsymbol{x}_j)$, 并将此核矩阵中心化为 $\boldsymbol{K}_c=\boldsymbol{K}-\frac{1}{m}\boldsymbol{K}\boldsymbol{1}-\frac{1}{m}\boldsymbol{1}\boldsymbol{K}+\frac{1}{m^2}\boldsymbol{1}\boldsymbol{K}\boldsymbol{1}$, 其中 $\boldsymbol{1}\in\mathbb{R}^{m\times m}$ 为分量全为 1 的矩阵. 进一步求出 $\boldsymbol{K}_c$ 的前 p 个最大特征值 λ_i 及其对应的正交特征向量 $\boldsymbol{u}_i(i=1,2,\cdots,p)$. 对 $\boldsymbol{u}_i$ 进行标准化使其满足 $\lambda_i\boldsymbol{u}_i^\top\boldsymbol{u}_i=1(i=1,2,\cdots,p)$ 来构造投影矩阵 $\boldsymbol{U}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p]$. 对任一样本 $\boldsymbol{x}$, 其投影特征为 $\boldsymbol{y}=\boldsymbol{U}^\top\boldsymbol{K}_{\boldsymbol{x}}$, 其中列向量 $\boldsymbol{K}_{\boldsymbol{x}}=[k(\boldsymbol{x}_1,\boldsymbol{x}),\cdots,k(\boldsymbol{x}_m,\boldsymbol{x})]^\top$. KPCA 的计算复杂度是 $\mathcal{O}(m^3)$, 仅与输入样本个数 m 密切相关.

2.3 线性 2 维主成份分析 (2DPCA)

PCA 和 KPCA 算法进行人脸识别时需要将二维图像矩阵拉成向量, 破坏了图像的二维结构, 图像重建效果较差. Yang 等 [5] 提出了 2DPCA 算法: 设 $\boldsymbol{A}_i\in\mathbb{R}^{s\times t}$ $(i=1,2,\cdots,m)$ 是训练集中的图像矩阵, 直接计算 $\boldsymbol{S}=\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{A}_i-\bar{\boldsymbol{A}})(\boldsymbol{A}_i-\bar{\boldsymbol{A}})^\top$ 前 p 个最大非零特征值及对应的单位正交特征向量为 $\boldsymbol{u}_i$ $(i=1,2,\cdots,p,p\le t)$, 得到投影矩阵 $\boldsymbol{W}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_p]\in\mathbb{R}^{s\times p}$, 其中 $\bar{\boldsymbol{A}}$ 为所有训练图像的均值. 对任一图像样本 $\boldsymbol{A}$, 其投影特征 $\boldsymbol{Y}=\boldsymbol{W}^\top\boldsymbol{A}$.

较 PCA 算法, 2DPCA 不需要将图像矩阵拉成向量, 能很好保留原始图像矩阵的 2D 信息, 且由原二维图像直接构造协方差矩阵, 维数仅为图像的行数或列数, 规模较小, 与样本个数无关, 可以高效计算特征向量来抽取图像的线性特征. 实验表明, 2DPCA 较 PCA 有较高的识别率以及较好的重建效果 [5]. 其主要运算量是求 $\boldsymbol{S}$, 计算复杂度仅为 $\mathcal{O}(ms^2t)$.

2.4 非线性核 2 维主成份分析 (K2DPCA)

Nhat 等 [14] 将 2DPCA 方法推广到非线性情形, 提出 K2DPCA, 进行非线性特征抽取. 较 KPCA 方法, K2DPCA 方法也不需将图像矩阵拉成向量, 而是将训练样本图像矩阵按行 (或列) 进行非线性映射. 对训练图像 $\boldsymbol{A}_i\in\mathbb{R}^{s\times t}$ $(i=1,2,\cdots,m)$, K2DPCA 本质上相当于对 $[\boldsymbol{A}_1\ \boldsymbol{A}_2\ \cdots\ \boldsymbol{A}_m]\in\mathbb{R}^{s\times mt}$ (即 mt 个 s 维训练样本) 或 $[\boldsymbol{A}_1^\top\ \boldsymbol{A}_2^\top\ \cdots\ \boldsymbol{A}_m^\top]\in\mathbb{R}^{t\times ms}$ (即 ms 个 t 维训练样本) 来执行 KPCA. 文献 [14] 在中小型数据集上的实验表明, K2DPCA 的识别率较 PCA/2DPCA/KPCA 有明显提高. 但 K2DPCA 需要计算核矩阵 $\boldsymbol{K}\in\mathbb{R}^{ms\times ms}$ 或 $\boldsymbol{K}\in\mathbb{R}^{mt\times mt}$ 及其特征值分解, 计算复杂度为 $\mathcal{O}(\min\{m^3t^3,m^3s^3\})$. 如当训练集为 160 个 $112*92$ 图像时, 至少需要计算核矩阵 $\boldsymbol{K}\in\mathbb{R}^{14720\times14720}$ 及其特征向量, 显然计算复杂度过大.

