

基于静电信号变分模态分解和随机森林的气路故障识别方法*

殷逸冰¹, 文振华², 左洪福³

- (1. 青岛理工大学 机械与汽车工程学院, 山东 青岛 266520;
2. 郑州航空工业管理学院 航空发动机学院, 河南 郑州 450015;
3. 南京航空航天大学 民航学院, 江苏 南京 210016)

摘要: 气路静电监测是面向航空发动机健康诊断的全新技术, 为使后续可应用于气路故障识别, 提出一种基于变分模态分解和随机森林的识别方法。首先分析了气路静电监测原理, 针对静电信号噪声干扰问题, 提出了一种基于变分模态分解和峭度-排列熵重构准则的静电信号增强方法, 并提出故障特征集构造方法, 帮助有效提取关键故障信息; 进一步通过开展燃烧室积碳、叶片-机匣碰摩、部件掉块等模拟实验, 获取故障静电信号和特征集, 并构建了基于随机森林的故障类型识别模型。结果表明: 所提方法测试集上识别精度达到90%以上, 且所提出新特征的归一化重要度达到0.2以上, 较传统特征更高, 能为基于静电监测的气路部件故障识别提供有效途径。

关键词: 航空发动机; 气路; 静电传感; 随机森林; 信号处理; 故障识别

中图分类号: V231.1 文献标识码: A 文章编号: 1001-4055 (2023) 05-2207017-13

DOI: 10.13675/j.cnki.tjjs.2207017

Gas-Path Fault Identification Method Based on Electrostatic Signal Variational Mode Decomposition and Random Forest

YIN Yi-bing¹, WEN Zhen-hua², ZUO Hong-fu³

- (1. School of Mechanical and Automotive, Qingdao University of Technology, Qingdao 266520, China;
2. School of Aeroengine, Zhengzhou University of Aeronautics, Zhengzhou 450015, China;
3. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Gas path electrostatic monitoring is a novel technology for aero-engine health diagnosis. To realize typical gas path fault identification by using this technology, a method based on variational mode decomposition and the random forest is proposed. Firstly, the principle and model of gas path electrostatic monitoring are analyzed. Aiming at the noise interference in electrostatic signal, a signal enhancement method based on variational mode decomposition and Kurtosis-Permutation Entropy reconstruction criterion are studied, and a features dataset construction method is proposed to extract the key fault information. The fault electrostatic signals and feature datasets are obtained through the simulated experiments, such as combustion chamber carbon deposi-

* 收稿日期: 2022-07-05; 修订日期: 2022-10-11。

基金项目: 国家自然科学基金 (51975539); 航空科学基金 (2018ZD55008); 山东省矿山机械工程重点实验室开放基金 (2022KLMM203)。

作者简介: 殷逸冰, 博士, 讲师, 研究领域为状态监测与故障诊断技术。

通讯作者: 文振华, 博士, 教授, 研究领域为航空器健康管理技术。E-mail: zhenhuawen@zua.edu.cn

引用格式: 殷逸冰, 文振华, 左洪福. 基于静电信号变分模态分解和随机森林的气路故障识别方法[J]. 推进技术, 2023, 44(5): 2207017. (YIN Yi-bing, WEN Zhen-hua, ZUO Hong-fu. Gas-Path Fault Identification Method Based on Electrostatic Signal Variational Mode Decomposition and Random Forest [J]. *Journal of Propulsion Technology*, 2023, 44(5):2207017.)

tion, blade-casing rubbing, and component drop, and the faults are identified based on the random forest model. The results show that the recognition accuracy of the proposed method is more than 90 % on the test dataset, and the normalized importance of the proposed new feature is more than 0.2, which is higher than the classical features, this method can provide an effective way for fault identification of gas path components based on electrostatic monitoring.

Key words: Aeroengine; Gas path; Electrostatics monitoring; Random Forest; Signal processing; Fault identification

1 引言

航空发动机作为国防战略核心装备,需具备极高的可靠性,而各气路工作部件作为发动机的核心,长期处于高温、高压、高应力的极端环境,容易发生燃烧室积碳、部件掉块、叶片-机匣碰摩等气路机械故障,进而引发发动机整机故障,甚至可能造成巨额经济损失和灾难性事故^[1]。另一方面,传统定时维护模式下运营成本居高不下,在确保气路部件安全运行的同时需考虑经济性,避免维修过度造成资源浪费。因此,通过对发动机气路进行有效监测感知气路健康状态十分重要^[2]。传统发动机气路故障监测技术大多着眼于发动机性能参数,侧重于尾气排气温度、转速、压力等气路热力学参数监控,但其弊端在于其不具备一定故障预警能力,只有当气路部件在故障较为严重时,故障信息才能在监测参数中反映出来,无法在故障早期进行在线诊断或提供预警信息。以燃烧室积碳故障为例,只有当积碳严重到一定程度造成发动机性能衰退和尾气温度超限时才能够被探测到^[3]。

近年来,一种基于静电感应的气路静电监测技术成为发动机故障预测领域的热点。相较于传统技术,静电监测通过非接触方式探测发动机尾气所含带电的故障早期产物,具有更强的故障征兆发现能力,能为常规方法提供有益补充^[3]。英美等国对气路静电监测进行了技术研发,并已形成成熟的产品应用于战斗机中。Powrie等^[4]利用航空发动机加速测试实验监测到了叶片碰摩及燃烧室材料丢失等故障,通过故障数据积累与分析实现了故障定位;Melissa等^[5]首次研究了静电监测技术在工业燃气轮机中的应用,并通过实验表明静电监测能够提高燃气轮机的异常状态监测能力。但由于静电监测技术涉及军事被禁止出口,国内难以获取更详细的公开文献。国内南京航空航天大学率先开展了发动机静电监测领域相关研究,相继对静电感应机理、静电监测技术(气路、滑油、磨损监测等方向)、信号处理以及相关

验证实验开展了针对性研究。

文献[6-8]开展了某型发动机静电监测实验,获取了在气路部件碰摩状态下静电信号的异常波形,并提出的基于粗糙集神经网络的异常颗粒识别方法,但仍未深入探究故障与静电信号的关联关系;Liu等^[9]针对碰摩故障研究了一种基于时频特征的初步判断方法,但并未讨论如何与其他气路故障进行具体区分;唐欣等^[10]开展了某型船用燃气轮机台架监测实验;殷逸冰等^[11]在涡扇发动机试车台架上开展了机载化监测实验,探测到滑油泄露故障时信号异常,静电信号和发动机性能参数的正相关关系,该次实验是发动机静电监测技术首次在民航发动机上成功验证。付宇等^[12]研究了一种基于自组织网络的方法对气路故障进行初步判别,但是所提取的特征并不全面,也未考虑到样本不平衡和竞争层神经元数量对模型泛化能力影响的问题。以往文献表明,静电监测在发动机故障预警具有较高研究价值和应用的可行性。但目前大部分研究仅证明了“在故障发生时会产生异常静电信号”这一浅层次问题,未对常发性气路故障的静电监测识别问题进行深入的探究。例如,由于燃烧室积碳、部件掉块、叶片-机匣碰摩这几类典型气路故障机理各异,存在故障时尾气静电信号的隐含信息必然不同,如何通过分析并挖掘出能够代表不同故障模式差异的信号特征,合理构建故障特征集,并形成具有强泛化能力的智能模型进行故障识别,这些问题仍亟待探索。

