

基于 ReliefF-PCA 和 SVM 的发动机故障诊断方法研究

朱丽娜¹ 王学杰² 张进杰^{1*}

(1. 北京化工大学 机电工程学院 高端机械装备健康监控与自愈化北京市重点实验室, 北京 100029;

2. 海军海南地区装备修理监修室, 海南 三亚 572021)

摘要: 针对柴油发动机机组振动信号非线性和非平稳性以及机组实际故障案例样本数据少的特点, 提出了一种基于 ReliefF、主成分分析(PCA)以及支持向量机(SVM)的柴油发动机故障诊断方法。首先提取发动机冲击信号的特征参数, 运用 ReliefF 选择出其中的敏感特征以降低处理过程的计算难度; 然后采用 PCA 进一步提取敏感特征, 消除各特征之间的相关性, 避免冗余; 最后利用 SVM 实现机组的故障分类, 诊断不同类型的故障。将本文方法应用于柴油机实际典型故障案例中, 结果表明该方法能有效提取柴油机缸盖振动信号中的故障敏感特征, 并实现多种典型故障的诊断。

关键词: 柴油发动机; 故障诊断; ReliefF; 主成分分析(PCA); 支持向量机(SVM)

中图分类号: TP206.3 **DOI:** 10.13543/j.bhxbzr.2018.01.009

引言

以柴油发动机为代表的活塞式发动机作为典型的往复机械, 广泛应用于工、农、国防以及交通运输业等领域。由于柴油机结构复杂, 存在点火冲击、气门冲击等多种激励源, 机组振动信号具有明显的非线性和非平稳性, 所以振动信号是故障监测诊断研究中的重要内容。

长期以来, 企业对柴油发动机故障监测与诊断技术的研究不够重视, 机组故障诊断大多依赖于专家经验以及对排温、缸压等热工参数的监测, 这些方法监测效果滞后, 难以实现机组故障预警。由于振动信号能反映机组的冲击特征, 目前已有越来越多的学者基于振动监测方法, 通过分析缸盖振动信号对柴油机进行特征提取和故障诊断方法的研究^[1-3]。由于柴油机振动信号特征众多, 如何从中提取出故障敏感特征以降低特征维度、提高诊断效率是研究的热点。常用的降维方法有遗传算法(GA)^[4]、ReliefF 系列算法^[5-6]和主成分分析(PCA)^[7-8], 这3种算法都存在一定的局限。其中

GA 易过早收敛, 运行效率相对较低; 利用 ReliefF 算法衡量往复压缩机气阀故障特征性能的结果表明, ReliefF 选择出的敏感特征之间存在相关性, 无法除去信号的冗余特征信息^[9]; 利用 PCA 进行船舶柴油机故障诊断的结果表明, 该方法可以消除各个故障特征之间的相关性, 但如果不将特征作任何处理就直接采用 PCA 降维, 会使计算效率降低^[10]。

因此, 本文针对柴油机的实际故障特征以及振动信号特征, 将 ReliefF 与 PCA 的优点结合, 提出一种基于 ReliefF-PCA 和支持向量机(SVM)的发动机故障诊断方法。首先使用 ReliefF 从柴油机振动信号中选取敏感特征, 再利用 PCA 进一步降维后, 将不同工况下的敏感特征用于 SVM 中进行故障分类。

1 基于 ReliefF-PCA 的敏感特征选择

1.1 利用 ReliefF 技术提取敏感特征

ReliefF 技术根据类间距离与类内距离大小来评估不同特征的敏感度。敏感特征满足以下条件: 某一特征同一类的类内距离越小, 不同类的类间距离越大, 则这一特征就越敏感。距离评估算法包括最早的 Relief 算法以及之后拓展开发出的 ReliefF 及 RReliefF 算法。其中 Relief 只适用于两类分类问题, RReliefF 适用于目标属性为连续值的回归问题, 只有 ReliefF 算法可用于多分类问题处理。利用 ReliefF 方法进行距离评估的步骤如下:

1) 初始化权值矩阵向量 $W=0$;

收稿日期: 2017-02-22

基金项目: 国家“863”计划(2014AA041806); 中央高校基本科研业务费(ZY1617)

