

基于决策树的涡轴发动机气路健康评估方法^{*}

倪波, 李秋红, 徐嘉伸, 刘鑫洋

(南京航空航天大学 能源与动力学院, 江苏 南京 210016)

摘要: 为解决涡轴发动机退化模式多、故障样本大、分类困难的问题, 将梯度提升决策树 (GBDT) 算法应用于涡轴发动机气路健康评估。首先对涡轴发动机故障标签进行了拆解, 针对每一类标签设计了一个分类器。通过特征重要度排序筛选了6个分类标签的输入特征, 有效降低模型复杂度。通过遍历确定树的最优数量和深度, 确保了分类的准确性。仿真结果表明, 特征筛选减少测试时间16%~35%; 相比广泛使用的支持向量机 (SVM) 算法, 在特征数量相同条件下, GBDT测试时间缩短了31.88%~65.28%, 相比极限学习机 (ELM), 误诊样本数量低于其千分之一。可见GBDT算法在涡轴发动机大样本故障评估中表现出更高的分类精度, 验证了其在发动机气路健康评估上的有效性。

关键词: 涡轴发动机; 健康评估; 梯度提升决策树; 特征筛选; 分类器

中图分类号: V231.1 文献标识码: A 文章编号: 1001-4055 (2022) 05-200608-10

DOI: 10.13675/j.cnki. tjjs. 200608

Gas Path Health Assessment Method of Turboshaft Engine Based on Decision Tree

NI Bo, LI Qiu-hong, XU Jia-shen, LIU Xin-yang

(College of Energy and Power, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In order to solve the problem of turboprop engine health assessment which is characterized by multi-degraded modes, tremendous fault samples and hard to classify, the gradient boosting decision tree (GBDT) algorithm was applied to the gas path health assessment of turboprop engine. Firstly, the fault labels of turboprop engine were disassembled, and a classifier was designed for each tag. The input features of six classification tags were screened out according to the feature importance ranking, which effectively reduced the complexity of the model. The traversal method was adopted to determine the optimal number and depth of the tree to ensure the accuracy of the classification. The simulation results show that feature selection reduces test time by 16% to 35%. Compared with the widely used support vector machine (SVM) algorithm, under the condition of the same number of features, the test time of GBDT is reduced by 31.88% to 65.28%, and compared with extreme learning machine (ELM), the number of misdiagnosed samples is less than 1/1000 that of ELM. It can be seen that the GBDT algorithm has higher classification accuracy in the health assessment of turboprop engine with a large amount of fault samples, which verifies its effectiveness in engine gas path health assessment.

^{*} 收稿日期: 2020-08-13; 修订日期: 2020-10-22。

基金项目: 国家科技重大专项 (2017-V-0004-0054)。

作者简介: 倪波, 硕士生, 研究领域为航空发动机建模与控制。

通讯作者: 李秋红, 博士, 副教授, 研究领域为航空发动机建模与控制。

引用格式: 倪波, 李秋红, 徐嘉伸, 等. 基于决策树的涡轴发动机气路健康评估方法[J]. 推进技术, 2022, 43(5): 200608. (NI Bo, LI Qiu-hong, XU Jia-shen, et al. Gas Path Health Assessment Method of Turboprop Engine Based on Decision Tree[J]. *Journal of Propulsion Technology*, 2022, 43(5):200608.)

Key words: Turboshift engine; Health assessment; Gradient boosting decision tree; Feature selection; Classifier

1 引言

涡轴发动机是由多个部件组成的大型复杂非线性系统,需要在高温、高压的条件下工作。随着使用寿命的增加,转动部件叶片容易发生侵蚀、积碳、打伤等问题,引起转动部件流量和效率的退化。当退化达一定程度,发动机工作状态就会远离设计点,使发动机性能受到影响,甚至引起超温和超压,威胁到发动机的安全性。因此,准确判断发动机的健康状态,有助于开展发动机健康评估与视情维修,避免资源的浪费和事故的发生,具有重要研究价值^[1-5]。

目前,国内外航空发动机的健康评估方法主要有三类,基于仿真模型^[4-6]、基于数据^[7-8]、基于知识规则^[9-10]的评估方法。基于模型的健康评估方法依赖于发动机部件级模型,评估精度受到模型精度的影响,且评估过程中需要调用部件级模型,实时性差。基于知识规则的诊断方法建立在已有知识规则的基础上,依赖专家经验,对专业知识要求比较高。

随着计算机技术以及智能算法的发展,基于数据驱动的健康评估方法得到广泛关注。神经网络(NN)^[11]、支持向量机(SVM)^[12]、极端学习机(ELM)^[13-14]等各种智能算法被用于发动机的健康评估。文献[12]研究了基于SVM的故障分类方法,通过采用不同的正则化参数提高了算法的诊断精度。文献[13-14]研究了基于ELM的故障分类算法,通过改进数据结构、重构标签值提高了算法的实时性。这些算法采取浅层网络结构,在取得成功的同时还存在一定缺陷,不能处理大量数据下的发动机故障诊断问题,因此算法均采用发动机部件流量和效率进行关联的方法进行诊断,针对包线内的少量数据进行了故障诊断方法验证,且只进行是否发生故障的分类,对故障程度未进行进一步分解。而发动机本身是强非线性系统,流量和效率之间的关联特性会随发动机工作状态而发生变化,且发动机个体之间也存在差异,因此不能确保这种关联的准确性。

对于涡轴发动机来讲,具有压气机、燃气涡轮、动力涡轮三个转动部件,在同一个工作点,按每个部件流量和效率均从0间隔0.5%到4%的退化情况进行计算,则会有 $9^6=531441$ 种组合,数据量非常大,需要能够进行大量数据处理的深度学习方法,才能很好的实现故障分类。对此本文将梯度提升决策树