此外,由于实际采集静电信号容易引入其各类噪声,使故障特征提取受到严重影响,因此需要在特征提取对信号进行处理。文振华等^[13]提出一种基于独立分量分析的消噪方法,但该方法需构建参考噪声解决信息欠定问题,灵活性有欠缺;Zhang等^[14]针对磨粒监测研究了一种基于谱插值和奇异值差分谱的降噪方法,但该方法对单一随机噪声去除效果较理想;付宇等^[15]提出一种基于稀疏分解的静电信号去噪算法,不足之处在于算法计算量较大,不适用于在线处理。考虑到后期实际工程应用对信号在线处

理要求,目前仍缺乏一种能够针对静电信号多类噪声,兼顾降噪效果与在线处理效率的信号处理方法,以支撑实际应用。

综上所述,针对静电监测技术工程应用前尚需探索的问题,本研究提出一种基于变分模态分解(Variational Mode Decomposition, VMD)和随机森林(Random Forest, RF)的气路故障识别算法。首先,基于VMD算法对原始信号进行分解,VMD是一种非递归计算方法,相较于传统经验模态分解及其改进方法采用递归运算方式,其计算速度得到明显提升;结合根据静电信号含随机噪声、毛刺干扰的特点,提出峭度-排列熵筛选准则对分解后模态分量进行自适应重构,综合去除噪声,最大程度保留信号中所含的有用信息,实现信号有用成分增强,满足快速处理和噪声综合去除需求;基于时域统计和小波方法,分别对增强信号和原始信号进行特征提取,完成信号特征集构造;进一步开展燃烧室积碳、叶片-机匣碰摩、部件掉块等气路故障模拟实验,获取故障静电信号和特征样本集;最后,考虑故障样本和正常样本数量不平衡的特点及模型泛化能力,基于随机森林模型对故障样本集进行训练和识别,并通过随机森林中的特征重要度评估所有特征贡献程度,帮助后续优化故障特征集构造方法。

2 静电监测原理与信号增强方法

2.1 监测原理

气路静电监测技术主要基于静电感应原理,监测原理如图1所示,具体如下:当静电传感器金属探极感应空间内经过带电颗粒时,探极与带电物形成静电场并相互作用,探极内部电子随着带电物移动重新分布以平衡位于附近的静电场,使得探极内部产生感应电流,经过传感器等效电路的转换输出感应电压,进一步被信号调理后产生静电信号。以积碳故障为例,发动机稳定燃烧排放一定大小的碳烟颗粒,这些颗粒吸附燃烧过程产生的带电离子和电子致使荷电,使尾气总体荷电水平保持平稳;随着碳沉积,逐渐使喷嘴燃油雾化性能退化和不完全燃烧加剧,生成较多具有更强电荷吸附能力的大粒径异常碳烟颗粒,使监测中频繁出现异常静电信号脉冲。而当存在气路部件掉块或碰摩等故障,尾气会携带磨损下来的具有异常粒径和电荷的金属颗粒物,这些金属颗粒由于材料离子键锻炼和摩擦而带上异常量电荷^[2],静电传感器也会实时捕捉此类异常电荷变化。

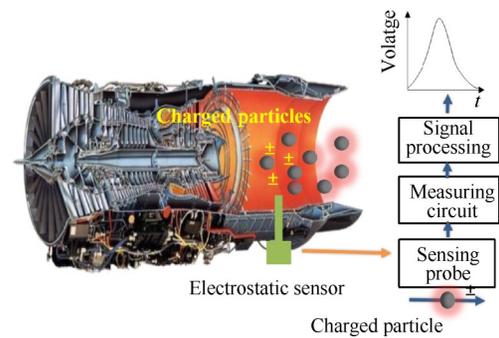


Fig. 1 Principle of electrostatic monitoring

2.2 测量模型

传感器测量等效电路如图2所示。图中 R 为等效测量电路中的等效电阻^[16]; C 为等效电路中等效电容; Q 为感应电荷量; $U(t)$ 为当前 t 时刻传感器输出电压。其输出电压和感应电荷间的拉式变换关系式为

$$U(s) = \frac{R \cdot s \cdot Q(s)}{R \cdot C \cdot s + 1} \quad (1)$$

式中 $U(s)$ 为输出电压的拉氏变换; $Q(s)$ 为感应电荷的拉氏变换。

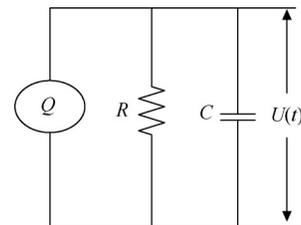


Fig. 2 Measurement circuit

当 $R \cdot C \cdot s \ll 1$ 时,联合式(4)计算得到输出电压为

$$u(t) = R \cdot Q'(t) \quad (2)$$

式中 $Q'(t)$ 为 t 时刻感应电荷量导数。最终,传感器的输出电压与感应电荷量的变化速率成正比。

2.3 基于变分模态分解的信号增强

航空发动机实测静电信号包含大量背景带来的噪声,主要噪声包括空间工频干扰、随机白噪声和随机信号毛刺等,这些噪声会对有用信号特征提取产生影响。以往研究中常采用类经验模态分解的方法。传统经验模态分解方法存在模态混叠且重构过程中本征模态分量(Intrinsic Mode Functions, IMF)难以选择等问题。后续互补集合经验模态分解(Complementary Ensemble Empirical Mode Decomposition, CEEMD)、自适应噪声完全集合经验模态分解(Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, CEEMDAN)等改进方法的出现对模

态混叠处理具有不错效果,但是重复添加白噪声的过程使得计算复杂度和时间大幅增加,客观上不适用工程应用;结合过往所获取静电实验数据发现,当发动机存在故障时,产生的有用故障静电信号存在一定稀疏冲击性特征,且主要集中于低频段。因此,原始信号噪声和有用信号可以被当为几类典型组成模态。

本研究考虑上述航空发动机故障静电信号特点、噪声特性和工程计算量问题,提出一种基于变分模态分解和峭度-排列熵准则的静电信号优化重构方法,通过VMD方法对信号进行快速分解,获取若干个模态分量,进而根据故障信号具有冲击性稀疏特征和随机噪声信息含量较多的特点,提出利用峭度-排列熵准则帮助从已获取的模态中进行剔除、筛选和重构,以进一步增强有用的故障信号,在帮助开展更为精准的特征提取。

VMD是一种非递归式方法,规定了最后信号分解的个数,相较于传统经验模态分解及其改进方法采用递归运算方式,其计算速度会得到明显的提升,结合根据静电信号含噪特点提出的筛选准则,可将信号中所含的有用信息最大程度保留,综合去除噪声,因此可以该方法满足快速处理和噪声综合去除这两个目标。VMD能够自适应分解的信号,过程实质是构造和求解约束变分问题。即是通过寻求若干个模态函数,使得分解后得到的信号模态的频谱估计带宽之和最小,约束条件为各模态之和等于输入信号^[17]。该算法较为适合内部包含几类主要模态的测量信号分解,而静电含噪信号具有类似特点^[15]。

该算法中,输入信号被分解为指定个数的IMF,IMF被定义为一个有带宽限制的调幅-调频函数^[18],其表达式为

$$u_k(t) = A_k(t) \cos[\phi_k(t)] \quad (3)$$

式中 $A_k(t)$ 是 $u_k(t)$ 的瞬时幅值,且 ≥ 0 ; $\phi_k(t)$ 是信号相位,是一个非负函数; $\omega_k(t)$ 是 $u_k(t)$ 的瞬时频率。

$$\omega_k(t) = \frac{d\phi_k(t)}{dt} \quad (4)$$

寻求 N 个模态函数 $u_k(t), k=1, 2, \dots, N$,使 N 个IMF估计带宽之和最小。通过以下步骤估计各IMF的带宽:

(1)对模态函数 $u_k(t)$ 进行希尔伯特变换,获取相应的边际谱;

$$S_k = (\delta(t) + \frac{j}{\pi t}) \cdot u_k(t) \quad (5)$$

(2)通过指数混合调制方法,预估每个模态分量

的中心频率,并将其频带调制到相应的基频带;

$$S_{\beta k} = \left[(\delta(t) + \frac{j}{\pi t}) \cdot u_k(t) \right] e^{-j\omega_{\beta k} t} \quad (6)$$

(3)计算基频带中解调信号梯度的平方 $L2$ 范数,估计每个模态带宽,构造的约束变分问题可表达为

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_k \left\| d_t \left[(\delta(t) + \frac{j}{\pi t}) \cdot u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} & \sum_k u_k(t) = x(t) \end{aligned} \quad (7)$$

式中 $\{u_k\} = \{u_1, u_2, \dots, u_k\}$ 为模态函数集合;中心频率集为 $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_k\}$; $\delta(t)$ 为单位脉冲函数; d_t 是对时间求导; s.t.为约束条件; $x(t)$ 为原始输入信号; N 为最终分解得到的模态数量。

为求解式(7)中变分约束模型最优解,引入二次惩罚因子 α 和拉格朗日乘法算子 $\lambda(t)$ 将其变为非约束变分问题,得到增广拉格朗日表达式为

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left\| d_t \left[(\delta(t) + \frac{j}{\pi t}) \cdot u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ & \left\| x(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), x(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (8)$$

(4)利用乘法算子交替方向法搜索拉格朗日函数极小值点,并迭代更新 u_k^{n+1} 和 ω_k^{n+1} ,以及每个IMF的带宽和中心频率,满足误差条件截止。最终输入信号被分解为 N 个IMF,即为最优分解结果。

由于VMD算法需预设模态数 N ,其取值影响最后分解结果的优劣。若 N 过小,其分解的模态分量信息会丢失,或者成频率混叠,若 N 过大,会造成过分解与计算量增加,分解效果欠佳。考虑采用中心频率方法对 N 值进行选择,惩罚因子采用默认值2000,选用实测静电信号进行分解,采样频率2kHz,不同 N 值对应IMF的中心频率示意如表1所示。

Table 1 IMF center frequency under different N values

N	Center frequency/Hz							
2	41	458	-	-	-	-	-	-
3	41	247	643	-	-	-	-	-
4	36	185	469	735	-	-	-	-
5	36	159	362	583	791	-	-	-
6	33	121	287	481	657	829	-	-
7	33	108	267	457	631	826	865	-

当 $N=7$ 时出现了中心频率相近的模态分量,导致过分解,因此确定模态分量个数 $N=6$ 。从表中可以看出, $N \geq 6$ 时,中心频率的最大值和最小值区域稳定因

此本文设置分解尺度 N 值为 6。某段静监测电信号的分解结果如图 3 所示。

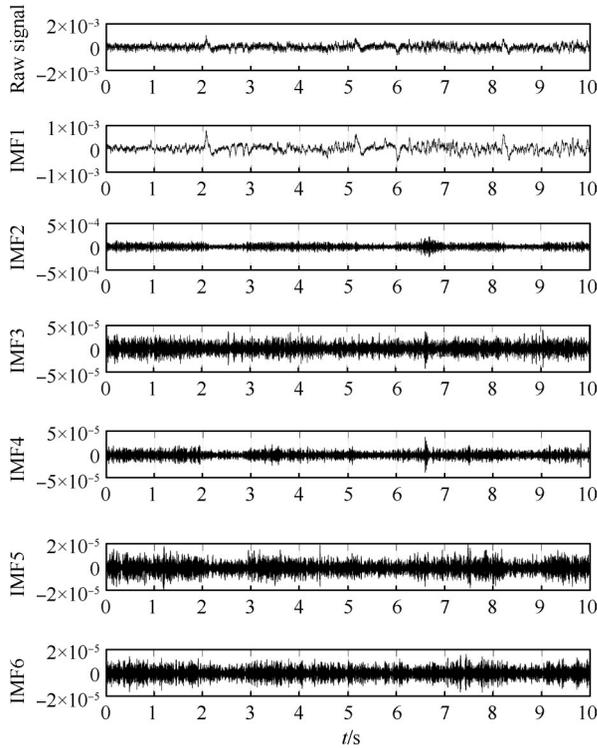
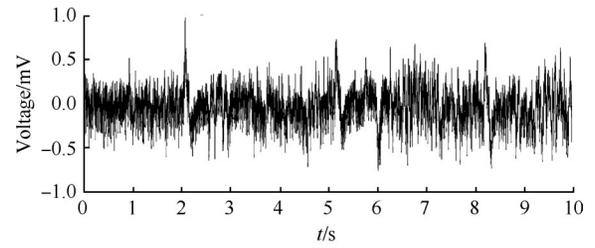


Fig. 3 Decomposition result by VMD

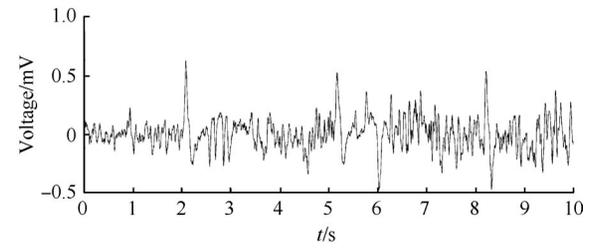
2.4 基于峭度-排列熵准则的信号重构增强

为更精准提取特征,需选取分解结果中合适的 IMF 分量进行信号重构增强。故障发生时,静电信号中会产生与故障相关的冲击脉冲,导致信号整体复杂程度和信息含量降低。基于这一特点,本文采用基于峭度准则和排列熵准则方法对 VMD 分解后的 IMF 进行选择。其中,排列熵值的大小代表输入信号的随机程度,熵值越大,随机噪声成分越多;通过计算不同 IMF 分量的排列熵,剔除排列熵较大的 IMF,达到进一步降噪的目标。而峭度值高的分量代表包含更多静电监测颗粒在时域上的冲击特征信息,排列熵和峭度的具体计算原理在此不再赘述。本研究拟首先将排列熵值最高的两个分量进行剔除,进一步在剩余 IMF 中选择峭度排名前二的分量作为重构成分。

某段实测静电信号经过 VMD 分解后得到各 IMF,基于本研究提出的重构准则,剔除具有高排列熵的 IMF,再于剩余 IMF 选择峭度值排名前二的分量进行重构,得到的重构信号如图 4 所示。从图中可以看出,原始静电信号毛刺和随机噪声被大幅度滤除,有用信号脉冲已被显著增强。



(a) Raw signal



(b) Reconstruction signal

Fig. 4 Reconstruction result

3 特征集构造和故障识别方法

3.1 特征提取

针对静电信号,主要提取时域特征和频域能量特征;时域特征主要基于第 2 节方法得到的重构信号进行提取。经典时域静电参数包括:代表信号活跃程度的活动率水平 (Active Level, AL);事件率 (Event Rate, ER)^[19]。频域特征则通过对原始信号频域能量分布分析后进行提取。本研究根据不同故障机理,提出负事件贡献率和低频能量贡献比这两个新特征,对静电信号进行更深层次的度量。

3.1.1 活动率水平

活动率水平 AL 指静电信号在某一段时间内的有效值,单位为皮库 (pc),计算公式为

$$AL = \sqrt{\frac{1}{T} \int_0^T C_i^2(t) dt} \quad (9)$$

式中 T 常取 1s 时间长度,实际可按照信号分析需求自行设定; $C_i(t)$ 为电荷量信号,通过对电压积分获取。

3.1.2 正负事件率和负事件贡献率

事件率 ER 物理意义指在 T 时间内,超过当前时间段下 AL 值若干倍 (通常为 3) 电荷量点的个数 M 所占该段时间内总采样点数 N_{all} 的百分比,其表达式为

$$ER = \frac{M}{N_{all}} \cdot 100\% \quad (10)$$

式中 N_{all} 为电荷信号在时间 T 内的采样点数,如采样频率为 2kHz,则取 $N_{all}=2000T$ 。

事件率包括正事件率 (Positive Event Rate, R_{PE}) 和负事件率 (Negative Event Rate, R_{NE})。通过监测信号

