



计算机工程

Computer Engineering

ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于 EnsembleBRB-SHAP 的航空发动机健康状态可解释预测方法
作者: 游雅倩, 闫辉, 苏耀峰, 王晓双, 鄢睿丞
DOI: 10.19678/j.issn.1000-3428.0070152
网络首发日期: 2024-12-11
引用格式: 游雅倩, 闫辉, 苏耀峰, 王晓双, 鄢睿丞. 基于 EnsembleBRB-SHAP 的航空发动机健康状态可解释预测方法[J/OL]. 计算机工程.
<https://doi.org/10.19678/j.issn.1000-3428.0070152>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于 EnsembleBRB-SHAP 的航空发动机健康状态可解释预测方法

游雅倩, 闫辉*, 苏耀峰, 王晓双, 鄢睿丞

(国防科技大学, 信息通信学院, 武汉 430030)

摘 要: 航空发动机健康状态预测作为发动机健康管理的重要环节之一, 能够为提升飞机可靠性、降低发动机维护成本等工作提供定量化依据。然而, 传统的航空发动机健康状态预测对可解释性关注度较低, 导致对发动机视情维修等决策的支撑性不足。为此, 本文面向发动机健康状态预测的可解释需求, 提出基于 EnsembleBRB-SHAP 模型的航空发动机健康状态可解释预测方法。首先, 采用数据驱动法训练多个航空发动机健康状态预测子置信规则库 (Belief Rule-Based, BRB) 模型。在此基础上, 构建航空发动机健康状态预测集成置信规则库 (EnsembleBRB) 模型, 在有效利用多源不确定数据的同时, 保证模型的预测准确性。然后, 基于沙普利加性解释 (SHapley Additive exPlanations, SHAP), 对 EnsembleBRB 模型进行分析解释, 定位影响发动机健康状态的关键因素, 实现航空发动机健康状态的可解释性预测。最后, 引入商用模块化航空推进系统仿真软件记录的发动机故障实验监测数据, 验证所提方法的可行性与有效性。结果表明, 该方法在航空发动机健康状态预测中的准确性为 $MSE=0.0122$, 通过局部可解释性与全局可解释性分析, 归纳得出低压涡轮机冷却液泄漏量、风扇转速等是决定发动机健康状态的关键参数, 进而更好地支撑航空发动机健康管理等决策工作。

关键词: 集成学习算法; 置信规则库 (Belief Rule-Based, BRB); 沙普利加性解释 (SHapley Additive exPlanations, SHAP); 可解释性; 发动机健康状态预测

DOI: 10.19678/j.issn.1000-3428.0070152

Health State Prediction of Aero-Engine based on EnsembleBRB-based SHapley Additive exPlanations Approach

YOU Yaqian, YAN Hui*, SU Yaofeng, WANG Xiaoshuang, YAN Ruicheng

(College of Information and Communication, National University of Defense Technology, Wuhan 430030, China)

【Abstract】 As one of the important aspects of engine health management, predicting the health state of aero-engines can provide quantitative basis for improving aircraft reliability and reducing engine maintenance costs. However, traditional aviation engine health state prediction lacks sufficient attention to interpretability, resulting in a decrease in support for engine decision-making such as condition-dependent maintenance. This paper proposes an interpretable prediction method for the health state of aero-engines based on the EnsembleBRB-SHAP approach, considering the demand for interpretability in engine health state prediction. Firstly, a data-driven approach is used to train multiple sub aero-engine health state prediction models based on the **Belief Rule Based (BRB)**. Then, the EnsembleBRB model for predicting the health state of aero-engines is constructed to utilizes multiple sources of uncertain data while ensuring the prediction accuracy. Based on the **SHapley Additive exPlanations (SHAP)** framework, the constructed EnsembleBRB model is analyzed and interpreted to identify the key features and achieve interpretable prediction of aero-engine health state. Finally, the feasibility and effectiveness of the proposed method are verified by introducing the experimental monitoring data of engine faults recorded by the Commercial Modular Aero-Propulsion System Simulation. The results show that the accuracy of the proposed method in predicting the health state of aero-engine is $MSE=0.0122$. Through the analysis of local interpretability and global interpretability, it is summarized that the LPT coolant bleed and physical fan speed are the key parameters determining the health status of the engine, which in turn can better support the decision-making for the health management of the aero-engine and other work.