(GBDT)方法用于涡轴发动机故障的分类中。

决策树是用于分类的主要技术之一,通过属性和类别间的关系实现预测,梯度提升则是利用集成学习的思想,能有效降低单个分类器的误差,提高分类的准确度,将决策树和梯度提升相结合的GBDT算法近年来在机器学习领域大放异彩,得到了广泛的关注^[15]。文献[16]使用GBDT多分类算法进行云平台软件健康评估,与常用于分类的随机森林、决策树、AdaBoost算法相比,预测精度有明显提高。文献[17]将GBDT算法应用于电动汽车行驶里程预测,效果同样优于多元线性回归以及分类和回归树算法(CART)。文献[18]基于GBDT算法和特征筛选对铁路事故类型进行了预测,证明了该算法在分类任务上相比随机森林、决策树等算法优越性。这些研究结果都表明,GBDT算法在分类问题上具有良好的应用前景。

本文将GBDT算法应用于大数据样本下的涡轴发动机健康评估,通过特征筛选确定评估网络的输入,在避免过拟合的同时提高了分类精度。同时采用多标签和多标签值的多元分类方法,相比只判断是否发生故障的二分类问题,能更加准确评估发动机健康状态。

2 气路故障分析及算法介绍

2.1 涡轴发动机气路故障分析

航空发动机的故障大部分由气路部件引起。对于涡轴发动机来说,压气机、涡轮等部件由于承受大的应力和载荷,是最有可能发生故障的部件。引起故障的原因分为两类:一是发动机长期使用过程中受到的腐蚀、积垢、磨损引起的性能缓慢退化;二是由于突发性外物吸入损伤等原因引起的性能突变。故障引起发动机转动部件特性的变化,如流量和效率的变化,进而引起发动机工作状态发生改变,表现在相关参数(转速、压力、温度)的传感器测量参数偏离额定状态。而不同程度的退化和损伤引起的偏离程度不同,因此,可以通过传感器测量参数的变化,对发动机故障程度进行评估。

目前的故障评估中对故障模式的分类还比较单一。文献[7-8,12-14]仅针对涡扇发动机4个转动部件效率各退化1%的情况开展故障诊断研究。文献[19]针对涡轴发动机气路故障,考虑了单个转动部

件流量、效率分别退化和同时退化情况,没有考虑多个部件同时退化,且退化量均为单一值,共考虑了 10 种可能的情况。而实际发动机不太可能只一个部件发生性能退化,且退化程度也不可能是单一情况。

对于涡轴发动机来说,其主要组成部件如图 1 所示,包括进气道、压气机,燃气涡轮和动力涡轮。其 3 个转动部件的流量和效率分别退化就有 6 类退化情况,如果每一类分为无退化、轻度退化和中度退化,则共有 $3^6=729$ 种组合情况。

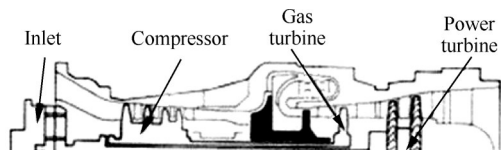


Fig. 1 Structural diagram of turboshaft engine

若采用一个诊断模型来实现 729 种故障的分类,无疑会使诊断问题变得非常复杂。而对于多标签分类问题,可以拆解标签,即对每一个标签构造一个分类器,以降低分类的复杂度。为此本文对故障标签进行拆分,针对 6 类故障设置 6 个标签,构造 6 个分类器,对 3 种不同程度的退化,按表 1 进行分类。表 1 中的 D 代表退化,下标 com,gt,pt 分别代表压气机、燃气涡轮和动力涡轮,下标 wa,et 分别代表流量和效率。当部件流量和效率偏离设计点小于 1% 时,考虑到部件制造安装公差也会引起的偏离,且偏离程度较低,归于无故障类别 I。当偏离在 1%~2.5% 时,发动机性能发生变化,归类于轻度故障类别 II,在 2.5%~4% 则认为故障影响较大,归类于中度故障类别 III。鉴于目前分类方式的样本数量已经很大($9^6=531441$ 组),本文没有对更大程度的故障进行采样。

发动机性能退化会对发动机输出参数产生影响,以涡轮部件为例,其静子叶片的弯曲和扭曲程度会随使用慢慢减小,造成高压涡轮流通能力增加,效率降低,流动过程中阻力减小,高压涡轮出口压力降低,高压涡轮做功减少,出口温度升高、转速降低,因此可以通过高压涡轮出口压力和温度的变化对标准高压涡轮流通能力的流量特性进行评估。同时发动

机部件间的共同工作,使得涡轮部件的退化同样会影响压气机部件的出口参数,进而对其流路下游参数产生影响。为此本文选择燃油流量 W_{fb} ,压气机换算转速 $N_{c,cor}$,压气机出口总温 T_3 ,总压 p_3 ,燃气涡轮进口总温 T_{42} ,总压 p_{42} ,燃气涡轮出口总温 T_{43} ,总压 p_{43} ,动力涡轮进口总温 T_{44} ,总压 p_{44} ,动力涡轮出口总温 T_{45} ,总压 p_{45} 共 12 个变量作为故障诊断的特征参数。不同参数对不同故障的敏感程度不同,因此本文针对不同类型的故障,通过特征筛选建立故障诊断模型。

鉴于本文研究中考考虑的故障模式多,用到大量的故障样本,因此采用具有大数据处理能力的 GBDT 算法来实现健康评估。

2.2 GBDT 算法

梯度提升决策树是一种基于 CART 回归树和提升方法 (Boosting) 的机器学习方法,提升方法可以将弱分类器提升为拟合能力较强的强学习器,实现了对弱分类器的集成学习,提高模型精度,其模型可以表示为

$$F(x) = \sum_{m=1}^M T(x; \theta_m) \tag{1}$$

式中 $T(x; \theta_m)$ 表示以 θ_m 为参数的树模型, M 表示树的个数。

决策树 $T(x; \theta_m)$ 的损失函数用 $L(\cdot)$ 表示,GBDT 算法对于多分类问题(假设有 K 个类别)常用的损失函数为^[14]

$$L\left(\left\{y_k; F_k(x)\right\}_1^K\right) = -\sum_{k=1}^K y_k \ln p_k(x) \tag{2}$$