正负极性区分并计算,例如 R_{NE} 参数代表阶段时间内异常负电信号发生的频次。典型的事件包括大的碳颗粒和由故障引起的异常大粒径颗粒。当出现积碳等性能退化类故障,会导致燃烧产物中的异常碳烟颗粒增加,由于是非金属颗粒主要带负电,从而使 R_{NE} 显著增加;而发生金属部件互相碰摩或掉块时,会引起 R_{PE} 参数变化。

在一段采集时间内,考虑通过计算信号中 R_{NE} 和总事件率的比值衡量异常负信号频次情况,并作为区分不同故障模式的一类参数。因此,在上述分析的基础上,定义静电信号负事件贡献率特征 R_{NN} 为

$$R_{NN} = \frac{\sum_{i=1}^{N_1} R_{NE}}{\sum_{i=1}^{N_1} R_{NE} + \sum_{i=1}^{N_1} R_{PE}} \quad (11)$$

式中 N_1 取值和时间间隔有关,一般 N_1 等于信号对应的总时间和 AL 中时间间隔 T 取值的商。

3.1.3 峭度

当金属件碰摩故障发生时,将会在时域形成典型的冲击特征,因此还考虑使用峭度准则作为其中一个特征。峭度是反映随机变量分布特性的数值统计量,是数据的归一化4阶中心矩,即

$$Kv = \frac{\frac{1}{N} \sum_{i=1}^N \left(X_i - \frac{1}{N} \sum_{i=1}^N X_i \right)^4}{\left(\frac{1}{N} \sum_{i=1}^N \left(X_i - \frac{1}{N} \sum_{i=1}^N X_i \right)^2 \right)^2} \quad (12)$$

式中 X 是输入信号的序列。峭度对冲击类的信号较为敏感,当故障颗粒增多时,会引发静电信号时域冲击,平均峭度值增加较快。

除时域特征外,信号频域也包含了较多信息,而静电信号较多集中于低频区间。正常和故障状态下,噪声能量和有用异常信号能量在总能量中的占比必定有较大差异,主要体现在能量分布。因此,噪声能量需要作为一种参考加入特征计算,频域特征也主要基于原始监测信号进行计算。综上,本研究基于小波和小波包分解方法,构建两个频域信息特征,低频能量贡献率和小波能谱熵,分别用于衡量信号低频成分比重和信号能量分布的有序程度。

3.1.4 低频能量贡献比

该特征是基于小波包分解,获取频域等尺度能量分布特征^[20]。分解理论本文不再赘述,小波包能量分解采用4层分解,获得频域上16个频带的能量分布($E_0 \sim E_{15}$),将前4个频带的能量($E_0 \sim E_3$)总和作为低频带的能量,记 E_{SUM} 为信号所有频带总能量,小波

包能量谱获取方法本文不再赘述。定义特征低频能量贡献比(Ratio of Low Frequency Band, R_{LFB})为:在某段时间内信号的小波包归一化能量谱中,设定的低频带能量和信号总能量比值,计算方法为

$$R_{LFB} = \frac{\sum_{i=0}^3 E_i}{E_{SUM}} \quad (13)$$

某段时间内信号的小波归一化能量谱中,设定的低频带能量和信号总能量比值。图5给出了分解计算示意图。

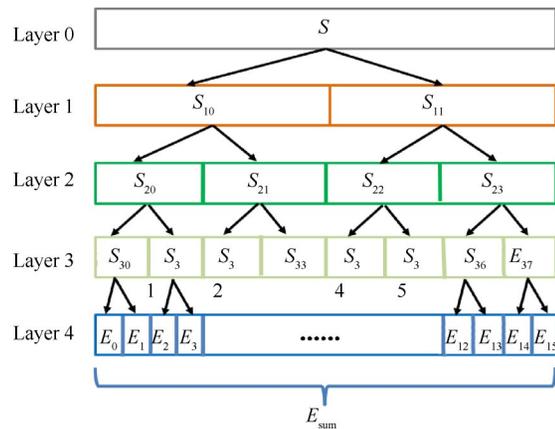


Fig. 5 Energy decomposition

3.1.5 小波能谱熵

小波能谱熵(Wavelet Energy Entropy, WEE)是一种将小波分解结果融入到信息熵计算过程中的度量方法,需将小波系数看作概率分布的序列^[21],通过计算序列熵值反映系数矩阵稀疏度。据小波分解过程对离散信号进行 M_1 层的小波分解。原始信号即可表示如下:

$$f(n) = \sum_{m=1}^{M_1} D_m(n) + A_J(n) = \sum_{m=1}^{M_1+1} D_m(n) \quad (14)$$

取不同离散尺度 M_1 , 可以对应得到不同的小波系数。在小波分解后,在某个尺度下的信号总能量为该尺度下对应的小波系数的能量之和:

$$E = \sum_{k=1}^{N_1} |D_m(k_s)|^2, m = 1, 2, 3, \dots, M_1 \quad (15)$$

式中 N_1 为目标信号的采样点数; $D_m(k_s)$ 为此尺度下的小波重构系数,并获取信号在各个尺度上能量谱结果 $[E_0, E_1, \dots, E_{M_1}]$, 形成对信号能量的一种划分。

某一时间窗内信号总能量等于各分量能量 E_m 之和, 设 $R_m = E_m/E$, 有 $\sum_{m=1}^M R_m = 1$, 那么, 根据信息熵的定义, 相应的小波能谱熵则定义为

$$WEE = - \sum_{m=1}^M R_m \log_2^{R_m} \quad (16)$$

该特征在实际信号中所代表的物理意义是信号的频率分布均匀性,信号频率越简单则能谱熵越小,信号频率越复杂则能谱熵越大。当发动机出现气路故障,经常伴随着异常带电颗粒排出,这些带电异常颗粒引发低频响应,这就使得原有信号能量分布结构被重新构筑。

一段静电监测信号的样本,经过VMD分解和重构后提取时域指标,首先获取时域特征。当信号样本的对应时间大于1s时,样本的活动率水平和正负事件率取需要取同一组信号内多个时间阶段计算结果的平均值;随后,对原始信号进行频域指标的提取,最终从单一样本中获取特征向量 $F=[AL, R_{PE}, R_{NE}, R_{NN}, Kv, R_{LFB}, WEE]$ 。后续每个信号的样本都需获取特征向量 F ,用于构建训练集和测试集。

3.2 故障识别算法

考虑到后续工程应用中故障样本数量相对于正常样本数量较少,存在样本不平衡问题,且需要适应不同的发动机类型,需要获取泛化能力较强故障识别模型,因此在信号处理和特征提取阶段后,本研究提出基于随机森林算法对故障识别模型进行训练。随机森林是近年来出现的一种能够较好解决分类问题的算法,通过训练多个弱分类器解决相同的问题,并将它们结合起来以获得更好结果,具有较强的模型泛化能力,不容易过拟合且训练速度比较快,近年来逐步应用于工程领域中^[22-24]。此外,随机森林模型能够对本研究特征集中不同特征的重要程度进行评估,帮助后续优化特征集构造方法,选择区分度更高的特征作为组成元素。

