

# 基于小波包变换和 GA-PLS 算法的故障特征选择\*

张培林, 李 胜

(军械工程学院七系 石家庄, 050003)

**摘要** 为了提高故障诊断的分类精度,减小分类运算时间等问题,需要从原始特征集中选择出更为优化的特征子集合,因此,提出了一种基于小波包变换和 GA-PLS 算法的特征选择方法。首先,采用小波包变换对提取出的振动信号进行分解,从而得到小波包的分解系数;其次,运用遗传算法-偏最小二乘法从原始信号和小波包系数的统计学特征中选择出最优特征集;最后,将最优特征集作为输入,输入到支持向量机中以实现对不同故障的诊断与识别。应用于轴向柱塞泵故障诊断中,与现有特征选择方法对比,实验结果验证了本研究特征选择方法的有效性。

**关键词** 小波变换; 遗传算法-偏最小二乘法; 轴向柱塞泵; 特征选择

**中图分类号** TH137

## 引 言

特征选择可以有效地降低特征集的整体维数,减少分类时的运算量,从而提高分类准确率<sup>[1]</sup>。采用不同的数据分析方法可以在原始信号中提取特征组成原始特征集,而不同的特征选择方法会对其分类结果产生比较大的影响。文献[2]提出了综合样本矢量方向离散度概念,以此作为故障特征参数的优选准则。文献[3]通过最大化特征-类信息量和最小化特征-特征信息量去除冗余特征,但是该算法必须对各个特征量进行离散化处理,需要有一定的先验知识,而且计算量较大。

小波包分析可根据高频调制特性,对振动信号进行小波包变换和包络解调,从而提取故障特征<sup>[4-6]</sup>。这种方法虽然可以直接把从时域或频域提取出的特征作为输入,输入到分类器中进行分类,但是提取的特征集中可能包含着部分噪声、冗余特征或不相关特征。

目前,常用的特征选择方法是主成分分析法(principle component analysis, 简称 PCA),该方法可以用少数的特征来拟合原始特征,但缺点是用于线性问题的求解,不能反映原始数据对分析结果的非线性影响。遗传算法-偏最小二乘法(genetic algorithm-partial least square, 简称 GA-PLS)是 Leardi R<sup>[7-8]</sup>于 1994 年为了结合遗传算法(GA)的全局优化搜索能力和偏最小二乘法(PLS)有效解决

变量间多重共线性问题的能力而提出的。该算法主要用于非线性问题的求解,具有较好的变量选择及模型优化效果,被广泛地运用到有机化合物定量结构-活性相关(QSAR)、近红外技术等方面<sup>[9-10]</sup>,取得了较好特征选择和分类效果。

笔者将遗传算法-偏最小二乘法引入到特征集的选择,设计了基于小波包变换和 GA-PLS 算法的特征选择方法,以轴向柱塞泵的振动信号为例进行特征选择,与其他特征选择方法相比,实验结果验证了该方法的有效性,为轴向柱塞泵的特征选择提供了一种新方法。

## 1 理论基础

### 1.1 小波包变换

小波包变换可同时分解振动信号的低频部分和 高频部分,并且可以依据振动信号的主要特征,自适应地选择相应的频带以匹配信号的频谱,从而提高时频的分辨率<sup>[11-12]</sup>。

若  $V_0$  包含被分析的信号,将  $V_0$  分别投影到  $V_1$  和  $W_1$  上,其中: $V_1$  为低频分量; $W_1$  为高频分量; $V_1$  空间进一步分解为  $V_2$  和  $W_2$ 。因此,将一个子空间  $V_j$  分解为空间  $V_{j+1}$  和  $W_{j+1}$ ,即

$$V_j = V_{j+1} + W_{j+1} \quad (1)$$

在小波包分解过程中,不仅将  $V_j$  分解,还将  $W_j$  也进行了分解。令  $V_0$  开始分解,则  $W_0^0 = V_0$  和

$\Psi_0^0(t) = \varphi(t)$ , 进而得到  $W_0^0$  分解为  $W_1^0$  和  $W_1^1$ , 且  $W_0^0 \oplus W_1^1 = W_0^0$ 。若继续分解下去, 在高层  $W_j^p$  被分解为  $W_{j+1}^{2p}$  和  $W_{j+1}^{2p+1}$  两个子空间, 对应  $W_j^p$  的基为  $\{\psi_j^p(t-2^{j+1}n)\}_{n \in \mathbb{Z}}$ ;  $W_{j+1}^{2p}$  和  $W_{j+1}^{2p+1}$  的基分别为  $\{\psi_{j+1}^{2p}(t-2^{j+1}n)\}_{n \in \mathbb{Z}}$  和  $\{\psi_{j+1}^{2p+1}(t-2^{j+1}n)\}_{n \in \mathbb{Z}}$ , 其中, 基函数满足二尺度方程