针对这一缺点, [15] 提出了分块计算核矩阵及其特征向量的方法, 通过计算 s (或 t) 个 $\mathbb{R}^{m\times m}$ 中的核矩阵及其特征向量, 来求出投影矩阵. 该技巧节省了核矩阵存储空间, 也缩短了训练时间. 但其本质是仅考虑了原核矩阵 $\boldsymbol{K}$ 的分块对角矩阵, 而省略了其他部分, 分类效果提高有限. 文献 [16] 首先对图像矩阵按列分块, 再对每一小块构造核矩阵, 计算核矩阵特征向量, 进而求出每一小块的投影方向, 同样将测试图像分块投影. 此方法降低了计算的复杂度, 但没有利用图像的全局信息. 文 [17] 是将图像矩阵先分块, 将每一块作为一行, 重新构造图像矩阵, 再执行 K2DPCA 方法. 此方法也降低了计算复杂度, 且提高了识别率, 但分块重排也一定程度破坏了图像本身的二维结构.

由于 K2DPCA 的计算复杂度依赖于样本个数与图像矩阵行数 (或列数), 特别当训练样本个数很大时, 上述几种修正方法仍存在存储需求过大的问题. 本文提出有效 Cholesky 分解方法来计算大规模核矩阵 $\boldsymbol{K}$ 的低秩近似分解, 实现大规模核矩阵的情形下的非线性主成分提取, 提高人脸识别的准确率.

3 基于非完全 Choesky 分解的非线性 PCA

为了克服 K2DPCA 处理大样本数据训练时面临的计算复杂度、存储复杂度过大的问题, 本节提出利用

Cholesky 分解得到大规模核矩阵的低秩近似, 来降低相关算法的计算/存储复杂度, 有效实现大规模样本的 KPCA 或 K2DPCA 训练.

3.1 核矩阵的低秩近似 —— 选主元的 Cholesky 分解方法

针对有 M 个样本的训练问题 (KPCA 时 $M=m$, 而 K2DPCA 方法 $M=ms$ 或 mt), 为了避免计算整个核矩阵, 文献 [18–20] 等研究了半正定核矩阵 $\boldsymbol{K}$ 的 Nyström 近似, 通过随机选择等方法确定基本集 $\mathbb{B}\subset\mathbb{M}:=\{1,2,\cdots,M\}$, 得到 $\boldsymbol{K}$ 的低秩近似矩阵 $\widetilde{\boldsymbol{K}}=\boldsymbol{K}_{\mathbb{MB}}\boldsymbol{K}_{\mathbb{BB}}^{-1}\boldsymbol{K}_{\mathbb{MB}}^{\top}$, 其中 $\boldsymbol{K}_{\mathbb{MB}}$ 为核矩阵 $\boldsymbol{K}$ 的相应于基本集 $\mathbb{B}$ 中的列形成的子矩阵, 而 $\boldsymbol{K}_{\mathbb{BB}}$ 为 $\boldsymbol{K}$ 对应于标号集 $\mathbb{B}$ 中的行和列组成的子矩阵. 这些随机选择集 $\mathbb{B}$ 的方法不能保证得到 $\boldsymbol{K}$ 的最优秩 r 低秩近似.

若令 $\boldsymbol{L}=\boldsymbol{K}_{\mathbb{MB}}\boldsymbol{K}_{\mathbb{BB}}^{-\frac{1}{2}}$, 则 $\widetilde{\boldsymbol{K}}=\boldsymbol{L}\boldsymbol{L}^{\top}$. 本文采用选主元的 Cholesky 方法迭代确定基本集 $\mathbb{B}$, 得到 $\boldsymbol{L}$ 以降低存储复杂度, 并且确保得到的 $\widetilde{\boldsymbol{K}}$ 是在矩阵迹范数意义下 $\boldsymbol{K}$ 的最优秩 r 低秩近似. 具体过程如下:

设第 i 次的迭代时得到近似矩阵 $\widetilde{\boldsymbol{K}}^i=\boldsymbol{L}^i{\boldsymbol{L}^i}^{\top}$, 记相应的误差矩阵为 $\boldsymbol{E}^i=\boldsymbol{K}-\widetilde{\boldsymbol{K}}^i$. 不妨设此时前 i 个指标构成基本集 $\mathbb{B}_i$(可通过改变训练样本的输入次序实现), 其余指标集记为 $\mathbb{N}_i$, 则 $\boldsymbol{E}^i$ 具有形式 $\begin{bmatrix}0&0\\0&\boldsymbol{T}\end{bmatrix}$, 其中半正定矩阵 $\boldsymbol{T}=\boldsymbol{K}_{\mathbb{N}_i\mathbb{N}_i}-\boldsymbol{K}_{\mathbb{N}_i\mathbb{B}_i}\boldsymbol{K}_{\mathbb{B}_i\mathbb{B}_i}^{-1}\boldsymbol{K}_{\mathbb{N}_i\mathbb{B}_i}^{\top}$ 是 $\boldsymbol{K}_{\mathbb{B}_i\mathbb{B}_i}$ 关于 $\boldsymbol{K}$ 的 Schur 补 [21−22], 故 $\boldsymbol{E}^i$ 也是一个半正定矩阵. 由矩阵理论 [21−22], 对半正定矩阵 $\boldsymbol{E}^i$ 有: $\max_{j,k}\boldsymbol{E}^i_{j,k}=\max_j\boldsymbol{E}^i_{j,j}$ 且 $\boldsymbol{E}^i=\boldsymbol{0}\Leftrightarrow\max_j\boldsymbol{E}^i_{j,j}=0$, 还有

$$\|\boldsymbol{E}^i\|_2\le\|\boldsymbol{E}^i\|_F\le trace(\boldsymbol{E}^i)\tag{1}$$

其中 $trace(\boldsymbol{E}^i)$ 为 $\boldsymbol{E}^i$ 对角元素的和. 这样最大程度减小 $\boldsymbol{E}^i$ 对角线的元素就可以最大程度的极小化 $trace(\boldsymbol{E}^i)$, 从而减小了近似误差.

