

文章编号: 1001-0920(2007)09-1044-05

一种 KPCA 的快速算法

赵峰^a, 张军英^{a,b}

(西安电子科技大学 a. 计算机学院, b. 雷达信号处理国家重点实验室, 西安 710071)

摘要: 针对大训练集时核主分量分析(KPCA)的计算代价大、特征提取速度慢等问题,提出一种 KPCA 的快速算法. 该算法通过训练样本在特征空间所张成的子空间的一组标准正交基,将训练集上的 KPCA 过程,转化为以所有核训练样本在这组基下的坐标为数据集的 PCA 过程. 其求解过程只需特征值分解一个阶数等于基的个数的矩阵,而且对某样本进行特征提取时,只需计算该样本与构成这组基的样本间的核函数. 实验结果验证了该算法的有效性.

关键词: 核主分量分析; 标准正交基; 核矩阵

中图分类号: TP391.4 **文献标识码:** A

Fast algorithm about KPCA

ZHAO Feng^a, ZHANG Jun-ying^{a,b}

(a. School of Computer Science and Engineering, b. National Key Lab of Radar Signal Processing, Xidian University, Xi'an 710071, China. Correspondent: ZHAO Feng, E-mail: zhaofeng1016@126.com)

Abstract: The kernel principal component analysis (KPCA) suffers from the large computation complexity and the slow speed of feature extraction for the case of large number of training samples. Therefore, a fast algorithm of KPCA is presented. Based on a orthonormal basis of the sub-space spanned by the training samples mapped onto the feature space, the algorithm translates the processing of KPCA on the training samples into that of PCA on the data which consists of the coordinate values of all the kernel training samples under the orthonormal basis. The proposed algorithm only needs the eigenvalue decomposition of a matrix whose size is equal to the number of the orthonormal basis. Moreover the feature extraction for one sample only needs to calculate the kernel functions between the sample and the training samples which compose the orthonormal basis. The experimental results show the effectiveness of the presented algorithm.

Key words: Kernel principal component analysis (KPCA); Orthonormal basis; Kernel matrix

1 引言

核主分量分析^[1] (KPCA) 作为主分量分析 (PCA) 的一种非线性推广,能有效捕捉数据的非线性特征,因此在模式识别、回归分析等领域受到广泛重视与应用^[2-4]. 但在 KPCA 的求解过程中,需要特征值分解一个 $M \times M$ 的核矩阵 (M 表示训练样本数),其计算复杂度是 $O(M^3)$,且对某样本进行特征提取时,需要计算该样本与所有训练样本间的核函数. 因此 M 大时, KPCA 面临计算代价大、特征提取效率低的问题. 针对这些问题,文献[5]将求解 PCA 的期望最大算法 (EM) 推广到求解 KPCA. 但该算法的缺陷是占用的存储空间较大,并且算法的收敛性不能保证. 文献[6]将训练集分成若干子集,然后

在特征空间中,将总体协方差矩阵用一些特殊向量近似表示,其中这些特殊向量是各个子集的协方差矩阵的一些较大特征值所对应的特征向量. 基于这种近似表示,使得 KPCA 在求解过程中,只需对一个阶数等于这些特殊向量数目的核矩阵进行特征值分解. 但关于训练集分成多少个子集与选取多少个特征向量作为特殊向量的问题,文献并没从理论上给出一个界定,因此算法本身有很大的不确定性. 另外,上述两种算法在 KPCA 的求解中,尽管都不同程度地降低了 KPCA 的计算复杂度,但它们都没有涉及如何改善测试阶段的特征提取速度问题.

本文基于训练集在特征空间中的空间结构,提出一种 KPCA 的快速算法. 该算法主要分两步:1)

收稿日期: 2006-06-05; 修回日期: 2006-08-29.

基金项目: 国家自然科学基金项目(60371044, 60574039); 十五国防预研项目(413070501).

作者简介: 赵峰(1974—),男,济南人,博士,从事智能信息处理、雷达目标识别等研究; 张军英(1961—),女,西安人,教授,博士生导师,从事人工神经网络、智能信息处理等研究.