$$\begin{cases} \Psi_{j+1}^{2p}(t) = \sum_{n=-\infty}^{+\infty} h(n)\Psi_j^p(t-2^jn) \\ \Psi_{j+1}^{2p+1}(t) = \sum_{n=-\infty}^{+\infty} g(n)\Psi_j^p(t-2^jn) \end{cases} \quad (2)$$

且满足  $W_{j+1}^{2p} \perp W_{j+1}^{2p+1}$  和  $W_{j+1}^{2p} \perp W_{j+1}^{2p+1}$ ,  $h[n]$  和  $g[n]$  为共轭镜像和滤波器,  $g[n] = (-1)^{1-n} \cdot h[1-n]$ 。

因此, 小波包分解的表达式如下

$$V_j = W_{j-k}^{2^k} \oplus W_{j-k}^{2^k+1} \oplus \dots \oplus W_{j-k}^{2^{k+1}-1} \quad (3)$$

随着尺度  $j$  的增加, 子空间  $W_j^p$  的数目增加, 频带划分愈细<sup>[11]</sup>。

图 1 列出了小波包分解对一个信号进行三层分解的树状图。

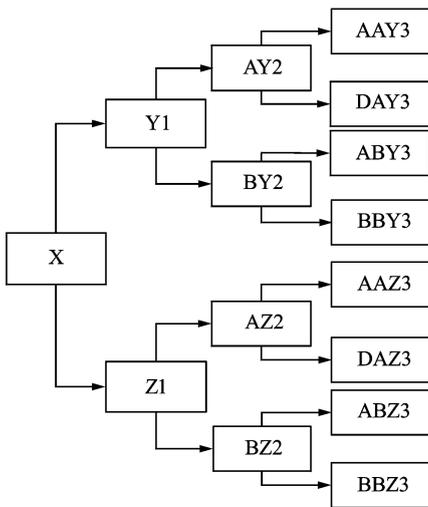


图 1 小波包分解图示

Fig. 1 The figure of wavelet packet decomposition

## 1.2 GA-PLS 算法

为了达到更好的特征选择效果, Leardi 等<sup>[7-8]</sup>提出了把遗传算法 (genetic algorithm, 简称 GA) 和偏最小二乘法 (partial least square, 简称 PLS) 结合在一起来, 组成了一个新的变量选择方法——遗传算法-偏最小二乘法 (GA-PLS), 并将该方法用于波谱数据中波长的特征选择中, 取得了较好的效果。由于该方法结合了遗传算法的全局优化搜索能力和偏最小二乘法, 有效地解决变量间多重共线性问题的能

力, 因此能够对特征进行有效的筛选。

GA-PLS 方法选择特征的基本过程<sup>[8]</sup>如下。

1) 随机初始化染色体, 染色体长度应等于特征的个数, 赋二进制值给每个染色体, 字符串中的字符对应一个特征, 其中选中相应的变量用 1 表示, 未被选中相应的变量用 0 表示。

2) 评价个体适应度, 适应度表示特征对模型性能的影响, 其中适应度值较高的特征被保留, 而适应度值低的特征就被删除。文中的适应度函数如下所示

$$F_{\text{fitness}} = \frac{1 - (n-1)(1-q^2)}{(n-c)} \quad (4)$$

$$q^2 = 1 - H_{\text{PRESS}}/h_{\text{SSY}} \quad (5)$$

其中:  $n$  为特征个数;  $q^2$  为通过留一法得到的交叉验证值;  $c$  为被选择出的特征个数;  $h_{\text{SSY}}$  为变量的平方和;  $H_{\text{PRESS}}$  为留一法交叉验证后的预测平方和。

