

基于特征优化和支持向量机的航空发动机 气路故障诊断

吴超¹, 陈磊², 刘渊¹, 周绮凤², 王奕首²

(1. 中国航发湖南动力机械研究所, 湖南株洲 412002; 2. 厦门大学 航空航天学院, 福建厦门 361005)

摘要: 针对现有数据驱动的航空发动机故障诊断算法易受飞行监控数据中冗余特征及噪声的干扰, 不能及时修正监测数据中不平衡样本分布对模型泛化性能影响等问题, 通过在支持向量机模型中引入特征增维和采用提取算法, 提出基于特征优化和支持向量机的航空发动机气路故障诊断方法, 并建立相应模型。将涡桨发动机及 CFM56-7B 发动机航后数据输入模型, 分析与预测实际故障发生时刻, 并将预测结果与真实结果进行比较, 同时将其结果与采用随机森林等 4 种故障诊断方法所得结果进行对比验证。结果表明: 特征优化算法的应用能显著缩短各类故障诊断方法运行时间 20% 以上; 基于特征优化和支持向量机的故障诊断方法使预测准确率达 99.8%; 针对非平衡实测数据, 特征优化算法和回归预测思想的引入能显著提高算法在不平衡数据集上的性能, 与非回归算法相比故障检测率提高到 91.67%。

关键词: 故障诊断; 特征优化; 支持向量机; 主成分分析; 深度自编码器; 航空发动机

中图分类号: V231.1

文献标识码: A

doi: 10.13477/j.cnki.aeroengine.2024.04.004

Aeroengine Gas-Path Fault Diagnosis Based on Feature Optimization and Support Vector Machine

WU Chao¹, CHEN Lei², LIU Yuan¹, ZHOU Qi-feng², WANG Yi-shou²

(1. AECC Hunan Aviation Powerplant Research Institute, Zhuzhou Hunan 412002, China;

2. School of Aerospace Engineering, Xiamen University, Xiamen Fujian 361005, China)

Abstract: Aiming at the problems that existing data-driven aeroengine fault diagnosis algorithms are susceptible to the disruptive effects of redundant features and noise in flight monitoring data, and unable to timely address the impact of imbalanced sample distribution in monitoring data on the model's generalization performance, by introducing feature augmentation and using extraction algorithms in support vector machine models, an aeroengine gas-path fault diagnosis method was proposed based on feature optimization and the support vector machine, and the corresponding model was established. A simulation dataset from a turboprop engine and a flight dataset from a CFM56-7B engine were input into the model to analyze and predict the fault occurrence time. The predicted results were compared with the actual results, and the latter compared with those obtained by four fault diagnosis methods such as random forest. The results show that the application of the feature optimization algorithm can significantly shorten the computational time of various fault diagnosing methods by more than 20%; the fault diagnosis method based on feature optimization and support vector machine achieves a prediction accuracy of 99.8%; for unbalanced measured data, the introduction of feature optimization algorithm and regression prediction can significantly improve the performance of the algorithm on imbalanced datasets, and the fault detection rate is improved to 91.67% compared with non-regression algorithms.

Key words: fault diagnosis; feature optimization; support vector machine; principal component analysis; deep autoencoder; aeroengine

0 引言

航空发动机是为飞机飞行提供推进动力的一种

复杂的热力机械, 在工作过程中, 很多部件长期处于高转速、高应力、高温、高压状态, 使得发动机发生故障的时间具有不确定性。作为飞行器的重要组成部分

收稿日期: 2022-01-13 基金项目: 国家级基础加强项目(2019-JCJQ-ZD-339-00)资助

作者简介: 吴超(1983), 男, 硕士, 高级工程师。

引用格式: 吴超, 陈磊, 刘渊, 等. 基于特征优化和支持向量机的航空发动机气路故障诊断[J]. 航空发动机, 2024, 50(4): 30-37. WU Chao, CHEN Lei, LIU Yuan, et al. Aeroengine gas-path fault diagnosis based on feature optimization and support vector machine[J]. Aeroengine, 2024, 50(4): 30-37.

分,发动机的运行状态直接影响其安全服役。因此,及时且准确地进行发动机的故障诊断与管理不仅能实现视情维护和减少维修成本,而且能提高飞行安全性。总之,通过研究和发展准确而可靠的发动机故障诊断技术,能有效地增强飞行器运行过程的安全性及经济性^[1]。