基于线性相关理论,设计出一种优化算法,快速确定训练集在特征空间中所张成的子空间的一组基,并由此得到一组标准正交基;2) 将所有训练样本对应的核样本向这组正交基上作投影,得到一个相应的投影向量集,并对投影向量集进行 PCA. 本文从理论上证明了对该投影向量集进行 PCA 等价于对训练集进行 KPCA,但该算法实现过程只需特征值分解一个 $r \times r$ 的矩阵(其中 r 表示基的个数),计算复杂度仅为 $O(r^3)$. 同时,对某样本进行特征提取时,也仅需计算该样本与构成这组正交基的样本间的核函数. 实验结果表明本文算法与 KPCA 特征提取能力相当,但计算复杂度小,特征提取速度快.

2 KPCA 的描述

设 $X = \{x_1, x_2, \dots, x_M\}$ 为 R^n 空间中的一组样本集. KPCA 通过一个非线性映射 ϕ 将样本映射到特征空间 F 中,即 $\phi: X \rightarrow F$ (其中称 $\phi(x)$ 为 x 对应的核样本);然后对数据集 $\phi(X) = \{\phi(x_1), \phi(x_2), \dots, \phi(x_M)\}$ 进行 PCA.

假设核样本已经中心化,即 $\sum_{i=1}^M \phi(x_i) = 0$, 其协

方差矩阵 $C = \frac{1}{M} \sum_{i=1}^M \phi(x_i) \phi(x_i)^T$. 首先需要求解 C 的特征值 λ 及相应特征向量 $v \in F$, 即

$$Cv = \lambda v. \tag{1}$$

根据再生核理论,存在一组系数 $\alpha_i (i = 1, 2, \dots, M)$ 使得

$$v = \sum_{i=1}^M \alpha_i \phi(x_i). \tag{2}$$

通过定义一个 $M \times M$ 的核矩阵

$$K_{i,j} = k(x_i, x_j) = \phi(x_i)^T \phi(x_j),$$

并结合式(1)和(2)得

$$K \alpha = \lambda \alpha, \tag{3}$$

其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)^T$. 所以为确定特征向量 v 而求取系数 $\alpha_i (i = 1, 2, \dots, M)$ 的问题就依赖于特征值分解核矩阵 K .

设 $\lambda_1 > \lambda_2 > \dots > \lambda_k > 0 (k \leq M)$ 为式(3)前 k 非零特征值,对应的特征向量为 $\alpha^1, \alpha^2, \dots, \alpha^k$. 则由式(2)可求得特征空间 F 中的主轴方向 v^1, v^2, \dots, v^k 的表达式.

另外,为归一化主轴方向 v^k , 需要 α^k 满足

$$\sum_{i=1}^M (\alpha^k)^T \alpha^k = 1. \text{ 考虑到 } \sum_{i=1}^M \phi(x_i) = 0 \text{ 的假设, 只需将}$$

核矩阵换成 $\bar{K} = K - I_M K - K I_M + I_M K I_M^{[1]}$, 其中 I_M 为 $M \times M$ 矩阵,且 $(I_M)_{i,j} = 1/M$.

由式(2)知,对任一测试样本 x , 它在 F 中的第 k 个主分量为

$$d_k(x) = (\alpha^k)^T \bar{k}_x. \tag{4}$$

其中

$$\begin{aligned} \bar{k}_x &= k_x - K I_M - I_M k_x + I_M K I_M, \\ k_x &= (k(x_1, x), k(x_2, x), \dots, k(x_M, x))^T, \\ I_M &= \underbrace{(1/M, 1/M, \dots, 1/M)}_M \end{aligned}$$

3 标准正交基的确定

事实上,数据集 $\phi(X)$ 在 F 中所张成的子空间 $\{\phi(x_i)\}_{i=1}^M$ 的空间结构可由该空间的一组基来捕获^[7],而且 $\{\phi(x_i)\}_{i=1}^M$ 的维数(即基的个数)等于核矩阵 K 的秩 r , 一般情况下 $r \ll M$. 下面从线性相关理论^[8] 推导出一种寻找 $\{\phi(x_i)\}_{i=1}^M$ 的一组基的优化算法.

定理 1 设 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bs})$ 为 s 个线性无关的核样本, x 为任意一个训练样本, 则 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bs}), \phi(x)$ 线性无关 $\Leftrightarrow k_{xx} - K_{sx}^T K_s^{-1} K_{sx} > 0$. 其中: $k_{xx} = k(x, x)$, $K_s = (k(x_{bj}, x_{bj}))_{1 \leq j \leq s}$, $K_{sx} = (k(x_{b1}, x), k(x_{b2}, x), \dots, k(x_{bs}, x))^T$.