第一作者: 女, 1993年生, 硕士生

* 通讯联系人

E-mail: zjj87427@163.com

2) 在特征矩阵样本中任取 1 个样本 \mathbf{R} , 并找出与 \mathbf{R} 最近的 k 个同类样本 \mathbf{H}_i 和 k 个不同类样本 \mathbf{M}_i ($i=1, 2, \dots, k$);

3) 对于样本 \mathbf{R} 的所有 j 个特征 F_j ($j=1, 2, \dots, J$) 依次按式(1)~(3)进行计算

$$W(F_j) = W(F_j) - \sum_{i=1}^k \text{diff}(F_j, \mathbf{R}, \mathbf{H}_i)/k + \sum_{i=1}^k \text{diff}(F_j, \mathbf{R}, \mathbf{M}_i)/k \quad (1)$$

$$\text{diff}(F_j, \mathbf{R}, \mathbf{H}_i) = |\text{Value}(F_j, \mathbf{R}) - \text{Value}(F_j, \mathbf{H}_i)| / (\max(F_j) - \min(F_j)) \quad (2)$$

$$\text{diff}(F_j, \mathbf{R}, \mathbf{M}_i) = |\text{Value}(F_j, \mathbf{R}) - \text{Value}(F_j, \mathbf{M}_i)| / (\max(F_j) - \min(F_j)) \quad (3)$$

其中 $\text{Value}(F_j, \mathbf{A})$ 是样本 \mathbf{A} 的第 j 个特征 F_j 的值, \mathbf{A} 可取值为 \mathbf{R}, \mathbf{H}_i 或 \mathbf{M}_i 。

4) 循环步骤 1)~3) m 次, 即随机选择 m 个样本进行权值向量的计算, 从而获得特征向量 \mathbf{W} 。

以二维特征为例, Relief 系列算法的原理如图 1 所示。

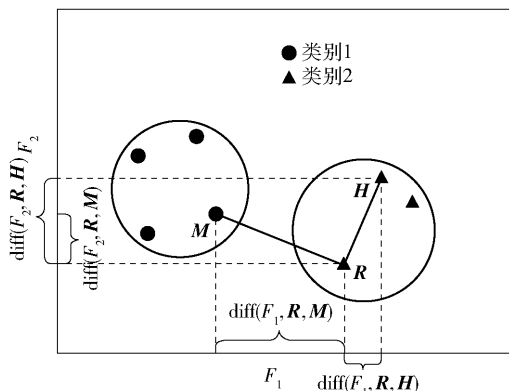


图 1 Relief 系列算法原理图

Fig. 1 Schematic of the Relief series algorithm

图中 \mathbf{R} 与 \mathbf{H} 在特征 F_1 上的距离比 \mathbf{R} 与 \mathbf{M} 在特征 F_1 上的距离要小, 即: $\text{diff}(F_1, \mathbf{R}, \mathbf{H}) < \text{diff}(F_1, \mathbf{R}, \mathbf{M})$, 这说明特征 F_1 对于样本的分类是有利的, 而特征 F_2 则相反。因此, 应该增大 F_1 的权值, 减小 F_2 的权值。保留权重值较大的特征参数来形成特征矩阵, 并除去权重值较小 (即故障敏感性低) 的特征参数, 便可获得最终的特征向量 \mathbf{W} 。

1.2 利用 PCA 技术降维

PCA 通过将原始数据映射到一个高维空间中使数据变得可分。在这个新空间的坐标系统下, 变换了的数据点的方差沿新的坐标轴得到了最大化, 而且数据点之间不再线性相关, 这组变换了的数据

点被称为主成分。由于主成分中包含了原始特征的绝大部分信息, 且互不相关也不存在冗余信息, 因此可利用主成分代替原始的特征参数, 从而实现数据降维, 大大降低数据处理的复杂程度, 也避免了维度灾难。PCA 降维前先利用式(4)进行归一化:

$$\mathbf{Y} = (\mathbf{X} - X_{\min}) / (X_{\max} - X_{\min}) \quad (4)$$

其中 \mathbf{X} 为通过 ReliefF 处理初步得到的一个特征矩阵, X_{\min} 是每个特征参数的最小值, X_{\max} 是最大值。由此得到新的特征矩阵 \mathbf{Y} , 再利用 PCA 进行敏感特征选择。

