

基于KPCA和CPSO的故障检测方法

唐勇波^{1,2}, 桂卫华¹, 欧阳伟³

(1. 中南大学信息科学与工程学院, 长沙 410083; 2. 宜春学院物理科学与工程技术学院, 江西 宜春 336000;

3. 中国瑞林工程技术有限公司, 南昌 330002)

摘要: 提出一种基于核主元分析(KPCA)和混沌粒子优化群(CPSO)算法的非线性故障检测方法。通过核函数完成非线性变换, 将变量由非线性的输入空间转换到线性的特征空间来计算主元, 构造平方预测误差统计量检测故障是否发生。为避免粒子群算法的早熟现象, 利用混沌优化的搜索特性, 将CPSO算法应用到KPCA核参数的优化中。变压器故障检测结果表明, 与基于PCA、KPCA和PSO-KPCA的故障检测方法相比, 该方法的检测正确率较高。

关键词: 核主元分析; 粒子群优化算法; 混沌优化; 故障检测; 溶解气体分析

Fault Detection Method Based on KPCA and CPSO

TANG Yong-bo^{1,2}, GUI Wei-hua¹, OUYANG Wei³

(1. School of Information Science and Engineering, Central South University, Changsha 410083, China;

2. School of Physical Science and Engineering, Yichun University, Yichun 336000, China;

3. China Nerin Engineering Co., Ltd., Nanchang 330002, China)

【Abstract】 A nonlinear fault detection method based on Kernel Principal Component Analysis(KPCA) and Chaos Particle Swarm Optimization(CPSO) algorithm is presented. KPCA performs nonlinear transformation by kernel function to map the nonlinear input space into linear feature space, computes principal component and detects faults by utilizing SPE statistics. The kernel parameters of kernel principal component are optimized in order to enhance the fault detection performance. For the premature convergence problem of the Particle Swarm Optimization(PSO) algorithm, the CPSO algorithm is adopted to utilize the chaos optimization's search properties. Experimental results of transformer show that the proposed method has better detection performance than PCA, KPCA and PSO-KPCA method.

【Key words】 Kernel Principal Component Analysis(KPCA); Particle Swarm Optimization(PSO) algorithm; chaos optimization; fault detection; Dissolved Gas Analysis(DGA)

DOI: 10.3969/j.issn.1000-3428.2012.24.057

1 概述

电力变压器是电力系统的重要设备之一, 变压器运行在复杂的强电磁场环境中, 因此, 变压器的故障诊断是一个非常重要且复杂的工作。油中溶解气体分析(Dissolved Gas Analysis, DGA)是电力系统诊断充油电气设备早期潜伏性故障的重要方法, 变压器运行过程中积累了大量的油中溶解气体数据, 对变压器的故障诊断具有重要价值。

主元分析(Principal Component Analysis, PCA)是一种线性的数据降维方法^[1], 在处理非线性问题时, 通常不能取得好的效果。实际工业过程往往是非线性的, Scholkopf 等人提出的核主元分析(Kernel Principal Component Analysis, KPCA)能有效地提取非线性特征^[2-5]。鉴于核参数的选择对检测性能有较大影响, 本文尝试对核参数进行寻优。粒子群优化(Particle Swarm Optimization, PSO)算法是非线性连续优化问题、组合优化问题和整数非

线性优化问题的有效优化工具, 但该算法具有易陷入局部极值点、进化后期收敛速度慢、精度较低等不足^[6]。由于混沌运动具有随机性、遍历性等特点, 为提高种群的多样性和粒子搜索的遍历性, 研究者将混沌思想引入粒子群优化算法中, 改善粒子群优化算法摆脱局部极值点的能力, 提高算法的收敛速度和精度^[7-10]。本文提出基于混沌粒子群优化(Chaos Particle Swarm Optimization, CPSO)算法和KPCA的故障检测方法(CPSO-KPCA), 并将其应用于变压器故障检测中。

2 CPSO-KPCA 故障检测方法

2.1 核主元分析

核主元分析的主要思想是通过引入非线性映射 ϕ 将输入空间 $X_{N \times M}$ 映射到一个高维特征空间 F , 然后在高维特征空间计算主元成分。映射数据 $\phi(x_i)$ 的协方差矩阵可以表示为:

作者简介: 唐勇波(1975—), 男, 讲师、硕士, 主研方向: 智能故障诊断; 桂卫华, 教授、博士生导师; 欧阳伟, 教授

收稿日期: 2011-11-02 **修回日期:** 2011-12-21 **E-mail:** tyb2003045@yahoo.com.cn

$$C^F = \frac{1}{N} \sum_{i=1}^N \varphi(x_i) \varphi^T(x_i) \quad (1)$$

解决在特征空间的特征值问题:

$$\lambda v = C^F v \quad (2)$$

C^F 的特征矢量 v 可表示为:

$$v = \sum_{i=1}^N a_i \varphi(x_i) \quad (3)$$

计算主元, 即:

$$t_k = \langle v_k, \varphi(x) \rangle = \sum_{i=1}^N a_i^k \langle \varphi(x_i), \varphi(x) \rangle \quad (4)$$

其中, $\langle x, y \rangle$ 表示 x 与 y 的点积。为避免直接计算非线性映射, 在特征空间定义核函数矩阵 $K = [k_{ij}]_{N \times N}$, 其中, $k_{ij} = \langle \varphi(x_i), \varphi(x_j) \rangle$ 。核函数主元分析的详细求解过程参见文献[4]。

对于核函数的选择, 本文选取高斯径向基函数:

$$k(x, y) = \exp(-\|x - y\|^2 / c) = \exp(-\|x - y\|^2 / (\gamma \cdot m \cdot \delta^2)) \quad (5)$$

其中, γ 为核参数; m 为原始输入空间维数; δ^2 为方差。

通过对测量数据进行平方预测误差(SPE)检验, 可判断过程是否发生变化, SPE 统计量定义为:

$$SPE = \|\varphi(x) - \varphi_p(x)\|^2 = \sum_{j=1}^n t_j^2 - \frac{\sum_{j=1}^n t_j^2}{n} \quad (6)$$

SPE 的控制限可根据其近似分布计算:

$$SPE_a \sim g \chi_h^2 \quad (7)$$

其中, $g = b/2a$, $h = 2a^2/b$, a 和 b 是 SPE 的估计均值和方差; p 为主元个数; n 为非零特征值的个数。

2.2 混沌粒子群优化算法

粒子群优化算法是一种群体的进化技术, 具有原理简单, 易于实现、全局收敛等特点。PSO 采用速度-位置搜索模型, 设粒子群由 m 个粒子构成, 第 i 个粒子在 D 维解空间的位置为 $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, $x_i \in [a_i, b_i]$, 速度 $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, 每个粒子都有一个适应值函数 $f(x_i)$ 来计算 x_i 当前的适应值, 衡量粒子位置的优劣。 x_i 所经历的最优位置, 称为个体极值 $pBest_i$, 而群体所经历的最优位置, 称为全局极值 $gBest$ 。在每次迭代中, 粒子跟踪个体极值、全局极值和自己前一时刻的状态来调整当前时刻的位置和速度, 迭代公式如下:

$$v_i(k+1) = wv_i(k) + c_1r_1(pBest_i - x_i(k)) + c_2r_2(gBest - x_i(k)) \quad (8)$$

$$x_i(k+1) = x_i(k) + v_i(k+1) \quad (9)$$

其中, w 为惯性权重; k 为迭代次数; r_1 和 r_2 为 $[0, 1]$ 之间的随机数; c_1 和 c_2 为学习因子, 通常取为 2。

由于 PSO 算法不具备遍历特性, 且存在由于某些粒子在迭代中出现停滞而导致的算法早熟现象, 引入具有遍历特性的混沌机制^[6-7]。产生混沌的规则很多, 式(13)为混沌变量 cx_i 的一种演变算式:

$$cx_i(k+1) = 4cx_i(1 - cx_i(k)), i = 1, 2, \dots, m \quad (10)$$

变量 x_i 可由式(14)~式(15)与混沌变量 $cx_i \in (0, 1)$ 进行往返映射:

$$cx_i = (x_i - a_i) / (b_i - a_i) \quad (11)$$

$$x_i = a_i + cx_i(b_i - a_i) \quad (12)$$

核参数 γ 的 CPSO 优化算法描述如下:

Step1 确定算法参数, 初始化种群: 给定群体规模 m , 随机产生每个粒子的初始位置和速度 r_i, v_i , 按式(13)计算各粒子的适应值 $f(r_i)$, 有 $pBest_i = r_i$, 经比较得到 $gBest$, 用预测误差作为适应度函数:

$$f(\gamma_i) = 1 - G_mean = 1 - \sqrt{tp \cdot tn} \quad (13)$$

预测误差最小时所对应的粒子即为最优解。其中,

$G_mean = \sqrt{tp \cdot tn}$ 为几何均值。

故障样本的检测正确率为:

$$tp = |TP| / (|TP| + |FN|) \quad (14)$$

其中, $|TP|$ 为实际故障而判别为故障的样本个数; $|FN|$ 为实际故障而判别为正常的样本个数。

正常样本的检测正确率为:

$$tn = |TN| / (|TN| + |FP|) \quad (15)$$

其中, $|TN|$ 为实际正常而判别为正常的样本个数; $|FP|$ 为实际正常而判别为故障的样本个数。

Step2 将 r_i 通过式(11)的变换, 映射为混沌变量 cr_i 。

Step3 通过式(8)、式(9)计算速度 v_i , 并调整至新位置 r_i , 按式(13)计算适应值 $f(r_i)$ 。

Step4 混沌变量 cr_i 按式(10)作混沌变换为 cy_i 。

Step5 将 cy_i 通过式(12)变换, 映射为普通变量 y_i , 并按式(13)计算 $f(y_i)$ 。

Step6 比较 $f(r_i)$ 、 $f(pBest_i)$ 与 $f(y_i)$, 以其间的最优值更新 $pBest_i$; 比较各 $f(pBest_i)$ 与 $f(gBest)$, 若最优, 更新 $gBest$, 否则保留原值。

Step7 判断是否满足终止条件(误差精度或迭代次数), 若是, 终止算法运行, 输出结果; 否则, 返回到 Step2, 继续运行。

2.3 CPSO-KPCA 方法

CPSO-KPCA 方法描述如下:

输入 建模数据集 $X_M = \{x_{M1}, x_{M2}, \dots, x_{Mi}, \dots, x_{Mn}\}$, 样本 $x_{Mi} = (x_{i1}, x_{i2}, \dots, x_{id})$, d 表示样本维数; 训练样本集 X_{TR} , 训练样本集类别集 C_{TR} ; 测试样本集 X_{TE}

输出 测试样本集 X_{TE} 的类别

Step1 将建模数据集进行标准化。

Step2 输入训练样本集 X_{TR} , 通过 CPSO 寻优核参数。

Step3 按式(5)计算建模数据集核矩阵 K , 并对核矩阵 K 作均值化处理, 得到 \bar{K} :

$$\bar{K}_{ij} = (K - I_n K - K I_n + I_n K I_n)_{ij} \quad (16)$$

其中, I_n 是元素为 1/n 的 $n \times n$ 常数方阵。

Step4 解决 $n\lambda a = Ka$ 的特征值问题, 并标准化 a^k , 使 $\lambda < a^k, a^k \geq 1$, 按式(7)计算控制限 SPE_a 。

Step5 输入测试样本集 X_{TE} , 按照正常状态下模型的均值和方差进行标准化。

Step6 对标准化后的一个测试样本 $x_{ie} \in \mathbf{R}^d$, 计算其内核向量 $k_{ie} = \langle x_{ie}, x_{Mi} \rangle$, 并进行均值中心化处理 \bar{k}_{ie} , $\bar{k}_{ie} = k_{ie} - I_n K - k_{ie} I_n + I_n K I_n$, 其中, I_n 是元素为 1/n 的 n 维

行向量。

Step7 按式(4)提取测试样本的非线性主元,并按式(6)计算测试样本的 SPE 统计量。

Step8 检测测试样本的 SPE 统计量是否超过控制限 SPE_a , 进行类别的判别。

3 实验及结果分析

油中溶解气体分析的主要特征气体有氢气(H_2)、甲烷(CH_4)、乙烷(C_2H_6)、乙烯(C_2H_4)、乙炔(C_2H_2)。正常样本来自江西某供电局所辖变压器 DGA 数据,共 520 例,故障样本取自公开发表文献,整理得到故障样本 260 例。实验时,先从正常工况 520 个样本中随机选择 100 个样本作为建模样本集,然后,从剩余 420 个正常样本中随机选择 210 个样本并 260 个故障样本中随机选择 130 个样本作为测试样本集,其余 340 个样本作为训练样本集。

评价指标如下:

(1)全局的检测正确率:

$$accuracy = \frac{|TP|+|TN|}{|TP|+|FP|+|TN|+|FN|} \quad (17)$$

(2)故障样本和正常样本的检测正确率见式(14)和式(15)。

(3)几何均值:

$$G_mean = \sqrt{tp \cdot tn} \quad (18)$$

几何均值综合考虑了 2 个类的检测性能。

图 1 为训练样本集采用 PSO 和 CPSO 方法对 KPCA 核参数优化收敛曲线。可以看出,PSO 方法早期收敛速度较快,后期收敛速度变慢,最终所得到的优化结果比 CPSO 方法的优化结果差。CPSO 方法借助混沌变量的特性,防止了某些粒子在迭代中出现停滞而导致的 PSO 方法早熟现象,其收敛速度较慢,但搜索范围较广,最终所得到的优化结果较理想。PSO 方法寻优核参数 $\gamma = 1.7247$, 最优适应度函数值 $f(\gamma) = 0.0471$; CPSO 寻优核参数 $\gamma = 0.8361$, 最优适应度函数值 $f(\gamma) = 0.0407$ 。