证明 令 K_{s+1} 表示 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bs}), \phi(x)$ 的核矩阵, 则

$$\det(K_{s+1}) = \begin{vmatrix} K_s & K_{sx} \\ K_{sx}^T & k_{xx} \end{vmatrix} = \begin{vmatrix} K_s & K_{sx} \\ 0 & k_{xx} - K_{sx}^T K_s^{-1} K_{sx} \end{vmatrix}.$$

因为 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bs})$ 线性无关, 故 $\det(K_s) > 0$. 因此 $\det(K_{s+1}) > 0 \Leftrightarrow k_{xx} - K_{sx}^T K_s^{-1} K_{sx} > 0$. 即 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bs}), \phi(x)$ 线性无关 $\Leftrightarrow k_{xx} - K_{sx}^T K_s^{-1} K_{sx} > 0$.

根据定理 1, 可用如下迭代算法寻找子空间 $\{\phi(x_i)\}_{i=1}^M$ 的一组基. 假设利用 $t (t \leq M)$ 个样本 $\{x_i\}_{i=1}^t$ 完成了训练, 得到子空间 $\{\phi(x_i)\}_{i=1}^t$ 的一组基 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bt})$, 对于一个新样本 x , 依据定理 1 判定 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bt}), \phi(x)$ 是否线性无关. 如果无关, 则令 $\phi(x_{b(s+1)}) = \phi(x)$, 得到一个新的线性无关组 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bt}), \phi(x_{b(s+1)})$. 则 t 遍历所有训练样本时, 所求的线性无关组 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bt})$ 即为子空间 $\{\phi(x_i)\}_{i=1}^M$ 的一组基.

利用定理 1 进行无关性判别时, 需要对 s 阶矩阵 K_s 求逆. 随着 t 的增大, s 可能逐渐增大, 因此求逆的运算量很大. 为此本文设计了一种优化算法, 将求逆运算变为乘法运算, 减小计算代价.

定理 2 设线性无关组 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{bs})$ 对应的核矩阵为 K_s , $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{b(s-1)})$ 对应的核矩阵为 $K_{(s-1)}$, 并令

$$\alpha = (k(x_{b1}, x_{bs}), k(x_{b2}, x_{bs}), \dots, k(x_{b(s-1)}, x_{bs}))^T,$$

$$= k(x_{bs}, x_{bs}), D = K_{(s-1)(s-1)}^{-1},$$

则

$$K_s^{-1} = \frac{1}{D} \begin{bmatrix} (D^{-T}D)D + D^{-T}D & -D \\ & -D^{-T}D & 1 \end{bmatrix}. \tag{5}$$

根据逆矩阵的定义易证上式成立^[8]. 由定理 2, 对 s 阶矩阵的求逆运算转化为对 $s-1$ 阶矩阵的求逆运算. 依次下去, 求逆运算变为乘法运算. 结合定理 1, 便可方便地寻求子空间 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的一组基 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{br})$.

算法 1 寻找子空间基的迭代算法. 具体步骤如下:

- 1) 初始化. 在训练样本集 $X = \{x_1, x_2, \dots, x_M\}$ 任一样本 x , 满足 $k(x, x) > 0$, 令 $S = \{x\}, D = \{x\}, G = 1/k(x, x), t = 1$.
- 2) 如果 $t = M$, 则输出 D , 终止程序; 否则, 转下一步.
- 3) 在 $\bar{S} = X - S$ 中任选一个样本 x^* , 令 $S = S \cup \{x^*\}, t = t + 1$; 并验证下式是否成立:

$$k_{tt} - K_{st}^T G K_{st} = 0. \tag{6}$$

其中

$$\bar{x}_i = D, (K_{st})_i = k(\bar{x}_i, x^*), k_{tt} = k(x^*, x^*).$$

- 4) 如果式(6)成立, 返回 2); 否则令

$$D = D \cup \{x^*\},$$

$$G = \frac{1}{k_{tt} - K_{st}^T G K_{st}} \begin{bmatrix} (k_{tt} - K_{st}^T G K_{st})G + & -GK_{st} \\ GK_{st} K_{st}^T G & \\ & -K_{st}^T G & 1 \end{bmatrix},$$

返回 2).