20世纪80年代,人工神经网络的发展为航空发动机的气路故障诊断提供了可靠的分析工具。通过充分利用发动机专业领域的知识、经验以及历史数据,主要包括基于神经网络、支持向量机、模糊逻辑、专家知识系统及深度核极限学习机^[2]等算法,使基于数据驱动的发动机气路故障诊断方法得已长足发展。Kyriazis等^[3]通过概率神经网络与模糊逻辑相结合以增强对气路故障的识别能力;Sarkar等^[4]认识到发动机的突发故障和衰退性故障存在不同的特性,提出了基于符号动态过滤模型的预测方法,并通过分析时间序列数据以区分2类故障;Lin等^[5]将原始故障信号转化为谱熵图像并输入卷积神经网络,提高了故障诊断算法的实时性;崔建国等^[6]利用无标签数据样本集对深度自编码网络进行预训练以辅助Softmax分类器进行诊断决策;曹惠玲等^[7]将高斯核函数支持向量机作为AdaBoost算法的弱分类器,实现了多分类气路故障诊断,有效解决了非线性和学习样本不足的问题;田静等^[8]引入了遗传算法并应用于正则化参数和核参数的调优,可自适应地选取最小二乘支持向量机参数,提高故障诊断模型的抗噪能力。但现存的问题限制了基于数据驱动的发动机气路故障诊断方法的推广与应用。一方面,受传感器精度及环境干扰的影响,飞行监控数据中常包含较多冗余特征及噪声,这类数据在增加模型训练负担的同时降低了故障诊断精度;另一方面,发动机实际运行时所采集的监测数据中的正常样本数量远大于故障样本数量,极不均衡的样本分布对故障诊断方法的学习过程造成影响,弱化模型的泛化能力。

本文提出了一种基于特征优化与支持向量机的航空发动机气路故障诊断方法,采取故障概率估计增加故障数据的方法,结合回归预测的思想,进行发动机航后数据的气路故障诊断预测。

1 故障诊断方法

结合特征优化和支持向量机(Support Vector

Machine, SVM)的气路故障诊断模型结构如图1所示。

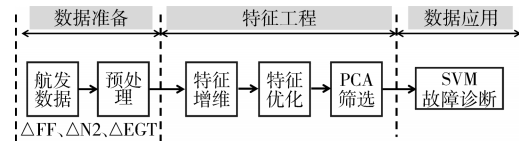


图1 故障诊断模型算法结构

由于直接利用原始数据集作为模型输入存在如下问题:原始特征之间的冗余性和相关性不利于机器学习模型的训练;在数据集中重要特征所构成的特征子集对故障问题的诊断起主导作用。因此对原始样本空间中的特征子集进行增维与最优子集选择,能优化模型运算效率并提升诊断精度。

本文提出的方法对经预处理后的发动机航后数据采用指数、对数、组合等多种方式进行特征增维。为从增维后的数据中选取具有区分度的特征组合,引入了近邻成分分析(Near Component Analysis, NCA)算法和深度自编码器(Auto Encoder, AE)等特征分析与提取方法,利用主成分分析法(Principal Component Analysis, PCA)对基于NCA和AE特征优化算法得到的结果进行比较,选择更具有可分性的故障特征进行故障诊断。

1.1 特征优化算法

本文采用特征优化算法对特征增维后的发动机状态特征进行评估分析,为后续特征选取提供依据。特征优化的基本任务是利用特征选择和特征提取算法对现有特征进行评估并找出最有效的特征。特征选择是从已有的 M 个特征中选出 N 个($N < M$)特征构成1个特征子集,通过优化特征信息达到后续任务指标的过程。特征提取则是在原有的特征集合基础上产生新的特征来帮助问题的求解。

1.1.1 近邻成分分析(NCA)算法

在特征空间中引入距离度量的学习能够极大地提高分类器的性能,有较高的实际应用价值。作为典型的距离测度学习算法,近邻成分分析法已在医疗检测^[9],语音识别^[10]和人脸识别^[11]等领域得到实际应用。NCA算法在数据样本中随机选择近邻,并通过优化留一法(Leave-One-Out, LOO)来搜索变换矩阵,并利用该矩阵获得原始数据的低维嵌入表示。

对数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$,算法流程如下所示,其中 $x_i \in R^d$ 代表第 i 个数据样本的特征向量, $y_i \in R$ 代表该样本的标签。

(1)先定义距离函数以计算样本之间的相对距离。

$$d_w(x_i, x_j) = \sum_{r=1}^p w_r^2 |x_{ir} - x_{jr}| \quad (1)$$

式中： w_r 为特征权重， x_{ir} 、 x_{jr} 分别是样本 x_i 、 x_j 的第 r 维特征。

(2)定义 k 是与 $d_w(x_i, x_j)$ 相关的相似函数。在相对距离较小时， $k(d_w(x_i, x_j))$ 取得较大值，1个典型的相似函数为 $k(z) = \exp(-z/\sigma)$ 。

(3)由此可得，对于样本点 x ，其近邻选择为样本点 x_j ，且 x 继承 x_j 的样本标签的LOO概率分布表示为

$$p_{ij} = \frac{k(d_w(x_i, x_j))}{\sum_{j=1, j \neq i}^n k(d_w(x_i, x_j))} \quad (2)$$

(4)计算样本 x_i 被正确分类的平均LOO概率

$$p_i = \sum_{j=1, j \neq i}^n p_{ij} y_{ij} \quad (3)$$

$$y_{ij} = I(y_i = y_j) = \begin{cases} 1, & \text{if } y_i = y_j \\ 0, & \text{otherwise} \end{cases}$$

