

基于粒子群优化 KFCM 的风电齿轮箱故障诊断*

李 状¹, 柳亦兵¹, 滕 伟¹, 林 杨^{1,2}

(1. 华北电力大学能源动力与机械工程学院 北京, 102206) (2. 北京华能新锐控制技术有限公司 北京, 102209)

摘要 针对基于有监督学习的方法无法识别未知类别故障, 提出了一种基于粒子群优化模糊核聚类(kernel fuzzy c-means clustering, 简称 KFCM)的风电机组齿轮箱故障诊断方法。首先, 建立以训练样本分类错误率为目标的聚类模型, 利用 KFCM 对训练样本进行分类; 然后, 以初始聚类中心和核函数参数作为优化变量, 利用粒子群优化算法求解聚类模型, 获得最优分类结果下每个类的类心; 最后, 根据新样本与各类心之间的核空间样本相似度判断新样本属于已知故障或者未知故障。以某风电机组齿轮箱为例, 对提出方法的有效性进行试验验证。结果表明, 与传统基于有监督学习的神经网络方法相比, 该方法能有效诊断已知和未知类别的故障。

关键词 模糊核聚类; 粒子群优化算法; 风电机组; 齿轮箱; 故障诊断

中图分类号 TH165.3; TH17

引 言

齿轮箱是风电机组传动链的关键设备, 由于风电机组运行环境恶劣, 导致齿轮箱故障频繁发生, 不仅事后维修工作难度大, 而且费用高。因此, 对风电机组齿轮箱进行有效的监测和故障诊断, 保障风电机组齿轮箱安全可靠运行具有重要意义。

故障诊断的本质是对故障信息的模式识别与分类^[1]。近年来国内外的许多学者将模式识别方法引入到风电齿轮箱故障诊断中, 目前大量的研究多采用基于有监督学习的模式识别方法, 例如 BP 神经网络、支持向量机、故障树分析法^[2-7]等, 通过对大量已知故障类别的训练样本的学习, 实现对待识别样本的分类与诊断。但是这种方法只能识别训练样本中已知故障类别, 如果对未知故障类别的样本进行分类识别, 则会被分类到已知故障类别中, 导致错误诊断。在实际故障诊断中, 获取所有已知故障的样本数据需要花费大量的人力和时间, 获取完备的故障样本十分困难, 这样使得基于有监督学习的模式识别方法难以开展有效的应用。

KFCM 是一种基于无监督学习的聚类算法, 利用核函数将原始输入空间中的样本数据映射到高维特征空间中再进行模糊聚类的方法, 能有效改善复

杂数据集的聚类性能^[8], 近年来在气动系统、汽轮机、变压器及卫星控制系统^[9-12]等故障诊断领域得到应用。但在实际工程应用中, KFCM 的分类效果依赖初始聚类中心和核函数参数的选择。针对此问题, 在 KFCM 聚类中引入粒子群优化算法求解最优解。粒子群优化(particle swarm optimization, 简称 PSO)算法是一种基于群智能的优化算法^[13], 该方法通过记忆与反馈机制实现高效的寻优搜索。

笔者提出一种基于粒子群优化 KFCM 的风电机组齿轮箱故障诊断方法。将该法应用于风电机组齿轮箱故障诊断中, 验证其在工程应用中的有效性。

1 KFCM 算法原理

KFCM 利用核函数通过非线性映射将原始特征空间 X 映射至高维特征空间 F 后再进行聚类。非线性映射 Φ 定义为

$$\Phi: x_k \rightarrow \Phi(x_k) \in F \quad (1)$$

其中: x_k 为原始特征空间样本, $x_k \in X$ 。

KFCM 算法的聚类目标函数表示为

$$J_m = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m \|\Phi(x_k) - \Phi(v_i)\|^2 \quad (2)$$