式中 $y_k = \{0, 1\}$ 表示是否属于第 k 类, $p_k(x)$ 为样本 x 属于第 k 个类别的概率。

$$p_k(x) = \frac{e^{F_k(x)}}{\sum_{l=1}^K e^{F_l(x)}} \tag{3}$$

式中 $p_k(x)$ 为第 k 个类别的 Softmax 函数。

GBDT 通过最小化损失函数来确定下一棵决策树参数的最优 $\hat{\theta}_m$

Tabel 1 Failure modes for different health parameters (%)

Mode	$D_{com, wa}$	$D_{com, et}$	$D_{gt, wa}$	$D_{gt, et}$	$D_{pt, wa}$	$D_{pt, et}$
I	0 ~ 1	0 ~ 1	0 ~ 1	0 ~ 1	0 ~ 1	0 ~ 1
II	1 ~ 2.5	1 ~ 2.5	1 ~ 2.5	1 ~ 2.5	1 ~ 2.5	1 ~ 2.5
III	2.5 ~ 4	2.5 ~ 4	2.5 ~ 4	2.5 ~ 4	2.5 ~ 4	2.5 ~ 4

$$\hat{\theta}_m = \arg \min_{\theta} \sum_{i=1}^n L(y_i, T_{m-1}(x_i) + T(x; \theta_m)) \quad (4)$$

在GBDT算法中,特征筛选占有重要的地位。通过特征筛选,可以从诸多特征形成的集合中,选出适合所研究问题的子集,从而降低模型规模,提高分类问题的精度和实时性。

文献[15]指出特征 j 的全局重要度 $I_{G,j}$ 通过特征 j 在单棵树 m 中的重要度 $I_{m,j}$ 的平均值来衡量,其计算方式为^[15]

$$I_{G,j} = \frac{1}{M} \sum_{m=1}^M I_{m,j}(T(x; \theta_m)) \quad (5)$$

$$I_{m,j}(T(x; \theta_m)) = \sum_{t=1}^{L-1} i_t^2 I(v_t = j) \quad (6)$$

式中 L 为树的叶子节点数量, v_t 是和节点 t 相关联的特征, i_t^2 是节点 t 分裂后平方损失的减少值。

特征筛选的流程如图2所示。在图2所示的特征筛选流程中,依次构建 M 棵树,每构建一棵树均对输入特征在单棵树中的重要度进行评估,完成所有树的构建后,再计算特征的全局重要度,依据全局重要度对特征进行排序和筛选。

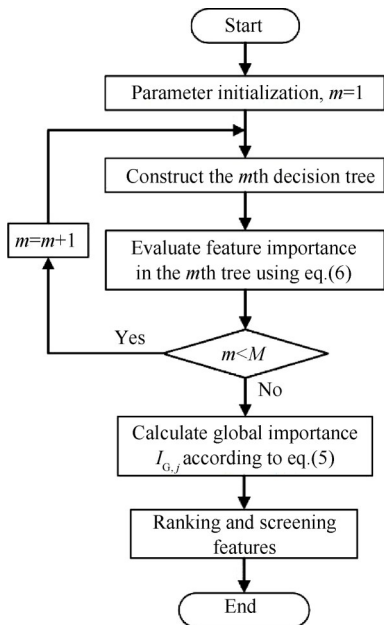


Fig. 2 Process of feature screening

非叶子节点 t 在分裂时的 i_t^2 越大,说明特征越重要。重要度排序靠前的特征是构建分类模型的主要成分,也是树模型分裂的主要依据。因此可以在模型训练中保留重要特征,而舍弃影响小的特征,在保证模型精度的同时,提高模型的实时性。

3 涡轴发动机健康评估

3.1 健康评估系统构建

将GBDT算法应用于涡轴发动机的健康评估中,首先基于历史数据训练GBDT分类器,获得诊断模型,经测试满足要求后,可用于对发动机当前健康状态的评估,评估系统构建过程如图3所示。第一步基于专家知识确定所研究分类问题的数据集,并将数据的标签进行拆解,对应表1中的6类退化,将多标签分类问题拆成若干个单标签分类问题,并对每一个标签进行划分,依据专家经验将每一类标签划分为三类,标签值分别记为(1,0,0),(0,1,0),(0,0,1),对应表1中的无故障、轻度故障和中度故障;第二步是进行特征重要度排序并进行特征筛选;第三步就是确定树的数目和深度以及采样率,基于GBDT算法在筛选的特征上建立分类树模型。

GBDT需要基于历史数据进行模型训练,虽然发动机在实际使用过程中积累了海量的飞行数据,但是由于保密等原因,大量真实的航空发动机故障数据是很难获取的。故本文参考文献[20-21],以高置信度的涡轴发动机/直升机一体化仿真平台为基础,生成发动机性能退化历史数据。选择前飞速度 $v_x=10\text{m/s}$,高度 $H=500\text{m}$ 处进行历史数据采集,通过循环模拟压气机、燃气涡轮和动力涡轮的不同程度的流量以及效率退化,如表1所示的每类退化,均采集了从0间隔0.5%到4%的9种情况,6类退化,共采集了 $9^6=531441$ 组不同组合的样本。相比其他文献,样本更全面,具有多退化程度、退化类型组合全面的特征,包含了所研究退化程度内的所有可能退化组合情况,更符合发动机实际工作情况中的各种不确定性。同时由于涡轴发动机本身是一个强非线性系统,使得不同退化程度引起的发动机输出变化也呈现非线性,但是特征参数对同一退化是单调的,而GBDT算法基于特征空间划分的思想,基于特征参数值的大小进行子空间划分,可以高效处理这类单调、非线性分类问题。

