

基于核主成分分析的铁谱磨粒特征提取方法研究*

李岳,温熙森,吕克洪

(国防科技大学 机电工程与自动化学院 湖南 长沙 410073)

摘要 :针对铁谱分析的磨粒识别过程中存在原始磨粒特征描述指标参数多、非线性突出的问题,提出基于核主成分分析的铁谱磨粒特征提取方法,介绍该方法的原理与算法。结合某柴油发动机故障检测与分析系统中铁谱磨粒自动识别的应用实例,并与传统主成份分析方法进行对比分析,结果表明该方法在进行样本非线性特征参数指标综合以及特征维数压缩方面具有可行性和有效性。

关键词 :核主成分分析(KPCA),铁谱磨粒,特征提取

中图分类号 :TP277 文献标识码 :A

KPCA-based Technique for Debris Feature Extraction

LI Yue, WEN Xi-sen, Lü Ke-hong

(College of Mechatronics Engineering and Automation, National Univ. of Defense Technology, Changsha 410073, China)

Abstract :To deal with the problem of Debris' feature extraction with the characteristic of large scale number in description parameter and nonlinear relationship among these parameters, a KPCA-based method is presented. Based on the detailed introduction of KPCA algorithm, a real case was studied for the purpose of constructing a debris automatic recognition sub-system which can be served for diesel engine fault detection and analysis system. The results of experimental research and a comparison with linear PCA demonstrate that KPCA-based approach is feasible and valid for synthesizing nonlinear characteristic parameter and reducing feature dimension.

Key words :KPCA ; debris ; feature extraction

铁谱分析是目前广泛应用于动力装置状态监测与故障诊断领域的一种有效的不解体诊断方法,铁谱分析的核心技术是磨粒识别,磨粒识别的准确性依赖于磨粒的特征描述是否有效。由于磨粒形貌特征非常复杂,对各类磨粒的描述往往需要采用多项参数指标,而这些参数指标中又可能存在较强的相关性,因此,去除参数间的相关性和降低磨粒参数维数有助于提高识别准确率和减少识别工作量。文献[1]提出采用多元统计分析中的主成分分析(Principal Components Analysis, PCA)进行铁谱的磨粒识别,取得一定效果,为动力装置状态识别与故障诊断提供了一种有效的特征选择与特征提取方法。但 PCA 方法的核心过程是采用一组线性变换进行空间映射,主要适合解决磨粒特征参数指标间的线性相关问题,属于线性特征提取方法范畴。由于磨粒各特征参数指标间在本质上存在非线性,在对这些参数指标进行特征提取(综合)时,线性 PCA 方法可能出现各指标的贡献率相对比较分散,不能有效得到综合性更强的特征指标参数,有时还可能会导致有用信息的丢失。

本文从研究非线性特征提取的目标出发,提出用于磨粒特征提取的非线性 PCA 方法,即基于核主成分分析(KPCA)的磨粒特征提取方法。KPCA 方法是对 PCA 的非线性拓展,在一个与输入样本空间有非线性关系的高维空间上计算主元,即将分析样本映射到某个高维特征空间上,在这个特征空间中使用线性理论,这种分析方法表现出了比线性 PCA 更有效的特征提取能力。

1 KPCA 非线性特征提取算法及应用步骤

核主成分分析(Kernel Based Principal Component Analysis, KPCA)是通过使用“核技巧”将线性 PCA 拓展为非线性 PCA 的一种方法,基本思路是将原始输入向量 x 映射到高维特征空间 $\Phi(x)$,然后在特征

* 收稿日期 2006 - 11 - 16

作者简介:李岳(1964—),男,教授,在职博士生。

空间中对 $\Phi(x)$ 进行线性 PCA 计算, 这样, 在特征空间中对 $\Phi(x)$ 进行的线性 PCA 就相当于在输入空间 x 中的非线性 PCA。可用图 1 简单描述其过程^[2]。