3) 根据求解的适应度值进行遗传操作。复制适应度值较高的个体到下一代, 适应度低的则删除。交叉操作交换互相配对的两个个体的部分基因以产生新的个体, 变异操作改变个体的基因造成基因突变达到产生新个体的目的。文中复制操作采用比例选择法, 交叉操作采用单点交叉法, 变异插座采用基本位变异。

4) 偏最小二乘法的回归模型确定通过遗传算法选择的特征, 以避免单独采用遗传算法而引起过拟合, 从而简化模型。

5) 终止条件。文中 GA-PLS 算法的终止条件是最大进化代数。

在本研究中, 根据经验, 对影响 GA-PLS 算法性能的参数设置如下: 种群大小为 30, 每条染色体平均由 5 个变量组成, 迭代次数为 100, 交叉概率为 0.5, 突变概率为 0.01, 遗传迭代次数为 100。

## 2 特征选择流程

对一个特征选择模型来说, 它能够在大量的特征中剔除不相关和冗余特征, 提取出有效的特征, 为分类奠定基础, 以达到减少运算时间、提高分类准确率的目的。因此, 特征选择的基本思想是: 原始振动信号的采集, 通过采用小波包变换对信号进行分解得到小波包系数; 同时, 对原始信号和小波包系数提取统计特征量, 组成原始特征集; 利用 GA-PLS 算法对原始特征集进行特征选择, 将选择出的最优特征集输入到支持向量机, 最后得到分类的最终结果。特征选择的流程图如图 2 所示。

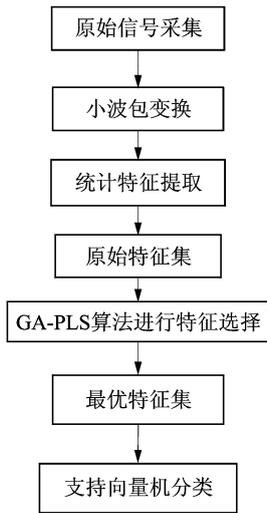


图 2 特征选择流程图

Fig. 2 The flow chart of feature selection

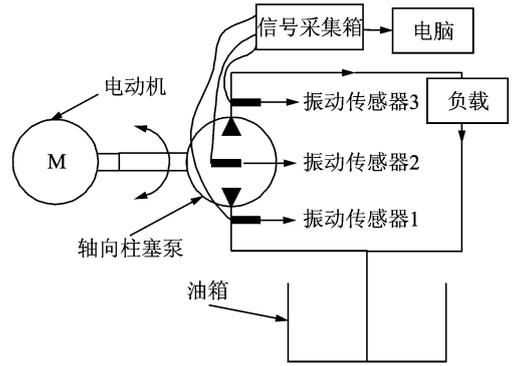


图 3 轴向柱塞泵实验装置简图

Fig. 3 Experiment device sketch of axial piston pump

### 3 实例分析

#### 3.1 特征提取和分类器的选择

为了验证该特征选择方法的有效性和实用性,对轴向柱塞泵的几种主要故障在液压系统综合检测与试验设备上进行了试验分析。文中的试验数据都来自于综合检测试验设备。

在试验中,轴向柱塞泵的型号为 25MCY14-1B,理论排量为 25 mL/r,公称压力为 31.5 MPa,转速  $n=1\ 500\text{ r/min}$ 。传感器型号为 CA-YD-185,安装在轴向柱塞泵的进油口管道、出油口管道和轴向柱塞泵壳体上,实验负载为 15 MPa。实验装置如图 3 所示。在综合检测试验设备上模拟了轴向柱塞泵的 3 种工作状态:a. 正常状态;b. 缸体与配流盘磨损;c. 柱塞滑履松动。在采集信号中,采样频率为 4 kHz,每个样本的长度为 1 024 个点,对每种工作状态采集 50 个样本,其中 20 个作为训练样本,30 个作为测试样本。

对于一个时域信号,文中提取的特征是统计特征参数,即选取的特征参数集是由信号的 10 个统计特征(峰峰值、均方值、方差、方根幅值、均方幅值、峭度、波形指标、峰值指标、脉冲指标、边界指标)组成。

对于分类器的选择,由于支持向量机(SVM)能够较好地解决小样本学习问题,又可以利用核函数以降低算法的复杂度,因此,在文中采用支持向量机作为最终的分类器。对支持向量机内的参数设置如下:采用径向基函数作为核函数,  $\gamma=0.003$ ,  $\epsilon=0.04$ ,  $C=500$ 。

