

基于 CART 决策树的柴油机故障诊断方法研究

江志农¹ 魏东海¹ 王磊² 赵志超² 茆志伟¹ 张进杰^{1*}

(1. 北京化工大学 机电工程学院 诊断与自愈工程研究中心, 北京 100029;

2. 中石化重庆天然气管道有限责任公司, 重庆 404100)

摘要:采用一种自适应局部有效值(RMS)计算方法提取柴油机缸盖振动信号时域特征,结合分类回归树(CART)算法建立故障分类模型并进行柴油机的状态识别。通过实验获取柴油机失火和撞缸两种故障工况及正常工况下的振动数据,计算出原始信号的局部RMS后,根据自适应阈值确定点火冲击区域和非点火上止点冲击区域提取局部特征,最后将特征输入CART算法中构建分类模型来验证所提取特征的有效性。结果表明:柴油机在3种状态下的识别率均达到100%,基于CART算法和局部特征提取的方法能够有效诊断柴油机故障。

关键词:分类回归树(CART)算法;柴油机故障诊断;局部有效值(RMS)计算;自适应阈值;特征提取

中图分类号:TP391.4 **DOI:** 10.13543/j.bhxbzr.2018.04.013

引言

柴油机是一种高速运转的往复式动力机械,其气缸盖振动信号中包含大量的柴油机运行状态信息。柴油机气缸盖振动信号是一种典型的宽频带混叠非平稳时变信号,频率成分十分复杂,难以发现与故障相对应的特征频率^[1];而且振动信号中含有大量噪声,有用信息往往被淹没在噪声里,给柴油机故障诊断造成障碍,难以保证对报警进行判断和分析处理的正确性^[2]。

随着设备复杂程度的提高,故障原因和种类越来越多样化,故障历史数据也越来越庞大,这使得柴油机故障诊断进入了海量数据处理与智能诊断的阶段,利用人工智能与数据挖掘技术对柴油机进行状态监测与故障诊断成为当前的研究热点^[3]。模式识别常用的方法有神经网络和支持向量机(SVM)等,但是神经网络的构建过程较为复杂,SVM的计算量较大,而决策树算法具有结构简单、计算量小等优点,在多个领域得到应用。Mu等^[4]利用模糊粗糙集和决策树算法对核电厂故障进行诊断,有效识别出核电厂常见故障;Gangadhar等^[5]基于决策树算

法,利用振动信号进行单点切割工具的故障诊断;Amarnath等^[6]利用决策树算法诊断了滚动轴承声音信号的故障,发现决策树算法具有结构简单、计算量较低等优点;Demetgul^[7]研究了决策树算法应用于工业生产系统的故障诊断;Gao等^[8]通过挖掘遥测数据研究航天器故障诊断方法,分析并设计出了故障树和专家系统。以上文献中决策树算法主要应用于核电厂、刀具、滚动轴承、工业生产系统和航天器等故障诊断领域,但少有将决策树算法用于柴油机故障识别方面的研究。

本文引进决策树方法对柴油机故障诊断进行研究,采用一种自适应局部有效值(RMS)计算方法提取柴油机缸盖振动信号时域特征,基于分类回归树(CART)算法建立了柴油机故障分类模型。首先采用滑动窗口方法计算振动信号的局部RMS值,然后使用自适应阈值进行冲击定位,最后提取局部区域的有效值和峰值特征,输入CART算法中对柴油机故障进行分类。实验结果表明,该方法能有效利用机组数据,对柴油机常见故障诊断的准确率达到100%。

1 CART 算法原理

CART算法又称CART决策树算法,是应用于分类的一种树结构,其中每个内部节点代表某一属性的一次测试,每条边代表一个测试结果,叶子节点代表某个类或类的分布。CART决策树生成规则具有容易理解、计算量较小、可以处理属性值连续的情

收稿日期:2017-12-26

基金项目:国家“863”计划(2014AA041806);国家重点研发计划(2016YFF0203305);中央高校基本科研业务费(JD1815)

