

基于PCA和SVM的内燃机故障诊断*

刘永斌^{1,2}, 何清波², 孔凡让², 张平²

(1. 安徽大学电气工程与自动化学院 合肥, 230039) (2. 中国科学技术大学精密机械与精密仪器系 合肥, 230027)

摘要 为有效对内燃机运行状态进行评估, 根据内燃机振动信号特征和故障样本较少的特点, 提出了基于主分量分析和支持向量机进行内燃机状态判别的故障诊断方法。提取内燃机振动特征参数, 利用主分量分析消除其信息冗余, 提取反映内燃机运行状态的主分量特征, 实现内燃机振动特征参数降维。通过选择适合内燃机振动信号的径向基核函数, 构造一对多的支持向量机多类分类器, 对主分量特征进行训练学习, 实现内燃机运行状态判别。通过对模拟内燃机不同运行状态的试验分析, 结果表明该方法可以有效识别内燃机不同的运行状态。

关键词 主分量分析; 支持向量机; 内燃机; 故障诊断

中图分类号 TP206.3; TH183.1; TK418

引言

机械设备内部发生异常时, 其振动特性会随之变化。由于振动信号中包含了丰富的设备状态信息, 对机械设备振动信号的测量和分析, 可判断出工作状态和故障性质。目前, 振动信号分析方法已被广泛应用于机械设备故障诊断^[1]。内燃机是由各零部件按照一定的关系组成的一个复杂的系统, 在运行过程中, 各部件相互关联并相互影响。内燃机具有多个环节的振动, 冲击、振动相互影响, 所采集到的内燃机振动信号是具有很大干扰噪声的复杂信号^[2-3]。在内燃机故障诊断中, 需要提取一些判别设备状态的特征参数。由于每个特征参数对状态的敏感程度各不相同, 为了全面准确地对设备进行诊断, 一般同时采用多种特征参数, 而多个特征参数的相关性往往又造成分析的困难。同时, 采用特征参数太多将会影响模式分类的计算代价甚至分类精度。另外, 由于内燃机系统的运动复杂, 既有旋转运动又有往复运动, 对其振动信号特征分类的难度大。一些基于数据学习的方法在故障诊断研究中逐渐得到了广泛应用, 如神经网络和贝叶斯分类方法等。但在实际应用中, 能得到的故障数据样本数量有限, 这些方法的使用受到了极大的限制。选用合适的分类器用于设备状态识别显得非常需要。目前, 国内外学者做了大量的研究, 李宏坤等^[2]利用KPCA-SVM方法对柴油机燃油系统的不同状态进行识别; 沈志熙等^[3]利用经

验模态分解与支持向量机相结合的办法对柴油机的常见故障进行了诊断; 贾继德^[4]针对内燃机振动信号的非平稳性的特点采用阶比跟踪、阶比分析和小波分析等方法对内燃机的常见故障进行了研究; Q. He等^[5]利用主分量分析提取内燃机噪声特征均取得了较好的效果。

主分量分析(PCA)是基于二阶统计特性的一种多元统计方法, 在信息压缩和消除数据间的相关性方面效果显著, 有效用于人脸、字符等特征提取^[6-7]。Malhi^[8]提出一种基于PCA的特征抽取方案来保证从多个特征中选择最有效的特征来进行轴承状态监测的故障分类。Baydar^[9]基于PCA的多元统计方法建立了一个正常状态模型进行齿轮箱状态监测。此外, 相关研究表明, PCA可以有效用于原始特征的降维^[5,10]。PCA方法通过变量变换法把相关变量变为少数不相关的主分量来概括原来多个参数的特征, 实现了多诊断参数的融合, 既提取出了有用信息, 又能使设备诊断工作简化, 给数据分析带来很大方便。Vapnik^[11-12]等人在统计学习理论的基础上提出了支持向量机算法(SVM), SVM基于结构风险最小化原则, 通过核函数实现从样本空间到高维特征空间的非线性映射, 利用支持向量来刻画因子与对象之间的非线性依赖关系。该方法对小样本条件下的非线性映射具有优势, 又能限制过学习, 特别适合于小样本集的数据处理, 广泛应用于模式识别^[12]和回归分析^[12-14]等领域, 在故障诊断方面具有一定的优越性^[15-16]。笔者基于PCA方法提取内燃机振动