#### 3.2 小波包变换和 GA-PLS 算法

轴向柱塞泵的 3 种工作状态的时域波形如图 4

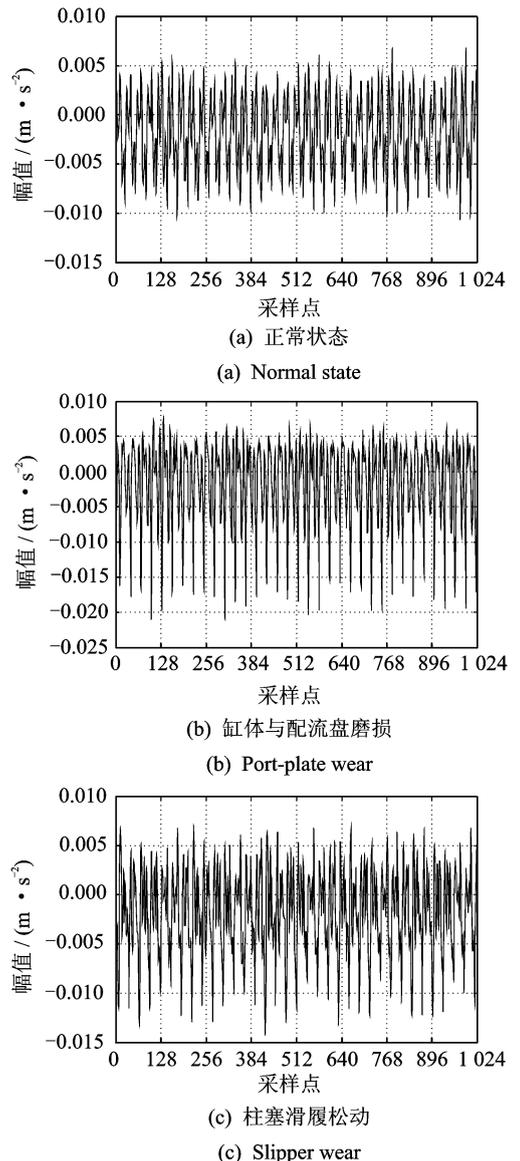


图 4 原始振动信号时域图

Fig. 4 Time domain of original vibration signals



表 2 柱塞滑履松动工作状态的原始特征集  
Tab. 2 Original feature set of slipper wear

参数	1	2	3	4	5	6	7	8	9
峰峰值	0.021 1	0.037 2	0.029 5	0.084 4	0.018 4	0.030 3	0.049 7	0.050 0	0.055 0
均方值	0.073 0	0.018 7	0.044 0	0.019 5	0.032 2	0.014 9	0.031 1	0.056 7	0.085 7
方差	0.072 9	0.018 4	0.043 4	0.019 2	0.031 7	0.014 6	0.030 6	0.055 2	0.084 4
方根幅值	0.026 1	0.088 0	0.042 8	0.028 5	0.036 6	0.024 9	0.035 9	0.055 5	0.018 9
均方幅值	0.854 4	0.136 6	0.663 5	0.441 3	0.567 2	0.386 1	0.557 5	0.153 0	0.292 7
峭度	-2.845 9	-2.869 1	-2.918 3	-3.159 2	-2.940 8	-2.894 7	-3.041 9	-3.071 8	-2.918 8
波形指标	32	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6
峰值指标	32	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6	11.417 6
脉冲指标	1 024	134	134	134	134	134	134	134	134
边界指标	1 048 491	17 854	17 854	17 854	17 854	17 854	17 854	17 854	17 854

从表 1 和表 2 可以看出,在这些特征中,存在着一些冗余和无关特征,需要去除出去,因此,采用 GA-PLS 方法可以剔除这些特征。

对原始特征集进行 GA-PLS 算法的特征选择,结果如图 7 和图 8 所示。图 7 和图 8 给出了各个特征的个体适应度值,依据经验,笔者选择适应度值大于 0.500 0 的特征作为最优特征集。

由于里面有冗余和不相关的特征,因此,经过合并和剔除个体适应度值不符合要求的特征,表 3 和表 4 列出了缸体与配流盘磨损工作状态和柱塞滑履松动工作状态的原始特征集用 GA-PLS 方法选择后的最优特征集。