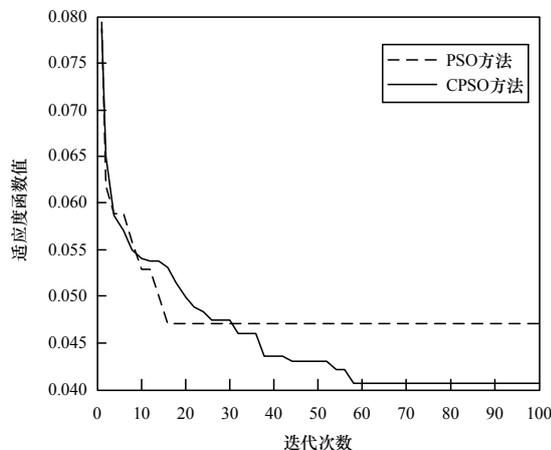


图 1 KPCA 核参数优化收敛曲线

表 1 为测试样本集应用到 PCA、KPCA、PSO-KPCA 和 CPSO-KPCA 方法的检测结果。可以看出,PCA 方法的

正常样本检测正确率 tn 最小,为 0.957 1,而 KPCA、PSO-KPCA 和 CPSO-KPCA 方法的 tn 均大于 97%,且相差不大,说明核参数的变化对正常样本的检测正确率 tn 的影响不大;对故障样本的检测正确率 tp ,CPSO-KPCA 方法的 tp 最大,为 0.9,大于 PCA 方法的 0.884 6 和 PSO-KPCA 方法的 0.869 2,核参数选择不当,如 $r=0.1$ 或 10,KPCA 方法的 tp 会显著降低,导致全局检测正确率和几何均值也大大下降,甚至低于 PCA 方法的检测效率;而核参数选择比较恰当,如当 $r=1$ 时,KPCA 方法的 tp 为 0.892 3,稍微高于 PCA 和 PSO-KPCA 方法,此时,全局检测正确率和几何均值均稍好于 PCA 和 PSO-KPCA 方法,这说明核参数的选取对故障样本的检测正确率 tp 有较大影响,进而影响全局检测正确率和几何均值;CPSO-KPCA 方法的全局检测正确率为 95%,几何均值为 93.96%,均高于 PCA、KPCA 和 PSO-KPCA 方法,说明该方法是有效的。

表 1 故障检测正确率比较

方法	tn	tp	$accuracy$	G_mean	
PCA	0.957 1	0.884 6	0.929 4	0.920 1	
KPCA	$r=0.1$	0.971 4	0.007 7	0.602 9	0.086 5
	$r=1$	0.981 0	0.892 3	0.947 1	0.935 6
PSO-KPCA	$r=10$	0.976 2	0.276 9	0.708 8	0.519 9
		0.976 2	0.869 2	0.935 3	0.921 1
CPSO-KPCA	0.981 0	0.900 0	0.950 0	0.939 6	

4 结束语

本文提出一种基于核主元分析和混沌粒子群优化算法的油浸式电力变压器故障检测方法,利用混沌运动的随机性、遍历性等特点,改善了粒子群优化算法摆脱局部极值点的能力,提高了检测精度。对 KPCA 核参数利用 CPSO 方法寻优,采用 SPE 统计量进行检测,可实现变压器故障检测的目的,该方法的检测效率达到 95%,几何均值为 93.96%。实验结果表明,与 PCA、KPCA 和 PSO-KPCA 方法相比,该方法具有更好的检测性能。下一步将研究如何解决其在 2 类数据不平衡条件下存在的问题,以提高故障诊断和检测效率。

参考文献

- [1] 肖应旺,徐保国.改进 PCA 在发酵过程监测与故障诊断中的应用[J].控制与决策,2005,20(5):571-574.
- [2] 张曦,阎威武,刘振亚,等.基于核主元分析和邻近支持向量机的汽轮机凝汽器过程监控和故障诊断[J].中国电机工程学报,2007,27(14):56-61.
- [3] 齐咏生,王普,高学金,等.改进 MKPCA 方法及其在发酵过程监控中的应用[J].仪器仪表学报,2009,30(12):2530-2538.
- [4] Lee J M, Yoo C, Choi S W, et al. Non-linear Process Monitoring Using Kernel Principal Component Analysis[J]. Chemical Engineering Science, 2004, 59(1): 223-234.

(下转第 250 页)