(5)推广至样本空间中，以 λ 为正则化项，所有样本的平均LOO概率为

$$F(w) = \frac{1}{n} \sum_{i=1}^n p_i - \lambda \sum_{r=1}^p w_r^2 \quad (4)$$

此时，利用NCA进行特征选择的问题转化为求使 $F(w)$ 取得最大值的权重向量 w

$$\hat{w} = \arg \min_w -F(w) \quad (5)$$

在确定权重向量后，可依此获得各样本特征所对应的权重值以供筛选。

1.1.2 深度自编码器(AE)

深度自编码器^[12]是神经网络的一种，是直接使用单层或多层神经网络对输入数据进行映射，得到转换后的输出向量作为从输入数据提取出的特征。传统的自动编码器一般应用于数据降维^[13]、文本提取^[14]、语音识别^[15]等任务，能同时表征线性和非线性变换，相对其他的降维方法具有更高的适应性。

深度自编码器特征压缩模型如图2所示。从图中可见，典型的自编码器模型是由1个简单的3层神经网络结构组成，包括1个输入层、1个隐藏层和1个输出层。输入层和隐藏层构成编码网络 $h = f(x)$ ，

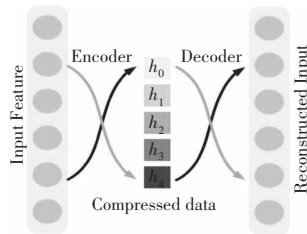


图2 深度自编码器特征压缩模型

隐藏层和输出层构成解码网络 $y = f(h)$ 。其中， $x \in \mathbb{R}^{n \times 1}$ 代表输入值； $h \in \mathbb{R}^{m \times 1}$ 代表隐藏值； $y \in \mathbb{R}^{n \times 1}$ 代表输出值； $f(\cdot)$ 为激活函数。

自编码器模型的训练目标是使用稀疏的高阶特征重新组合来重构输入，其输入输出层具有相同的维数。作为一种数据压缩算法，自动编码器通过最小化输入 x 与输出 y 的误差进行学习，其特征表达误差函数 J_{AE} 定义为输入与输出的二范数

$$J_{AE} = \frac{1}{2N} \sum_{i=1}^N \|x^{(i)} - y^{(i)}\|_2^2 \quad (6)$$

1.2 支持向量机

支持向量机(Support Vector Machines, SVM)是在分类与回归分析中常用的监督式学习模型，目前已应用于蛋白质结构预测^[16]、故障诊断^[17]等领域。SVM将实例表示为空间中的点，并在空间中搜索1个能够将某个值最大化的超平面，使得单独类别的实例被尽可能宽的间隔分割开。若存在新的实例，将其映射到同一空间并基于所在间隔的位置预测所属类别。超平面和所有实例间的最小距离被称为间隔。SVM算法在空间中搜索分离超平面如图3所示。 $w \cdot x + b = 0$ 即为分离超平面，对于线性可分的数据集，此类超平面有无穷个，但是几何间隔最大的分离超平面却是唯一的。

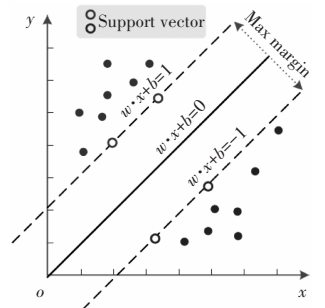


图3 SVM算法在空间中搜索分离超平面

另外，在解决线性不可分问题时，需引入核函数使高维特征空间上的计算映射到低维，避免在高维空间中向量内积的计算量过大。对于不同的任务需根据数据的先验知识选用符合数据分布的核函数。目前常用的核函数包括线性核函数、多项式核函数、高斯核函数和Sigmoid核函数。本文采用高斯核函数作为SVM的核函数，其表达式为

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (7)$$

式中： σ 为高斯核函数参数。

非线性支持向量机的学习过程可视为在训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 中搜寻最优超平面和分类决策函数的过程。其中 $x_i \in \mathbb{R}^n$ ， $y_i \in \{+1, -1\}$ ，

$i = 1, 2, \dots, N$ 。具体计算过程如下:

(1)选取适当的核函数 $K(x, z)$ 和 $C > 0$, 构造并求解凸二次规划问题:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i$$

$$\text{s.t.} \sum_{i=1}^N \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, N$$

计算得到最优解 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ 。

(2)选择 α^* 的分量 α_j^* , 满足条件 $0 < \alpha_j^* < C$, 计算 $b^* = y_i - \sum_{i=1}^N \alpha_i^* y_i K(x_i, y_j)$; 分类决策函数为 $f(x) = \text{sign}(\sum_{i=1}^N \alpha_i^* y_i K(x, x_i) + b^*)$ 。

2 数据预处理与特征优化

本文涉及的航空发动机气路性能测量参数来自涡桨发动机气路性能模型生成的仿真数据和 CFM56-7B 发动机航后数据以及维修记录^[18-19]。气路性能仿真模型模拟了涡桨发动机处在正常状态、单部件故障、放气活门故障以及组合故障等多种故障情境下的运行状况,并在不同故障模式下获取的数据样本与健康基线^[18]的对比,以记录发动机状态的异常波动。本文涉及的气路测量参数及相关变量见表 1、2。单独故障(部件故障或放气活门故障)及其相应的单一性能参数变化量(其他性能参数保持不变)见表 3。组合故障及其相应的性能参数变化量见表 4。为使仿真数据更接近真实飞行数据,模型仿真过程中加入了一定数量级的噪声。