表 3 缸体与配流盘磨损工作状态经 GA-PLS 方法选择后的最优特征集

Tab. 3 Optimal feature set of port-plate wear by GA-PLS

最优特征集					
0.028 6	0.032 1	0.053 3	0.011 3	0.026 0	0.008 4
0.006 8	0.010 2	0.498 4	0.339 1	0.012 1	0.029 6
0.013 3	0.011 0	0.024 8	0.065 9	0.011 5	0.042 3
0.011 9	0.029 1	0.013 1	0.365 0	0.104 7	0.064 9
0.205 6	0.041 7	0.070 9	0.035 1	0.023 5	0.067 5
0.032 1	0.052 3	0.021 9	0.810 2	0.346 6	0.110 0
0.543 9	134	11.569 4	17 951	-3.000 1	

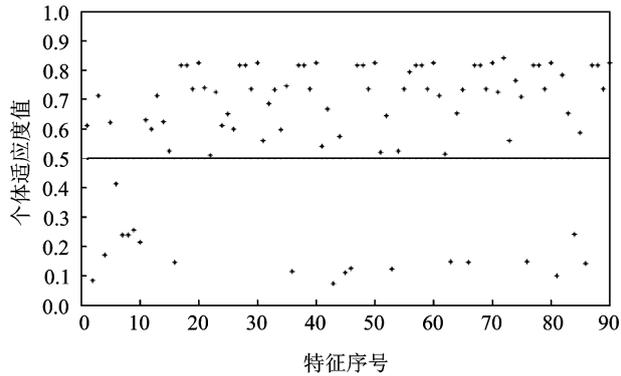


图 7 缸体与配流盘磨损工作状态各个特征的个体适应度值  
Fig. 7 The value of each feature in port-plate war

表 4 柱塞滑履松动工作状态经 GA-PLS 方法选择后的最优特征集

Tab. 4 Optimal feature set of slipper wear by GA-PLS

最优特征集					
0.021 1	0.037 2	0.029 5	0.084 4	0.018 4	0.073 0
0.026 1	0.088 0	0.042 8	0.032 2	0.036 6	0.024 9
0.854 4	0.136 6	0.663 5	0.441 3	0.567 2	0.386 1
0.031 1	0.056 7	0.085 7	0.049 7	0.055 0	0.055 5
0.018 4	0.043 4	-3.159 2	0.084 4	0.014 6	0.153 0
-2.845 9	-2.869 1	0.035 9	11.417 6	-3.071 8	0.019 2
1 048 491	17 854	0.557 5	0.292 7	134	

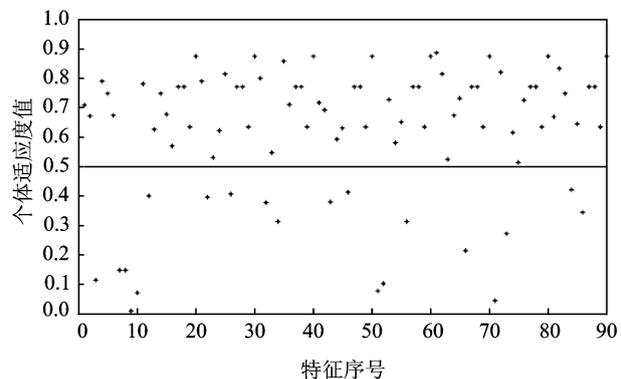


图 8 柱塞滑履松动工作状态各个特征的个体适应度值  
Fig. 8 The value of each feature in slipper wear

将表 1 和表 3 对比,表 2 和表 4 对比后可以得出,经过特征选择后,特征集的维数大大减少了,对缸体与配流盘磨损工作状态和柱塞滑履松动工作状态而言,由最初的 90 个特征减少为 41 个特征,这些

特征就能作为最优特征集。选择出的特征对以后的故障分类很有效,大大减轻了分类器的工作量,同时有助于提高分类器的分类精度。

为了突出本方法的优越性,采用主成分分析方法来进行实验对比。主成分分析(principle component analysis, 简称 PCA)方法是一种常用的特征选择方法。在文中,根据经验,当前  $s$  个主成分包含了全部变化方差的 85% 以上时,取  $s$  作为保留的主成分个数比较合适,用这  $s$  个主成分可以很好地拟合原始特征数据集。首先,对原始特征集中的冗余特征去掉,得到剩下的 60 个特征;然后,采用 PCA 方法对缸体与配流盘磨损工作状态和柱塞滑履松动工作状态的原始特征集进行特征选择;最后,根据 85% 的全部方差变化得到优化特征,如表 5,6 所示。