第一作者:男,1967年生,教授

*通信联系人

E-mail: zjj87427@163.com

况等优点,其决策过程需要从决策树的根节点开始,将待测数据与决策树中的特征节点进行比较,并按照比较结果选择下一个比较分支,直到叶子节点作为最终的决策结果。

决策树类型有多种,CART 决策树是一种典型的二叉结构树,已被广泛应用到统计学领域。相比 ID3 决策树和 C4.5 决策树,CART 决策树能够有效地处理数据样本的非线性问题,不用对样本作出假设,所以适用于解决数据样本和目标变量之间关系不明确的复杂性问题;另外,区别于 ID3,CART 决策树能够划分连续的属性和离散的属性,基于 Gini 指标求出作为根节点的属性,然后自上而下以递归的方法构建模型,直至每个样本集在划分后都是完全纯净的,则停止建树。Gini 指标用于度量样本的杂质度,对于样本集合 D ,定义如式(1)所示^[9]

$$\text{Gini}(D) = 1 - \sum_{i=1}^N p_i^2 \quad (1)$$

式中, N 为类标签的数目, p_i 为类别 i 在样本集合 D 中出现的概率。Gini 指标的值越小,说明样本的杂质度越低,样本越纯净,所以建树过程中优先选择 Gini 指标小的属性,当属性的种类多于两种时,对各种两类划分的组合分别计算 Gini 指标,并且自动选用最小化 Gini 指标的划分组合。

CART 决策树算法在满足下述条件之一时停止建树:①所有叶子节点的样本数为 1,样本都属于同一类或者样本数小于某个给定的最小值;②训练数据集不再具有属性向量作为分支选择;③决策树的高度达到了用户所设定的阈值,或者分支后的叶节点中的样本属性都属于同一类^[10]。

2 特征参数提取

2.1 算法流程

目前用于分析柴油机状态的信号主要有振动信号、噪声信号、压力信号、转速信号、温度信号、排放信号等(图 1)。本文主要研究柴油机失火和撞缸两类故障。检测柴油机失火故障最直接、最有效的参数是气缸内的压力信号,但是柴油机气缸内的压力和温度都非常高,不利于传感器的布置和监测^[11],故本文利用柴油机气缸盖振动信号来间接监测气缸点火的实时燃烧情况。振动加速度传感器安装在柴油机气缸盖上,在气缸内点火燃烧的相应时刻,传感器监测到的振动信号会突然增强;而在该气缸发生失火故障时,相应时刻的振动信号比正常情况下大

大减小,所以点火段的局部时域信号可以较好地反映失火故障的状态。当柴油机发生撞缸故障时,在气缸点火上止点和非点火上止点两个时刻,振动信号幅值较正常情况均增大,且非点火上止点时刻冲击明显增大。

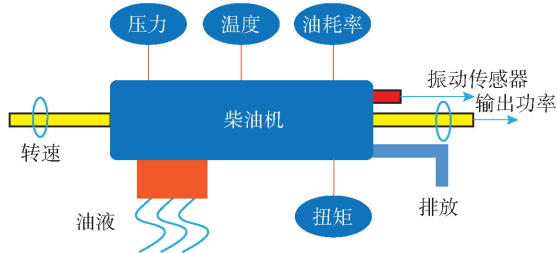


图 1 柴油机状态监测的各类信号

Fig. 1 Various signals of diesel condition monitoring

实验对柴油机正常、失火和撞缸 3 种工况进行模拟实验,在参考相关文献并且对故障机理进行研究后,最终选定点火上止点有效值、点火上止点峰值、非点火上止点有效值和非点火上止点峰值作为故障识别的 4 个敏感特征参数,特征参数用于模型树的构建。特征提取的算法流程如图 2 所示。

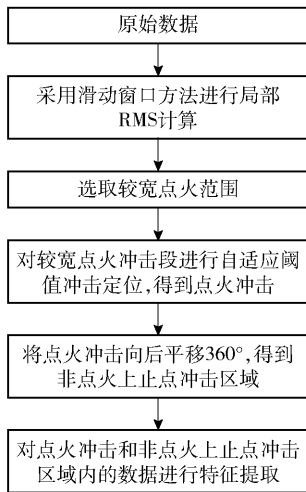


图 2 特征提取的算法流程

Fig. 2 Algorithm flow of feature extraction

2.2 特征提取

应用一种滑动窗口方法计算原始信号局部 RMS 值,对每一个点在一定宽度的窗内计算其有效值,将原始信号转换为值域大于零的信号,以达到量化振动强度的目的。本文以柴油机单周期正常信号为例进行分析,原始信号的时域波形如图 3 所示。固定宽度的滑动窗口依次通过信号中每一个点,第 i 个点的局部 RMS 值由公式(2)计算^[12]

$$y_{\text{rms}}(i) = \sqrt{A + B + \cdots + C} \quad (2)$$

式中, $A = \frac{1}{2N_w + 1} y(i + N_w)^2$, $B = \frac{1}{2N_w + 1} y(i + N_w - 1)^2$, $C = \frac{1}{2N_w + 1} y(i - N_w)^2$, y 代表原始信号, $y_{\text{lms}}(i)$ 是原始信号中第 i 个点的局部 RMS 值。设 $N_w = 2N_w + 1$ 为滑动窗口的大小, N_w 根据经验取值 20 为最佳^[12]。利用公式(2)进行相应计算后, 原始信号转换为局部 RMS 信号, 冲击区域便很容易被划分出来。利用 MATLAB 软件对原始数据进行相应处理, 转换后振幅随角度变化关系如图 4 所示。