* 国家自然科学基金资助项目(编号:51075379, 51005221); 中央高校基本科研业务费专项基金资助项目
收稿日期: 2010-07-14; 修改稿收到日期: 2010-09-28

信号主要特征,结合支持向量机的优越分类性能实现内燃机的故障诊断。

1 内燃机振动信号特征分析

内燃机的组成部件众多,形状各异,且结构复杂。运行过程中既有旋转运动又有多缸往复运动,既有点火、爆发又有其他不平衡冲击。不同工作循环之间也存在一定的差异与影响,且激励源众多,干扰成分复杂。信号经过多个零部件的衰减、调制传至外部。因此,采集到的振动信号是有很大干扰噪声的复杂振动信号。为了减少环境、测试系统以及信号复杂性对数据的影响,使信号分析结果具有客观、统一的标准,笔者采用均值-方差标准化对信号进行归一化预处理。

设所测内燃机振动信号为一维离散数据 $\{x_1, x_2, \dots, x_n\}$, n 为样本长度,则用均值-方差标准化对信号进行归一化预处理数可得

$$x'_i = \frac{x_i - \bar{x}}{\sqrt{\sigma}} \quad (1)$$

其中: \bar{x} 为样本均值; σ 为样本方差。

经过归一化预处理后,振动信号具有零均值和单位方差。为了全面准确地对设备进行诊断,基于时域与频域分析,提取连杆轴承振动信号时域与频域特征参数。

时域特征^[15]:提取连杆轴承振动信号的绝对均值、最大峰值、均方根值、方根幅值、方差、峰-峰值、歪度、峭度、峰值指标、波形指标、脉冲指标、裕度指标和峭度指标13个时域特征参数。

频域特征:将预处理后的4种状态振动信号变换为频域信号,提取平均频率和谱峰稳定指数,并根据连杆轴承不同状态的频率特点,将频域平分成5个频带,分别计算每个频带的相对功率谱能量,共提取7个频域特征参数。

上述共计提取20种特征参数表征内燃机振动特性,参数定义与计算方法见文献[15]。分别抽取内燃机正常、气门响、挺杆响、销子响、敲缸响和连杆轴承故障6种状态振动信号,每种状态测取50组数据进行计算,提取其状态特征参数。图1给出了内燃机6种状态的20个特征参数。其中:横坐标为样本序号,从1开始每50个样本属于一种状态;纵坐标为幅值。由于每个参数所表征的物理意义各不相同,参数的尺度大小也不一样,为了评价时各参数之间能够

相互参照,对每个参数进行了归一化处理。由图1可知,这些特征参数对连杆状态的表征各不相同、规律不一,有的参数对状态反映规律比较稳定,有的则对状态反映不敏感。

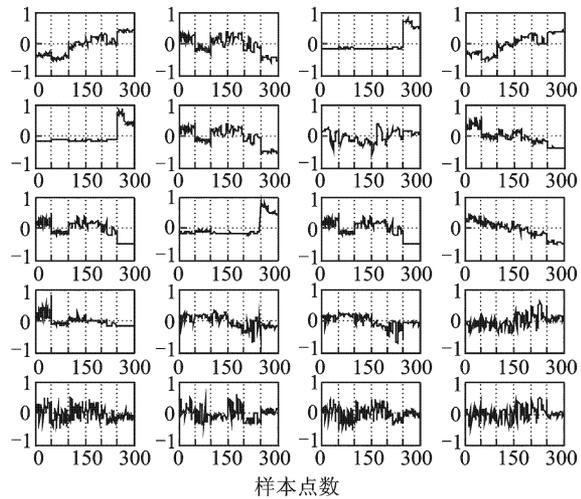


图1 内燃机振动信号特征参数表示

2 内燃机振动特征提取

2.1 主分量分析(PCA)

设内燃机振动信号模式向量矩阵为 \mathbf{X} ,其列向量 $\mathbf{X}_k = (x_{1k}, x_{2k}, \dots, x_{nk})^T$ 为 n 维模式向量,设备的某一状态可由 \mathbf{x}_k 描述, \mathbf{x}_k 的协方差矩阵为

$$\mathbf{R}_x = \sum_{k=1}^N (\mathbf{X}_k - \bar{\mathbf{x}})(\mathbf{X}_k - \bar{\mathbf{x}})^T \quad (2)$$