表 5 缸体与配流盘磨损工作状态经 PCA 方法选择后的最优特征集

Tab. 5 Optimal feature set of port-plate wear by PCA

最优特征集					
1 048 576	0.498 4	0.065 9	0.029 1	0.011 9	-2.999 7
17 951	0.365 0	0.064 9	0.028 6	0.011 5	-2.999 8
1 024	0.346 6	0.053 3	0.026 0	0.011 3	-2.999 9
134	0.339 1	0.052 3	0.0248	0.011 0	-3.000 1
32	0.205 6	0.042 3	0.024 5	0.010 8	-3.000 2
11.575 7	0.110 0	0.041 7	0.023 5	0.010 6	
11.569 4	0.104 7	0.035 1	0.021 9	0.010 2	
0.810 2	0.070 9	0.032 1	0.013 3	0.008 2	
0.543 9	0.067 5	0.029 6	0.012 1	0.006 8	

表 6 柱塞滑履松动工作状态经 PCA 方法选择后的最优特征集

Tab. 6 Optimal feature set of slipper wear by PCA

最优特征集					
1 048 491	0.557 5	0.073 0	0.037 2	0.021 1	-2.918 3
17 854	0.441 3	0.072 9	0.035 9	0.019 5	-2.918 8
1 024	0.386 1	0.056 7	0.032 2	0.018 9	-3.041 9
134	0.292 7	0.055 5	0.031 1	0.018 4	-3.071 8
32	0.153 0	0.055 2	0.030 3	0.014 9	-3.159 2
11.417 6	0.136 6	0.050 0	0.029 5	0.014 6	
0.854 4	0.088 0	0.049 7	0.028 5	-2.845 9	
0.663 5	0.085 7	0.044 0	0.026 1	-2.869 1	
0.567 2	0.084 4	0.042 8	0.024 9	-2.894 7	

从表 5 和表 6 可以看出,全部方差变化大于 85% 的特征为 50 个,也就是说,可以用 50 个主成分的数据信息来描述变换前的 90 维特征数据的变化状态。但是,与表 3 和表 4 相比,多了 9 个特征,数

量增加了 22%,这说明本特征选择方法要比 PCA 方法要有效得多。

### 3.3 实验结果分析

将表 3~表 6 的结果作为分类器的输入,选用分类能力较好的支持向量机作为分类器。在分类准确率和分类时间上,对文中方法、原始特征集和 PCA 方法进行对比,比较结果如表 7 所示。

表 7 支持向量机分类准确度和执行时间

Tab. 7 Classification accuracy and executing time

特征集	分类准确率/%	运行时间/s
原始特征集	90.1	154.9
经 PCA 选择后的特征集	94.4	89.1
经 GA-PLS 选择后的特征集	98.7	47.6

从表 7 可以看出,经过 GA-PLS 算法选择后的特征集的分类准确率比原始特征集高 8% 左右,且运行时间缩短了 3 倍,而且,与经 PCA 选择后的特征集相比,分类准确率也提高了 4.3%,分类时间缩短了 2 倍。这说明经过 GA-PLS 特征选择后,随着特征维数的数量大大减少,有利于进行分类,有效地提高了分类准确率并缩短了运算时间。

## 4 结 论

1) 小波包变换能够有效地把轴向柱塞泵的振动信号分解成各个频段的小波包分解系数,使得时域信息更加明显,为有效地提取特征奠定基础。

2) 利用 GA-PLS 算法能有效地剔除原始特征集中的冗余或不相关特征,选择出了优化的特征集,与 PCA 方法相比,大大提高分类准确率和缩短了运行时间。

### 参 考 文 献

- [1] 曹建军,张培林,任国全,等. 基于蚁群优化的振动信号特征选择[J]. 振动与冲击,2008,27(5):24-27.  
Cao Jianjun, Zhang Peilin, Ren Guoquan, et al. Vibration signal feature selection based on ant colony optimization[J]. Journal of Vibration and Sound, 2008, 27(5):24-27. (in Chinese)
- [2] 谢涛,张育林. 基于遗传算法与最大最小原理的故障模式特征选择[J]. 国防科技大学学报,1998,20(2):17-21.