其中: $v_i (i=1, 2, \dots, c)$ 为原始特征空间的聚类中心; c 为类别数; n 为原始特征空间样本数量; μ_{ik} 为

* 国家自然科学基金资助项目(51305135); 中国华能集团科技资助项目(HNKJ13-H20-05); 中央高校基本科研业务费专项资金资助项目(2014XS15)

收稿日期: 2015-03-31; 修回日期: 2015-08-14

第 k 个样本 x_k 对第 i 类的隶属度; m 为加权指数;

$$\mu_{ik} \in [0,1], 0 < \sum_{k=1}^n \mu_{ik} < n.$$

式(2)的约束条件为

$$\sum_{i=1}^c \mu_{ik} = 1 \quad (k=1,2,\dots,n)$$

定义核函数 $K(x,y) = \Phi^T(x)\Phi(y)$, 核空间的欧式距离为

$$\|\Phi(x_k) - \Phi(v_i)\|^2 = K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i) \quad (3)$$

常见的核函数有高斯核函数、多项式核函数及 Sigmoid 核函数等, 笔者选用高斯核函数

$$K(x,y) = \exp[-\|x-y\|^2/(2\sigma^2)] \quad (4)$$

其中: σ 为高斯核参数。

根据约束条件, 结合式(3)和式(4), 运用拉格朗日乘法求式(2)的极小值, 可求得隶属度和聚类中心

$$u_{ik} = \frac{[1/(1 - K(x_k, v_i))]^{1/(m-1)}}{\sum_{j=1}^c [1/(1 - K(x_k, v_j))]^{1/(m-1)}} \quad (5)$$

$$v_i = \sum_{k=1}^n \mu_{ik}^m K(x_k, v_i) x_k / \sum_{k=1}^n \mu_{ik}^m K(x_k, v_i) \quad (6)$$

2 粒子群优化算法原理

定义在 D 维空间中的微粒群体规模为 N , 其中的一个粒子 x_i 的坐标位置可以表示为 $x_i = (x_i^1, x_i^2, \dots, x_i^D)$, $i=1,2,\dots,N$ 。在第 t 次迭代中, 粒子 x_i 移动速度为 $v_i(t) = (v_i^1, v_i^2, \dots, v_i^D)$, 当前粒子 x_i 的最优位置为 $l_i(t) = (l_i^1, l_i^2, \dots, l_i^D)$, 整个粒子群的全局最优位置表示为 $p_i(t) = (p_i^1, p_i^2, \dots, p_i^D)$ 。粒子群中每个粒子的移动速度和位置更新公式为

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 (l_i(t) - x_i(t)) + c_2 r_2 (p_i(t) - x_i(t)) \quad (7)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (8)$$

其中: ω 为惯性权重因子; c_1 和 c_2 为加速常数; r_1 和 r_2 为区间 $[0,1]$ 内的随机数。

3 粒子群优化 KFCM 故障诊断方法

3.1 粒子群算法求解 KFCM 聚类模型

首先利用 KFCM 方法对已知类别的训练样本进行分类。定义一个含有 c 类、样本特征维数为 d 的训练样本 X , 以训练样本的分类错误率评价聚类有效性, 并以此为聚类目标建立聚类模型, 根据文献[14], 训练样本的分类错误率 W 为

$$W = 1 - \frac{1}{c} \sum_{j=1}^c \frac{|C_j \cap C_j^L|}{|C_j^L|} \quad (9)$$

其中: C_j 为数据集 X 经过 KFCM 聚类运算后分在第 i 类的样本集; C_j^L 和 $|C_j^L|$ 分别为数据集 X 中第 j 类的样本集和所含样本数量; $|C_j \cap C_j^L|$ 为 C_j 和 C_j^L 交集所含有的元素数量。

