

基于支持向量机的发动机故障诊断

唐浩, 屈梁生

(西安交通大学机械工程学院, 710049, 西安)

摘要: 针对发动机的故障特点, 提出了一种基于主分量分析和支持向量机的发动机故障诊断方法. 利用小波包对声级计采集到的解放 CA141 型汽车发动机声音信号进行特征提取, 应用主分量分析方法在不损失有效信息的情况下, 将原始特征向量中的冗余信息约简, 在此基础上通过支持向量机对发动机故障进行分类. 诊断结果表明, 该方法在保证较高诊断精度的同时, 可将支持向量机的训练时间缩短 1/3, 从而提高了故障诊断效率.

关键词: 支持向量机; 主分量分析; 小波包; 故障诊断

中图分类号: TB533 **文献标识码:** A **文章编号:** 0253-987X(2007)09-1124-03

Fault Diagnosis of Engine Based on Support Vector Machine

Tang Hao, Qu Liangsheng

(School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

常见的发动机异响故障包括活塞敲缸响、活塞销响、气缸漏气、挺杆响、曲轴轴承响、连杆轴承响和正时齿轮响等. 本文对采集到的声音信号进行 3 层小波包分解, 结合时域信息, 构成原始特征向量. 针对该向量中可能含有的冗余信息, 利用主分量分析 (PCA) 进行维数约简, 最后利用支持向量机 (SVM) 进行故障分类, 从而提高了故障分类精度, 缩短了故障诊断时间.

1 理论基础

小波包变换是在多分辨率分析基础上构成的一种更为精细的正交分解方法, 它在全频带对信号进行多层次的频带划分, 不仅继承了小波变换所具有的良好时频局部化优点, 而且继续对其没有再分解的高频频带做进一步分解, 从而提高了频率的分辨

率. 一个三层小波包分解的结构如图 1 所示. 图中信号的分解关系为

$$S = S_7 + S_8 + S_9 + S_{10} + S_{11} + S_{12} + S_{13} + S_{14}$$

主分量分析^[1] 又称 K-L (Karhunen-Loeve) 变换, 是一种典型的对随机变量进行统计描述的统计学分析方法. 假设有一个 n 维特征向量 $\mathbf{X} = (x_1, x_2, \dots, x_n)$, 相应于每个特征有 m 个样本, 于是形成一个 $n \times m$ 维的特征矩阵 $\mathbf{A}$, $\mathbf{A}$ 的每一列可看成是 n 维特征空间上的点, 这样在 n 维空间上共有 m 个样本点. 现在需要寻找一种变换矩阵 $\mathbf{D}_{n \times n}$, 使得变换后的矩阵 $\mathbf{B} = \mathbf{D} \times \mathbf{A}$, 且在 $\mathbf{B}$ 的某个 p ($p < n$) 维子空间 E 上, 当 m 个样本点投影到此子空间的坐标上以后, 投影分量的方差最大, 这就是 K-L 变换的出发点. 换句话说, 通过 K-L 变换, 希望消除 $\mathbf{X}$ 中各分量的相关性, 通过去除那些带有较少信息的坐标轴来

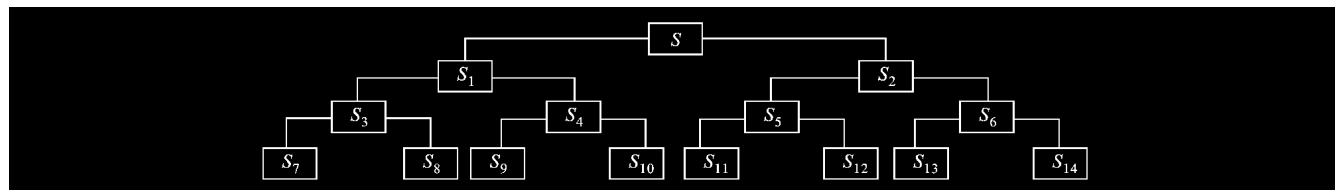


图1 小波包分解结构图

降低特征空间的维数,且不会产生很大的信息损失,从而有效地实现信息压缩。

支持向量机^[2-4]是由线形可分模式发展而来的,主要思想是建立一个超平面作为决策曲面,使得正、反例的边缘间隔最大化。将 n 个 d 维样本表示为

$$(x_i, y_i), i = 1, \dots, n; \quad x \in \mathbf{R}^d, y \in \{+1, -1\}$$

把建立线性支持向量机的问题转化为求解一个二次凸规划的问题,即

$$\left. \begin{aligned} \min \quad & \frac{1}{2}(\omega\omega) + C \sum_{i=1}^n \xi_i \\ \text{约束条件: } & y_i[(\omega x_i) + b] - 1 + \xi_i \geq 0 \end{aligned} \right\} \quad (1)$$