选取372008组作为训练样本,159433组作为测试样本,进行健康评估系统设计。首先进行特征筛选。第2节给出的12个待选特征参数涵盖了涡轴发动机三大转动部件的出口参数和热端部件的进口参数、燃油和压气机转速。由于涡轴发动机控制目标为动力涡轮转速恒定,因此动力涡轮转速不作为待选特征向量。如果将全部12个变量均作为GBDT的输入,会由于特征向量维数较高而增加模型的复杂

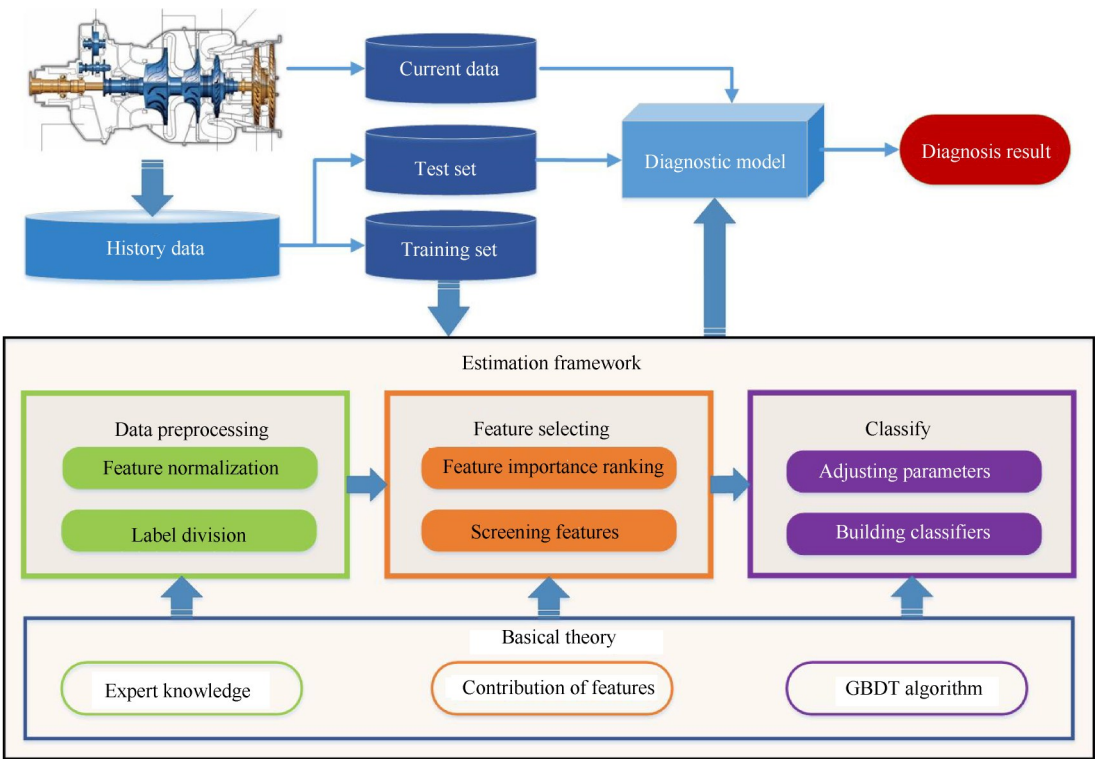


Fig. 3 Construction process of turbine shaft engine health assessment system

度和训练耗时,甚至引起过拟合,因此本文基于特征重要性分析进行特征筛选,从而降低模型复杂度,减少训练和预测时长。

为了提高分类精度,本文对采集的数据进行了相对增量式的处理,以设计点无退化情况下的发动机输出为基准点,记为 X_{base} ,计算其他状态相对基准状态的偏离,基于偏离量进行特征筛选和分类器的训练。相对偏离量为

$$\Delta X = \frac{X - X_{base}}{X_{base}} \quad (7)$$

3.2 特征筛选

根据图 2 所示的特征筛选流程,进行特征重要度计算,得到图 4(a)~(f)分别为压气机流量、效率,燃气涡轮流量、效率以及动力涡轮流量、效率退化作为标签值时,发动机状态参数的全局重要度排序。

如图 4(a)所示,压气机流量发生退化,会影响压气机转速以及压气机出口压力,反之,这些状态量是反映压气机流量退化的重要特征量,这也符合发动机气动热力学原理,其他部件退化同理,这里不做赘述。

根据图 4 特征重要度排序,筛选每个标签下,相对特征重要度在 0.005 以上的特征,得到各分类器所选取的特征参数如表 2 所示。

Table 2 Classifier labels and corresponding feature inputs

Label	Feature input
$D_{com, wa}$	$N_{C, cor}, p_3, T_{42}, T_{43}, T_{44}$
$D_{com, et}$	T_3, p_3, T_{44}, T_{43}
$D_{gl, wa}$	$p_3, p_{42}, p_{43}, p_{44}, W_{Fb}$
$D_{gl, et}$	$T_3, W_{Fb}, p_{42}, T_{42}, p_{43}, p_{44}$
$D_{pt, wa}$	$T_{45}, p_{43}, T_{43}, T_{44}, p_{44}, p_{42}, T_{42}, W_{Fb}$
$D_{pt, et}$	$T_{45}, p_{43}, T_{43}, T_{44}, p_{44}, p_{42}, T_{42}, p_{45}$

3.3 建模过程

GBDT算法是基于回归树的集成算法,树的数目(即迭代次数 Iteration)以及每棵树的最大深度(Max depth)对分类结果的精度影响很大,同时采样率也对分类结果有较大影响,较小的采样率会影响模型的泛化能力,较大的采样率可能引起过拟合。为此在涡轴发动机健康评估的多分类问题中,本文通过如下步骤建立 GBDT 评估模型:

- (1) 将树的深度固定为 5,采样率固定为 1。
- (2) 改变迭代次数,得到不同的树模型。
- (3) 以测试误诊最少的为最佳迭代次数后,再改变树的深度建立分类模型。
- (4) 比较不同深度下模型在测试集上的误诊样本数量,得到最佳深度。