2.1 数据预处理与特征增维

由于发动机实际故障种类较多,且对不同测量参数具有不同程度的影响,直接利用现有的测量参数作

表 3 部件单独故障、放气活门故障与性能参数变化量的对应关系

故障类型	参数变化	故障类型	参数变化
低压压气机	效率系数	低压压气机	流量系数
效率降低	-0.017	流量增加	+0.018
高压压气机	效率系数	高压压气机	流量系数
效率降低	-0.011	流量降低	-0.012
低压涡轮	效率系数	低压涡轮	流量系数
效率降低	-0.012	流量增加	+0.018
高压涡轮	效率系数	高压涡轮	流量系数
效率降低	-0.012	流量增加	+0.008
低压活门位置异常	PCBLP22=0.05	高压活门漏气	PCBLP22=0.02

表 4 组合故障与性能参数变化量的对应关系

故障种类	低压压气机	低压压气机	高压压气机	高压压气机
	流量系数	效率系数	流量系数	效率系数
外物损伤	+0.02	-0.050	+0.02	-0.030
叶片腐蚀	-0.02	-0.025	-0.02	-0.020
叶片积垢	-0.01	-0.015	-0.01	-0.015
叶片磨损	-0.04	-0.030	-0.02	-0.015

为故障特征建模可能无法进行准确的故障诊断。因此采取特征增维方式增加故障特征,进而优选故障特征以挖掘故障的隐藏信息,从而提升发动机故障诊断的精度,提升算法的有效性。

特征增维一般是在已有特征的基础上,通过线性组合、对数变换和指数变换等形式增加故障特征的数量。本文对已有的故障特征做多种类型的特征增维,包括指数增维、对数增维、差分增维和组合增维。2类航空发动机具体的特征增维方式分别见表 5、6。组合增维一般要选择具有相关关系的参数进行组合。

表 5 涡桨发动机特征增维方式

特征类型	增维特征
指数增维	$ALTP^{-1}, MA^{-1}, ANV^{-1}, POWER^{-1}, \Delta WFB^{-1}, \Delta PCNC^{-1}, \Delta PCNF^{-1}, \Delta T8M^{-1}, ALTP^{-2}, MA^{-2}, ANV^{-2}, POWER^{-2}, \Delta WFB^{-2}, \Delta PCNC^{-2}, \Delta PCNF^{-2}, \Delta T8M^{-2}, ALTP^2, MA^2, ANV^2, POWER^2, \Delta WFB^2, \Delta PCNC^2, \Delta PCNF^2, \Delta T8M^2$
	$\log(ALTP+1), \log(MA+1), \log(ANV+1), \log(POWER+1), \log(\Delta WFB+1), \log(\Delta PCNC+1), \log(\Delta PCNF+1), \log(\Delta T8M+1)$
对数增维	$\log(ALTP+1), \log(MA+1), \log(ANV+1), \log(POWER+1), \log(\Delta WFB+1), \log(\Delta PCNC+1), \log(\Delta PCNF+1), \log(\Delta T8M+1)$
组合增维	$\Delta WFB/\Delta PCNC, \Delta WFB/\Delta PCNF, \Delta PCNF/\Delta T8M, \Delta WFB/\Delta T8M, \Delta PCNC/\Delta PCNF, \Delta PCNC/\Delta T8M$

表 1 涡桨发动机测量参数

变量名称	变量描述
ALTP	飞行高度
MA	飞行马赫数
ANV	转速恒定调节量
POWER	功率
ΔWFB	燃油流量增量
$\Delta T8M$	发动机排气温度增量
$\Delta PCNC$	高压涡轮转子转速增量
$\Delta PCNF$	低压转子转速增量

表 2 CFM56-7B 发动机测量参数

变量名称	变量描述
ΔEGT	发动机排气温度增量
$\Delta N2$	高压压气机转子转速增量
ΔFF	燃油流量增量

表6 CFM56-7B发动机特征增维方式

特征类型	增维特征
指数特征	$\Delta N2^{-1}$, ΔFF^{-1} , ΔEGT^{-1} , $\Delta N2^{-2}$, ΔFF^{-2} , ΔEGT^{-2} , $\Delta N2^2$, ΔFF^2 , ΔEGT^2
对数特征	$\log(\Delta N2+1)$, $\log(\Delta FF+1)$, $\log(\Delta EGT+1)$
差分特征	$\text{Diff}(\Delta EGT)$, $\text{Diff}(\Delta N2)$, $\text{Diff}(\Delta FF)$
组合特征	$\Delta FF/\Delta N2$, $\Delta EGT/\Delta FF$, $\Delta EGT/\Delta N2$

2.2 特征选择与提取

特征增维方法在增强原始特征的表达能力的同时,也为数据集引入了较多的冗余特征。为了消除冗余特征的影响、减少数据的存储空间、提升模型性能,利用特征选择方法和特征降维方法对所获得的特征进行优化选择。