随机森林模型本质是由若干个决策树模型集成为分类能力更强的决策群。该方法基于一个重要假设:当弱模型被正确组合时,可以得到更精确、更鲁棒的模型。进行诊断时,将测试集输入训练好的模型,然后让每一棵树独自进行决策,即“投票”,最后综合投票结果,将所有决策树中分类结果最多的那类为最终结果。假设RF模型 $\{h(x, \theta_s), S = 1, 2, \dots\}$ 是由若干个决策树组成的集合,集合 $h(x, \theta_s)$ 中是利用CART算法构造的未剪枝的分类决策树, x 为输入样本, θ_s 是独立同分布随机向量,决定决策树生长,算法步骤如下:

(1)自助重采样。利用Bootstrap重抽样方法从含有若干样本的原始数据集中进行 S 次随机且有放回的抽样,抽取样本集大小为原始数据集的63.2%,剩

余样本作为测试数据。形成 S 个Bootstrap子样本集,每个子样本集中均含有 n_1 个样本^[18]。

(2)决策树生成。针对每个子集,采用CART算法进行决策树构建。定义训练样本输入特征个数为 M_f ,从 M_f 个特征中随机选择 m_{try} 个特征 (m_{try} 保持不变) 作为候选特征,按照节点不纯度最低原则从 m_{try} 个特征中选择最优特征进行分裂生长;然后递归调用上述过程构建各分枝,使树充分生长,直到所有特征都已经使用^[19]。

(3)重复步骤(2),利用 S 个Bootstrap样本集构建 S 个决策树形成随机森林。

(4)投票分类。用集成的多颗决策树进行分类,分类标签是由所有分类树的结果综合而成。随机森林使用的是投票原则,即少数服从多数原则。投票公式为

$$c_p = \arg \max \left(\frac{1}{n_{tree}} \sum_{i=1}^{n_{tree}} I\left(\frac{n_{h_i,c}}{n_{h_i}}\right) \right) \quad (17)$$

式中 n_{tree} 是森林中决策树的数目, $I(*)$ 是示性函数; $n_{h_i,c}$ 是树对类别 C 的分类结果, h_i 是树的叶子节点数。

3.3 总体算法流程

根据前文分析处理方法,本节给出VMD-随机森林的气路故障静电监测辨识方法总体实现步骤如下,算法流程如图6所示。

- (a)输入若干组原始静电信号;
- (b)对每组原始信号进行VMD分解,基于峭度-熵准则重构后获取增强信号;
- (c)对原始信号进行小波包分解和小波分析获取信号能量分布和小波能谱;
- (d)对重构增强信号提取时域特征参数 $AL, R_{PE}, R_{NE}, R_{NN}, Kv$, 基于(3)获取的能量特征提取 R_{LFB} 和 WEE 参数,构建特征向量样本集;
- (e)重复(a)~(d)过程,共获取若干组特征向量集合,构建模型训练样本集。
- (f)进行特征重采样和随机森林中多个决策树模型训练,形成树模型群;
- (g)重复(a)~(d)过程,构建模型测试集,输入到已训练完成的识别模型,并最终输出样本所属的故障类别。

4 模拟实验与分析

4.1 故障模拟实验方法

本研究针对的主要气路故障为燃烧室积碳、叶片-机匣碰摩、部件烧蚀掉块故障三类常见的气路异常问题。三类故障监测机理如下:燃烧室积碳故障

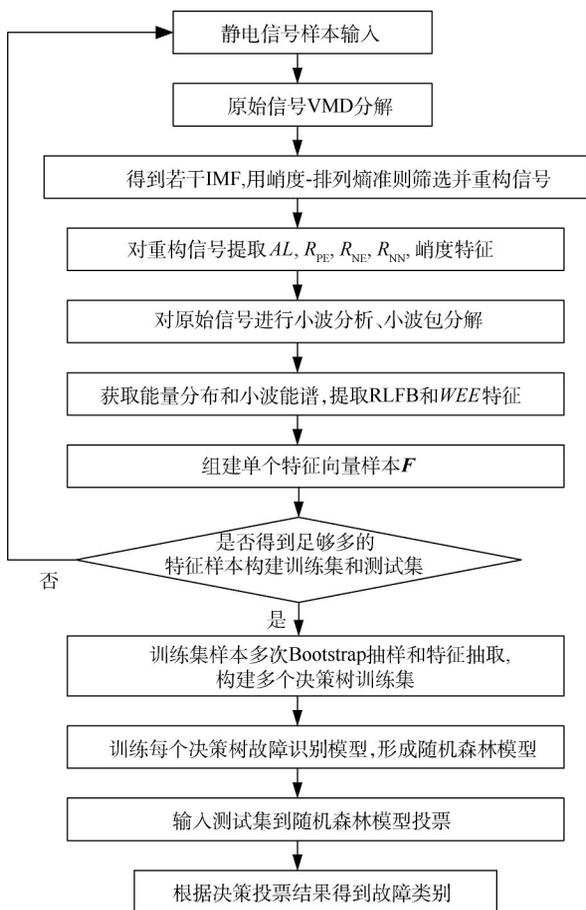


Fig. 6 Fault identification algorithm flow of VMD-RF

是燃油燃烧不充分形成尾部碳烟或碳烟在喷嘴处积聚形成碳块,造成发动机燃烧室出口燃气温度不稳定,燃烧效率降低,尾气中带电颗粒会剧增;叶片-机匣碰摩故障是发动机的涡轮叶片与机匣相互摩擦,引起叶片材料耗损,磨损下的带电颗粒随尾气排放出;部件烧蚀掉块故障是指燃烧室燃烧不稳定导致的后端气路部件烧蚀掉落并带上电荷,带电的掉块常引起后级部件的打伤。

因此,需用合理实验方案进行故障模拟,本研究在静电监测故障模拟试验台上开展实验,该实验台能够完整模拟碰摩故障、燃烧故障和部件烧蚀掉块故障,并装有小型涡喷发动机,具体如图 7 所示。

用激振器-转盘碰摩实验、过量燃油工况实验、流场颗粒注入实验分别模拟叶片-机匣碰摩、燃烧室积碳和部件烧蚀掉块等三类故障,获取相关实验数据。碰摩模拟实验通过相关子系统实现,包括激振往复装置、碰摩转子盘和控制装置等子装置,通过激振装置上的碰摩头往复撞击转子盘进行模拟;积碳故障模拟通过控制发动机功率与过量燃油喷入进行模拟,使其处于富油状态,模拟积碳故障时的尾气环

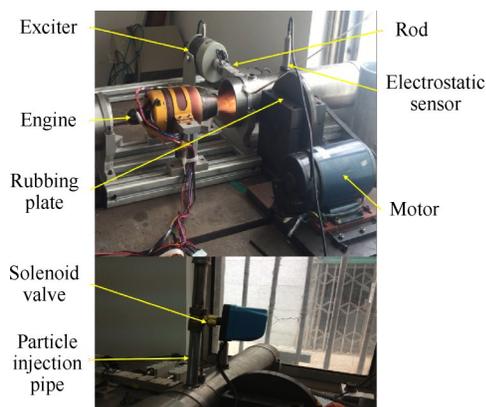


Fig. 7 Fault simulation test rig

境;烧蚀掉块故障则通过利用电磁阀门开关进行颗粒物注入尾气流道中实现模拟。静电传感器安装于尾气管道中,在故障模拟实验开启后采集尾流环境下的静电信号。

4.2 实验样本获取

故障模拟实验下和正常状态实验下的样本获取过程与方法如下:

(1)从实验平台分别开展三类目标故障模拟实验和正常工况静电监测实验,每类实验均开展若干次,并都采集整个实验过程的静电信号数据;