程序终止时, 集合 D 中的向量所对应的核样本即为 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的一组基. 下面将这组基标准正交化.

定理 3 设子空间 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的一组基为 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{br})$, 则 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的一组标准正交基为

$$(\phi_1, \phi_2, \dots, \phi_r) = (\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{br})) C. \tag{7}$$

其中: $C = (u_1/\sqrt{\lambda_1}, u_2/\sqrt{\lambda_2}, \dots, u_r/\sqrt{\lambda_r})$, u_1, u_2, \dots, u_r 为核矩阵 $K_{ss} = (k(x_{bi}, x_{bj}))_{1 \leq i, j \leq s}$ 的一组标准正交基, $\lambda_1, \lambda_2, \dots, \lambda_r$ 为对应的特征值.

证明 令 $U = (u_1, u_2, \dots, u_r)$, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$, 则 $U^T K_{ss} U = \Lambda$. 令 $C = (\phi_1, \phi_2, \dots, \phi_r)$, $\bar{C} = (1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, \dots, 1/\sqrt{\lambda_r})^T$, 则 $C^T = \bar{C}^T K_{ss} C = \bar{C}^T U^T K_{ss} U = \bar{C}^T \Lambda = I$.

4 基于正交基的快速算法

设 $\phi_1, \phi_2, \dots, \phi_r$ 为 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的一组标准正交

基, 由于这组基不仅可以表征该空间的结构, 而且如果核样本 $\phi(x_i)$ 与 $\phi(x_j)$ 正交, 则它们在 $\phi_1, \phi_2, \dots, \phi_r$ 的投影坐标向量 y_i 与 y_j 也正交. 基于这种考虑, 本文提出一种 KPCA 的快速算法.

设 $\phi(x_i)$ ($1 \leq i \leq M$) 在 $\phi_1, \phi_2, \dots, \phi_r$ 下的坐标向量记为 y_i . 由式(7)得

$$y_i = (\phi_1, \phi_2, \dots, \phi_r)^T \phi(x_i) = C^T (k(x_{b1}, x), k(x_{b2}, x), \dots, k(x_{br}, x))^T. \tag{8}$$

由式(8), 得到一个新的数据集 $Y = \{y_1, y_2, \dots, y_M\}$. 其满足下面的定理:

定理 4 分别对 $\phi(X) = \{\phi(x_1), \phi(x_2), \dots, \phi(x_M)\}$ 与 $Y = \{y_1, y_2, \dots, y_M\}$ 进行主分量分析, 则二者有相同的非零特征值, 并且对应的主轴方向具有下列关系:

$$v = (\phi_1, \phi_2, \dots, \phi_r) = ((\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{br})) C), \tag{9}$$

其中 v_k 分别为 $\phi(X), Y$ 的第 k 个主轴方向.

证明 令 $v = (\phi_1, \phi_2, \dots, \phi_r)$, 由式(8)知

$$\phi(X) = Y. \tag{10}$$

又 $\phi(X)$ 中心化后的协方差矩阵为

$$C = \phi(x) H_M H_M^T \phi(x)^T, \tag{11}$$

其中

$$(H_M)_{i,j} = \begin{cases} (M-1)/M, & i=j, 1 < i < M; \\ -1/M, & i \neq j, 1 < j < M. \end{cases}$$

联合式(10)和(11), 不难得到

$$C = (Y H_M H_M^T Y^T)^T. \tag{12}$$

令 $\bar{C} = Y H_M H_M^T Y^T$, 则其为 Y 中心化后的协方差矩阵. 即 $\phi(X)$ 与 Y 中心化后的协方差矩阵有下列关系:

$$C = \bar{C}^T. \tag{13}$$

由 C 及 \bar{C} 的定义及关系知 $\text{rank}(C) = \text{rank}(\bar{C}) = r$, 这就意味着 C 为半正定矩阵, \bar{C} 为正定矩阵. 下面推导二者的特征值与对应特征向量的关系:

1) 设 $\lambda = 0$ 为 C 的特征值, 对应特征向量为 v , 则 $\lambda = \bar{C}^T v$ 是 \bar{C} 的特征值为 λ 的特征向量, 即 $\bar{C} = \lambda v$.