本文算法思想: 从空集 $\mathbb{B}_0$ 出发, 将 $\boldsymbol{E}^i$ 的对角线上最大元素对应的列号加到基本集 $\mathbb{B}_i$ 中得到 $\mathbb{B}_{i+1}$, 来构造 $\boldsymbol{K}$ 的近似矩阵 $\widetilde{\boldsymbol{K}}_{i+1}=\boldsymbol{K}_{\mathbb{MB}_{i+1}}\boldsymbol{K}_{\mathbb{B}_{i+1}\mathbb{B}_{i+1}}^{-1}\boldsymbol{K}_{\mathbb{MB}_{i+1}}^{\top}$, 这样 $\boldsymbol{E}^{i+1}=\boldsymbol{K}-\widetilde{\boldsymbol{K}}_{i+1}$ 就是当前具有最小迹范数的误差矩阵. 由 (1) 式, 这同时也减小了误差矩阵的其它范数的上界. 选择 $\boldsymbol{E}^i$ 对角线上最大元素的方法本质上等价于文献 [21, 23] 的选主元技巧.

由于矩阵迹运算具有线性性质, 故该方法得到秩 r 的 Nyström 型低秩近似是矩阵迹范数意义下最优近似, 即在迹范数意义下, 最终得到的近似误差矩阵 $\boldsymbol{E}$ 的迹不超过其它任何秩 r 的这种 Nyström 型近似的误差矩阵的迹.

具体实现步骤见算法 1, 这里利用了 [21, 23] 的选主元技巧, 并给出便于核矩阵计算的精巧实现过程.

Algorithm 1 核矩阵的选主元 Cholesky 分解低秩近似算法

输入: 训练数据集 $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M\}$, 及核函数 $k(\boldsymbol{x},\boldsymbol{y})$, 误差上界 ϵ, 低秩上界 r;

输出: 基本集 $\mathbb{B}$ 和低秩矩阵 $\boldsymbol{L}$, 满足 $trace(\boldsymbol{K}-\boldsymbol{L}\boldsymbol{L}^{\top})\le\epsilon$ 或满足低秩限制 $rank(\boldsymbol{L})=r$.

$\mathbb{B}_0:=\emptyset$, $\mathbb{N}_0:=\mathbb{M}$, $\boldsymbol{d}^0:=[k(\boldsymbol{x}_1,\boldsymbol{x}_1),\cdots,k(\boldsymbol{x}_M,\boldsymbol{x}_M)]$, $\delta_0=\|\boldsymbol{d}^0\|_1$; $i:=0$;
while $\delta_i>\epsilon$ 且 $i<r$ **do**
  $j_0=\arg\max_{j\in\mathbb{N}_i}\boldsymbol{d}^i_j$, 计算 $\boldsymbol{k}=[k(\boldsymbol{x}_1,\boldsymbol{x}_{j_0}),\cdots,k(\boldsymbol{x}_M,\boldsymbol{x}_{j_0})]^{\top}$;
  $\mathbb{B}_{i+1}:=\mathbb{B}\cup\{j_0\}$, $\mathbb{N}_{i+1}:=\mathbb{N}_i\setminus\{j_0\}$;
  if $i=0$ **then**
    Set $\boldsymbol{L}^{i+1}:=\boldsymbol{k}/\sqrt{\boldsymbol{k}_{j_0}}$;
  else
    $\boldsymbol{l}:=\frac{1}{\sqrt{\boldsymbol{k}_{j_0}-\boldsymbol{u}^{\top}\boldsymbol{u}}}(\boldsymbol{k}-\boldsymbol{L}^i\boldsymbol{u})$, $\boldsymbol{L}^{i+1}:=\begin{bmatrix}\boldsymbol{L}^i&\boldsymbol{l}\end{bmatrix}$, 其中 $\boldsymbol{u}^{\top}$ 为 $\boldsymbol{L}^i$ 的第 j_0 行;//Cholesky 分解更新
  end if
  $\boldsymbol{d}^{i+1}_j=\boldsymbol{d}^i_j-\boldsymbol{l}_j^2, j\in\mathbb{N}_{i+1}$; $\delta_{i+1}=\sum_{j\in\mathbb{N}_{i+1}}\boldsymbol{d}^{i+1}_j$;
  $i:=i+1$.
end while
return $\mathbb{B}=\mathbb{B}_i$, $\boldsymbol{L}=\boldsymbol{L}^i$.

该算法单次循环的主要计算量是进行 Cholesky 更新时计算 $\boldsymbol{L}^i\boldsymbol{u}$, 其复杂度为 $\mathcal{O}(Mr_i)$, $r_i=1,2,\cdots,r$. 故算法总的计算复杂度为 $\mathcal{O}(Mr^2)$, 且只需计算核矩阵的最多 r 列及对角线上的所有元素, 存储复杂度仅为 $\mathcal{O}(Mr)$.