首先使用NCA算法对特征增维后的2类航空发动机数据集进行特征选择,该算法利用机器学习为每个特征打分,避免特征选择结果与实际模型结果脱节。在利用NCA算法获得权重向量前,需要对正则化项 λ 进行寻优并选定平均损失的最小参数。通过选取目标函数在极小值点时的 λ 值,求取此时对应的权重向量 w 作为特征排序结果。在涡桨发动机和CFM56-7B发动机上利用NCA算法获得的部分特征的重要性评分(评分>0.1)分别如图4、5所示。

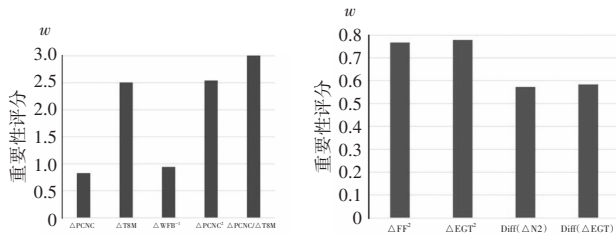


图4 在涡桨发动机上利用NCA算法获得的部分特征的重要性评分(评分>0.1) 图5 在CFM56-7B发动机上利用NCA算法获得的部分特征的重要性评分(评分>0.1)

依上述图中标示的特征权重值,筛选出最终应保留的增维后特征组合见表7。

表7 2种类型发动机特征选择后保留的特征子集

数据集	特征类型	特征名称
涡桨发动机	原始特征	$\Delta PCNC$, $\Delta T8M$
	指数特征	ΔWFB^{-1} , $\Delta PCNC^2$
	组合特征	$\Delta PCNC/\Delta T8M$
CFM56-7B 发动机	差分特征	$\text{Diff}(\Delta EGT)$, $\text{Diff}(\Delta N2)$
	指数特征	$\Delta N2^{-1}$, ΔFF^{-1} , ΔEGT^{-1}

此外,与特征选择中的过程类似,选用了深度自编码器模型作为特征降维工具,同样以特征增维后的

2类航发特征作为输入。通过建立1个典型的3层神经网络,自编码器能够利用模型的隐层结构进行特征降维与特征提取,将输入的特征映射成低维特征。

2.3 主成分分析算法评估特征优化结果

为了在各类特征选择和特征降维方法中选择最适合所选数据集进行特征优化,通过在模型中引入PCA法^[20]对各类特征优化方法的结果进行评估。PCA是针对模式分类的特征提取的最典型工具。作为经典的特征提取方法,PCA在不减少原始数据所包含的信息前提下,将原始数据集转化为由维数较少的“有效”特征成分来表示,使其在统计均方意义下达到方差最优的目的。同时,借助PCA易于可视化的优势,在2维平面上分别展示NCA算法和深度自编码器处理后的数据集进而评估故障样本的分类能力及其优劣。

利用2种特征优化方法对在不同发动机气路性能仿真数据集上的结果进行分析。以涡桨发动机数据集为例,原始数据集和经过2种方法分别优化后的2类数据集的PCA可视化结果如图6~8所示。

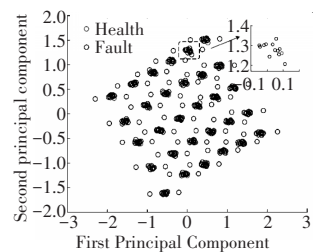


图6 数据集原始特征的PCA可视化结果

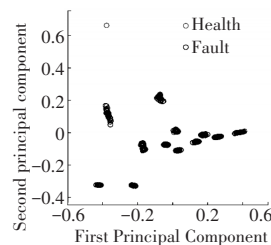


图7 NCA算法筛选后特征子集的PCA可视化结果

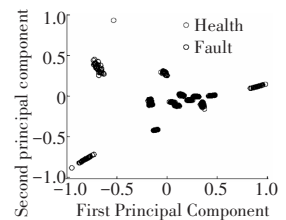


图8 AE算法压缩所得的特征PCA可视化结果

对比3类数据集的样本分布状况,发现相较于原始涡桨发动机的仿真数据集,利用NCA算法和深度自编码器优化后的特征子集均能实现故障样本与正常样本的较好分离。由于NCA算法所得特征子集对不同种类的气路故障具有更好的分离性,因此更适用于气路故障诊断。

3 故障诊断与结果分析

基于上述特征优化后的航发数据集,构建支持向量机模型,开展气路故障诊断,并通过与其他常见分类器的对比,验证了所提出的故障诊断方法的有

效性。

3.1 基于特征优化和支持向量分类模型的故障诊断

使用NCA算法获得的重要性评分 >0.1 的增维后,在特征输入SVM分类器中进行故障诊断。对于涡桨发动机,其输入特征为 $\Delta PCNC$ 、 $\Delta T8M$ 、 ΔWFB^{-1} 、 $\Delta PCNC^2$ 和 $\Delta PCNC/\Delta T8M$,5种特征(表7)。模型中的SVM分类器使用径向基函数核(Radial Basis Function, RBF),数据集按照7:3的比例划分训练集和测试集。