由题意知: $Cv = \lambda v, v = \dots$. 又 $C = \bar{C}^T$, 故 $(\bar{C}^T)(\lambda v) = \lambda v$, 即 $\bar{C} = \lambda v$. 故 $\bar{C}^T \bar{C} = \lambda^T \lambda$, 即 $\bar{C} = \lambda v$.

2) 设 $\lambda \neq 0$ 为 \bar{C} 的特征值, 对应特征向量为 v , 则 $\lambda v = \bar{C} v$ 是 C 的特征值为 λ 的特征向量, 即 $Cv = \lambda v$.

由 1) 和 2) 知 C 与 \bar{C} 具有相同的非零特征值, 而且对应的特征向量具有 $v = \dots$ 的关系. 因此, 对 C 进行特征值分解可以转化为对 \bar{C} 进行相应的处理. 又因为 $v/\sqrt{\lambda} = \dots = (v/\sqrt{\lambda})$, 即特征向



量的单位化以后,式(9)仍成立,所以定理4成立.

由定理4,对数据集 $\Phi(X) = \{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_M)\}$ 的主分量分析,可以转化为对 $Y = \{y_1, y_2, \dots, y_M\}$ 的主分量分析. Y 的协方差矩阵只是一个 $r \times r$ 矩阵,因此对其进行特征值分解的计算复杂度仅为 $O(r^3)$. 进一步,对于任一测试样本 x ,依据式(9),它在 F 中的第 k 个主分量为

$$d_k(x) = (v^k)^T \Phi(x) = (C^k)^T k_b(x). \quad (14)$$

其中: k 表示 Y 的第 k 个主轴方向, $k_b(x) = (k(x_{b1}, x), k(x_{b2}, x), \dots, k(x_{br}, x))^T$. 显然,运用式(14)计算主分量,只包含 r 个核函数的计算,而传统的 KPCA 中,却包含 M 个核函数的计算(见式(4)),因此本文算法加快了特征提取速度.

由以上结论,给出本文算法(FKPCA)的主要步骤.

算法2 FKPCA 的主要步骤如下:

- 1) 选择核函数 $k(x, y)$ 及相应参数;
- 2) 根据算法1确定子空间 $\{\Phi(x_i)\}_{i=1}^M$ 的一组基 $\Phi(x_{b1}), \Phi(x_{b2}), \dots, \Phi(x_{br})$;
- 3) 根据式(7)确定一组标准正交基 v^1, v^2, \dots, v^r , 并求解 $\Phi(x_i)$ 在该组基下的坐标向量 y_i (见式(8));
- 4) 对数据集 $Y = \{y_1, y_2, \dots, y_M\}$ 执行 PCA, 求得其主轴方向(按降序排列) v^1, v^2, \dots, v^r ;
- 5) 根据式(9),得到 $\Phi(X)$ 的主轴方向 v^1, v^2, \dots, v^r , 并选取前 p ($1 \leq p \leq r$) 个用于特征提取;
- 6) 对一个新样本 x ,按式(14)计算其主分量.

5 实验分析

为了验证本文算法的效果,本文选用了一组模拟数据与实测数据进行实验,并同 KPCA 进行了比较. 需要说明的是在确定子空间的基时,考虑到计算误差的客观存在,把式(6)替换为 $k_n = K_n^T G K_n$, 本文实验取 $\alpha = 0.1$.

模拟数据是一组二维数据,生成方式为 x 服从 $[-1, 1]$ 均值分布, y 按方程 $y_i = (x_2/e + l) +$ 生成,其中 $e = 2, l = -1$, 是服从均值为0,方差为0.2的高斯分布的噪声. 实验中,对训练样本数目 $M = 200, 300, 400, 500, 600$ 时,分别用两种方法提取数据的特征,其中核函数均采用 $d = 2$ 的多项式形式 $k(x, y) = (x^T y)^d$.

图1显示的是 $M = 200$ 时,两种方法所得到的主分量在输入空间中的等高线. 其中离散点表示模拟数据,虚曲线表示等高线. 从上到下,分别为 KPCA 与 FKPCA;从左到右为特征值按降序排列的前3个主分量,特征值标在图的上方. 图中同一曲线上的点表示:在特征空间中,这些点在该主轴方向上

的投影值相同. 该曲线称为等高线. 从图中可以看出,两种方法得到的等高线基本一致,说明二者提取的数据特征基本相同. 当 M 取其他数值时,可得到类似的结果.