(2)从四种类型实验中获取的数据中进行随机的片段性抽取,设定单一信号抽取片段长度为 5s(即 10000 个数据点)。为模拟实际情况中可能出现的样本不平衡情况,在每个故障类型实验采集信号中随机抽取 30 次,在正常工况实验采集信号中随机抽取 90 次,共计获取 180 段原始信号。

(3)依次对每段原始信号进行 VMD 分解,得到 K 个模态分量 IMF,并基于所提出的筛选准则获取重构信号样本;

(4)分别对对应的重构信号提取时域特征,对原始信号提取频域特征,形成信号特征向量,构建最终样本集合;

(5)将构建信号特征样本集进行归一化处理,输入随机森林中进行训练,得到故障类型识别模型;

(6)重新进行步骤(2),在每个类型模拟实验所采集信号中再次随机抽取 15 组信号,共计获取 60 组信号;继续对 60 段信号进行步骤(3)~(4),得到样本量为 60 的集合,作为识别模型的测试集;

(7)将步骤(6)中获取测试集进行归一化处理,输入到随机森林模型分类器中进行类型辨识,得到测试数据的故障分类诊断结果。

4.3 故障静电数据分析

经过4.2节模拟实验,分别得到了四种类型的信号,包括三种前文提及的故障发生时信号以及正常状态信号。图8分别给出了这四种状态时的典型原始气路静电信号。

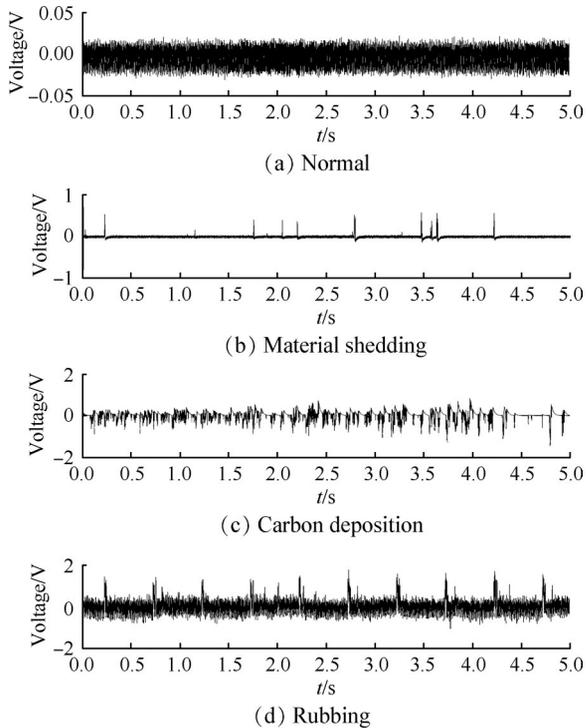


Fig. 8 Electrostatic signal under normal and fault simulation condition

在掉块故障模拟实验中进行了多次金属掉块投放活动,发现有多次较为明显的异常静电信号,且脉冲方向为正;燃烧室积碳模拟实验中开启后,静电信号在时域的振幅明显变大,信号峰值在负极性部分较多,且一直持续,说明积碳故障模拟中持续感应到了尾气中不断出现的异常带电颗粒,如碳烟和未燃尽的油滴;碰摩模拟实验中,激振碰摩杆撞击碰摩盘的频率设为2Hz,在原始信号中出现了周期性的脉冲信号,其频率与碰摩频率保持一致。从时域信号来看,可以看出正常信号与其他三类故障信号的时域表现各有差异。在对信号进行降噪处理后,进行相关特征提取。

图9和图10分别给出了四类样本获取的平均 R_{NE} 和 R_{NN} 的对比,其中平均 R_{NE} 按如下方式获取:对信号样本每隔400个数据点依次计算 R_{NE} 后再进行平均得到最终 R_{NE} ,作为样本最后提取的一个特征。

可以看出,积碳故障情况下, R_{NE} 和 R_{NN} 特征最高,而掉块故障和碰摩故障情况下, R_{NE} 和 R_{NN} 特征的值较

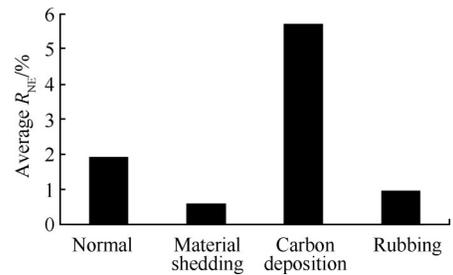


Fig. 9 Average R_{NE} of different signal samples

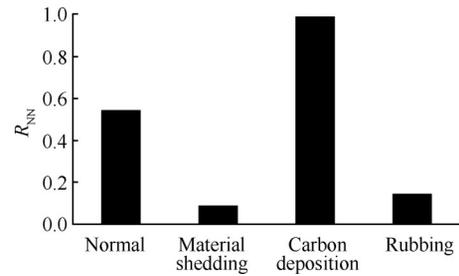


Fig. 10 R_{NN} of different signal samples

低,符合原始信号的时域波形特点。

图11给出了这四种不同类型静电信号经过小波包分解后得到的归一化能量谱分布具体情况。可以看出四种不同类型的信号样本中,归一化能谱具有较大差异,积碳故障和碰摩故障发生时,低频段能量所占总能量的比重更高,当处于正常状态时,能量分布似乎更为均匀,低频段信号和高频段信号所占的比重差异不大。

图12是通过对多个样本获取归一化能量谱,计算得到的不同类型信号样本的平均 R_{LFB} 特征值对比图。可以看出,总体上积碳故障的平均 R_{LFB} 特征的值最高,碰摩故障的信号样本 R_{LFB} 值排在第二,而掉块故障和正常状态下信号的 R_{LFB} 故障维持在较低水平。通过故障数据初步分析,能够发现不同故障类型和正常状态下,信号部分特征已经表现出了一定的差异性。

4.4 故障识别与特征分析

经过样本处理过程,可将叶片-机匣碰摩、燃烧室积碳、部件掉块和正常状态四种类型分别记为F1~F4类。在获取样本后,需首先对故障模拟实验中获取的样本进行数据归一化;归一化完成后,利用3.2节中获取的训练样本集进行RF模型训练,初次训练模型设定决策树颗数 n_{tree} 为100,特征子集的个数 m_{try} 按照经验为总特征个数 M 的根号值,向下取整为2;测试集中F1~F4每类15组测试样本,依次按照顺序投入模型进行测试,初步得到的故障类型识别结果如图13所示。