3.2 低秩近似核矩阵 $\boldsymbol{K}$ 的特征向量计算

利用选主元 Cholesky 分解方法, 可将核矩阵近似分解为:

$$\boldsymbol{K} \approx \boldsymbol{L}\boldsymbol{L}^{\top} \tag{2}$$

其中 $rank(\boldsymbol{K}) = r$ $(r \ll M)$, $\boldsymbol{L} \in \mathbb{R}^{M\times r}$. 而相应的的中心化近似核矩阵为

$$\boldsymbol{K}_c = \boldsymbol{K} - \frac{1}{m}\boldsymbol{K}\mathbf{1} - \frac{1}{m}\mathbf{1}\boldsymbol{K} + \frac{1}{m^2}\mathbf{1}\boldsymbol{K}\mathbf{1} \approx (\boldsymbol{L} - \bar{\boldsymbol{L}})(\boldsymbol{L} - \bar{\boldsymbol{L}})^{\top},$$

其中 $\mathbf{1} \in \mathbb{R}^{M\times M}$ 为分量全为 1 的矩阵, $\bar{\boldsymbol{L}} = \frac{1}{M}\boldsymbol{L}\mathbf{1}$ 是以 $\boldsymbol{L}$ 的行均值向量为列排成的矩阵. 实际计算中, 无需直接计算大规模矩阵 $\boldsymbol{K}$ 或 $\boldsymbol{K}_c$ 及其特征向量. 若令 $\boldsymbol{P} = \boldsymbol{L} - \bar{\boldsymbol{L}}$, 可通过计算 $\boldsymbol{P}^{\top}\boldsymbol{P} \in \mathbb{R}^{r\times r}$ 的特征值和特征向量来得到投影矩阵. 具体地, 若 $\boldsymbol{P}^{\top}\boldsymbol{P}\boldsymbol{u} = \lambda\boldsymbol{u}$, 则 λ 必为 $\boldsymbol{K}_c$ 的特征值, 且相应的特征向量为 $\boldsymbol{P}\boldsymbol{u}$. 这样通过计算 $\boldsymbol{P}^{\top}\boldsymbol{P}$ 的前 p 个最大特征值相应的特征向量, 相应得到 $\boldsymbol{K}_c$ 的前 p 个最大的特征值及其特征向量, 再进行适当的标准化, 得出特征投影矩阵.

综上, 本文提出基于选主元 Cholesky 分解的 K2DPCA 算法, 记为 Chol+K2DPCA. 同样若 KPCA 的核矩阵规模过大时, 该方法也可用来该提高其计算效率, 相应算法记为 Chol+KPCA.

4 实验结果及分析

本节通过实验比较相关算法, 包括 KPCA/K2DPCA 和提出的方法 Chol+KPCA 以及 Chol+K2DPCA, 其中 KPCA 采用文 [6–7] 的典型实现, 而 K2DPCA 采用文献 [14] 的算法. 所有程序运行在 CPU 为 Intel Core i3 2100GHz、内存总共 4.00GB 的普通计算机上, 操作系统为 Win7, 编程环境为 Matlab2010b. 对于所有人脸识别实验, 使用基于欧氏距离的 1-NN 算法来分类.

4.1 中小规模训练数据的去噪效果比较

本实验比较 KPCA, K2DPCA 与提出的 Chol+K2DPCA 在 ORL 标准人脸数据库数据库上的实验分析, 选用高斯核 $k(\boldsymbol{x},\boldsymbol{y}) = \exp(-\gamma\|\boldsymbol{x}-\boldsymbol{y}\|^2)$. 由于核参数选择不是本文的重点, 故核参数 γ 采用简单网格法确定, 即对不同方法选择 $\{2^{-12}, 2^{-11}, \cdots, 2^{1}\}$ 中效果最优的参数. 针对该数据库, K2DPCA 方法 $\gamma = 2^{-1}$, 而 KPCA 算法 $\gamma = 2^{-11}$.

ORL 数据库包含 40 个人, 每人包含 10 张图像. 部分图像是在不同时间拍摄的, 面部表情和细节都有所不同. 所有灰度图像分辨率为 112×92, 这里为了计算方便, 将图像下采样为分辨率 $28 * 23$ 的灰度图. 实验中随机选取每个人的四张图片做训练, 剩余六张图片做测试, 即每次实验 160 张训练图像, 240 张测试图像. 这时 KPCA 的输入为 160 个 644 维的训练数据, 仅需计算 $\mathbb{R}^{160\times 160}$ 中的核矩阵, 故无需 Cholesky 分解; 而 K2DPCA 输入为 $M = 160 \times 23 = 3680$ 个 28 维的训练数据, 需要计算 $\mathbb{R}^{3680\times 3680}$ 中的核矩阵及其特征值分解, 可以直接实现 K2DPCA, 也可以采用本文分解算法提高其的效率.

同时考虑给人脸数据加上不同的噪声, 来比较各算法对噪声的适应性. 本文考虑两种噪声: 一种均值为 0, 方差为不同参数 v 的高斯噪声; 一种为密度为不同参数 v 的椒盐噪声 (salt & pepper), 加噪过程采用 Matlab 中的 imnoise 函数实现. 实验中原图及加噪示例如图 1.