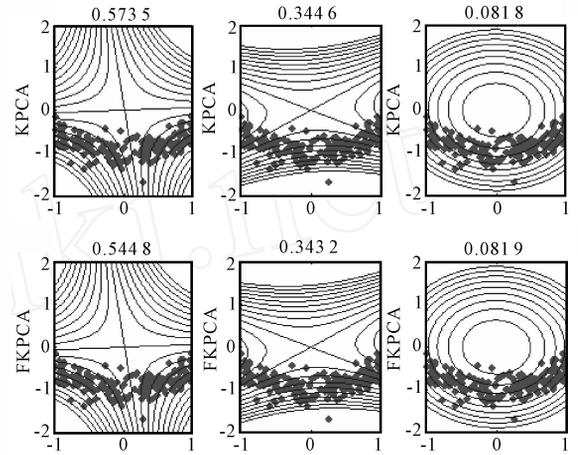


图1 主分量在输入空间中的等高线

在 FKPCA 的求解过程中,训练样本在特征空间中所张成的子空间的维数(即基的个数) r , 当 M 取上述不同值时,均有 $r = 4$. 说明这组模拟数据映射后,在特征空间中具有较强的线性相关性,因此随着 M 的增加,训练样本映射后所张成的子空间的维数却没有增加. 所以,对于不同的 M , FKPCA 的计算复杂度均为 $O(4^3)$, 相比 KPCA 的计算复杂度 $O(M^3)$, 计算代价大大降低. 特别是随着 M 的增大,优势更加明显. 图2描绘的是 M 不同时, KPCA 与 FKPCA 对同一组测试样本提取特征所花费的时间比值(即速度比)曲线. 比值越大,说明 FKPCA 相比于 KPCA 的特征提取速度快. 从图2可以看出,无论 M 取何值, FKPCA 的特征提取速度都有不同程度的提高,特别是随着 M 的增加,优势更加明显. 比如 $M = 600$ 时,二者的比值高达 182.

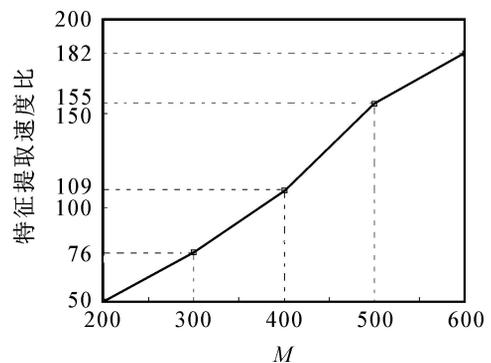


图2 特征提取速度比随 M 的变化曲线

实测数据集选自 USPS 手写体数据库. 该数据库包含了手写体数字 0 ~ 9 的训练集和测试集,每个样本 256 维. 本文针对双数数字做降噪实验. 每个数字随机选取 500 个训练样本, 50 个测试样本,并对

测试数据按如下方式加噪声:1) 均值为 0, 方差为 0.2 的高斯噪声;2) 概率为 0.3 的椒盐噪声. 本文首先用两种方法对训练集进行特征提取, 并保留前 16 个特征向量(相应的特征值按降序排列), 然后对噪声测试样本做降噪处理. 其中:核函数采用高斯形式 $k(x, y) = \exp(-\|x - y\|^2 / nc)$, n 为输入空间维数, 实验中 $c = 0.1$. 与 PCA 不同, KPCA 不能直接重构图像. 为此, 本文采用文献[9]提出的一个迭代算法进行降噪处理.

设 x 是原始图像, 则重构图像 z 可由如下迭代公式得到^[9]:

$$z^{new} = \frac{\sum_{i=1}^m \exp(-\frac{\|\bar{x}_i - z\|^2}{2\sigma^2}) \bar{x}_i}{\sum_{i=1}^m \exp(-\frac{\|\bar{x}_i - z\|^2}{2\sigma^2})} \quad (15)$$

其中 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ 表示用以表达 F 中主轴方向 v^1, v^2, \dots, v^k 的训练样本. 对 KPCA 而言, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ 就是所有的训练样本 x_1, x_2, \dots, x_m ; 对 FKPCA 而言, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ 是子空间 $\{\phi(x_i)\}_{1 \leq i \leq m}$ 的基 $\phi(x_{b1}), \phi(x_{b2}), \dots, \phi(x_{br})$ 所对应的训练样本 $x_{b1}, x_{b2}, \dots, x_{br}$. $i = \sum_{j=1}^k d_j(x) \frac{v^j}{\|v^j\|}$, 这里 $\frac{v^j}{\|v^j\|}$ 满足 $v^j = \sum_{i=1}^m \phi(\bar{x}_i) d_j(x)$ 则表示 $\phi(x)$ 在 v^j 上的主分量, 即 $d_j(x) = \sum_{i=1}^m \phi(\bar{x}_i) k(\bar{x}_i, x)$.