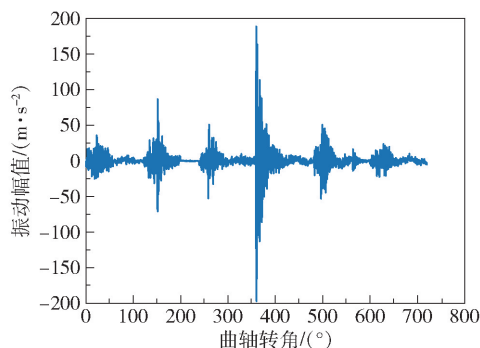


图3 原始信号的时域波形图

Fig. 3 The time domain waveform of the original signal

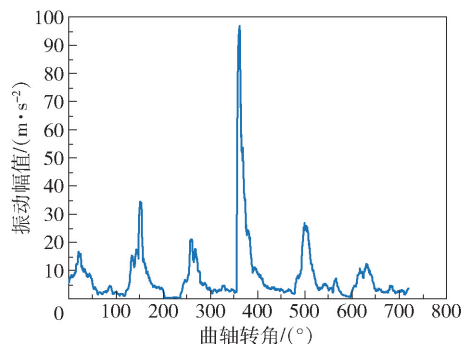


图4 局部 RMS 计算后的加速度信号

Fig. 4 Acceleration signal after partial RMS calculation

实验只对 6#缸进行故障模拟。6#缸初始相位滞后于 1#缸, 故将 6#缸振动信号的初始相位调整为与 1#缸一致, 使键相信号的脉冲时刻正好是气缸内活塞运动到上止点位置的时刻。另外, 通过编程对 6#缸振动信号的跑相位问题进行处理, 将点火上止点冲击调整至 360°附近, 非点火上止点冲击位于 720°附近。通过编程调整使整个周期过程清晰, 便于特征参数提取。当柴油机处于失火状态时, 周期信号在点火上止点冲击幅值较小, 如果直接选取阈值来定位冲击可能无法定位到点火范围, 而且噪声干扰较大时可能会导致定位错误; 所以本文先截取较宽的点火冲击信号, 程序设定截取角度为 320°~400°

(图 5), 然后采用自适应阈值方法定位局部冲击, 避免了选取全局阈值可能带来的定位错误。

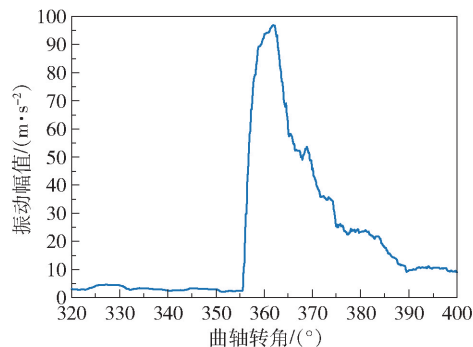


图5 提取宽范围的点火冲击

Fig. 5 Extraction of a wide range of ignition shocks

阈值处理是信号处理中常用的技术, 可以通过检查数据点是否低于给定阈值 Y_{rms} 来对数据点进行分类。本文使用一个自适应的阈值来区分局部 RMS 信号上的特定点的属性: 当局部 RMS 信号高于阈值时, 为冲击点; 当低于阈值时, 为非冲击点。自适应阈值法截取点火冲击的结果如图 6 所示, 用数学方式描述为

$$\text{impact}(i) = \begin{cases} \text{冲击点}, y_{\text{lms}}(i) > Y_{\text{rms}} \\ \text{非冲击点}, y_{\text{lms}}(i) < Y_{\text{rms}} \end{cases} \quad (3)$$

式中, $\text{impact}(i)$ 代表第 i 个点的属性, 阈值 Y_{rms} 通过公式(4)计算

$$Y_{\text{rms}} = S_f \sqrt{\frac{1}{N_d} \sum_{i=1}^{N_d} y(i)^2} \quad (4)$$

式(4)中, N_d 是点的个数; S_f 是比例因子, 其值根据不同的周期情况在一定范围内取值, 取值过小可能会将非冲击区域划分进来, 取值过大可能错过冲击, 本文经过多次试验, 发现 S_f 取值为 1 较为合适, 此时 $Y_{\text{rms}} = 33.14$ 。