为了验证基于特征优化和支持向量机的故障诊断算法的有效性,本节在涡桨发动机仿真数据集上,将所提算法与几种常见的故障诊断算法进行比较。

随机森林(Random Forest, RF)^[21]、感知机模型(Perceptron Model, PM)^[22]、决策树(Decision Tree, DT)^[23]和支持向量机^[24]4种故障诊断算法在涡桨发动机数据集上的统计分析结果见表8。根据本文的特征优化算法,设计了上述模型的4种变体共同进行试验。其中,ORIG表示该算法使用原始特征来训练模型;FO表示算法使用了特征优化算法筛选后的特征子集训练模型。

表8 基于原始特征和优化特征的故障诊断结果

算法名称	FDR/%	运行时间/s
ORIG SVM	99.65	0.013
ORIG RF	94.67	0.150
ORIG PM	63.84	0.010
ORIG DT	93.34	0.007
FO SVM	99.80	0.007
FO RF	95.71	0.143
FO PM	36.70	0.008
FO DT	96.60	0.003

注:表中结果均是虚警率FAR为0.658%时的诊断结果。

从表中的试验结果可见,特征优化方法可显著减少各类方法的运行时间,且提高了除感知机以外的所有故障诊断算法的准确性。其中,从故障诊断性能分析:基于特征优化和支持向量机的故障诊断算法,其结果得出了99.80%的最高故障检测率(False discovery rate, FDR)。在试验中,各方法对非故障样本的诊断精度均较高,预测过程中均只有1个正常样本被错分,因此虚警率(False alarm rate, FAR)均为0.658%。此外,试验中感知机模型发生了较明显的欠拟合现象。ORIG PC算法在涡桨发动机数据集上的训练集和测试集准确率均小于65%,表明算法陷入

欠拟合状态,难以学习到数据信息;而在FO PC中,特征选择算法对输入数据的特征项做了简化,使得其欠拟合现象更为严重,导致了FDR指标的进一步下降。

从模型运行效率分析,与采用原始特征的诊断模型相比,基于特征优化的各类方法运行时间均有缩短。在特征优化后的模型中,支持向量机模型的运行效率仅低于决策树算法的。由于决策树算法是一种基于树形分支的方法,是不需要计算特征之间的相似性等。本文选用的支持向量机算法是RBF函数核将输入的低维样本特征映射到高维空间,通过衡量样本之间的相似度做出分类决策,其训练过程中参数较多,需略长的训练时间。综合模型准确率和运行效率,本文提出的算法具有较好的故障诊断性能。

为了评价PCA算法评估结果的可靠性,本文探究了特征优化算法的选择对评估结果产生的影响。对比模型分别使用基于NCA和AE算法优化后的特征子集进行训练后的分类结果见表9。其中,AE表示使用自编码算法优化后的特征来训练模型;NCA表示算法使用了近邻成分分析算法筛选后的5维特征训练模型。相较于NCA算法,基于自编码AE算法优化后的各故障诊断模型需要更长的运行时间,且4种模型中的3种的准确率都很低。因此,PCA算法评估的各特征优化算法的优化质量与实际故障诊断中的表现是一致的。

表9 基于NCA和AE优化特征的故障诊断结果

算法名称	FDR/%	运行时间/s
AE SVM	99.65	0.019
AE RF	52.79	0.186
AE PM	54.17	0.012
AE DT	70.40	0.011
NCA SVM	99.80	0.007
NCA RF	95.71	0.143
NCA PM	36.70	0.008
NCA DT	96.60	0.003

注:表中结果均是虚警率FAR为0.658%时的诊断结果。

3.2 基于特征优化和支持向量回归模型的非平衡故障诊断算法

由于现实中获取的数据集中故障与正常2个类别的样本数目相差巨大,正常样本的数量远大于故障样本的数量,即为典型的不平衡数据集。以CFM56-7B型涡扇发动机数据为例,在1055条有效样本中,仅包含13条单一气路故障样本,即发动机流路因污垢

积累导致了其总体性能指标(排气温度裕度)下降而进行水洗。水洗故障标签的设定即发动机故障概率与维修记录的关系如图9所示,从图中可见,在发动机水洗前的一段时间内认为发动机已产生故障征兆,将其设定为故障状态。而更之前的状态和水洗之后的状态则视为健康状态。设定水洗前 N 个飞行循环为故障状态,作为后续的故障诊断算法所需样本。一般该循环数是通过反复测试确定的,并以该循环数为变量,测试每个循环数下的故障诊断精度,最后选择精度较高的循环数。