其中: N 为训练样本的数量; $\bar{\mathbf{x}}$ 为各样本模式向量的均值

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k \quad (3)$$

求解 \mathbf{R}_x 的全部特征值 $\lambda_i (i=1, 2, \dots, n)$ 和特征向量 \mathbf{v}_i ,将特征值 λ_i 按照从大到小的顺序排列: $\lambda_1 > \lambda_2 > \dots > \lambda_n$,则对应的特征向量为 $\mathbf{v}_i (i=1, 2, \dots, n)$ 。样本 \mathbf{x}_i 投影到特征向量 \mathbf{v}_i 得到该方向对应的主分量为

$$y_{ij} = \mathbf{v}_i^T (\mathbf{x}_j - \bar{\mathbf{x}}) \quad (4)$$

所有的特征向量张成一个 n 维正交空间, \mathbf{x} 投影到该正交空间得到相应的 n 维主分量。特征向量所对应的特征值越大,它在重构时的贡献也越大,特征值越小的特征向量在重构时的贡献就越小。设正交空间中前 m 个主分量为 y_1, y_2, \dots, y_m ,其累计方差贡献率为

$$\varphi(m) = \sum_{i=1}^m \lambda_i / \sum_{k=1}^n \lambda_k \quad (5)$$

当前面少数几个主分量的累计方差贡献率足够大,如 $\varphi(m) > 95\%$,即95%以上的原始数据信息保留在前面几个主分量中,可只取前 $m(m < n)$ 个主分量来表征原始信息,在保证信息完整的情况下,达到数据降维。

2.2 基于PCA的特征提取

PCA的基本思想是寻找高维数据的一些主分量表示,这些分量具有最大方差,用它们表示原数据具有最小的均方误差,是一种有效的多元统计分析工具。内燃机故障诊断需要获取对状态反映规律性好、敏感性强的特征来进行样本分类,而图1所示内燃机振动信号的20个特征参数所反映的特征状态各不相同,很难直接看出哪些特征参数规律性好且敏感性高;因此,在上述特征分析基础上来提取这样的特征尤为重要。笔者采用PCA方法来提取反映内燃机运行状态的主要特征参数。

分别选取内燃机正常(C1)、气门响(C2)、挺杆响(C3)、销子响(C4)、敲缸响(C5)和连杆轴承故障(C6)6种状态振动信号各50组进行分析。图2是内燃机6种工作状态的时域振动波形,从图中很难看出它们的差别。图3是6种工作状态振动信号的频域图,可以看出除了连杆轴承故障状态外,其他5种状态都是在1800 Hz左右频谱能量较大,且谱峰较分散,不便于分类识别。基于前述分析,采用特征参数的低维主分量来表示内燃机运行状态。对所有信号进行均值-方差标准化预处理,提取的样本特征参数构成特征参数矩阵,求解其协方差矩阵的全部特征值 $\lambda_i(i=1,2,\dots,n)$ 并从大到小排序,则可得与之对应的特征向量 $v_i(i=1,2,\dots,n)$ 。将图1所示特征参数矩阵按照式(4)投影到特征向量 v_i 上,可得 n 个特征参数的主分量 y_1, y_2, \dots, y_n 。为确保特征参数主分量的累计方差贡献率 $\varphi(m) > 95\%$,由式(7)计算,提取前4个主分量表示内燃机的状态。图4为利用PCA方法提取的前4个特征参数主分量。由图可以看出,第1个特征参数主分量PC1具有最大的特征值,且有数据变化的最大方差,抓住了各特征参数对状态表征的优点,较好地反映出了不同状态的区别特征,而其他特征参数主分量反映的则是局部变化特征,反映不同状态特征的能力相对较弱。