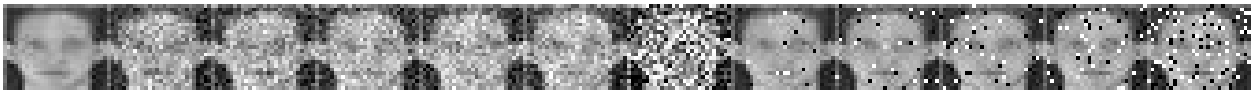


图 1 依次为原图, 参数 $v = 0.01, 0.02, 0.03, 0.04, 0.05, 0.08$ 的高斯噪声和 $v = 0.02, 0.04, 0.07, 0.09, 0.15$ 的椒盐噪声图像

实验所有结果均为十次随机实验的平均值, 且表格数据给出均方差.

实验 1 考虑 Cholesky 分解算法中参数 r 的选取. 进行 Cholesky 分解时, 一般不必直到满足 $trace(\boldsymbol{K} - \boldsymbol{L}\boldsymbol{L}^{\top}) \le \epsilon$ (ϵ 为很小的误差界), 故通常用核矩阵低秩近似的秩 r 终止分解算法. 这样有必要分析秩 r 对算法效果的影响. 本实验针对原始无噪声和加参数 $v = 0.03$ 高斯噪声的 ORL 数据, 分析当主成分个数 $p = 10$ 给定时, Cholesky 分解时的核矩阵的列数 r 对 Chol+K2DPCA 算法检测效果的影响, 结果如图 2 所示. 图 2 中将 KPCA 算法 ($p = 100$) 和 K2DPCA 算法 ($p = 10$) 在同样随机选择的数据集上的结果也列出来参照, 由于这两个算法与 r 无关, 用水平线表示.

图 2 表明: 1) 无论是否加噪声, K2DPCA 和 Chol+K2DPCA 算法明显优于 KPCA 方法, 和文献 [14–17] 的结果一致; 2) 无噪声情形, 随着 r 增加, 近似精度提高, Chol+K2DPCA 和 K2DPCA 趋于一致; 3) 有噪声

情形, 提出的 Chol+K2DPCA 算法明显优于 K2DPCA 和 KPCA, 这说明提出的分解方法对噪声具有鲁棒性 (尽管有小幅波动); 4) 针对 Chol+K2DPCA, 算法测试识别率随着低秩近似的秩 r 增加而增加, 而一般在秩 r 约为主成分 p 的 10~30 倍左右时趋于稳定. 下文实验中分解算法参数 $r = 20p$.

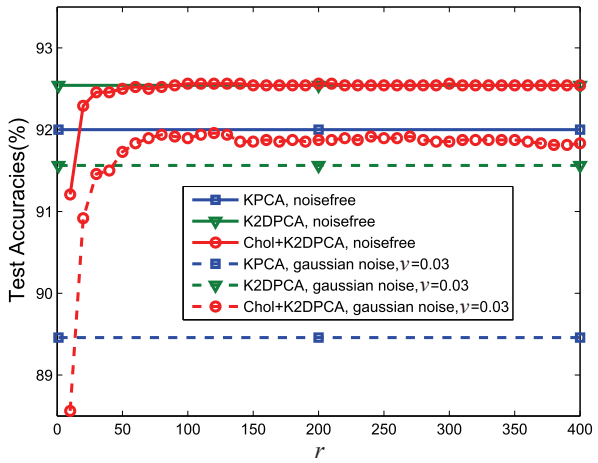


图 2　Chol+K2DPCA 算法参数 r 对测试识别率影响示意图. 实线为无噪声情形、虚线为参数 $v = 0.03$ 的高斯噪声情形. 原 KPCA 和 K2DPCA 用水平线表示

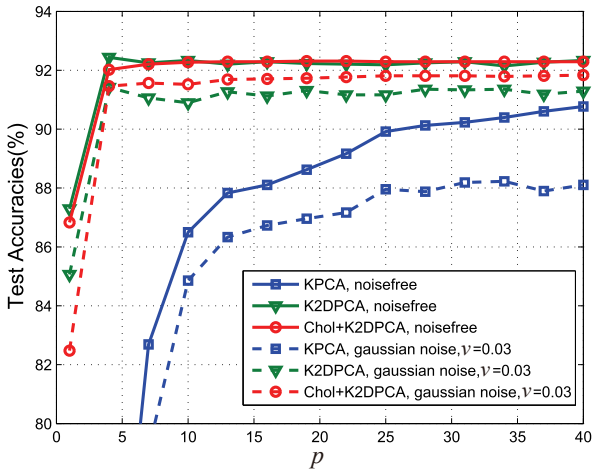


图 3　KPCA, K2DPCA 及 Chol+K2DPCA 算法的主成分个数 p 与测试识别率变化关系示意图 ($r = 15p$). 实线为无噪声情形、虚线为高斯噪声参数 $v = 0.03$ 的情形

实验 2 分析主成分的个数 p 对实验效果的影响. 对相关算法 KPCA, K2DPCA 和 Chol+K2DPCA, 实验分析当主成分个数 p 变化时, 测试识别率的变化情况, 结果如图 3 所示.