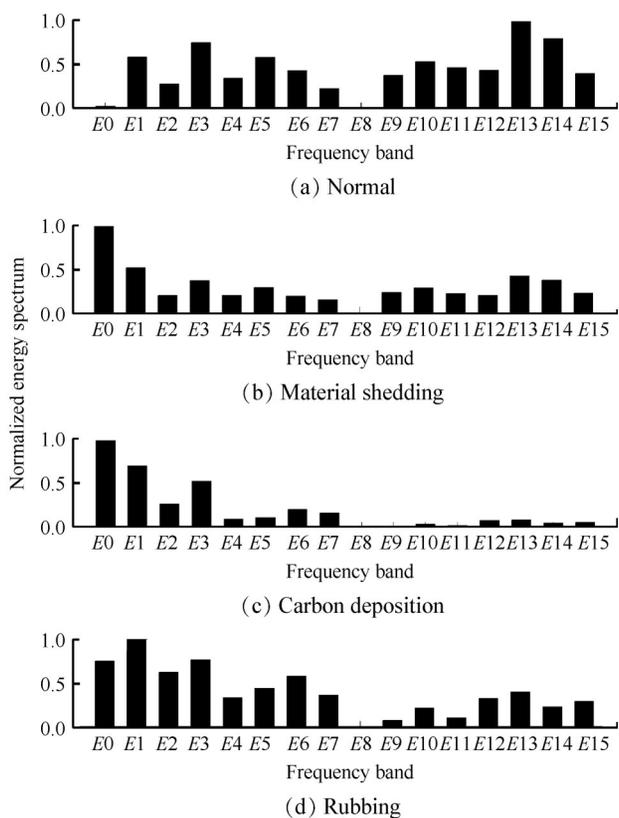


Fig. 11 Normalized energy spectrum of different signal samples

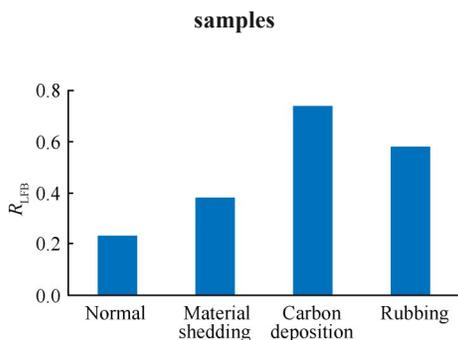


Fig. 12 R_{LFB} of different signal samples

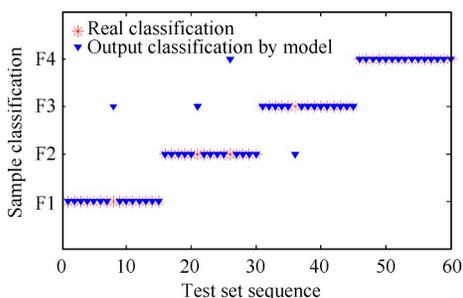


Fig. 13 Fault identification results

从图13中可知,在决策树颗数为100和特征子集个数为2的情况下,对总体60个测试样本进行识别,正确的个数为56个,识别准确率为93.3%;表2给出了本次故障识别过程中对应测试数据的故障分类混淆矩阵表。

Table 2 Confusion matrix of test data

Output	Output				Error rate/%
	F1	F2	F3	F4	
F1	14	0	1	0	6.66
F2	0	13	1	1	13.33
F3	0	1	14	0	6.66
F4	0	0	0	15	100

由表2可知,利用RF进行故障辨识时,F1被错分为F3的有1个,被分类错误率为6.66%;F2被错分为F3的有1个,被错分为F4的有1个,分类错误率为被分类错误率为6.66%,F3被错分为F2的有1个,被分类错误率为13.33%。

在整个辨识算法过程中涉及到两个参数的选择,分别为最优决策树的棵数和特征子集个数。本研究中特征子集个数保持不变,设为3个,研究不同决策树颗数下对模型识别正确率的影响。分别设 n_{tree} 为25, 50, 75, 100, 125, 150, 200, 300, 400, 600,依次训练随机森林模型,并通过车测试集进行模型测试,测试结果如图14所示。

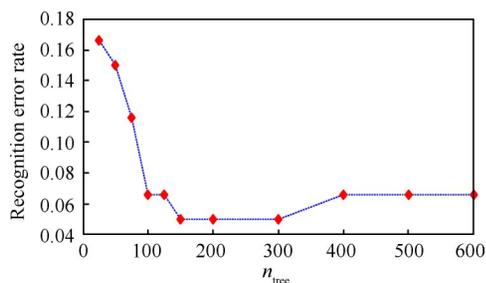


Fig. 14 Error rate under different n_{tree}

由图14可知,当 n_{tree} 取值为25~75时,错误识别率仍处于较高水平,随着决策树数量增加,故障辨识的错误率趋于稳定,在150~300时达到最低水平。随着决策树数量的增加,识别错误率反而略有增加,且训练时间将增加,因此在实际工程中,需要测试最佳的决策树颗数,可以设为150~200。

为衡量每个特征在模型中所有决策树中所做贡献,利用RF模型和GINI指数方法进行特征重要度排序,其原理是逐次对每个特征加噪,观察对结果正确率或基尼指数的影响,重要度详细计算过程可参考文献[25],文中不再赘述。经过标准化处理后,计算得出的每个特征重要性介于0~1且总和为1,7个所构建特征的重要度结果如图15所示。

从重要度分析结果可知,本文所构建的特征 R_{LFB}, R_{NN}, WEE, Kv 参数对训练模型的贡献度较高,达

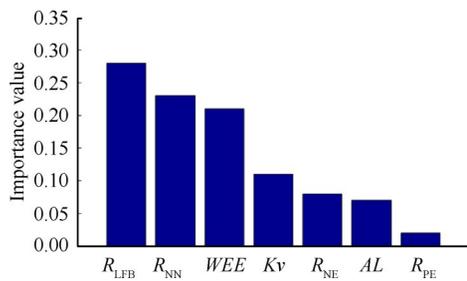


Fig. 15 Feature importance ranking

到0.2以上,而时域特征中的AL和 R_{NE} 对模型的贡献度偏低,从而说明本文所提出的新特征对故障识别过程具有更好的区分度。根据重要度分析结果,后续研究中可以继续开展对特征集构造筛选优化研究,进一步提高故障类型识别准确率。

4.5 算法对比

为对比VMD-RF算法识别效果,开展两种类型的对比实验测试,分别对比识别模型和信号处理方法的效果。

第一类:采用本文所提出的相同信号处理方法,同时采用4.3中相同数据集训练其他两类识别模型径向基神经网络模型(Radial Basis Function, RBF)和支持向量机(Support Vector Machine, SVM)模型,进行信号增强后的故障识别效果对比。其中,神经网络输入层节点数设为20,隐含层节点数为10,输出层节点数为2。训练中,学习率设为0.04,目标误差为 10^{-4} ;支持向量机模型的核函数选择径向基函数,惩罚系数设为50;将训练样本数据集输入创建的SVM和RBF神经网络进行训练,通过测试样本进行效果测试,最终对比结果如表3所示。从表3中结果可见,在识别准确率方面,随机森林模型和SVM模型比RBF神经网络模型的故障识别准确率更高,但是SVM模型的训练时间和测试时间都较长,运算量较大,验证随机森林方法在面向本文的故障识别任务时具有一定的优势。

Table 3 Comparison of different fault recognition algorithms

Algorithm	Accurate rate/%	Training time/s	Test time/s
VMD-RBF	86.6	0.543	0.167
VMD-SVM	93.3	1.812	0.106
VMD-RF	95.0	0.311	0.083

第二类:为对比本文提出信号处理方法,对实验过程中获取的信号样本分别做EMD, CEEMD, CEEMDAN分解,同时对分解后的IMF进行观察,人

工选择合适的IMF进行重构,后续进行同样步骤特征提取,分别训练随机森林故障识别模型,以对比VMD方法和筛选准则有效性,最终测试对比结果如表4所示。

Table 4 Comparison of different signal processing algorithms

Algorithm	Accurate rate/%	Signal processing time/s	Signal processing by VMD time/s
CEEMD-RF	91.6	5.783	0.782
CEEMDAN-RF	93.3	11.933	0.782
EMD-RF	86.6	1.648	0.782
VMD-RF	95.0	0.782	0.782

从对比结果看出,在单样本平均信号处理计算时长方面,VMD方法在处理单个信号样本时的平均耗时最优,仅花费0.78s左右,而采用递归分解计算方式的CEEMDAN, CEEMD, EMD方法在处理单个文本过程中评价耗时较长,其中CEEMDAN方法达到10s以上,在工程实践中VMD方法更具优势;在识别准确率方面,采用相同随机森林模型,经过CEEMD, CEEMDAN, VMD方法处理的样本提取相同特征后,测试得到识别准确率都较高,达到90%以上,而EMD方法处理后的样本识别准确率明显不如其他三种方法。综合两类对比实验结果和考虑计算时长及识别效果等因素,说明本文提出方法具有一定的优势,适用于后续工程实践的快速处理需求。