在线性 PCA 的情况下, 非线性 PCA 首先引入一个非线性映射 $\Phi: x \in R^n \mapsto \Phi(x) \in F \subseteq R^N$, 将输入空间的一个样本集 $x_k (x_k \in R^n, k = 1, 2, \dots, M, \sum_{k=1}^M x_k = 0)$ 变换到特征空间 (这里的特征空间 F 的维数可以任意大, 可以是无限维), 假设 $\sum_{k=1}^M \Phi(x_k) = 0$, 则特征空间中样本的协方差矩阵为

$$\bar{C} = \frac{1}{M} \sum_{j=1}^M \Phi(x_j) \Phi(x_j)^T \quad (1)$$

此时 PCA 是通过求解下列方程

$$\lambda V = \bar{C} V \quad (2)$$

获得特征值 $\lambda (\lambda \geq 0)$ 以及所对应的特征向量 $V (V \in F \setminus \{0\})$ 。

由于式 (2) 的所有解均在 $\Phi(x_1), \dots, \Phi(x_M)$ 张成的子空间内, 因此,

$$\lambda (\Phi(x_k) \cdot V) = (\Phi(x_k) \cdot \bar{C} V), \quad k = 1, 2, \dots, M \quad (3)$$

并且存在系数 $\alpha_i (i = 1, 2, \dots, M)$, 使得

$$V = \sum_{i=1}^M \alpha_i \Phi(x_i) \quad (4)$$

合并式 (3) 和式 (4), 得到

$$\lambda \sum_{i=1}^M \alpha_i (\Phi(x_k) \cdot \Phi(x_i)) = \frac{1}{M} \sum_{i=1}^M \alpha_i (\Phi(x_k) \cdot \sum_{j=1}^M \Phi(x_j) (\Phi(x_j) \cdot \Phi(x_i))), \quad k = 1, 2, \dots, M \quad (5)$$

定义一个 $M \times M$ 阶矩阵 K , 其中 $K_{ij} = \Phi(x_i) \cdot \Phi(x_j)$, 可以得到

$$M \lambda K \alpha = K^2 \alpha \quad (6)$$

其中 $\alpha = (\alpha_1, \dots, \alpha_M)^T$ 。通过下列特征方程

$$M \lambda \alpha = K \alpha \quad (7)$$

来求解非零特征值。由于

$$1 = \sum_{i,j=1}^M \alpha_i \alpha_j^k (\Phi(x_i) \cdot \Phi(x_j)) = \sum_{i,j=1}^M \alpha_i^k \alpha_j^k K_{ij} = (\alpha^k \cdot K \alpha^k) = \lambda_k (\alpha^k \cdot \alpha^k)$$

因此可通过标准化 α^k 来使特征空间相应的特征向量标准化, 即 $(V^k \cdot V^k) = 1$ 。

为提取主成分, 需要计算 F 中的样本在特征向量 V^k 上的投影, 若 x 为某一测试样本向量, 在特征空间的映射为 $\Phi(x)$, 则

$$(V^k \cdot \Phi(x)) = \sum_{i=1}^M \alpha_i^k (\Phi(x) \cdot \Phi(x_i)) \quad (8)$$

为测试样本对应于 Φ 的非线性主成分。

前面的讨论都是基于一个前提, 即所考察的观测值都是经过中心化处理, 这个前提条件在输入空间容易实现, 但在特征空间中我们不能明确地计算出样本映射在特征空间后的平均值。这个问题的一种解决办法是用 \bar{K} 取代 K 来求解特征值, \bar{K} 按下式计算

$$\bar{K}_{ij} = K_{ij} - \frac{1}{M} \sum_{m=1}^M 1_{im} K_{mj} - \frac{1}{M} \sum_{n=1}^M K_{in} 1_{nj} + \frac{1}{M^2} \sum_{m,n=1}^M 1_{im} K_{mn} 1_{nj} \quad (9)$$

通过上述分析, KPCA 特征提取算法的步骤可归结为:

- (1) 初始化输入样本 x , 计算矩阵 K, \bar{K} ;
- (2) 在特征空间计算其特征值并标准化特征值;
- (3) 找出最大的特征值和对应的特征向量;