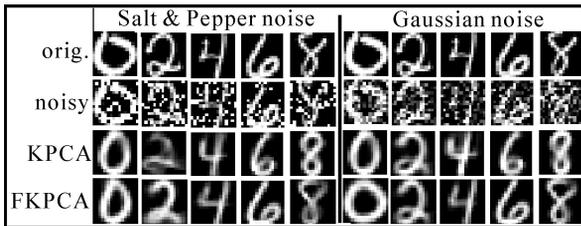


图 3 两种方法对 USPS 数据的降噪结果

表 1 两种方法在不同数字上的特征提取速度比

符号	0	2	4	6	8
训练样本数	500	500	500	500	500
基的个数	313	492	365	266	371
特征提取速度比	7.06	4.62	12.68	14.6	6.54

图 3 显示了两种方法对 USPS 数据的降噪结果. 从图中可以看出, 无论 KPCA 还是 FKPCA, 都可以达到较好的降噪效果. 说明两种方法提取的特征基本相同. 表 1 列出了两种方法用于不同数字的特征提取速度比(特征提取速度比是指运用 KPCA 与 FKPCA 对同一组测试样本提取特征所花费的时

间比). 从表 1 中可以看出, 对于不同数字, 其基的个数不同, 说明不同数字的训练样本映射后的子空间维数不同. 但基的个数均小于训练样本数, 说明该样本间存在线性相关关系. 反映到特征提取速度上, 从图 3 中可以看出 FKPCA 比 KPCA 快.

6 结 论

数据集 $\phi(X)$ 在特征空间 F 中所张成的子空间 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的空间结构可由一组标准正交基来反映, 而且 $\phi(x_i)$ 与 $\phi(x_j)$ 正交, 则它们在这组基下的投影坐标向量 y_i 与 y_j 也正交. 基于上述性质, 提出一种 KPCA 的快速算法——FKPCA. 即将 $\phi(X)$ 中的所有核样本向一组标准正交基上做投影, 则对数据集 $\phi(X)$ 的主分量分析, 也就是原输入空间的 KPCA, 转化为对投影坐标向量所构成的数据集 Y 的主分量分析. 本文从理论上保证了上述结论成立. 该算法的计算复杂度仅为 $O(r^3)$ (r 为基的个数), 相比于 KPCA 的计算复杂度 $O(M^3)$, 计算代价大大降低, 同时测试阶段的特征提取速度得到较大提高, 解决了大训练集对 KPCA 的计算代价以及特征提取速度的制约. 而且基于向量间的线性相关性理论, 本文专门设计了一个优化算法以“加速”子空间 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 的基的确定.

本文的算法设计思想可以推广到其他的核方法中, 特别是用以寻找子空间 $\{\phi(x_i)\}_{1 \leq i \leq M}$ 基的优化算法. 如何将本文算法推广到其他核方法中, 简化计算, 是一个值得探讨的问题.

参考文献(References)

- [1] Schölkopf B, Smola A J, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(6): 1299-1319.
- [2] Kim K I, Park S H, Kim H J. Kernel principal component analysis for texture classification[J]. IEEE Signal Process Letters, 2001, 8(2): 39-41.
- [3] Koutsogiannis G S, Soraghan J J. Selection of number of principal components for de-noising signals [J]. Electron Letters, 2002, 38(13): 664-666.
- [4] Darchapak C, Kanae S, Yang Z J. Kernel principal component regression with application to nonlinear prediction[R]. Research Reports on Information Science Kyushu University, 2003, 8(1): 19-23.
- [5] Rosipal R, Grolami M. An expectation maximization approach to nonlinear component analysis [J]. Neural Computation, 2001, 13(1): 505-510.
- [6] Zheng W M, Zou C R, Zhao L. An improved algorithm for kernel principal component analysis [J]. Neural Processing Letters, 2005, 22(1): 49-56.