PCA 的降维过程如下: 1) 按列的方式对特征矩阵 \mathbf{Y} 进行去均值处理得到新矩阵 \mathbf{Z} ; 2) 计算协方差矩阵 \mathbf{M} : $\mathbf{M} = \mathbf{ZZ}^T$, 矩阵 \mathbf{M} 的特征值和对应的特征向量可以通过奇异值分解 (SVD) 得到; 3) 对 \mathbf{M} 的特征值及相应的特征向量按特征值降序重新排序, 即获得新的特征向量 \mathbf{P} ; 4) 计算得分矩阵 \mathbf{S} : $\mathbf{S} = \mathbf{ZP}$ 。

2 基于 SVM 的故障诊断

SVM 最初针对的是模式识别中的线性可分问题。对线性可分问题 $D = \{(x_i, y_i) | i=1, 2, \dots, n\}$ (其中 $x_i \in \mathbf{R}, y_i \in \{+1, -1\}$ 为类别编号), 存在一个超平面 (或最优线性分类面) $\omega x + b = 0$ (ω 是 n 维向量, $x \in \mathbf{R}, b$ 为偏移量), 使得两类样本能完全区分。根据相关理论, 该线性可分问题能转化为求解二次型规划

$$\min \varphi(\omega) = \frac{1}{2} \|\omega\|^2$$

$$\text{s. t. } y_i(\omega x_i + b) - 1 \geq 0, i=1, 2, \dots, m \quad (5)$$

对于线性不可分问题, SVM 通过定义适当的内积核函数 K , 应用非线性变换将输入空间映射到一个高维特征空间, 因此在这个新空间能够获得超平面。支持向量是离分界线最近的点, 它们支撑着分类面, 即超平面是由支持向量确定的。支持向量数目越多, 说明超平面越复杂, SVM 的分类效率也越低。因此, 支持向量数越少代表分类效果越好。本文选用最常用的径向基核函数

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2), \gamma > 0 \quad (6)$$

本文中的柴油发动机故障诊断是一个多分类问题, 因此采用一对一分类法用于识别柴油机故障。

3 实验部分

3.1 实验装置

实验选用一台直列 6 缸柴油机 (图 2), 机组转



图 2 实验用柴油发动机

Fig. 2 Experimental diesel engine

速 800 r/min, 信号采样频率 25600 Hz。通过在该柴油机上安装 BH5000E 发动机在线监测系统获得机组的振动冲击信号。本文对 6#缸在柴油机上的几类典型故障进行实验研究, 获得机组在以下 5 种工况下的样本数据: 正常状态、6#缸小头瓦磨损、6#缸拉缸、6#缸小头瓦磨损状况下 6#缸失火以及 6#缸小头瓦磨损状况下 1#缸失火故障。首先对不同工况下的缸盖振动信号进行特征提取, 再利用本文提出的方法进行敏感特征提取和故障诊断。

3.2 数据采集及特征提取

基于直列 6 缸柴油机采集了 5 种状态下的机组样本数据, 其中 5 种状态下 6#缸缸盖振动数据如图 3 所示。

为了反映机组的运行工况, 提取机组振动信号在时域和频域上的特征共计 79 个。时域上包括缸盖振动加速度信号的直流流量、最大峰值、峭度、歪度

以及曲轴箱速度信号的有效值、峭度、歪度。频域上选取 db5 小波作为基函数数量, 对缸盖振动信号进行 4 层小波包分解可获得 16 个频段内的能量比值, 再分别对第一段和最后一段的加速度重构信号求取直流流量、最大峰值、峭度、歪度, 并在角域上将其均分为 24 段(信号总长为一个周期, 720°), 分别求取各段内信号的真峰值。其中各工况下小波包分解的前 8 个频段内的能量比值特征如图 4 所示。很明显, 各工况下的能量成分差异很大, 可用于区分机组不同的故障类型。