- Xie Tao, Zhang Yulin. Max-min principle based-selection for the optimal feature parameters in fault diagnosis using genetic algorithms[J]. Journal of National University of Defense Technology, 1998, 20 (2): 17-21. (in Chinese)
- [3] 史东锋, 屈梁生. 遗传算法在故障特征选择中的应用研究[J]. 振动、测试与诊断, 2000, 20 (3): 171-176.  
Shi Dongfeng, Qu Liangsheng. Application of genetic algorithm to fault feature selection[J]. Journal of Vibration, Measurement & Diagnosis, 2000, 20 (3): 171-176. (in Chinese)
- [4] 荆双喜, 方佳雨, 孟惠荣. 小波包-神经网络在斜轴泵故障诊断中的应用研究[J]. 振动、测试与诊断, 2000, 20 (2): 97-101.  
Jing Shuangxi, Fang Jiayu, Meng Huirong. Fault diagnosis of a bend axis piston pump by wavelet packet and neural networks[J]. Journal of Vibration, Measurement & Diagnosis, 2000, 20(2): 97-101. (in Chinese)
- [5] 张玲玲, 赵懿冠, 肖云魁, 等. 基于小波包-AR 谱的变速器轴承故障特征提取[J]. 振动、测试与诊断, 2011, 31 (4): 492-495.  
Zhang Lingling, Zhao Yiguan, Xiao Yunkui, et al. Automobile transmission bearing fault feature extraction based on wavelet packet-auto regressive model spectrum[J]. Journal of Vibration, Measurement & Diagnosis, 2011, 31 (4): 492-495. (in Chinese)
- [6] 牛雪梅, 冀鹏飞, 杨阳, 等. 双供油轴向柱塞泵压力流量的脉动特性[J]. 振动、测试与诊断, 2012, 32 (1): 151-156.  
Niu Xuemei, Ji Pengfei, Yang Yang, et al. Pressure and flow pulsation characteristics of dual discharging axial piston pump[J]. Journal of Vibration, Measurement & Diagnosis, 2012, 32 (1): 151-156. (in Chinese)
- [7] Leardi R. Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection [J]. Journal of Chemometrics, 1994, 8 (1): 65-79.
- [8] Gonzalez A, Lupiaffiez L R. Genetic algorithms applied to feature selection in PLS regression: how and when to use them [J]. Chemometrics and Intelligent Laboratory Systems, 1998, 41 (2): 195-207.
- [9] Wanchana S, Yamashita F, Hashida M. Quantitative structure/property relationship analysis of Caco-2 permeability using a genetic algorithm-based partial least squares method [J]. Journal of Pharmaceutical Sciences, 2002, 91 (10): 2230-2239.
- [10] Han S H, Yang H. Screening important design variables for building a usability model: genetic algorithm-based partial least-squares approach [J]. International Journal of Industrial Ergonomics, 2004, 33 (2): 159-171.
- [11] 邓卫强, 王跃钢, 杨颖涛, 等. 一种改进的 ARMA 模型参数估计方法[J]. 振动、测试与诊断, 2011, 31 (3): 377-380.  
Deng Weiqiang, Wang Yuegang, Yang Yitao, et al. Improved estimation method for ARMA model parameters[J]. Journal of Vibration, Measurement & Diagnosis, 2011, 31 (3): 377-380. (in Chinese)
- [12] 董长虹, 高志, 余啸海, 等. Matlab 小波分析工具箱原理与应用[M]. 北京: 国防工业出版社, 2004: 108-110.



**第一作者简介:**张培林,男,1955年12月生,博士、教授、博士研究生导师。主要研究方向为车辆维修理论与技术。曾发表有《基于蚁群算法的支持向量机参数优化》(《南京理工大学学报:自然科学版》2009年第33卷第4期)等论文。

E-mail: ZPL1955@163.com

**通信作者简介:**李胜,男,1986年12月生,硕士研究生。主要研究方向为车辆液压系统监测和智能检测与故障诊断。

E-mail: bcako@163.com