利用粒子群优化算法求解聚类模型, 以初始聚类中心和核函数参数作为优化变量, 定义初始聚类中心为 $Z_j = (z_j^1, z_j^2, \dots, z_j^k, \dots, z_j^D)$, 核函数参数为 σ , $j=1,2,\dots,c, k=1,2,\dots,D$, 粒子的编码为

$$x_i = (z_1^1, z_1^2, \dots, z_j^k, \dots, z_c^D, \sigma) \quad (10)$$

以式(9)作为目标函数, 搜索其极小值。粒子群算法求解 KFCM 聚类模型流程如图 1 所示。

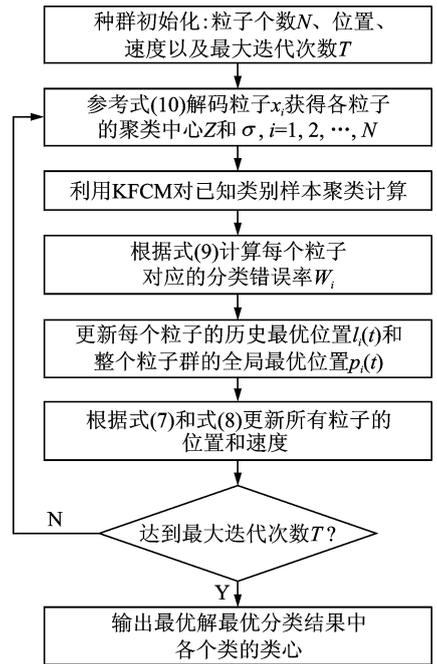


图 1 粒子群算法求解 KFCM 聚类模型流程图

Fig. 1 Flowchart of PSO algorithm optimizing KFCM clustering model

3.2 故障诊断流程

基于粒子群优化 KFCM 的故障诊断流程如图 2 所示, 具体步骤如下。

- 1) 获取已知 c 类故障的历史训练样本集 S 。
- 2) 利用 KFCM 对训练样本集进行分类, 粒子群算法求解 KFCM 聚类模型, 获得最优分类结果的每个类的类心 $o_j, j=1,2,\dots,c$ 。

3) 对于待诊断的新样本 s_{new} , 根据式(11)和式(12)判断 s_{new} 是否属于已知类别样本

$$\rho_j = \exp\left[-\frac{d_h(s_{new}, o_j) - \text{mean}\{d_h(s, o_j)\}}{\text{mean}\{d_h(s, o_j)\}}\right] \quad (11)$$

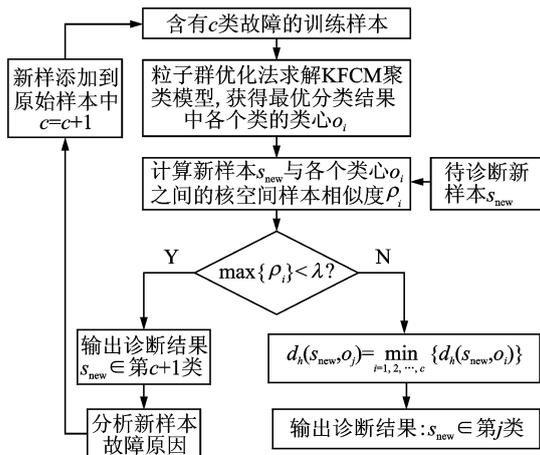


图2 基于粒子群优化 KFCM 的故障诊断流程图

Fig. 2 Flowchart of fault diagnosis based on KFCM optimized with PSO

$$\begin{cases} \max\{\rho_j\} \geq \lambda & (s_{\text{new}} \in \text{已知故障}) \\ \max\{\rho_j\} < \lambda & (s_{\text{new}} \in \text{未知故障}) \end{cases} \quad (12)$$

其中: ρ_j 为新样本 s_{new} 与类心 o_j 之间的核空间样本相似度 ρ_j [11]; d_h 为核空间上的欧式距离; C_j 为第 j 类中的样本; λ 为阈值常数, 取值范围为 $0 \sim 0.5$ [14], $j=1, 2, \dots, c$ 。