3 支持向量机分类器设计

支持向量机(SVM)是具有较强泛化能力的新

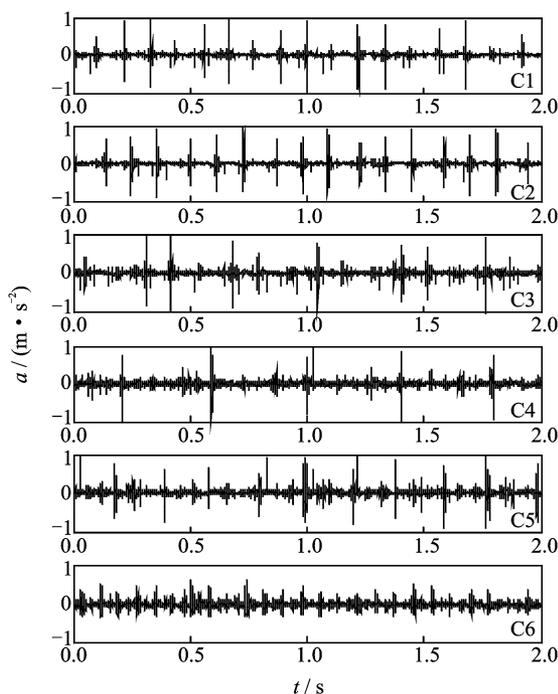


图2 内燃机6种工作状态的振动信号

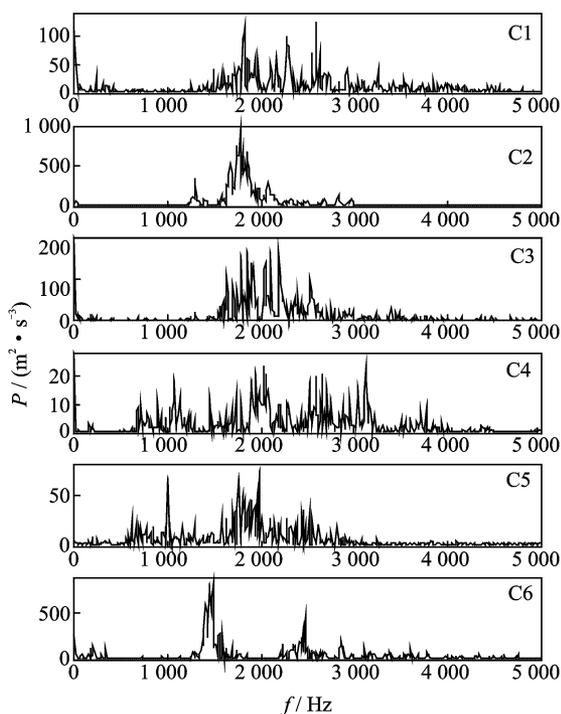


图3 内燃机振动信号频谱

机器学习方法,适合小样本分类。SVM的基本思想是通过非线性变换,将原始空间的数据映射到高维的特征空间,在高维空间中寻找具有最大边界的最优分类超平面来逼近分类函数。常用的SVM分类器主要是一对一和一对多两种方法,而在设备故障诊断中常常需要同时辨识多种状态,笔者采用基于一对多的多类分类器,具体描述如下。

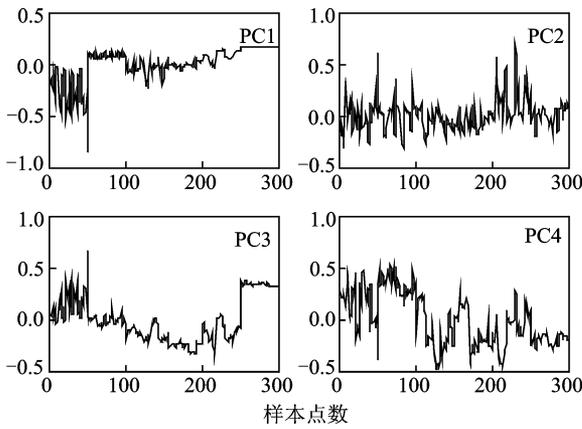


图4 内燃机前4个主分量特征

设有训练数据集 $(x_i, y_i), i=1, 2, \dots, l, x_i \in R^n$ 为输入数据, $y_i = \{-1, +1\}$ 为输出类别。寻求最优分类超平面可以归结为求解下列二次规划问题

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \quad (6) \\ \text{s. t.} \quad & \begin{cases} y_i(w^T \Phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \end{aligned}$$