由图 3 可看出, 无论是否加入噪声, 当选取的主成分个数相同时, Chol+K2DPCA 与 K2DPCA 算法识别率相近, 且显然均高于 KPCA 算法识别率. 而且随着主成分的增加, Chol+K2DPCA 与 K2DPCA 的识别率均逐渐增大, 都在 $p = 20$ 左右时都趋于稳定值, 故而下文实验选取 $p = 20$. 而 KPCA 的结果还有上升的趋势, 下文实验中主成分个数取更大值 $p = 100$. 通过对比实线与虚线, 也可以看出分解方法有较好的抗噪效果.

实验 3 采取上文选定的相关参数, 分析比较在原图像数据以及加上不同类别、不同参数的噪声后, KPCA/K2DPCA 方法与提出 Chol+K2DPCA 算法的识别率. 其中 K2DPCA 与 Chol+K2DPCA 算法 $p = 20$, 且 Chol+K2DPCA 算法中 $r = 20p = 400$, KPCA 算法取 $p = 100$, 结果如表 1 所示. 所有结果均为 10 次随机实验的平均值, 括号里的数字为均方差. 由于是否加噪声不影响算法的计算复杂度, 相应于表 1, 算法 KPCA、K2DPCA 和 Chol+K2DPCA 的平均训练时间分别为 0.037 秒、2.83 秒和 0.86 秒.

表 1　KPCA, K2DPCA 和 Chol+K2DPCA 在 ORL 数据集上测试识别率比较, 其中每人随机选取 4 张图像用来训练, 其余 6 张用来测试; 所有结果均为 10 次随机实验的平均值, 括号里的数字为均方差. 相应平均训练时间 KPCA 为 0.037 秒, K2DPCA 为 2.83 秒, Chol+K2DPCA 为 0.86 秒

	KPCA	K2DPCA	Chol+K2DPCA
无噪声 ORL 数据集	91.56(2.35)%	91.96(2.34)%	**92.56**(2.34)%
高斯噪声, $v = 0.01$	90.31(1.62)%	91.25(1.97)%	**91.67**(1.94)%
高斯噪声, $v = 0.02$	89.25(2.26)%	91.13(1.81)%	**91.44**(1.99)%
高斯噪声, $v = 0.03$	89.29(1.97)%	90.92(1.70)%	**92.19**(1.67)%
高斯噪声, $v = 0.04$	90.25(1.57)%	91.52(1.86)%	**92.38**(1.99)%
高斯噪声, $v = 0.05$	90.02(2.54)%	90.92(2.25)%	**91.69**(2.30)%
高斯噪声, $v = 0.08$	52.17(4.15)%	65.87(2.76)%	**77.21**(2.71)%
椒盐噪声, $v = 0.02$	88.79(1.92)%	90.48(1.52)%	**90.98**(1.67)%
椒盐噪声, $v = 0.04$	84.42(3.07)%	89.13(2.31)%	**91.02**(2.14)%
椒盐噪声, $v = 0.07$	73.67(2.78)%	83.44(2.81)%	**88.85**(2.38)%
椒盐噪声, $v = 0.09$	70.60(3.39)%	83.17(2.17)%	**88.00**(2.32)%
椒盐噪声, $v = 0.15$	51.94(4.32)%	70.00(3.56)%	**83.77**(2.85)%

从表 1 可以看出, 在无噪声情形下, 两种 2D 方法平均识别率较 KPCA 稍优; 在加入两种类型的噪声后, 利用本文方法的 Chol+K2DPCA 算法识别率明显高于原 K2DPCA 算法识别率, 两者均高于 KPCA 算法识

别率. 且噪声越大, 提出算法的优势越明显. 而在同一类噪声下, 随着噪声参数的增大, 三种算法识别率都逐渐减小, 但采用本文选主元的 Cholesky 技术的方法, 识别率下降幅度小, 这说明该方法具有一定的抗噪性能. 同时注意到 chol+K2DPCA 算法较原 K2DPCA 算法, 在测试识别率相当或更好的情形下, 节省了约 2/3 训练时间.

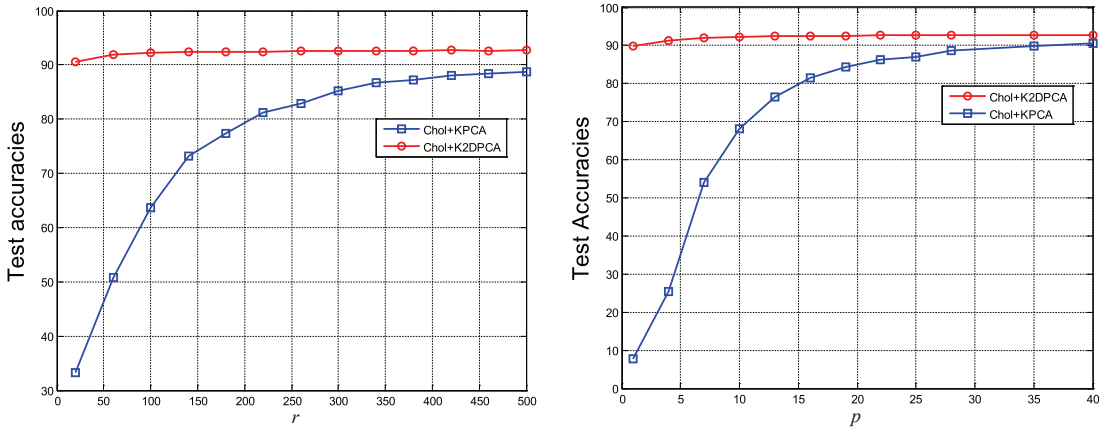
4.2 大规模数据实验

本实验选取 Extended YaleB 大型人脸数据库, 说明提出算法对大规模数据训练的有效性.