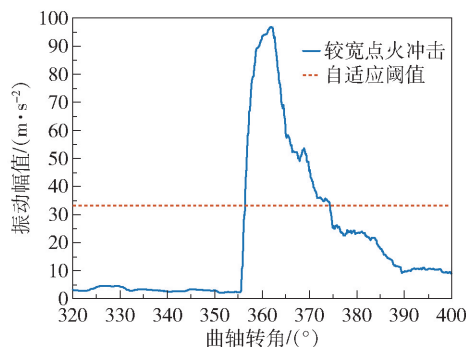


图6 自适应阈值法截取点火冲击

Fig. 6 Adaptive threshold method for intercepting ignition shock

由图 6 可知,点火冲击处于 $356^{\circ} \sim 374^{\circ}$, 由于理论点火上止点和理论非点火上止点间隔 360° , 故将点火冲击的角度范围向后平移 360° , 得到非点火上止点的冲击区域。提取点火上止点冲击有效值、点火上止点冲击峰值、非点火上止点冲击有效值、非点火上止点冲击峰值 4 个局部特征, 用于柴油机故障诊断模型的建立。

3 诊断及结果分析

3.1 模型建立

利用 MATLAB 软件整理所提取的特征数据用于构建 CART 决策树模型。实验的样本数据为正常 40 组、失火 40 组、撞缸 25 组, 训练样本集包括 75% 的正常数据, 75% 的失火数据以及 60% 的撞缸数据, 用剩余特征数据建立测试样本集来检测柴油机故障诊断树的识别效果。由训练集构建的 CART 决策树诊断模型如图 7 所示。

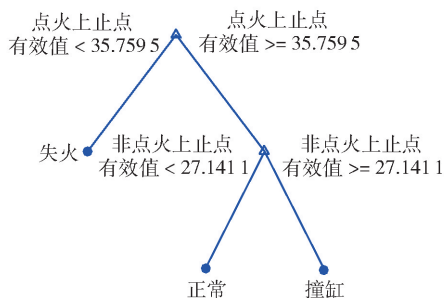


图 7 CART 决策树诊断模型

Fig. 7 CART decision tree diagnostic model

从图 7 树根节点开始遍历整个 CART 决策树, 可以得到以下分类规则:

- 1) 如果点火上止点冲击有效值 $< 35.759\ 5$, 柴油机发生失火故障;
- 2) 如果点火上止点冲击有效值 $> 35.759\ 5$, 且非点火上止点冲击有效值 $< 27.141\ 1$, 柴油机状态正常;
- 3) 如果点火上止点冲击有效值 $> 35.759\ 5$, 且非点火上止点冲击有效值 $> 27.141\ 1$, 柴油机发生撞缸故障。

以上规则体现了柴油机状态判断的参数特征取值情况, 可以发现分类过程只用到两个有效值特征而未用到峰值特征, 说明 CART 决策树可以从多个特征参数中提取合适的特征, 有效简化了模型。

3.2 结果分析

本文通过对测试样本集计算识别率来评估柴油

机故障诊断树的分类性能

$$P = \frac{A}{A+B} \times 100\% \quad (5)$$

式中, P 为识别率, A 为识别正确组数, B 为识别错误组数。利用测试样本集对柴油机故障诊断树进行测试, 测试结果如表 1 所示。由表 1 可以看出, CART 决策树对于柴油机正常、失火和撞缸 3 种工况的识别率都达到了 100%; 并且模型高度只有 3 层, 复杂程度比较低, 不用进行剪枝处理。

表 1 测试结果

Table 1 Test results

工况	测试组数	正确识别组数	识别率/%
正常	10	10	100
失火	10	10	100
撞缸	10	10	100

4 结束语

本文提出了一种基于缸盖振动信号的柴油机故障诊断方法, 采用自适应阈值的局部 RMS 计算方法获取特征并构建 CART 决策树模型。诊断结果表明, 本文方法对正常、失火和撞缸 3 种工况的识别率均达到 100%, 说明本文基于 CART 决策树和局部特征提取的方法能够准确诊断柴油机故障。

在本文后续的研究中, 将考虑解决以下两个问题: ①寻求对于更多故障类型和特征参数的分类处理方法; ②同时发生多种故障时, 如何保证模型仍具有较好的分类效果。

参考文献:

- [1] 吴定海, 张培林, 张英堂, 等. 基于时频奇异谱和 RVM 的柴油机故障诊断研究[J]. 机械强度, 2011, 33(3): 317-323.
WU D H, ZHANG P L, ZHANG Y T, et al. Study on diesel engine faults diagnosis based on time-frequency singular value spectrum and RVM[J]. Journal of Mechanical Strength, 2011, 33(3): 317-323. (in Chinese)
- [2] 姚浩, 李鹏, 郭晓斌, 等. 基于大数据的告警信息处理和故障设备定位技术研究[J]. 电网与清洁能源, 2014, 30(12): 72-77.
YAO H, LI P, GUO X B, et al. Research on alarm information disposal and fault device location based on mass data[J]. Power System and Clean Energy, 2014, 30(12): 72-77. (in Chinese)

- [3] 张翼飞,赵凯. 基于决策树的转子故障自动诊断研究[J]. 计算机测量与控制,2013,21(9):2375-2377,2381.
ZHANG Y F,ZHAO K. Rotor fault diagnosis based on decision tree[J]. Computer Measurement & Control, 2013,21(9):2375-2377,2381. (in Chinese)
- [4] MU Y,SHENG G M,XIA H, et al. Study on fault diagnosis method for nuclear power plant based on fuzzy rough sets and decision tree[J]. Advanced Materials Research, 2014,981:701-705.
- [5] GANGADHAR N, KUMAR H, NARENDRANATH S, et al. Fault diagnosis of single point cutting tool through vibration signal using decision tree algorithm[J]. Procedia Materials Science,2014, 5:1434-1441.
- [6] AMARNATH M, SUGUMARAN V, KUMAR H. Exploiting sound signals for fault diagnosis of bearings using decision tree[J]. Measurement,2013, 46(3):1250-1256.
- [7] DEMETGUL M. Fault diagnosis on production systems with support vector machine and decision trees algorithms[J]. International Journal of Advanced Manufacturing Technology,2013, 67:2183-2194.
- [8] GAO X, ZHANG T, LIU H J, et al. Spacecraft fault diagnosis based on telemetry data mining and fault tree analysis and design of expert system[J]. Advanced Materials Research,2013, 760/761/762:1062-1066.
- [9] 杜春蕾,张雪英,李凤莲. 改进的 CART 算法在煤层底板突水预测中的应用[J]. 工矿自动化,2014,40(12):52-56.
DU C L,ZHANG X Y,LI F L. Application of improved CART algorithm in prediction of water inrush from coal seam floor[J]. Industry and Mine Automation,2014,40(12):52-56. (in Chinese)
- [10] 刘年平,胡慧慧. 基于 CART 算法的煤与瓦斯突出判别分析[J]. 中国矿业,2015,24(6):128-131.
LIU N P,HU H H. Discriminate analysis of coal and gas outburst based on CART algorithm[J]. China Mining Magazine,2015,24(6):128-131. (in Chinese)
- [11] 高洪滨,欧阳光耀. 基于缸盖振动信号分析的柴油机失火故障检测[J]. 车用发动机,2005(1):32-34.
GAO H B,OUYANG G Y. Misfire detection of diesel engine based on analysis of cylinder head vibration signal[J]. Vehicle Engine,2005(1):32-34. (in Chinese)
- [12] JUSTIN F, GARY M B. Fault detection and diagnosis of diesel engine valve trains[J]. Mechanical Systems and Signal Processing,2016,72/73:316-327.

Fault diagnosis of diesel engines based on a classification and regression tree (CART) decision tree

JIANG ZhiNong¹ WEI DongHai¹ WANG Lei² ZHAO ZhiChao² MAO ZhiWei¹ ZHANG JinJie^{1*}

(1. Diagnosis and Self-Recovery Engineering Research Center, College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029; 2. Sinopec Chongqing Natural Gas Pipeline Co. Ltd., Chongqing 404100, China)

Abstract: An adaptive local root mean square (RMS) calculation method has been used to extract the time-domain characteristics of a diesel engine cylinder head vibration signal. The classification model of the diesel engine was established based on the classification and regression tree (CART) algorithm. The vibration data for diesel engine misfiring, crashing cylinders and the normal operating conditions were obtained through experiments. After calculating the local RMS of the original signal, the local characteristics were extracted based on the self-adaptive thresholds to determine the ignition impact area and non-ignition top dead center impact area. Finally, these features were input into the CART algorithm to build a classification model in order to verify the validity of the extracted features. The results show that the recognition rate of the diesel engine reaches 100% in all three states. Thus the CART algorithm and local feature extraction method can effectively diagnose diesel engine faults.

Key words: classification and regression tree (CART) algorithm; diesel engine fault diagnosis; local root mean square (RMS) calculation; adaptive threshold; feature extraction