(下转第 1057 页)

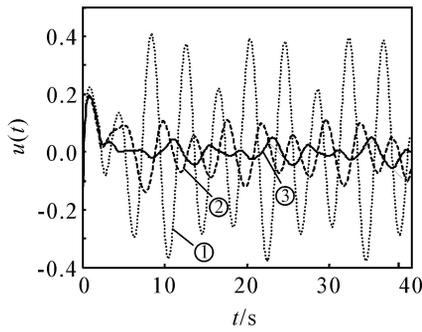


图 4 当 $\tau = 2$ 时控制律 $u(t)$ 曲线

数 k 下仿真曲线如图 3 和图 4 所示. 其中: 为 $k = 1$; 为 $k = 3$; 为 $k = 7$.

当 $\tau = 2$ 时性能指标及控制精度如表 2 所示.

表 2 当 $\tau = 2$ 时不同迭代次数时的性能指标

k	1	2	4	6	7
J_k	0.062 5	0.023 5	0.012 7	0.010 7	0.010 3
J	/	0.655 4	0.490 2	0.415 1	0.044 7

从表 2 也可看出,随着迭代次数增加,性能指标相对误差逐渐减小. 当 $k = 7$ 时,满足控制精度要求,因此可以把 $u_7(t)$ 作为近似最优减振控制律.

由两个表对照可以得出,随着时滞的增大,本文给出方法的迭代次数增加,控制质量也随之降低,但系统的性能指标仍趋于一稳定值,同时对扰动仍有很好的抑制作用.

7 结 论

本文研究时滞非线性系统在外部正弦干扰作用下的最优减振控制问题,给出了最优减振控制律的设计过程,并给出了一种设计实际近似最优减振控制律的无滞后方法. 仿真结果表明该方法对外部正弦扰动在不同的时滞下都有良好的鲁棒性.

参考文献(References)

[1] Bodson M. Performance of an adaptive algorithm for sinusoidal disturbance rejection in high noise [J]. Automatica, 2001, 37(7): 1133-1140.

[2] Mulgund S S, Stengel R F. Optimal nonlinear estimation for aircraft flight control in wind shear [J]. Automatica, 1996, 32(1): 3-13.

[3] Wang W, Tang G Y. Feedback and feedforward optimal control for offshore jacket platforms [J]. China Ocean Engineering, 2004, 18(4): 515-526.

[4] Beard R W, McLain T W. Successive Galerkin approximation algorithms for nonlinear optimal and robust control [J]. Int J of Control, 1998, 71(5): 717-743.

[5] Kurina G A. Asymptotic expansion of solutions of optimal control problems for discrete weakly controllable systems [J]. J Applied Mathematics and Mechanics, 2002, 66(2): 201-213.

[6] Garrard W L, Enns D F, Snell S A. Nonlinear feedback control of highly manoeuvrable aircraft [J]. Int J Control, 1992, 56(4): 799-812.

[7] 唐功友,高德欣. 带有持续扰动非线性系统的前馈-反馈最优控制[J]. 控制与决策, 2005, 20(4): 366-371. (Tang Gong-you, Gao De-xin. Feedforward and feedback optimal control for nonlinear systems with persistent disturbances [J]. Control and Decision, 2005, 20(4): 366-371.)

[8] Cimen T, Banks S P. Nonlinear optimal tracking control with application to super-tankers for autopilot design [J]. Automatica, 2004, 40(11): 1845-1863.

[9] McCaffrey D, Banks S P. Lagrangian manifolds and asymptotically optimal stabilizing feedback control [J]. Systems and Control Letters, 2001, 43(3): 219-224.

(上接第 1048 页)

[7] Baudat G, Anouar F. Kernel-based methods and function approximation [C]. Proc of the International Joint Conf on Neural Networks. Piscataway: Institute of Electrical and Electronics Engineers Inc, 2001: 1244-1249.

[8] Cheng Y P. Theory of matrix [M]. Xi'an: Northwest

University Press, 1989.

[9] Mike S, Scholkopf B, Smola A J. Kernel PCA and denoising in feature space [C]. In Advances in Neural Information Processing Systems. Cambridge: MIT Press, 1999: 524-536.