Extended YaleB 人脸数据库 [24][1]包括 28 个人的, 每人不同表情和光照下的 576 张人脸图像, 共 16128 张图像, 每张分辨率为 $192 * 168$, 实验中将每张图像下采样为分辨率 $32 * 28$. 仍选用高斯核函数 $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\gamma\|\boldsymbol{x} - \boldsymbol{y}\|^2)$, 算法 KPCA 和 K2DPCA 相应的核参数分别为 $\gamma = 2^{-4}$ 与 2^{-1}.

在实验中, 随机选取每个人的 300 张图像训练, 200 张测试, 即训练图像样本总数为 $8400(= 300 * 28)$ 张, KPCA 需要计算核矩阵 $\boldsymbol{K} \in \mathbb{R}^{8400\times 8400}$ 及其特征向量, 而 K2DPCA 需要计算核矩阵 $\boldsymbol{K} \in \mathbb{R}^{268800\times 268800}$ 及其特征向量, 这都超出了实验所用电脑的计算和存储能力. 但是采用本文提出的选主元的 Cholesky 分解方法, 这两种算法都可以快速有效实现训练.

首先分析识别率随 r 和 p 的变化情况, 结果如图 4 所示: 对 Chol+KPCA 和 Chol+K2DPCA 算法, 图 4(a) 描述当主成分个数 $p = 20$ 给定时, 算法识别率随核矩阵低秩近似的秩 r 变化情况, 图 4(b) 表示识别率随主成分个数 p 的变化曲线 (此时 $r = 15p$).



(a) (b)

图 4 测试识别率随核矩阵低秩近似的秩 r 和分解主成分个数 p 的变化示意图. (a) 测试识别率关系随 Cholesky 分解时低秩近似的秩 r 变化示意图; (b) 测试识别率随 PCA 主成分个数 p 变化示意图

由图 4 可以看出, 在同样条件下, Chol+K2DPCA 算法识别率均高于 Chol+KPCA 的识别率, 和前文实验结果一致. 由图 4(a) 看出, 当主成分个数 p 给定时, 算法的识别率随 Cholesky 方法中分解核矩阵列数 r 增加而增加, 且当 $r \geq 300 = 15p$ 时, Chol+K2DPCA 算法测试识别率均趋于稳定, 且图 4(b) 中当 $p \geq 20$ 时算法性能基本趋于稳定, 故下文实验中对 Chol+K2DPCA 设置参数 $r = 15p$ 且 $p = 20$. 而对 Chol+KPCA 算法随 r、p 增加, 算法识别率虽还有提高的趋势, 但出于计算复杂度考虑, 下文对 KPCA 设置参数 $r = 20p$ 且 $p = 40$. 表 2 给出采用本文提出方法在大规模 Extended YaleB 数据库上的实验结果.

表 2 两种算法在对规模数据集上的识别率和训练时间比较. 所有结果为 10 次随机实验的平均值, 括号里为均方差

	Chol+KPCA	Chol+K2DPCA
训练图像个数	8,400	8,400
核矩阵大小	8,400	268,800
测试识别率	91.12(0.13)%	92.67(0.06)%
训练时间	24.90(9.13) 秒	80.95(0.18) 秒

表 2 表明, Chol+K2DPCA 与 Chol+KPCA 两种算法, 由于采用本文的快速选主元 Cholesky 分解方法, 均可在较短时间内实现大规模数据的训练, 有效进行人脸识别检测. 所提方法拓宽了 KPCA/K2DPCA 等特征抽取方法的适用范围.

5 结束语

本文提出了一种基于 Cholesky 分解低秩近似的核主成分提取算法, 有效地得到大规模核矩阵的迹范数

1. 详细参阅: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.

意义下的最佳 Nyström 型低秩近似. 实验结果也表明, 提出算法不仅对于中小规模训练集有一定去噪效果, 而且可有效提取大规模训练样本非线性特征, 弥补了 KPCA 与 K2DPCA 算法对大规模训练数据因核矩阵过大而无法有效训练的缺点. 但研究中发现, K2DPCA 用于训练大规模数据时还面临着检测时复杂度过大的问题. 若待测试图像 $\boldsymbol{A} \in \mathbb{R}^{s\times t}$, 训练 K2DPCA 得到投影矩阵 $\boldsymbol{U} \in \mathbb{R}^{ms\times p}$ 后, 需计算测试核矩阵 $\boldsymbol{K} \in \mathbb{R}^{ms\times s}$, 再进行投影 $\boldsymbol{Y} = \boldsymbol{U}^{\top}\boldsymbol{K}$ 提取特征 $\boldsymbol{Y}$, 此时 $\boldsymbol{K}$ 为非对称的, 所提出的分解方法不适用, 故测试计算效率很低. 如何提高 K2DPCA 测试阶段的计算效率还是有待研究的问题.