(5) 改变采样率,比较不同采样率下误诊样本数量,得到最佳采样率。

(6) 以最佳迭代次数、最佳深度和最佳采样率建立分类模型。

建模过程中,该算法在测试集上误诊样本个数随迭代次数和树的深度的变化如图 5 所示,表 3 给出了不同采样率下构建的分类器模型在测试集上的误诊样本数。

如图 5 所示,在测试集上,误诊样本数量随迭代次数和树的深度增加总体呈下降趋势,达到最小误诊数量后,又会随迭代次数和树的深度波动,甚至增加。这是因为随着迭代次数和深度的增加,模型越趋复杂,训练难度增大,甚至过拟合,引起测试集上精度的下降。表 3 说明较低的采样率不能充分反映发动机的故障特征,引起测试样本较高的误

Table 3 Number of misdiagnosed samples with different sample ratio

Label	Sample ratio				
	0.6	0.7	0.8	0.9	1.0
$D_{com,wa}$	27	30	21	14	12
$D_{com,et}$	22	19	16	9	11
$D_{gt,wa}$	11	4	5	2	4
$D_{gt,et}$	12	5	2	1	1
$D_{pt,wa}$	61	65	50	46	47
$D_{pt,et}$	55	59	47	43	41

诊,采样率高也不一定能提高测试精度,反而可能由于训练过程中的过拟合,引起测试样本误诊率增大。根据图 5 和表 3 的结果,针对不同退化标签,选择的最优迭代次数、树的深度以及采样率如表 4 所示。