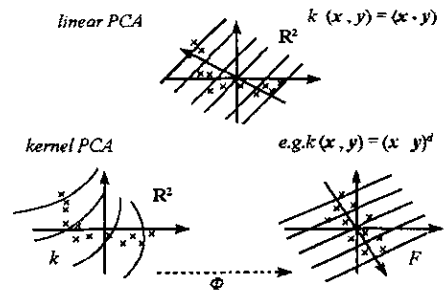


图 1 PCA 与 KPCA 示意图
Fig. 1 The graph of PCA and KPCA

(4)根据(8)式对样本特征指标参数进行综合,提取出样本的分类特征,实现样本的特征提取。

值得一提的是,KPCA最大主成分数目为样本数 M ,即KPCA的主成分是基于样本的;而在线性PCA中主成分数目决定于样本维数大小,即线性PCA的主成分是基于样本维数的。与PCA相同的是,如果只考虑排在前面的几个特征向量,KPCA同样可以对 x 降维。通过使用核方法实现了PCA的非线性化,并且线性PCA所具有的特性,在KPCA的 $\Phi(x)$ 中同样得到体现。KPCA通过引入非线性映射函数 $\Phi(x) \in R^M (M > N)$,将输入空间中非线性问题转换到高维特征空间采用线性方法处理,这里是利用“核函数”来巧妙地解决这个问题。根据泛函有关理论,只要某个核函数 $K(x, y)$ 满足Mercer条件,它就对应某一变换空间的内积,即 $K(x, y) = \Phi(x) \cdot \Phi(y)$,这样在高维空间实际上只需进行内积运算,而这种内积运算是可以用原空间中的函数实现的,而无需知道变换 Φ 的具体形式^[3]。常用的核函数有:线性核函数、多项式核函数、高斯核函数、多层感知机核函数等^[4]。由于高斯核函数适应范围广,算法相对简单,尤其对于复杂问题,优点更加明显,因此本文选取高斯核函数进行案例研究。

2 KPCA的应用研究与对比分析

本文研究工作主要针对某型履带式车辆的柴油发动机故障检测与分析系统中铁谱磨粒自动识别问题,核心是进行磨粒特征提取。

2.1 磨粒的特征参数描述与特点分析

铁谱磨粒是由摩擦副间的相互作用产生的,其形态各异、不尽相同。因此,仅用几个物理参数描述很难系统全面反映出磨粒的特征信息,实际应用中,对磨粒描述的特征参数往往多达数十个,主要包括:(1)尺寸参数:面积(A)、当量直径(D)、长轴长度(a)、短轴长度(b)、磨粒周长(L)等;(2)形状参数:圆形度(R_d)、长宽比(R_l)、凹度(C_v)、边界粗糙度等;(3)结构参数:孔洞率(E'')等;(4)灰度参数:灰度平均值(G_m)、灰度变化率(G_p)等;(5)边缘参数:边缘结构函数($H(r)$)、边缘转向度(A_n)等;(6)不规则形状参数:灰度纹理(F_{gt})、灰度结构(F_{gs})、边缘细节(F_{pt})、边缘结构(F_{ps})等^[5]。显然,与此相应的磨粒识别工作就不得不在高维参数空间中进行,识别难度高,复杂性大,准确性难于保证,因此对原参数空间进行有效的降维处理将有助于减少识别工作量和提高识别的准确率。由于磨粒特征参数之间关系的非线性特点,传统的特征提取方法(如主成分分析方法)在处理非线性样本时存在一定的局限,本文采用非线性主成分分析(KPCA)方法,解决磨粒特征提取中存在的上述问题。

2.2 实例研究

从铁谱自动分析系统的标准磨粒样本库中选取12组磨粒样本用于本文研究,每组磨粒样本原始特征描述参数空间为53维,如表1所示。

表1 磨粒样本的特征数据

Tab.1 Characteristic data of the debris samples

样本	参数1	参数2	...	参数25	...	参数52	参数53
序号	面积	等效直径	...	傅氏凹度	...	孔隙率	对比度
1	1067.5	36.8671	...	1.330549	...	0.012984	4.560697
2	6318.5	89.69372	...	1.129762	...	0.0027619	0.849903
3	71.25	9.524616	...	1.022309	...	0.0371622	18.78088
4	1450.25	42.9711	...	5.291819	...	0.0273926	4.161927
5	2047.75	9.5544	...	1.152386	...	0.0054638	1.556391
6	7	2.985411	...	1.5073	...	0.2222222	37.5
7	8.75	3.337791	...	0.6303674	...	0.125	41.37755
8	4214.25	73.25128	...	1.504672	...	0.0005964	0.091634
9	3128.5	63.11363	...	1.169952	...	0.000479	1.223147
10	1974.25	50.13675	...	3.53534	...	0.0012640	2.862975
11	142	13.44619	...	1.252497	...	0.025729	18.75843
12	40.5	7.180961	...	1.437836	...	0.0526315	16.99758