其中: $w \in R^n$ 和 $b \in R$ 为权重因子; ξ_i 为松弛变量; C 为惩罚参数。

原始空间中的数据 x_i 通过函数 Φ 映射到高维特征空间, 在特征空间中, 通过最大化边界 $2/\|w\|$ 寻找最优分类超平面, 决策函数可写为

$$\text{sgn}\left(\sum_{i=1}^l y_i \alpha_i \Phi(x_i)^T \Phi(x_j) + b\right) \quad (7)$$

映像函数 $\Phi(x_i)$ 可以用更加灵活的核函数 $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$ 代替, 引入核函数后, 决策函数改写为

$$\text{sgn}\left(\sum_{i=1}^l y_i \alpha_i K(x_i, x_j) + b\right) \quad (8)$$

常用的核函数有线性核函数、多项式核函数和径向基核函数, 分别描述如下。

线性核函数

$$K(x_i, x_j) = x_i^T x_j \quad (9)$$

多项式核函数

$$K(x_i, x_j) = (x_i^T x_j + 1)^d \quad (10)$$

其中: d 为多项式核函数的阶数。

径向基核函数

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (11)$$

其中: $\gamma > 0$ 为核参数。

对于可分的 k 类问题, 可以组合 k 个“一对多”分类器构造一个多类分类器实现分类, 其结构如图5所示。

鉴于内燃机振动信号特征, 在进行故障识别时, 先将测试样本 X_k 输入分类器1, 若判别式输出为1, 则判定测试样本属于类别1, 分类结束; 否则输入给分类器2, 进行上述判断, 以此类推, 直到分类结束, 若最终输出为-1, 则说明测试样本所属类别不在这 k 类之中。

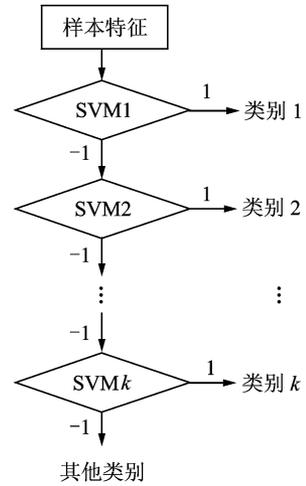


图5 一对多SVM多类分类器示意图

4 内燃机故障诊断

4.1 试验

试验是在一台全新的EQ6100六缸汽油发动机上进行的, 采用预先设置故障的方式来模拟内燃机可能发生的异常状态。在EQ6100型六缸汽油发动机的第五缸上先后设置气门响、挺杆响、销子响、敲缸响和连杆轴承故障5种异常状态, 所设各种故障状态的间隙大小如表1所示。为了所设故障状态逼真, 除了正常状态外, 每次保证内燃机仅有一处有一个故障, 并尽可能消除其他因素影响。振动信号的测取是在第五缸的缸体外放置B&K4384加速度传感器进行采集, 信号经B&K2635电荷放大器放大后送入计算机中的数据采集卡, 并在计算机中进行信号保存与分析, 振动信号采集试验如图6所示。整个测试过程中, 发动机空载, 转速保持在1000转左右, 采样频率为10kHz。

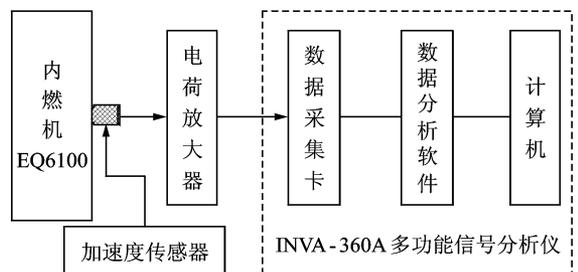


图6 试验装置示意图

表1 故障设置 mm

故障名称	正常状态	故障状态
气门响	0.200	0.800
挺杆响	0.080	0.380
销子响	0.005	0.095
敲缸响	0.03~0.050	0.500
连杆轴承故障	0.04~0.098	0.500