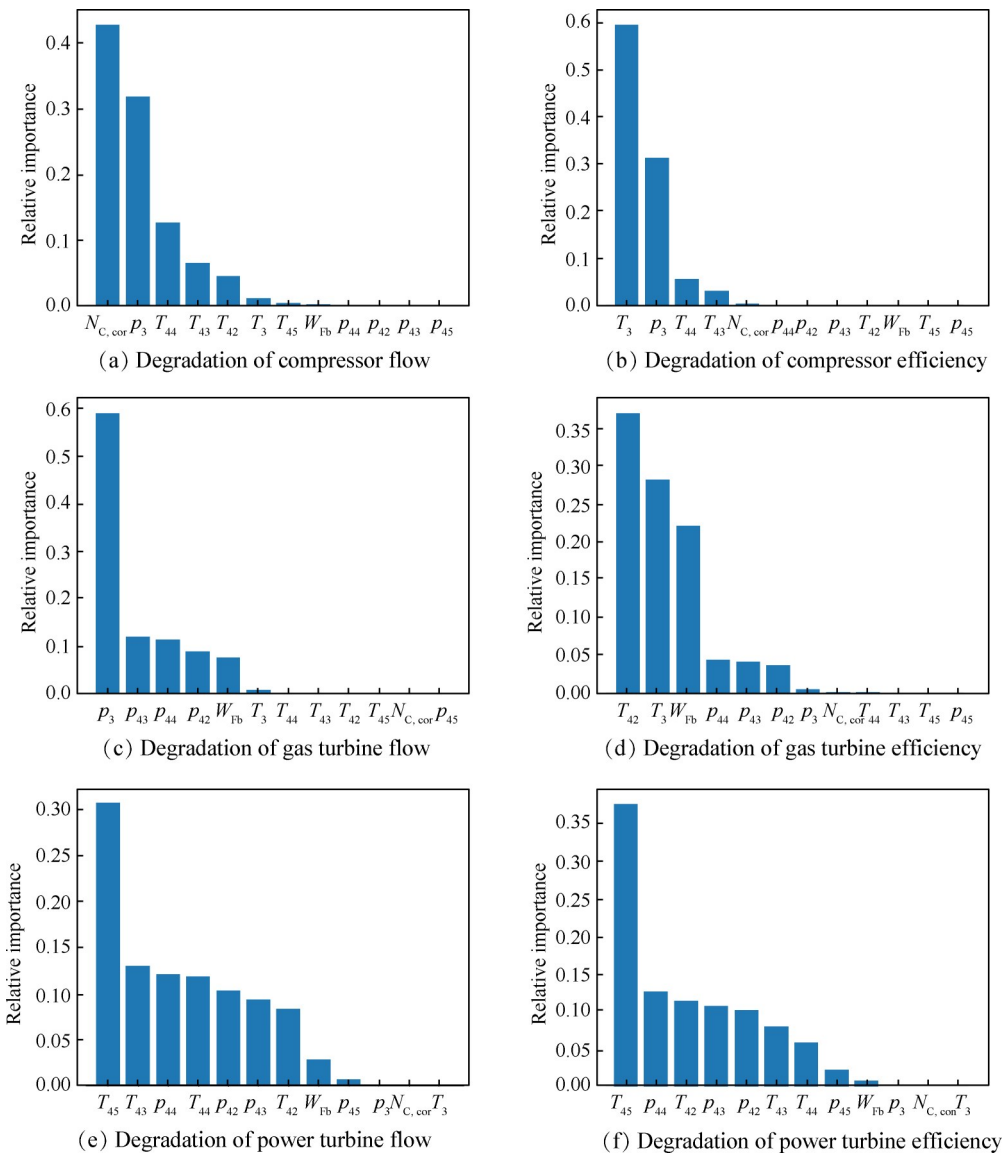


Fig. 4 Ranking of feature importance in each degradation case

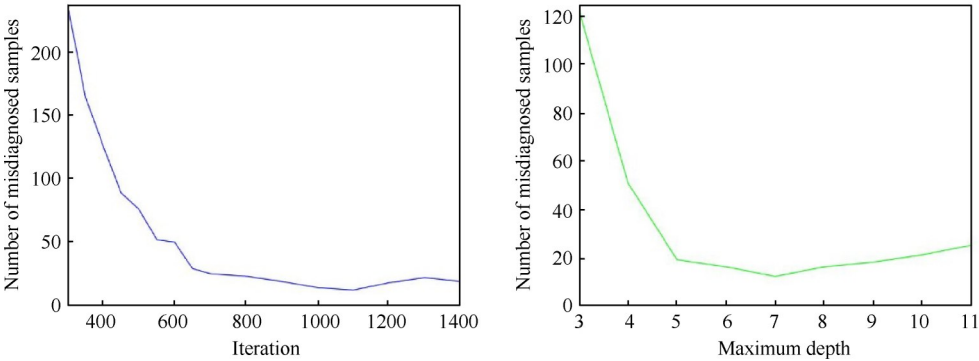
为了对比不同模型参数获得评估系统的精度,以 $D_{gl,et}$ 分类器为例,选择不同模型进行对比,具体参数组合及在测试集上的误诊样本个数如表 5 所示。从表中可以看出,参数组合 I 为本文遍历后得到的最优参数组合,其在测试集上的误诊样本数量最少,其它参数设置组合的误诊样本均多于本文所选参数。值得指出的是,由于 iteration 参数范围较大,而 GBDT 训练过程耗时较长,本文仅进行了大间隔的遍历,未在所得结果附近,进一步进行 iteration 和采样率的细致遍历。

3.4 分类结果测试与分析

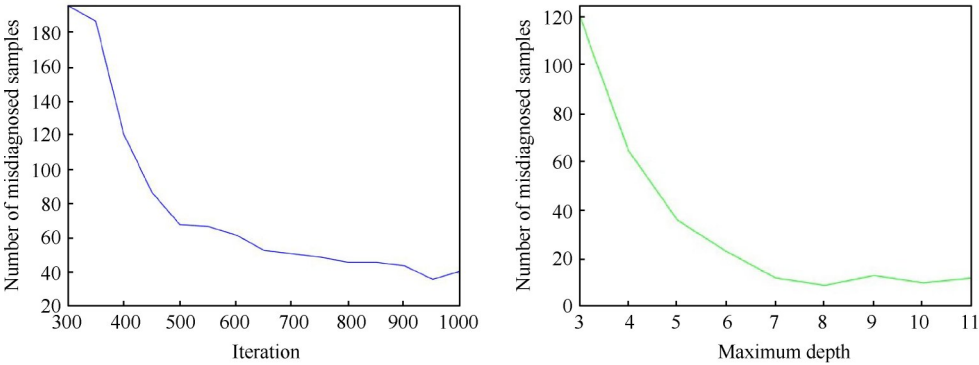
基于最优参数进行 GBDT 多分类模型训练,得益于 GBDT 算法的大数据分类能力,在训练集上达到

100% 的分类精度。用 159433 组测试数据进行测试,总测试耗时和误诊样本数量见表 6。为了便于对比,表 6 中,同时给出了常用于分类的 SVM(支持向量机)以及 ELM(极限学习机)的分类测试结果,SVM 和 ELM 的参数设置见表 7。表 7 中,Rbf 代表高斯核函数,核函数宽度是其特有参数,Linear 代表线性核函数,Logistics 表示逻辑回归函数。表 6 中,特征数量较少的模型为经过特征筛选之后的训练模型,未经筛选的则全部为 12 个特征。

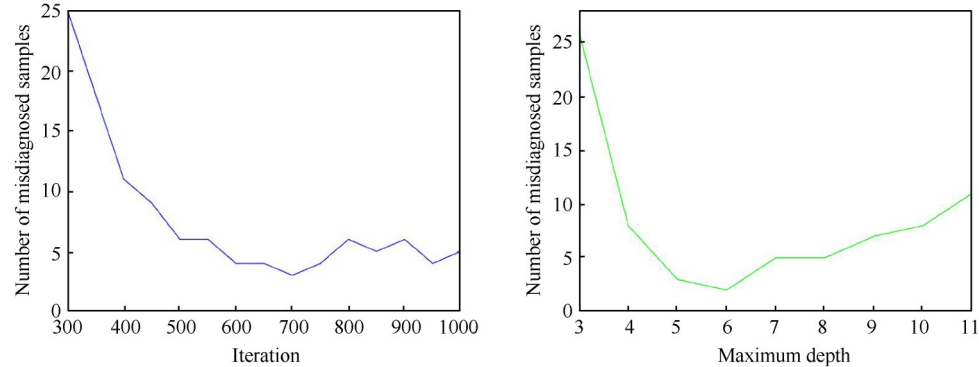
由表 6 可以看出,特征筛选在降低模型复杂度的同时显著缩短了测试时间,缩短测试样本整体评估时间 16%~35%,特征筛选后,模型的精度并没有下降,相反,大部分情况下,因为冗余信息被剔除,在测