4) 如果 s_{new} 属于已知故障, 则根据 s_{new} 与各类心之间核空间上的欧式距离进一步诊断, 若 $d_h(s_{\text{new}}, o_m) = \min_{j=1, 2, \dots, c} \{d_h(s_{\text{new}}, o_j)\}$, 则 $s_{\text{new}} \in$ 第 m 类, $m=1, 2, \dots, c$ 。

若 s_{new} 不属于已知故障, 则定义 $s_{\text{new}} \in$ 第 $c+1$ 类, 可以经过后续分析给出故障原因, 然后作为已知故障样本添加在训练样本中, 并返回步骤 2。

4 实例分析

4.1 风电机组齿轮箱测试描述

以某风电场 1.5 MW 双馈风力发电机组为研究对象, 风轮工作转速范围为 $11 \sim 21$ r/min, 齿轮箱结构采用一级行星轮与两级平行轴结合的传动方案, 结构简图如图 3 所示。所选的风电机组齿轮箱在运行过程中曾出现中速级小齿轮裂纹故障(记为 F_1)、高速级小齿轮点蚀故障(记为 F_2)和高速输出轴轴承内圈故障(记为 F_3)。在齿轮箱高速轴轴承座位置安装了压电加速度传感器, 采集了风轮工作转速下的正常运行状态(记为 N)和 3 种故障状态的加速度信号, 信号采样频率为 8 192 Hz。图 4 为 4 种状态下的部分原始时域波形图。为了验证提出的方法, 状态 N, F_1 和 F_2 作为已知类别的状态, 并分

别从 3 种状态信号中选取 30 组样本作为训练样本, 3 组样本作为测试数据; 状态 F_3 作为未知故障, 从振动信号中选取 3 组样本作为测试数据, 每个样本采样点数为 2 048。

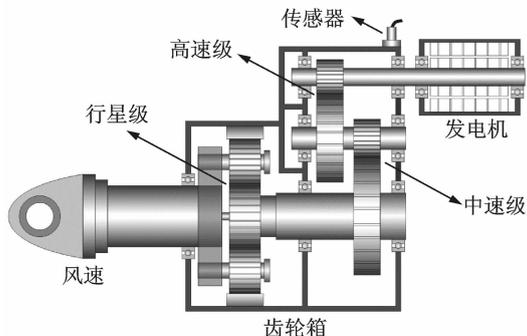


图3 风电机组齿轮箱结构示意图

Fig. 3 Structure of wind turbine gearbox

4.2 特征值提取

考虑风电齿轮箱振动信号具有非平稳性特点, 选用相对小波包能量作为风电机组齿轮箱振动信号特征值。小波包变换能够对非平稳信号进行有效地分析, 相对小波包能量能够反应信号在不同频带内的能量分布, 文献[15]给出了相对小波包能量特征值提取方法, 定义如下。

离散时间信号 $x(t)$ 经过 J 层小波包变换后得到 2^J 个频率段的小波包系数 $C_m(k)$, $m=0, 1, \dots, 2^J-1$, 某一频段上的小波包能量定义为该频段的小波系数的平方和

$$E_m = \sum_k |C_m(k)|^2 \quad (13)$$

因此, 所有频段的总能量为

$$E_{\text{total}} = \sum_{m=0}^{2^J-1} E_m \quad (14)$$

相对小波包能量为

$$\rho_m = E_m / E_{\text{total}} \quad (15)$$

对每个样本进行 3 层小波包变换, 然后根据式(13)~式(15)计算每个频段上的相对小波包能量 $[\rho_0, \rho_1, \dots, \rho_7]$, 用这 8 个频段的相对小波能量作为每个样本信号的特征值。