4.2 结果与分析

分别抽取内燃机气门响、挺杆响、销子响、敲缸响、连杆轴承故障及正常状态6种工况的振动信号各50组,采用前述的PCA方法,提取前4个主分量特征,如图4所示。4个主分量特征构成一个 300×4 的特征矩阵,内燃机的每种状态对应一个 50×4 的特征矩阵,每种状态的前25组数据用于训练,后25组数据用于测试。支持向量机分类的性能在很大程度上取决于核函数的选取。训练中,分别选用线性核函数、多项式核函数和径向基核函数对样本进行支持向量分类训练,训练结果如表2所示。表中,SVs为支持向量数,多项式核函数次数参数 $d=3$,径向基核函数的核参数 $\gamma=2.17$,其中,惩罚参数 C 与核函数参数采用交叉验证的方法选择。由表2可知,线性核与多项式核对内燃机挺杆响的状态识别率较径向基核函数低,且在挺杆响与销子响状态时,线性核与多项式核所需的支持向量较多。而总体上看,径向基核函数的识别率较高,支持向量数也适中。同时,径向基核与多项式核相比具有参数少的优点,这直接影响到模型选择的复杂性,因此本文采用径向基核函数构造SVM多类分类器进行状态识别。

表2 不同核函数的SVM诊断结果

内燃机状态	线性核		多项式核($d=3$)		径向基核	
	SVs	识别率/%	SVs	识别率/%	SVs	识别率/%
正常	20	99.67	32	99.00	104	99.00
气门响	20	100.00	48	100.00	104	98.67
挺杆响	484	76.67	220	85.33	128	92.00
销子响	244	92.00	168	92.00	76	92.33
敲缸响	80	96.00	118	99.33	108	98.00
连杆轴承故障	20	100.00	20	99.33	116	100.00

设备状态监测和故障诊断的实质是模式识别的问题^[7],正确识别设备状态特征是判别设备状态的基础;因此,分类器的性能的优劣直接影响设备故障诊断。表3为3层BP神经网络与SVM方法对内燃机6种状态的训练时间与识别率。考虑到小样本对神经网络的不利,在进行BP神经网络训练时,将内

燃机每种状态的前40个特征参数作为训练样本,后10个特征参数作为测试样本。其中,网络传递函数选择log-sigmoid函数,训练次数设为200次,网络性能目标设为0.001。在SVM分类训练时,将内燃机每种状态的前25个特征参数作为训练样本,后25个特征参数作为测试样本。SVM分类则使用LS-SVM算法,核函数选择如式(11)所示的径向基函数,核参数 $\gamma=2.17$ 。由表3可知,3层BP神经网络分类耗时较多,且识别率也不高。SVM分类器的平均识别率明显高于BP神经网络,综合分类性能优于3层BP神经网络,这说明支持向量机在内燃机故障诊断中具有良好性能。

表3 BP神经网络与SVM的结果对比

内燃机状态	BP神经网络		SVM	
	训练时间/s	识别率/%	训练时间/s	识别率/%
正常	1.629 4	84.15	1.083 1	99.00
气门响	1.669 8	88.56	1.115 8	98.67
挺杆响	1.699 1	86.57	0.992 8	92.00
销子响	1.646 5	78.87	1.066 9	92.33
敲缸响	1.626 9	92.20	1.133 9	98.00
连杆轴承故障	1.661 1	88.76	1.018 0	100.00

5 结束语

在EQ6100型六缸汽油发动机第五缸上先后模拟设置气门响、挺杆响、销子响、敲缸响和连杆轴承故障内燃机可能发生的5种异常状态,通过采集内燃机第五缸缸体上的振动信号,提取内燃机6种状态在时域和频域上的多维特征参数。针对内燃机振动信号的复杂性,采用主分量分析实现内燃机振动信号特征参数降维,提取内燃机运行状态特征参数的主分量表示。采用的主分量分析方法,既能降低维数减少计算的复杂性,又能有效表征内燃机的运行状态。

在故障诊断领域中,故障样本十分有限,这极大地限制了神经网络等基于传统统计学的模式识别方法的应用。笔者以PCA方法提取特征参数的主分量作为分类训练样本,分别选择线性、多项式和径向基核函数构造支持向量机分类器,对振动特征进行训练。通过对比选择适合内燃机振动信号的径向核函数与相关参数,实现内燃机的故障诊断。通过对内燃机振动信号主分量特征进行训练与测试,与BP神经网络进行比较。结果表明,基于统计学习理论的支持向量机可以较好地解决小样本问题,综合分类性能优于3层BP神经网络。这对内燃机故障诊断具有较强的实