(a) Degradation of compressor flow



(b) Degradation of compressor efficiency



(c) Degradation of gas turbine flow

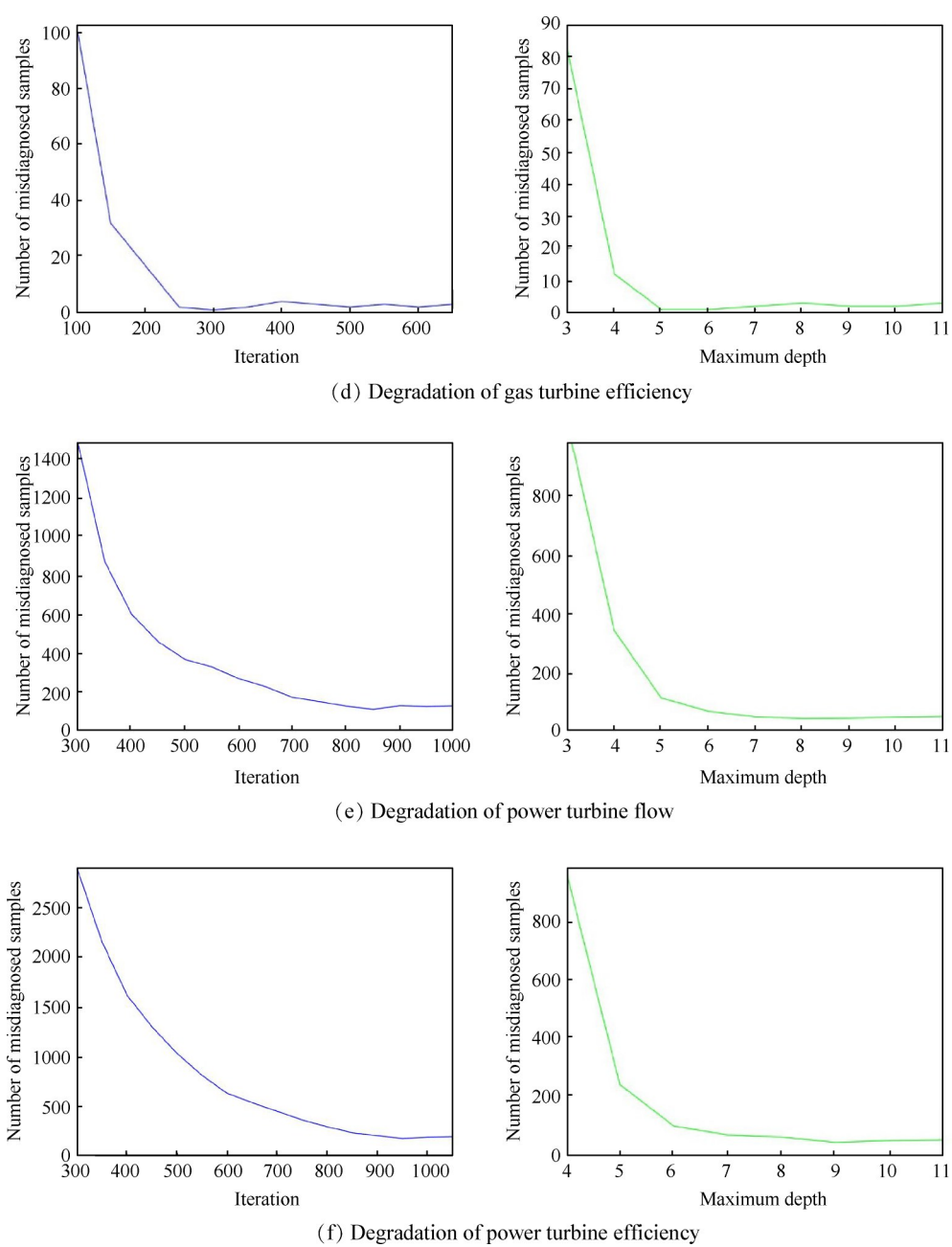


Fig. 5 Variation of number of misdiagnosed samples with iteration times and depth of the tree

Table 4 Optimal GBDT parameters for health estimation			
Label	Iteration	Depth of tree	Sample ratio
$D_{com, wa}$	1000	7	1.0
$D_{com, et}$	950	8	0.9
$D_{gt, wa}$	700	6	0.9
$D_{gt, et}$	300	5	0.9
$D_{pt, wa}$	850	8	0.9
$D_{pt, et}$	950	9	1.0

Table 5 Comparison of diagnostic results of different parameter combinations for $D_{gt, et}$				
Parameter combination	Iteration	Depth of tree	Sample ratio	Number of misdiagnosed samples
I	300	5	0.9	1
II	250	4	0.9	69
III	350	4	0.8	11
IV	400	6	1.0	3

试集上的分类精度反而有微小程度提高,误诊样本数量降低。与SVM算法相比,无论特征筛选前后,GBDT算法的测试时间均明显缩短,在特征数量相同

条件下,测试时间缩短了31.88%~65.28%,诊断系统的实时性得到了大大提高。在诊断精度上,SVM误诊样本比GBDT算法多出几十甚至上百倍,GBDT算

Table 6 Comparison of multi-classification diagnosis results

Label	Algorithm	Number of feature	Test time/s	Number of misdiagnosed sample	Accuracy/%
$D_{com,wa}$	GBDT	12	27.58	30	99.98
		5	23.14	12	99.99
	SVM	5	33.97	9652	93.95
	ELM	5	6.79	12827	91.95
$D_{com,et}$	GBDT	12	14.57	6	99.99
		4	9.26	9	99.99
	SVM	4	17.35	6346	96.02
	ELM	4	6.24	16998	89.34
$D_{gt,wa}$	GBDT	12	12.43	1	99.99
		5	9.92	2	99.99
	SVM	5	28.57	504	99.68
	ELM	5	6.45	7560	95.26
$D_{gt,et}$	GBDT	12	12.88	1	99.99
		6	8.95	1	99.99
	SVM	6	17.84	222	99.86
	ELM	6	6.22	7499	95.30
$D_{pt,wa}$	GBDT	12	27.49	83	99.95
		8	17.9	46	99.97
	SVM	8	44.90	7463	95.32
	ELM	8	6.41	16120	89.89
$D_{pt,et}$	GBDT	12	24.77	45	99.97
		8	19.07	41	99.97
	SVM	8	37.58	1216	99.24
	ELM	8	6.76	20236	87.31

Table 7 Main parameters of SVM and ELM classification

Label	SVM				ELM	
	Kernel function	Regularization parameter	Width of kernel function	Number of iterantions	Hidden layer node number	Activation function
$D_{com,wa}$	Rbf	0.6	0.6	1200	50	Logistics
$D_{com,et}$	Rbf	1.6	0.2	800	30	Logistics
$D_{gt,wa}$	Rbf	1	0.6	1000	20	Logistics
$D_{gt,et}$	Rbf	11	0.7	1300	20	Logistics
$D_{pt,wa}$	Linear	0.8	—	2000	40	Logistics
$D_{pt,et}$	Linear	1	—	1800	50	Logistics