4.3 KFCM 和粒子群算法参数设置

根据图 1 的粒子群算法求解 KFCM 聚类模型流程, 结合训练样本中故障类型数量以及特征值的数量, KFCM 和粒子群算法的参数设置如下: 类别数量 $c=3$; 特征维数 $d=8$; 加权指数 $m=2$; 群体粒子个数 $N=50$; 最大迭代次数 $T=100$; 每个粒子的初始速度 $v_i=0$, $i=1, 2, \dots, N$; 根据文献[3], 加速

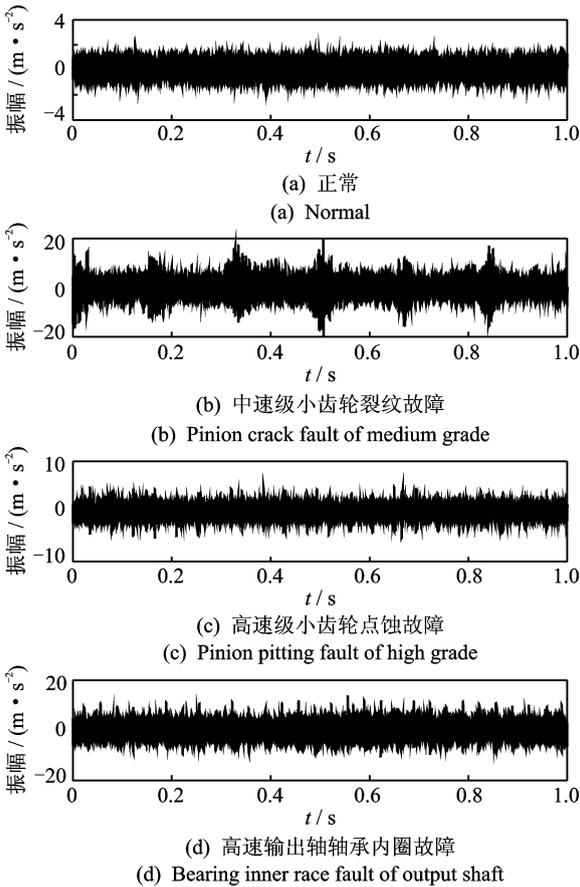


图 4 4 种状态振动加速度信号时域波形

Fig. 4 Vibration signals of 4 wind turbine gearboxes

表 1 最优分类结果下各状态的类心和核函数参数 σ

Tab. 1 The centers and the parameter σ of the optimized classification

状态类型	特征值($\times 10^{-2}$)								核函数参数 σ
N	38.70	36.24	4.08	16.47	0.90	0.66	1.76	1.17	
F ₁	18.35	35.58	4.81	35.42	1.35	0.55	2.77	1.18	25.798 3
F ₂	18.51	35.19	4.81	35.58	1.33	0.56	2.83	1.18	

献[14]给出的 λ 取值范围,笔者选取阈值常数 $\lambda = 0.2$ 。为了进一步证明提出方法的有效性,分别采用 BP 神经网络和 KFCM 方法进行故障诊断,并且与提出的方法分析结果进行对比,3 种方法的诊断结果如表 2 所示。

从表 2 可以看出,对已知类别的测试样本(样本 1~9)进行分类诊断时,提出的方法和基于有监督学习的 BP 神经网络均可以正确地对其分类与识别;对未知类别的测试样本(样本 10~12)进行分类时,BP 神经网络将测试样本分类在已知的故障类别中,与实际结果不符,这是因为 BP 神经网络只记忆了训练样本中的类别,因此给出了错误的诊断结果。所提出的方法将未知类别样本分在第 4 类,即不属于原有训练样本中的已知故障类别。直接利用 KFCM 算法由于受到初始聚类中心和核函数参数选取的影响,在对已知类别和未知类别的测试样本分类

常数 $c_1 = 2, c_2 = 2$; 最大限制速度 $v_{max} = 1$ 。

惯性权重因子调整公式为

$$\omega(t+1) = \omega(t) - t(\omega_{max} - \omega_{min}) / (T - 1) \quad (16)$$