经过上述的特征提取过程后, 采用 ReliefF-PCA 方法进行降维处理。先用 ReliefF 进行数据处理, 并根据权重降序选择前 60 维的特征, 再用 PCA 去冗余, 并选择前 3 维主成分作为敏感特征。用于 SVM 中进行训练和测试的样本数量如表 1 所示。SVM 经过训练后获得分类模型, 可用于测试数据的诊断。图 5(a)、(b) 分别为经 ReliefF-PCA 降维后的训练数据样本与测试数据样本。将图 5(a) 中的训练样本输入 SVM 中可确定最优分类超平面, 再用该超平面对之后输入的测试样本进行分类。由图 5 可知, 除少数样本外, 其余不同故障类型的样本均可被 ReliefF-PCA 明显地区分开。

3.3 故障诊断

为了验证ReliefF-PCA降维方法对故障诊断的

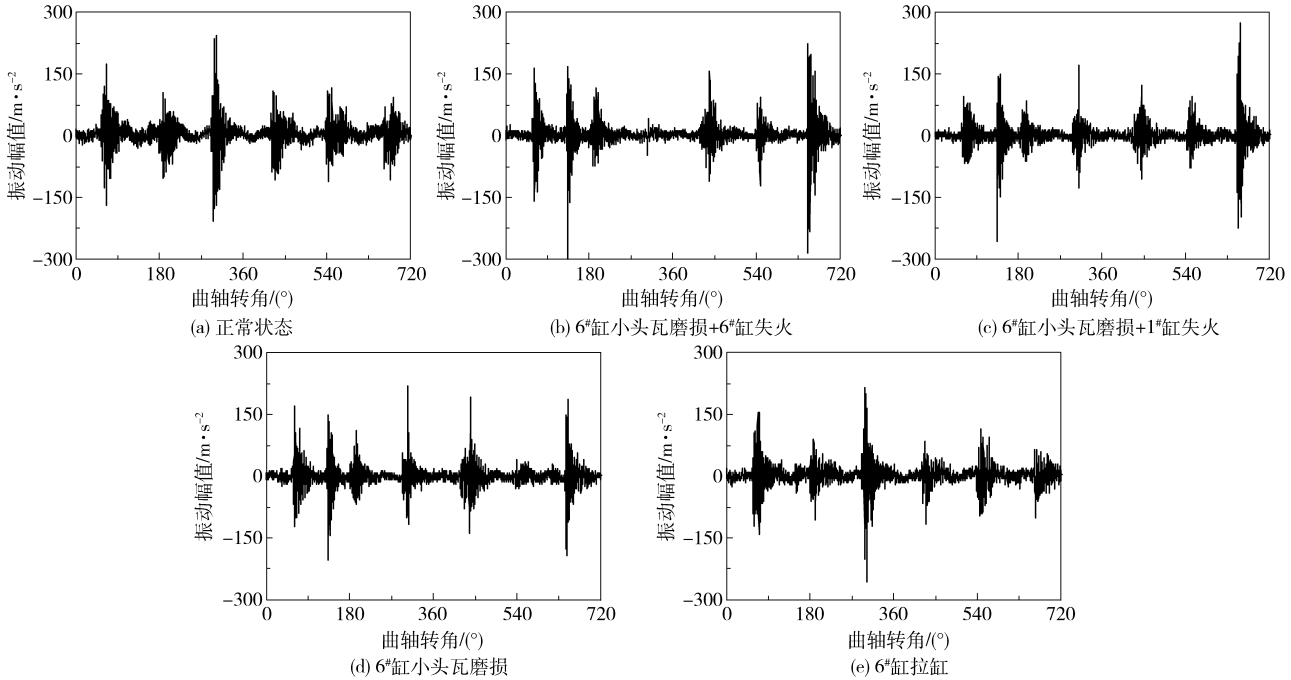


图 3 5 种工况下 6#缸缸盖振动数据

Fig. 3 The 6# cylinder head vibration signal under five different conditions

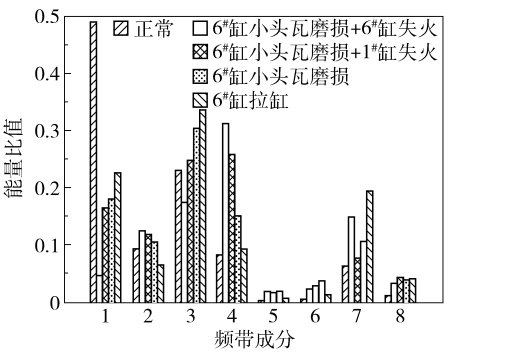


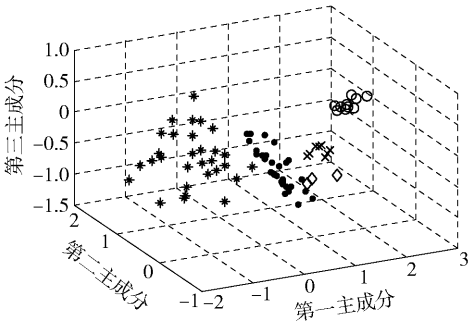
图4 能量比值特征向量对比图

Fig.4 Energy ratio feature vector

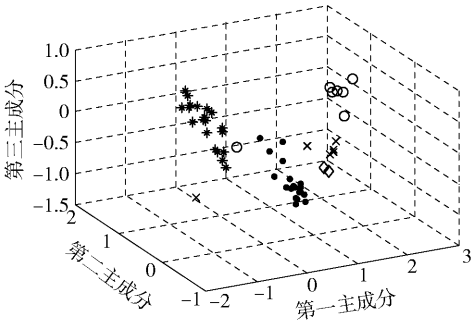
表1 训练样本和测试样本数量

Table 1 The number of training samples and test samples

工况	训练样本数	测试样本数
正常工况	30	20
6#缸小头瓦磨损+6#缸失火	10	7
6#缸小头瓦磨损+1#缸失火	8	6
6#缸小头瓦磨损	3	2
6#缸拉缸	30	20



(a) 训练样本



(b) 测试样本

* 正常; ○ 6#缸小头瓦磨损+6#缸失火; × 6#缸小头瓦磨损+1#缸失火; ◇ 6#缸小头瓦磨损; ● 6#缸拉缸。

图5 ReliefF-PCA 降维后的训练样本和测试样本三维图

Fig.5 3D plot of raining samples and testing samples after dimensionality reduction with ReliefF-PCA

有效性,使用多种不同的降维方法对样本数据加以

处理,并对比它们的 SVM 分类结果,如表 2 所示。由表 2 可知,本文方法与 PCA 方法的分类准确率比另外两种情况都高;而 ReliefF-PCA 法的支持向量总数最少,所以 ReliefF-PCA 法的分类效率最高。