式中: ω 、 b 为待定系数; ξ_i 为松弛因子,表示存在不能被超平面正确分类的样本; C 为惩罚系数,表示在分类间隔和错误率之间的折中。

由于目标函数和约束条件都是凸的,根据最优化理论,这一问题存在惟一的全局最小解。应用 Lagrange 乘子,并考虑满足 Kuhn-Tucker 条件 $a_i(y_i(\omega x_i + b) - 1) = 0$, 可求得最优超平面的决策函数为

$$f(x) = \text{sgn}((\omega^*x) + b^*) = \text{sgn}\left(\sum_{i=1}^n a_i^* y_i (x_i x) + b^*\right) \quad (2)$$

对于非线性问题,利用核函数将其从低维转化成高维空间的线性问题,在特征空间中求最优分类面。此时,相应的分类函数变为

$$f(x) = \text{sgn}\left(\sum_{i=1}^n a_i^* y_i K(x_i x) + b^*\right) \quad (3)$$

常用的核函数主要有多项式核函数、高斯核函数和 Sigmoid 核函数等,支持向量机最初是针对二分类问题提出的,而解决多分类问题目前普遍采用二分类组合的方法,主要有 1 对 1 算法、1 对多算法和 DAGSVM 算法。

2 发动机故障诊断

对解放 CA141 型汽车发动机进行试验,用声级计和磁带记录仪采集了发动机的声音信号。采样频率为 6 kHz,低通滤波为 2 kHz,数据长度为 8 192 点。采集了 14 组正常发动机的声音信号,还有 24 组活塞敲缸响、23 组活塞销响、14 组气缸漏气响和 11 组气门挺杆响信号,总计为 86 组信号。对所采集到的信号用三层小波包分解重构后,计算各频带信号的总能量,然后再分别计算出信号的波形指标、峰值指标、脉冲指标、裕度指标和峭度指标,归一化处理后最终得到了一组 13 维的原始特征向量。

对该组特征向量分别用 PCA、KPCA 算法进行维数约简,取前 4 个主分量(此时 PCA 算法的均方重构误差为 0.002 2)构成一组新的 4 维特征向量。其中,KPCA 算法中采用高斯核函数,取核参数 $\sigma = 4$ 。

SVM 在小样本情况下建立的分类器具有较好的推广能力。试验的原始输入共有 5 类(包括正常信号和 4 种故障信号),随机抽取每一类数据中的 2/3 作训练,另 1/3 作测试,将这一过程反复进行 50 次,取平均值作为在该参数下 SVM 分类的正确率。采用 1 对多算法选择高斯核函数,对于参数的选择,目前还没有十分好的方法,本文采用交叉验证优化参数法,即令 $\sigma = 1 : 4 : 100$, $C = 1 : 4 : 100$,这样 σ 与 C 分别可取 25 个值,即有 25×25 种组合。分别计算其正确率,将正确率最高的参数作为最优参数。试验流程和试验结果见表 1 和图 2。

表 1 SVM 检测结果

属性约简	σ	C	运行时间*/s	正确率/%
无	1	53	90	97.97
PCA	1	45	55	97.80
KPCA	1	93	60	95.73

*: SVM 运行 50 次的时间。

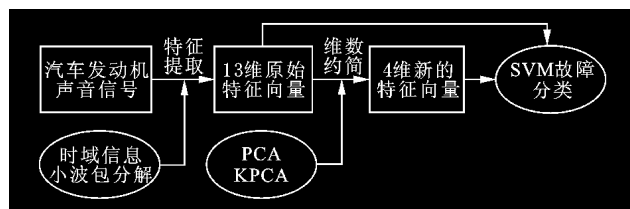


图 2 发动机诊断试验流程图

发动机的故障诊断结果表明,利用 PCA 算法进行属性约简后,SVM 的诊断精度变化不大,这是由于 SVM 对未知样本进行决策的计算复杂度主要取决于样本的数量,而与样本的维数关系不大。因此,SVM 在处理高维样本时可以获得较高的精度,当其经过属性约简后,训练时间缩短了 1/3。

3 结 论

本文讨论了支持向量机在机械故障诊断领域中的应用,提出将 PCA 算法和 SVM 技术相结合,用于发动机声音信号的故障诊断中。试验证明,该方法在保证高诊断精度的同时,明显地提高了诊断速度,是一种可行的故障诊断方法。

参考文献:

- [1] Jolliffe I T. Principal component analysis[M]. Berlin: Springer-Verlag, 2002:10-76.
- [2] Vapnik V N. The nature of statistical learning theory [M]. Berlin: Springer-Verlag, 2000:17-167.
- [3] Shawe-Taylor J, Cristianini N. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge: Cambridge University Press, 2000:8-107.
- [4] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998,2(2):121-167.

(编辑 管咏梅)