其中: t 为当前迭代数; $\omega_{max} = 1; \omega_{min} = 0.2$ 。

4.4 诊断结果与分析

利用 KFCM 对训练样本进行分类,并利用粒子群算法求解聚类模型,训练样本的分类错误率 W 与迭代次数关系如图 5 所示。从图中可以看出,当粒子群算法迭代到第 7 步时分类错误率达到最小值,此时分类错误率为 5.6%,最优分类结果下每个状态的类心和核函数参数 σ 如表 1 所示。

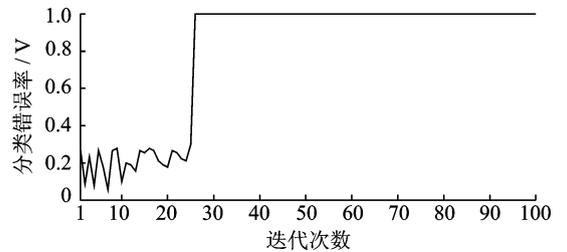


图 5 分类错误率变化曲线

Fig. 5 Curve of the classification error rate

3 种已知状态 N, F₁ 和 F₂ 分别编号为 1, 2, 3。根据图 2 所示的诊断流程,对测试样本(包含 3 类已知类别状态和 1 类未知故障)进行分类诊断。根据文

时均出现了错误分类。

表 2 不同方法诊断结果

Tab. 2 Diagnosis results of different methods

样本序号	类型(编号)	诊断结果		
		BP	KFCM	PSO+KFCM
1		1	1	1
2	N(1)	1	1	1
3		1	1	1
4		2	2	2
5	F ₁ (2)	2	3	2
6		2	3	2
7		3	3	3
8	F ₂ (3)	3	3	3
9		3	3	3
10		3	3	3
11	F ₃ (未知)	3	4	4
12		2	3	4

5 结束语

提出了一种基于粒子群优化 KFCM 的风电齿轮箱故障诊断方法。结合实际风电齿轮箱故障样本数据对提出的方法进行验证,并与传统 BP 神经网络和 KFCM 分类方法进行比较。结果表明,传统 BP 神经网络只能识别训练样本中已有的故障类别,而无法识别未知的故障;KFCM 方法受到初始聚类中心和核函数参数选取的影响,分类准确度不高;提出的方法通过粒子群优化算法求解 KFCM 聚类模型获得最优分类结果对应的各个类的类心以及核函数参数,提高了分类准确度,不仅能够准确地识别诊断出已知类别的故障样本,而且能有效地识别判断出未知类别的故障样本,为风电机组齿轮箱状态监测与故障诊断提供了一种新的思路。

参 考 文 献

- [1] Liu Wenyi, Wang Zhenfeng, Han Jiguang, et al. Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM [J]. *Renewable Energy*, 2013, 50: 1-6.
- [2] 司景萍,郭丽娜,牛家骅. 基于小波包神经网络的发动机振动信号分析[J]. *振动、测试与诊断*, 2013, 33(5): 763-769.
Si Jingping, Guo Lina, Niu Jiahua. Based on wavelet packet neural network of engine vibration signal analysis [J]. *Journal of Vibration, Measurement & Diagnosis*, 2013, 33(5): 763-769. (in Chinese)
- [3] 龙泉,刘永前,杨勇平. 基于粒子群优化 BP 神经网络的风电机组齿轮箱故障诊断方法[J]. *太阳能学报*, 2012, 33(1): 120-125.
Long Quang, Liu Yongqian, Yang Yongping. Fault diagnosis method of wind turbine gearbox based on bp neural network trained by particle swarm optimization algorithm [J]. *Acta Energetica Solaris Sinica*, 2012, 33(1): 120-125. (in Chinese)
- [4] 肖剑,周建中,张孝远,等. 基于 Levy-ABC 优化 SVM 的水电机组故障诊断方法[J]. *振动、测试与诊断*, 2013, 33(5): 839-844.
Xiao Jian, Zhou Jianzhong, Zhang Xiaoyuan, et al. Fault diagnosis for hydroelectric generator unit based on levy flight-artificial bee colony and support vector machine [J]. *Journal of Vibration, Measurement & Diagnosis*, 2013, 33(5): 839-844. (in Chinese)
- [5] 刘永斌,何清波,孔凡让,等. 基于 PCA 和 SVM 的内燃机故障诊断[J]. *振动、测试与诊断*, 2012, 32(2): 250-255.
Liu Yongbin, He Qingbo, Kong Fanrang, et al. Fault diagnosis of internal combustion engine using PCA and SVM [J]. *Journal of Vibration, Measurement & Diag-*