表2 SVM 故障分类效果对比

Table 2 Comparison of SVM fault classification effects table

工况	分类准确率/%			
	ReliefF-PCA	PCA	ReliefF	未降维
正常工况	100	100	100	100
6#缸小头瓦磨损+6#缸失火	86	86	86	86
6#缸小头瓦磨损+1#缸失火	83	83	17	17
6#缸小头瓦磨损	100	100	100	100
6#缸拉缸	100	100	100	65
支持向量总数目	42	54	66	80
总分类精度	96	96	89	76

4 结束语

由于发动机缸盖振动信号蕴含信息较多,采用 ReliefF 和 PCA 方法提取振动信号中的故障敏感特征能显著降低计算难度。通过实验分析可知,本文方法不仅能保证 SVM 对于柴油机典型故障的高精度诊断,而且能有效提高计算效率,验证了基于 ReliefF-PCA 和 SVM 的方法对于柴油机多类冲击故障诊断的有效性和可靠性。

参考文献:

[1] 李敏通. 柴油机振动信号特征提取与故障诊断方法研究[D]. 陕西 杨凌: 西北农林科技大学, 2012.
Li M T. Research on diesel engine vibration signal feature extraction and fault diagnosis methods [D]. Yangling, Shaanxi: Northwest Agriculture and Forestry University, 2012. (in Chinese)

[2] 段礼祥, 张来斌, 王朝晖. 柴油机状态监测与故障诊断特征参数研究[J]. 车用发动机, 2007(1): 76-79.
Duan L X, Zhang L B, Wang Z H. Characteristics research on condition monitoring and fault diagnosis of diesel engine[J]. Vehicle Engine, 2007(1): 76-79. (in Chinese)

[3] Arroyo J, Munoz M, Moreno F, et al. Diagnostic method based on the analysis of the vibration and acoustic emission energy for emergency diesel generators in nuclear plants[J]. Applied Acoustics, 2013, 74(4): 502-508.