法最低分类精度达到 99.97%,具有更好的分类效果。而与 ELM 相比,虽然测试时间较长,但是精度要高出很多。因此 GBDT 算法在涡轴发动机大数据样本下的健康评估中取得了良好的效果。

4 结 论

本文以涡轴发动机为对象,利用数值仿真模型模拟发动机性能退化,使用梯度提升回归树算法进行了大数据样本下的涡轴发动机健康评估,得到的结论如下:

- (1) 涡轴发动机退化模式多、故障样本大,适合采用标签分解的方式,对不同的标签分别设计分类器。
- (2) 通过特征重要度排序,提取不同分类标签下涡轴发动机的故障特征,降低了评估模型复杂度并缩短测试样本整体评估时间 16% ~ 35%。
- (3) 所建立的 GBDT 分类器模型与 SVM,ELM 算法的相比,误诊样本个数是另两种算法的百分之一甚至千分之一,表明 GBDT 算法在涡轴发动机大样本健康评估上具有较高的精度。

致谢:感谢国家科技重大专项和智能发动机控制项目的资助。

参考文献

- [1] 刘艳春, 杨德辉, 刘艳丽, 等. 基于神经网络的某型飞机发动机健康评估研究[J]. 电子设计工程, 2012, 20(11): 89-92.
- [2] 王 施, 王荣桥, 陈志英, 等. 航空发动机健康管理综述[J]. 燃气涡轮试验与研究, 2009, 22(1): 51-58.
- [3] 李业波, 李秋红, 黄向华, 等. 航空发动机气路部件故障融合诊断方法研究[J]. 航空学报, 2014, 35(6): 1612-1622.
- [4] Feng Lu, Jin-Quan Huang. Gas Path Health Monitoring for a Turbofan Engine Based on a Nonlinear Filtering Approach[J]. *Energies*, 2013, 6(1): 492-513.
- [5] Lu Feng, Jiang Chun-Yu, Huang Jin-Quan, et al. Aero Engine Gas Path Performance Tracking Based on Multi-Sensor Asynchronous Integration Filtering Approach[J]. *IEEE Access*, 2018, 55(5): 28305-28317.
- [6] 潘 阳, 李秋红, 王 元. 基于Broyden算法的航空发动机气路故障诊断[J]. 推进技术, 2017, 38(1): 191-198. (PAN Yang, LI Qiu-hong, WANG Yuan. Aero-Engine Gas Path Fault Diagnosis Based on Broyden Algorithm[J]. *Journal of Propulsion Technology*, 2017, 38(1): 191-198.)
- [7] Zhao Yong-Ping, Chen Yao-Bin, Hao Zhao, et al. Imbalanced Kernel Extreme Learning Machines for Fault Detection of Aircraft Engine[J]. *Journal of Dynamic Systems, Measurement, and Control*, 2020, 142(10): 1-17.
- [8] Zhao Yong-Ping, Wang Jian-Jun, Li Xiao-Ya, et al. Extended Least Squares Support Vector Machine with Applications to Fault Diagnosis of Aircraft Engine[J]. *ISA Transactions*, 2020, 97(8): 189-201.
- [9] 段隽喆, 李华聪. 基于故障树的故障诊断专家系统研究[J]. 科学技术与工程, 2009, 9(7): 1914-1917.
- [10] 黄嵩勋. 某发动机故障分析及专家系统知识库设计[D]. 大连: 大连理工大学, 2016.
- [11] 陈 恬, 孙健国, 杨蔚华, 等. 自组织神经网络航空发动机气路故障诊断[J]. 航空学报, 2003, 24(1): 46-48.
- [12] Xi Peng-Peng, Zhao Yong-Ping, Wang Pei-Xiao, et al. Least Squares Support Vector Machine for Class Imbalance Learning and Their Applications to Fault Detection of Aircraft Engine[J]. *Aerospace Science and Technology*, 2019, 84(8): 56-74.
- [13] Zhao Yong-Ping, Song Fang-Quan, Pan Ying-Ting, et al. Retargeting Extreme Learning Machines for Classification and Their Applications to Fault Diagnosis of Aircraft Engine[J]. *Aerospace Science and Technology*, 2017, 71(10): 603-618.
- [14] Zhao Yong-Ping, Huang Gong, Hu Qian-Kun, et al. Soft Extreme Learning Machine for Fault Detection of Aircraft Engine[J]. *Aerospace Science and Technology*, 2019, 91(5): 70-81.
- [15] Friedman J H. Greedy Function Approximation: A Gradient Boosting Machine[J]. *The Annals of Statistics*, 2001, 29(5): 1189-1232.
- [16] Zhu K, Ying S, Zhang N, et al. A Performance Fault Diagnosis Method for SaaS Software Based on GBDT Algorithm[J]. *Computers, Materials and Continua*, 61(3): 1161-1185.
- [17] Sun S, Zhang J, Bi J, et al. A Machine Learning Method for Predicting Driving Range of Battery Electric Vehicles[J]. *Journal of Advanced Transportation*, 2019(1): 1-14.
- [18] 钟敏慧, 张婉露, 李有儒, 等. 基于GBDT的铁路事故类型预测及成因分析[J/OL]. 自动化学报, 2020, DOI:doi.org/10.16383/j.aas.c190630.
- [19] 王召广, 杨宇飞, 黄 兴, 等. 基于ReliefF-LMBP算法的涡轴发动机气路故障模式识别[J]. 推进技术, 2021, 42(1): 220-229. (WANG Zhao-guang, YANG Yu-fei, HUANG Xing, et al. Gas Path Fault Mode Identification of Turboshift Engine Based on ReliefF-LMBP Algorithm[J]. *Journal of Propulsion Technology*, 2021, 42(1): 220-229.)
- [20] 罗启军. 涡喷发动机气路健康评估技术研究[D]. 南京: 南京航空航天大学, 2012.
- [21] 徐田镇. 航空发动机退化状态参数估计与控制技术研究[D]. 南京: 南京航空航天大学, 2017.

(编辑:梅 瑛)