参考文献

[1] Jolliffe I T. Principal component analysis[M]. 2nd ed. New York: Springer, 2002.

[2] 常雷雷, 李孟军, 鲁延京, 等. 基于主成分分析的置信规则库结构学习方法 [J]. 系统工程理论与实践, 2014, 34(5): 1297–1304.
Chang L L, Li M J, Lu Y J, et al. Structure learning for belief rule base using principal component analysis[J]. Systems Engineering — Theory & Practice, 2014, 34(5): 1297–1304.

[3] 刘志强, 吕学, 张利. 基于多分类 GA-SVM 的高速公路 AID 模型 [J]. 系统工程理论与实践, 2013, 33(8): 2110–2115.
Liu Z Q, Lü X, Zhang L. Highway automatic incident detection based on multi-class classification and GA-SVM[J]. Systems Engineering — Theory & Practice, 2013, 33(8): 2110–2115.

[4] 郁雪, 李敏强. 基于 PCA-SOM 的混合协同过滤模型 [J]. 系统工程理论与实践, 2010, 30(10): 1850–1854.
Yu X, Li M Q. Effective hybrid collaborative filtering model based on PCA-SOM[J]. Systems Engineering — Theory & Practice, 2010, 30(10): 1850–1854.

[5] Yang J, Zhang D, Frangi A F, et al. Two-dimensional PCA: A new approach to appearance based face representation and recognition[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2004, 26(1): 131–137.

[6] Scholkopf B, Smola A J, Muller K R. Nonlinear component analysis as a kernel eignvalue problem[J]. Neural Computation, 1998, 10(5): 1299–1319.

[7] Smola A J, Schökopf B. Learning with kernels[M]. Massachusetts: The MIT Press, 2002.

[8] Zheng W M, Zou C R, Zhao L. An improved algorithm for kernel principal component analysis[J]. Neural Processing Letters, 2005, 2(2): 49–56.

[9] Vapnik V N. An overview of statistical learning theory[J]. IEEE Trans on Neural Network, 1999, 10(5): 988–999.

[10] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 2000.

[11] 邓乃扬, 田英杰. 数据挖掘中的新方法 —— 支持向量机 [M]. 北京: 科学出版社, 2004.
Deng N Y, Tian Y J. New methods of data mining-support vector machines[M]. Beijing: Science Press, 2004.

[12] 余乐安, 汪寿阳. 基于核主元聚类的股票分类 [J]. 系统工程理论与实践, 2009, 29(12): 1–8.
Yu L A, Wang S Y. Kernel principal component clustering methodology for stock categorization[J]. Systems Engineering — Theory & Practice, 2009, 29(12): 1–8.

[13] 蒋少华, 桂卫华, 阳春华, 等. 基于核主元分析与多支持向量机的监控诊断方法及其应用 [J]. 系统工程理论与实践, 2009, 29(9): 153–159.
Jiang S H, Gui W H, Yang C H, et al. Monitoring model based on kernel principal component analysis and multiple support vector machines and its application[J]. Systems Engineering — Theory & Practice, 2009, 29(9): 153–159.

[14] Nhat D M, Lee S Y. Kernel-based 2DPCA for face recognition[J]. IEEE Symposium Signal Processing and Information Technology, 2007, 12(15): 35–39.

[15] Sun N, Wang H, Ji Z, et al. An efficient algorithm for kernel two-dimensional principal component analysis[J]. Neural Computation and Application, 2008, 17(1): 59–64.

[16] Eftekhari A, Forouzanfar M, Moghaddam H A, et al. Block-wise 2D kernel PCA/LDA for face recognition[J]. Information Processing Letters, 2010, 110(17): 761–766.

[17] Wang L, Zhou X. Approximation kernel 2DPCA by mixture of vector and matrix representation[J]. Computation Intelligence and Security, 2011, 12(3): 1298–1302.

[18] Smola A J, Schökopf B. Sparse greedy matrix approximation for machine learning[C]// Proceedings of the Seventeenth International Conference on Machine Learning (ICML'00), San Francisco, CA, USA, 2000: 911–918.

[19] Drineas P, Mahoney M W. On the Nyström method for approximating a Gram matrix for improved kernel-based learning[J]. Journal of Machine Learning Research, 2005, 6: 2153–2175.

[20] Higham N. Analysis of the Cholesky decomposition of a semi-definite matrix[C]// Cox M G, Hammarling S J. Reliable Numerical Computation. Oxford: Oxford University Press, 1990: 161–185.

[21] Golub G H, Loan C F V. Matrix computations[M]. Maryland: The John Hopkins University Press, 1996.

[22] 徐树方. 矩阵计算的理论与方法 [M]. 北京: 北京大学出版社, 1995.
Xu S F. Theory and method of matrix computation[M]. Beijing: Peking University Press, 1995.

[23] Harbrecht H, Peters M, Schneider R. On the low-rank approximation by the pivoted Cholesky decomposition[J]. Applied Numerical Mathematics, 2012, 62(4): 428–440.

[24] Georghiades A S, Belhumeur P N, Kriegman D J. From few to many: Illumination cone models for face recognition under variable lighting and pose[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(6): 643–660.