- [4] Hoell S, Omenzetter P. Optimal selection of autoregressive model coefficients for early damage detect ability with an application to wind turbine blades [J]. Mechanical Systems & Signal Processing, 2016, 70/71: 557-577.
- [5] 黄莉莉, 汤进, 孙登第, 等. 基于多标签 ReliefF 的特征选择算法[J]. 计算机应用, 2012, 32(10): 2888-2890/2898.
- Huang L L, Tang J, Sun D D, et al. Feature selection algorithm based on multi-label ReliefF [J]. Journal of Computer Applications, 2012, 32(10): 2888-2890/2898. (in Chinese)
- [6] 姜万录, 王友荣, 王振威, 等. 基于 ReliefF 算法和相关度计算结合的故障特征降维方法及其应用[J]. 液压与气动, 2015(12): 18-24.
- Jiang W L, Wang Y R, Wang Z W, et al. Fault feature dimension reduction method combined ReliefF algorithm with correlation calculation and its application[J]. Chinese Hydraulics & Pneumatics, 2015(12): 18-24. (in Chinese)
- [7] 苑宇, [马孝江]. 基于主分量分析的柴油机振动信号特征提取[J]. 中国机械工程, 2007, 18(8): 971-975.
- Yuan Y, [Ma X J]. Feature extraction from vibration signals of diesel based on PCA[J]. China Mechanical Engineering, 2007, 18(8): 971-975. (in Chinese)
- [8] 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 Y B, He Q B, Kong F R, et al. Fault diagnosis of internal combustion engine based on PCA and SVM[J]. Journal of Vibration, Measurement & Diagnosis, 2012, 32(2): 250-255. (in Chinese)
- [9] 马嘉俊. 基于 SVM 的气阀故障诊断研究[D]. 成都: 电子科技大学, 2016.
- Ma J J. Valve fault diagnosis research based on SVM [D]. Chengdu: University of Electronic Science and Technology of China, 2016. (in Chinese)
- [10] 彭秀艳, 柴艳有, 满新江. 基于 PCA-KFCM 的船舶柴油机故障诊断[J]. 控制工程, 2012, 19(2): 311-315.
- Peng X Y, Chai Y Y, Man X J. Fault diagnosis of marine diesel engine based on PCA-KFCM[J]. Control Engineering of China, 2012, 19(2): 311-315. (in Chinese)

An engine fault diagnosis method based on ReliefF-PCA and SVM

ZHU LiNa¹ WANG XueJie² ZHANG JinJie^{1*}

(1. Beijing Key Laboratory of Health Monitoring Control and Fault Self-Recovery for High-end Machinery, College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029;
2. Navy Supervision Office for Equipment Repair in Hainan Area, Sanya, Hainan 572021, China)

Abstract: Taking into account the nonlinear and non-stationary nature of diesel engine vibration signals, and the limited amount of available actual fault sample data, a fault diagnosis method for diesel engines based on ReliefF, principal component analysis (PCA) and support vector machines (SVM) is proposed. First of all, the characteristic parameters of the engine impulse signal are extracted. Then the ReliefF technology is used to select the sensitive features in order to reduce the computational difficulty of data processing. PCA is then used to further extract the sensitive features, and the correlation between the features is removed to avoid redundancy. Finally, the SVM is used to realize the unit fault classification, and different types of faults can be diagnosed. The method proposed in this paper is applied to actual typical faults occurring in diesel engines, and the results show that it can effectively extract the fault-sensitive features of diesel engine cylinder head vibration signals, and realize the diagnosis of many typical faults.

Key words: diesel engine; fault diagnosis; ReliefF; principal component analysis(PCA); support vector machine (SVM)

(责任编辑:汪 琴)