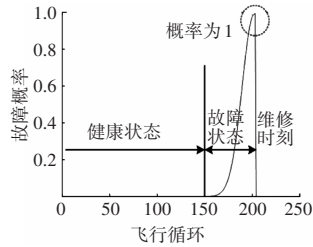


图9 发动机故障概率与维修记录的关系

本文考虑不平衡数据集的特性,同时发动机气路故障一般可视为发动机气路性能退化到一定程度的结果,即发动机的故障产生过程视为不同的衰退状态^[24]。在前述故障诊断算法中,根据发动机故障维修记录对故障样本数据进行了扩展生成,对于发动机气路故障,除了外来物击伤外,在实际飞行过程中,离故障维修点越近的时刻发生故障的可能性越大。因此在对原始数据进行特征优化后,将故障点及故障点前的部分数据,设置一定的置信区间,用正态分布进行拟合,以起到扩充故障样本数据的作用。然后将NCA算法抽取的4维特征数据包括 $\text{Diff}(\Delta \text{EGT})$ 、 $\text{Diff}(\Delta \text{N2})$ 、 $\text{Diff}(\Delta \text{EGT})$ 和 $\text{Diff}(\Delta \text{N2})$ 4种特征,利用支持向量回归机(Support Vector Regression, SVR)进行故障诊断,并依照故障诊断模型在训练集中的结果设置合适的分类阈值。

基于特征优化和SVM的诊断算法及结合回归预测的诊断算法在CFM56发动机数据集上的故障诊断结果见表10。从试验数据可知,结合回归预测的故障诊断算法能显著提高不平衡数据集上的故障检测准确率。

表10 基于不同算法的CFM56航空发动机故障诊断结果

算法种类	FAR/%	FDR/%
Regression Method ¹	0.23	91.67
Non-regression Method ²	0	40

注:1. 结合回归预测即基于特征优化和SVR的诊断算法;2. 基于特征优化和SVM的诊断算法。

为进一步论证结合回归预测的故障诊断算法对样本故障率的分析准确度,选取测试数据中1个故障点以及故障点前的80个点进行观察,并绘制故障概

率变化曲线如图10所示。从图中可见,当时间离故障点越近时,SVR模型所预测的故障概率随之增大。该可视化故障概率曲线表明结合回归预测的诊断算法对故障概率的预测有较好的拟合效果。

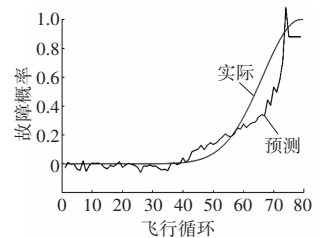


图10 CFM56发动机故障概率变化曲线

4 结论

(1)相比其他的故障诊断算法,本文建立的基于特征优化和SVM的故障诊断模型,具有最佳的准确性及较短的运行时间。

(2)针对常见的2类不平衡数据集,本文结合维修记录,扩展故障数据样本,并同回归预测思想相结合,提高了此类情况下诊断算法的准确性。

参考文献:

- 潘鹏飞. 基于试飞数据的航空发动机状态监测与故障诊断[J]. 推进技术, 2021, 42(12): 2826-2837.
PAN Pengfei. Flight data based condition monitoring and fault diagnosis of aero-engine[J]. Journal of Propulsion Technology, 2021, 42(12): 2826-2837. (in Chinese)
- 逢珊, 杨欣毅, 张勇, 等. 应用深度核极限学习机的航空发动机部件故障诊断[J]. 推进技术, 2017, 38(11): 2613-2621.
PANG Shan, YANG Xinyi, ZHANG Yong, et al. Application of deep kernel extreme learning machine in aeroengine components fault diagnosis[J]. Journal of Propulsion Technology, 2017, 38(11): 2613-2621. (in Chinese)
- Kyriazis A, Mathioudakis K. Gas turbine fault diagnosis using fuzzy-based decision fusion [J]. Journal of Propulsion and Power, 2015, 25(2): 335-343.
- Sarkar S, Mukherjee K, Ray A, et al. Fault diagnosis and isolation in aircraft gas turbine engines[C]// American Control Conference. Seattle: IEEE Press, 2008: 2166-2171.
- Li H, Gou L, Zheng H, et al. Intelligent fault diagnosis of aeroengine sensors using improved pattern gradient spectrum entropy[J]. International Journal of Aerospace Engineering, 2021, 10(1155): 1-20.
- 崔建国, 李国庆, 蒋丽英, 等. 基于深度自编码网络的航空发动机故障诊断[J]. 振动测试与诊断, 2021, 41(1): 85-89, 201-202.
CUI Jianguo, LI Guoqing, JIANG Liying, et al. Aero-engine fault diagnosis based on deep self-coding network[J]. Journal of Vibration Measurement and Diagnosis, 2021, 41(1): 85-89, 201-202. (in Chinese)
- 曹惠玲, 高升, 薛鹏. 基于多分类AdaBoost的航空发动机故障诊断