际价值,为内燃机智能诊断提供了一种新途径。

参 考 文 献

- [1] 何正嘉, 訾艳阳, 孟庆丰, 等. 机械设备非平稳信号的故障诊断原理及应用[M]. 北京: 高等教育出版社, 2001: 1-8.
- [2] 李宏坤, 马孝江. 基于KPCA-SVM的柴油机状态识别方法的研究[J]. 振动、测试与诊断, 2009, 29(1): 42-45.
Li Hongkun, Ma Xiaojiang. Pattern recognition of diesel engine by using kernel principle component analysis and support vector machine [J]. Journal of Vibration, Measurement & Diagnosis, 2009, 29(1): 42-45. (in Chinese)
- [3] 沈志熙, 黄席樾, 马笑潇. 基于EMD和支持向量机的柴油机故障诊断[J]. 振动、测试与诊断, 2010, 30(1): 10-22.
Shen Zhixi, Huang Xiyue, Ma Xiaoxiao. Fault diagnosis of diesel engine based on empirical mode decomposition and support vector machine [J]. Journal of Vibration, Measurement & Diagnosis, 2010, 30(1): 10-22. (in Chinese)
- [4] 贾继德. 往复机械非平稳信号的特征提取及诊断研究[D]. 合肥: 中国科学技术大学, 2004.
- [5] He Q, Yan R, Kong F, et al. Machine condition monitoring using principal component representations [J]. Mechanical Systems and Signal Processing, 2009, 23(2): 446-466.
- [6] Jolliffe I J. Principal component analysis [M]. New York: Springer, 1986: 29-59, 363-365.
- [7] Perlibakas V. Distance measures for PCA-based recognition [J]. Pattern Recognition Letters, 2004, 25: 711-724.
- [8] Malhi A, Gao R. PCA-based feature selection scheme for machine defect classification [J]. IEEE Transaction on Instrumentation and Measurement, 2004, 53: 1517-1525.
- [9] Baydar N, Chen Q, Ball A, et al. Detection of incipient tooth defect in helical gears using multivariate statistics [J]. Mechanical Systems and Signal Processing, 2001, 15: 303-321.
- [10] Hu Qiao, He Zhengjia, Zhang Zhousoo, et al. Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble [J]. Mechanical Systems and Signal Processing, 2007, 21: 688-705.
- [11] Vapnik V N. Statistical learning theory [M]. Beijing: Publishing House of Electronics Industry, 2004: 93-110, 123-167.
- [12] Vapnik V, Golowich S, Alex S, et al. Support vector method for function approximation regression estimation, and signal processing [J]. Advances in Neural Information Processing Systems, 1996, 9: 281-287.
- [13] Liu Y, You Z, Cao L, et al. A novel and quick SVM-based multi-class classifier [J]. Pattern Recognition, 2006, 39: 2258-2264.
- [14] 曹冲锋, 杨世锡, 周晓峰, 等. 改进支持向量机模型的旋转机械故障诊断方法[J]. 振动、测试与诊断, 2009, 29(3): 270-273.
Cao Chongfeng, Yang Shixi, Zhou Xiaofeng, et al. Fault diagnosis of rotating machinery based on an improved support vector machines model [J]. Journal of Vibration, Measurement & Diagnosis, 2009, 29(3): 270-273. (in Chinese)
- [15] 陈克兴, 李川奇. 设备状态监测与故障诊断技术 [M]. 北京: 科学技术出版社, 1991: 40-87.
- [16] Yuan Shengfa, Chu Fulei. Support vector machines-based fault diagnosis for turbo-pump rotor [J]. Mechanical Systems and Signal Processing, 2006, 20: 939-952.

第一作者简介:刘永斌,男,1971年9月生,博士、工程师。主要研究方向为设备状态监测与故障诊断、智能材料驱动。曾发表《一种非平稳、非线性振动信号检测方法的研究》(《振动与冲击》2007年第26卷第12期)等论文。
E-mail: lyb@ustc.edu.cn

通讯作者:何清波,博士、副教授。
E-mail: qbhe@ustc.edu.cn