5 结论

本文提出了一种基于静电信号变分模态分解和随机森林的气路故障识别方法,进一步通过开展典型气路故障模拟实验,相关实验及分析结果表明:

(1)基于VMD方法能够快速有效地对静电信号进行分解和重构,使信号有用成分得到明显的增强,能够为准确提取故障特征奠定可靠基础。

(2)VMD-RF算法能够较好地完成故障类型识别任务,对测试集样本识别准确率能够达到90%以上;随着RF模型决策树数量增加,故障识别准确率逐渐趋于稳定,在实际应用中需要通过调试确定具体树模型的数量。

(3)在所提出的特征中,低频能量贡献比、负事件贡献率特征以及小波能谱熵特征在RF模型中的重要度值较高,归一化重要度达到0.2以上,

后续可以围绕特征重要度指标对特征集构造进一步优化。

(4)综合正确率和计算量考虑,本文提出的VMD-RF故障识别算法具有一定优势,能够为静电监测技术后续工程应用提供了有力的支撑。

由于尾气静电监测原理是通过异常带电颗粒感知发动机的健康状况,因此其主要面向的故障是能够产生带电颗粒的类型,例如异常燃烧等,对于一些常见的喘振和失速工况,可能会通过影响燃烧状态而使得尾气带电情况异常,但单一的静电监测方式仍无法总体判定,需要结合发动机振动和温度监测综合识别,具体的结合方法有待后续进一步研究。同时,静电监测作为一种感知故障直接产物的技术,其优点在于灵敏度高,能够通过监测信号挖掘更早期的预警信息,但目前不足在于仍需要找到持续优化的信息处理方式和体系化的信息挖掘方法,并与其他监测手段相互结合,从而实现发动机多源监测信息的融合。

致谢:感谢国家自然科学基金、航空科学基金、山东省矿山机械工程重点实验室开放基金的资助。感谢南京航空航天大学航空安全与保障研究所的支持。

参考文献

- [1] Powrie H E G, Fisher C E. Engine Health Monitoring: Towards Total Prognostics[C]. *Washington: IEEE Aerospace Conference Proceedings*, 1999: 1112-1123.
- [2] Powrie H E G, Fisher C E. Monitoring of Foreign Objects into the Intake of a Gas Turbine Aero-Engine[C]. *Washington: IEEE Aerospace Conference Proceedings*, 1999: 907-920.
- [3] 付宇,殷逸冰,冯正兴,等.融合静电信号和气路参数的发动机性能评估方法[J].*推进技术*, 2019, 40(2): 215-221. (FU Yu, YIN Yi-bing, FENG Zheng-xing, et al. A Method for Engine Performance Evaluation by Fusing Electrostatic Signal and Gas Path Parameters [J]. *Journal of Propulsion Technology*, 2019, 40(2): 215-221.)
- [4] Powrie H E G, Novis A. Gas Path Debris Monitoring for F-35 Joint Strike Fighter Propulsion System PHM[C]. *Big Sky: IEEE Aerospace Conference Proceedings*, 2006: 314-322.
- [5] Melissa W, David R. Engine Distress Detection in Gas Turbines with Electrostatic Sensors[C]. *Glasgow: Proceedings of ASME Turbo Expo 2010: Power for Land, Sea, and Air*, 2010: 39-51.
- [6] Wen Z H, Zuo H F, Pecht M G. Electrostatic Monitoring of Gas Path Debris for Aero-Engines[J]. *IEEE Transactions on Reliability*, 2011, 60(1): 33-40.
- [7] 文振华.基于静电感应的航空发动机气路监测技术研究[D].南京:南京航空航天大学,2009.
- [8] Wen Z H, Ma X J, Zuo H F. Characteristics Analysis and Experimental Verification of Electrostatic Sensor for Aero-Engine Exhaust Gas Monitoring[J]. *Measurement*, 2014(47): 633-644.
- [9] Liu P P, Zuo H F, Sun J Z. The Electrostatic Sensor Applied to the Online Monitoring Experiments of Combustor Carbon Deposition Fault in Aero-Engine[J]. *IEEE Sensors*, 2014(14): 686-694.
- [10] Tang X, Chen Z H, Li Y. Compressive Sensing-Based Electrostatic Sensor Array Signal Processing and Exhaust Abnormal Debris Detecting of Gas Turbines[J]. *Mechanical Systems and Signal Processing*, 2018(105): 404-426.
- [11] Yin Y B, Cai J, Zuo H F, et al. Experimental Investigation on Electrostatic Monitoring Technology for Civil Turbofan Engine[J]. *Journal of Vibro-Engineering*, 2017, 19(2): 967-987.
- [12] 付宇,殷逸冰,冒慧杰,等.基于静电监测和神经网络的航发典型故障诊断[J].*计算机集成制造系统*, 2019, 25(11): 11-19.
- [13] Wen Z H, Hou J X, Atkin J. A Review of Electrostatic Monitoring Technology: The State of the Art and Future Research Directions[J]. *Progress in Aerospace Sciences*, 2017, 10(94): 1-11.
- [14] 张营,左洪福,佟佩声,等.基于谱插值和奇异值差分谱的滚动轴承静电监测信号去噪方法[J].*航空动力学报*, 2014, 29(8): 1996-2002.
- [15] 付宇,殷逸冰,左洪福.一种基于稀疏分解的静电信号去噪方法[J].*航空动力学报*, 2018, 33(11): 10-19.
- [16] 钟志荣,左洪福,郭家琛,等.基于阵列式静电传感器的颗粒带电量估计方法[J].*仪器仪表学报*, 2020, 41(7): 80-90.
- [17] 赵昕海,张术臣,李志深,等.基于VMD的故障特征信号提取方法[J].*振动、测试与诊断*, 2018, 38(1): 11-19.
- [18] 李奕江,张金萍,李允公.基于VMD-HMM的滚动轴承磨损状态识别[J].*振动与冲击*, 2018, 37(21): 61-67.
- [19] 文振华,侯军兴,左洪福.航空发动机静电监测信号的特征分析及提取[J].*振动、测试与诊断*, 2015, 35

- (3): 6-13.
- [20] 赵光权, 姜泽东, 胡 聪, 等. 基于小波包能量熵和DBN的轴承故障诊断[J]. 电子测量与仪器学报, 2019, 33(2): 32-38.
- [21] 何宽芳, 黎 祺, 李学军, 等. 小波能谱熵和优化神经网络的滚动轴承诊断方法[J]. 机械科学与技术, 2013, 32(11): 1670-1674.
- [22] Ma S, Chen M, Wu J, et al. Intelligent Fault Diagnosis of HVCB with Feature Space Optimization-Based Random Forest[J]. *Sensors*, 2018, 18(4): 1221-1232.
- [23] Qin X, Li Q, Dong X, et al. The Fault Diagnosis of Rolling Bearing Based on Ensemble Empirical Mode Decomposition and Random Forest [J]. *Shock and Vibration*, 2017(1): 1-9.
- [24] Panigrahy P S, Santra D, Chattopadhyay P. Decent Fault Classification of VFD Fed Induction Motor Using Random Forest Algorithm [J]. *Artificial Intelligence for Engineering Design Analysis and Manufacturing*, 2020 (2): 1-13.
- [25] 丁学辉, 许海林, 罗颖婷, 等. 基于随机森林特征优选与MAEPSO-ELM算法的变压器DGA故障诊断[J]. 电力科学与技术学报, 2022, 37(2): 181-187.

(编辑:朱立影)