- [J].北京航空航天大学学报,2018,44(9):1818-1825.
CAO Huiling, GAO Sheng, XUE Peng. Aeroengine fault diagnosis based on multi-classification AdaBoost[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44 (9): 1818-1825. (in Chinese)
- [8] 田静,胡鹤翔.基于GA-LSSVM的航空发动机气路故障诊断[J].中国民航大学学报,2021,39(3):29-33.
TIAN Jing, HU Hexiang. Fault diagnosis of aero-engine gas path based on GA-LSSVM[J]. Journal of Civil Aviation University of China, 2021, 39 (3): 29-33. (in Chinese)
- [9] Djerioui M, Brik Y, Ladjal M, et al. Neighborhood component analysis and support vector machines for heart disease prediction[J]. Ingénierie des Systèmes d'Information, 2019, 24(6):591-595.
- [10] Singhmiller N, Collins M, Hazen T J. Dimensionality reduction for speech recognition using neighborhood components analysis[C]// Eighth Annual Conference of the International Speech Communication Association. Antwerp:DBLP, 2007: 1158-1161.
- [11] Wang F, Zhang H, Zuo W, et al. Face recognition using fast neighborhood component analysis with spatially smooth regularizer[C]// International Conference on Intelligent Science and Intelligent Data Engineering. Berlin: Springer, 2012: 450-457.
- [12] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313 (5786): 504-507.
- [13] 胡昭华,宋耀良.基于Autoencoder网络的数据降维和重构[J].电子与信息学报,2009,31(5):1189-1192.
HU Zhaohua, SONG Yaoliang. Dimensionality reduction and reconstruction of data based on autoencoder network[J]. Journal of Electronics and Information Technology, 2009, 31 (5): 1189-1192. (in Chinese)
- [14] 刘勘,袁蕴英.基于自动编码器的短文本特征提取及聚类研究[J].北京大学学报(自然科学版),2015,2(2):282-288.
LIU Kan, YUAN Yunying. Short texts feature extraction and clustering based on auto-encoder[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2015, 2 (2): 282-288. (in Chinese)
- [15] 王山海,景新幸,杨海燕.基于深度学习神经网络的孤立词语音识别的研究[J].计算机应用研究,2015,32(8):2289-2291,2298.
WANG Shanhai, JING Xinxing, YANG Haiyan. Study of isolated speech recognition based on deep learning neural networks[J]. Application Research of Computers, 2015, 32 (8): 2289-2291, 2298. (in Chinese)
- [16] Guo J, Chen H, Sun Z, et al. A novel method for protein secondary structure prediction using dual-layer SVM and profiles[J]. Proteins-structure Function and Bioinformatics, 2004, 54 (4): 738-743.
- [17] 徐启华,师军,耿帅.应用快速多分类SVM的航空发动机故障诊断方法[J].推进技术,2012,33(6):961-967.
XU Qihua, SHI Jun, GENG Shuai. Aero-engine fault diagnosis by a new fast multi-class support vector algorithm[J]. Journal of Propulsion Technology, 2012, 33 (6): 961-967. (in Chinese)
- [18] 刘渊,余映红,田彦云,等.航空发动机排气温度基线建模新方法研究[J].推进技术,2022,43(4):16-25.
LIU Yuan, YU Yinghong, TIAN Yanyun, et al. Investigation on new method for baseline modelling of aeroengine exhaust gas temperature [J]. Journal of Propulsion Technology, 2022, 43 (4): 16-25. (in Chinese)
- [19] 王奕惟,莫李平,王奕首,等.基于全航段QAR数据和卷积神经网络的航空发动机状态辨识[J].航空动力学报,2021,36(7):1556-1563.
WANG Yiwei, MO Liping, WANG Yishou, et al. Aero-engine status identification based on full-segment QAR data and convolutional neural network[J]. Journal of Aerospace Power, 2021, 36 (7): 1556-1563. (in Chinese)
- [20] Wold S, Esbensen K, Geladi P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2 (1-3): 37-52.
- [21] 庄进发,罗键,彭彦卿,等.基于改进随机森林的故障诊断方法研究[J].计算机集成制造系统,2009,15(4):777-785.
ZHUANG Jinfa, LUO Jian, PENG Yanqing, et al. Fault diagnosis method based on modified random forests[J]. Computer Integrated Manufacturing Systems, 2009, 15 (4): 777-785. (in Chinese)
- [22] 王志,艾延廷,沙云东.基于BP神经网络的航空发动机整机振动故障诊断技术研究[J].仪器仪表学报,2007(S1):4.
WANG Zhi, AI Yanting, SHA Yundong. Research on fault diagnosis technology of whole body vibration of aero-engine based on BP neural network[J]. Chinese Journal of Scientific Instrument, 2007(S1): 4. (in Chinese)
- [23] 关雪飞,杨军,马小兵.决策树在航空发动机状态数据分析中的应用[C]//中国现场统计研究会学术年会.中国现场统计研究会第十三届学术年会论文集.北海:中国学术期刊电子出版社,2007:159-162.
GUAN Xuefei, YANG Jun, MA Xiaobing. Application of decision tree in aero-engine state data analysis[C]//Annual Conference of China Field Statistics Research Association. Proceedings of the 13th Annual Academic Conference of the China Field Statistics Research Association. Beihai: China Academic Journal Electronic Publishing House, 2007: 159-162. (in Chinese)
- [24] 徐启华,师军.基于支持向量机的航空发动机故障诊断[J].航空动力学报,2005,20(2):298-302.
XU Qihua, SHI Jun. Aero-engine fault diagnosis based on support vector machine [J]. Journal of Aerospace Power, 2005, 20 (2): 298-302. (in Chinese)

(编辑:贺红井)