nosis, 2012, 32(2): 250-255. (in Chinese)

- [6] Yang Zhiling, Wang Bin, Dong Xinghui, et al. Expert system of fault diagnosis for gear box in wind turbine [J]. *Systems Engineering Procedia*, 2012, 4: 189-195.
- [7] 程珩,黄超勇,张永刚. 基于粒子群优化决策树的齿轮箱故障诊断[J]. *振动、测试与诊断*, 2013, 33(1): 153-156.
Cheng Hang, Huang Chaoyong, Zhang Yonggang. Constructed of SVM decision tree based on particle swarm optimization algorithm for gear box fault diagnosis [J]. *Journal of Vibration, Measurement & Diagnosis*, 2013, 33(1): 153-156. (in Chinese)
- [8] Lin Kuoping. A novel evolutionary kernel intuitionistic fuzzy c-means clustering algorithm [J]. *Fuzzy Systems IEEE Transactions on*, 2014, 22(5): 1074-1087.
- [9] 蒋全胜,贾民平,胡建中,等. 一种基于人工免疫的模糊核聚类算法 [J]. *中国机械工程*, 2008, 5: 594-597.
Jiang Quansheng, Jia Mingping, Hu Jianzhong, et al. A new artificial immunity based fuzzy kernel clustering algorithm [J]. *China Mechanical Engineering*, 2008, 5: 594-597. (in Chinese)
- [10] 黄保海,李岩,王东风,等. 基于 KPCA 和 KFCM 集成的汽轮机故障诊断 [J]. *电力自动化设备*, 2010, 7: 84-87.
Huang Baohai, Li Yan, Wang Dongfeng, et al. Steam turbine fault diagnosis based on KPCA and KFCM ensemble [J]. *Electric Power Automation Equipment*, 2010, 7: 84-87. (in Chinese)
- [11] Ma Hui, Ekanayake C, Saha T K. Power transformer fault diagnosis under measurement originated uncertainties [J]. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2012, 19(6): 1982-1990.
- [12] Hu Di, Sarosh A, Dong Fengyun. A novel KFCM based fault diagnosis method for unknown faults in satellite reaction wheels [J]. *ISA Transactions*, 2012, 51: 309-316.
- [13] Kennedy J, Eberhart R C. Particle swarm optimization [C] // *IEEE International Conference on Neural Networks*. Piscataway: IEEE, 1995: 1942-1948.
- [14] Li Chaoshun, Zhou Jianzhong. Semi-supervised weighted kernel clustering based on gravitational search for fault diagnosis [J]. *ISA Transactions*, 2014, 53: 1534-1543.
- [15] Shao Renping, Hu Wentao, Wang Yayun, et al. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform [J]. *Measurement*, 2014, 54: 118-132.



第一作者简介:李状,男,1987年8月生,博士生。主要研究方向为旋转机械设备故障诊断方法。曾发表《Crack fault detection for a gearbox using discrete wavelet transform and an adaptive resonance theory neural network》(《Strojnicki Vestnik-Journal of Mechanical Engineering》2015, Vol. 61, No. 1)等论文。
E-mail: lizhuang@ncepu.edu.cn