根据KPCA方法对磨粒样本组进行主成分分析,采用高斯核函数($k(x, y) = \exp(-\|x - y\|^2 /$

$2\sigma^2$) 其参数 σ^2 分别选取 1000、2000, 对磨粒样本的分析结果如表 2 所示。从表 2 不难看出, 按照累计贡献率大于 85% 的原则^[6] 选取特征值, 则对应于 $\sigma^2 = 1000$ 的条件下, 需选取前 3 个特征值 $\lambda_1, \lambda_2, \lambda_3$, 而采用 $\sigma^2 = 2000$, 仅需选取前 2 个特征值 λ_1, λ_2 便可满足要求。由此可知, 在获得主要信息的基础上, 采用 KPCA 方法可以获得很好的降维效果。即通过计算与特征值所对应的特征向量, 计算特征空间中的样本映射在特征向量上的投影, 就可对磨粒样本的 53 维原特征描述空间进行有效的综合, 获得维数较小的样本特征综合描述指标, 进而为磨粒识别工作提供有效的分类特征。

表 3 为采用线性主成分分析(PCA)对同一组磨粒样本计算的结果, 与表 2 对比可知, PCA 各主元贡献率相对比较分散, 在截取相同的信息量的基础上(即累计贡献率相同), KPCA 具有主成份特征明显, 贡献率集中, 主成份参数维数较少等优点, 其性能明显高于 PCA 的分析结果。

3 结论

文中采用核主成分分析(KPCA)方法对具有非线性高维数特点的铁谱磨粒原始特征描述指标进行特征提取, 并与主成分分析(PCA)方法进行比较。结果表明, KPCA 方法能够较好地解决磨粒描述参数指标间存在的非线性相关问题, 同时能够在保留原始数据主要信息的基础上获得较好的降维效果, 达到了对铁谱磨粒特征进行有效提取的目的, 为后续磨粒识别的准确性和效率的提高奠定了了的基础。KPCA 方法为一种新型非线性特征提取方法, 在铁谱的磨粒识别领域里有着较好的应用前景。

表 2 磨粒样本 KPCA 处理结果

Tab. 2 KPCA result of the debris samples

特征值序号	KPCA(高斯核函数)			KPCA(高斯核函数)		
	$\sigma^2 = 1000$			$\sigma^2 = 2000$		
	特征值 λ_i	贡献率	累计贡献率	特征值 λ_i	贡献率	累计贡献率
1	0.2559	53.9799	53.9799	0.1882	75.6805	75.6805
2	0.1434	30.2475	84.2273	0.0546	21.9632	97.6437
3	0.0589	12.4308	96.6581	0.0048	1.9303	99.5740
4	0.0113	2.3839	99.0420	0.0007	0.2881	99.8622
5	0.0028	0.5880	99.6300	0.0002	0.0791	99.9413
6	0.0010	0.2126	99.8426	0.0001	0.0346	99.9759
7	0.0005	0.0964	99.9390	0	0.0125	99.9884
8	0.0001	0.0295	99.9685	0	0.0054	99.9938
9	0.0001	0.0202	99.9887	0	0.0027	99.9965
10	0	0.0085	99.9972	0	0.0025	99.9989
11	0	0.0028	100	0	0.0011	100
12	0	0	100	0	0	100

表 3 磨粒样本 PCA 处理结果

Tab. 2 PCA result of the debris samples

特征值序号	PCA		
	特征值 λ_i	贡献率	累计贡献率
1	25.26068	47.5600	47.5600
2	7.4128	13.9865	61.5465
3	662764	11.8423	73.3888
4	3.7759	7.1243	80.5131
5	3.0863	5.232	86.3362
6	2.3096	4.3578	90.6941
7	1.6585	3.1292	93.8233
8	1.4230	2.6848	96.5081
9	0.9362	1.7665	98.2746
10	0.5131	0.9682	99.2428
11	0.4013	0.7572	100
12	0	0	100
⋮	⋮	⋮	⋮
53	0	0	100

参考文献:

- [1] 李岳, 吕克洪. 主成分分析在铁谱磨粒识别中的应用研究[J]. 国防科技大学学报, 2004, 26(1).
- [2] Schölkopf B, Smola B A, Müller K R. Kernel Principal Component Analysis. Advances in Kernel Methods-support Vector Learning[M]. Cambridge, MA: MIT Press, 1999: 327-352.
- [3] Shawe-Taylor J, Nello Cristianini. Kernel Method for Pattern Analysis[M]. Cambridge University Press, 2004.
- [4] David V, Sanchez A. Advanced Support Vector Machines and Kernel Methods[R]. Neurocomputing, 2003, 55: 5-20.
- [5] Li Y J, Zuo H F, Wu Z F. Debris Monitoring and Analyzing System and Its Application in Aeroenging[J]. Transaction of Nanjing University of Aeronautics & Astronautics, 2001, 18(2).
- [6] 王江萍. 机械设备故障诊断技术及应用[M]. 西安: 西北工业大学出版社, 2001.