【Key words】 ensemble learning algorithm; Belief Rule-Based (BRB); shapley additive explanations (SHAP); interpretability; engine health state prediction

0 引言

目前, 我国航空业正值快速发展时期。作为航

空飞行器的核心动力源, 航空发动机是一种对安全和可靠性要求极高、设计生产和维修保养花费极大

基金项目: 国家自然科学基金“基于可解释置信规则库的装备作战试验评估方法研究”(72401287); 国家自然科学基金“动态数据驱动的装备试验评估方法”(72471238)

E-mail: yanhui13@nudt.edu.cn

的大型复杂设备。然而,航空发动机在运转过程中,其内部元器件受本身制造工艺和工作环境等多种因素的影响,不可避免地会发生健康状态退化,一旦恶化至故障状态,将对航空飞行器造成严重影响,尤其当故障发生在飞行时,可能造成重大人员伤亡,带来不可挽回的损失。如果准确地预测发动机健康状态,就能在发动机性能退化的初期,尤其在还没有造成重大危害时,确定维护的最佳时机,不仅能够极大地提高安全性,避免故障的发生,同时能够有效的降低停机时间、减少维修周期、简化维修步骤、延长发动机使用寿命,从而降低维修成本。因此,准确有效地预测航空发动机的健康状态,进而制定合理的维修保养策略,对于提升飞行器可靠性、降低运营维护成本具有重要的意义。

目前,发动机健康状态预测的方法主要分为基于物理模型(physics-based models, PbMs)的预测方法和基于数据驱动模型(data-driven models, DDMs)的预测方法两大类^{[1][2]}。基于物理模型的方法往往需要在深入分析发动机工作原理的基础上构建数学模型,其建模成本较高,模型可移植性差。随着数据科学的蓬勃发展,基于数据驱动模型的预测方法在预测准确性和效率上取得了一定的优势,逐渐成为当前健康状态预测的主流。然而,在实际应用中仍存在以下问题:一是航空发动机组成部件复杂,传感器收集的监测参数较多,且受到环境变化、传感器磨损等因素的影响,监测数据难免存在不确定性。二是健康状态预测作为发动机健康管理(Prognostic and Health Management, PHM)^[3]的重要环节,是视情维修、故障诊断、调整改型等众多决策工作的依

据和支撑,要求预测结果具备良好的可解释性。目前多数基于数据驱动的预测方法未能有效利用多元不确定数据,且模型可解释性较弱,难以为航空发动机健康状态预测工作提供有效支撑。

本文结合集成置信规则库(Ensemble Belief Rule-Based, EnsembleBRB)^[4]和沙普利加性解释框架(SHapley Additive exPlanations, SHAP)^[5]来应对上述问题,主要工作包括:1)充分置信规则库在不确定信息、不完备信息的描述与处理方面的优势^{[6]-[7]},同时引入集成学习策略,降低其面向多元监测数据时的建模成本,实现多元不确定监测数据在发动机健康状态预测中的有效利用;2)基于SHAP框架提出面向模型预测过程与结果的解释流程,对样本预测结果进行归因分析,实现从模型决策空间向解释空间的映射^{[8]-[10]},辅助业务人员精准判断导致发动机健康状态恶化的核心因素,指导发动机视情维修等决策工作的开展,3)以商用模块化航空推进系统仿真软件(C-MAPSS)^[11]产生的一系列涡轮发动机退化数据为例开展应用研究,分析影响发动机健康状态的核心因素,验证本文所提方法的可行性与有效性。

1 相关工作

本节首先对常用的发动机健康状态预测方法进行梳理,进而对本文采用的置信规则库和沙普利加性解释进行介绍。

1.1 发动机健康状态预测方法

基于物理模型的预测方法需要深入分析发动机的理论力学、热力学、流体力学等,构建物理失效模型,进而根据设备的运行状态和负荷对其运行及

健康状态进行预测。例如,文献[12]建立了单轴燃气涡轮发动机的综合非线性热力学模型,描述了工况与循环参数之间的关系,通过热损失指数和功率亏损指数等指标对发动机性能状态进行监测与评估。文献[13]基于零维系统分析方法构建了单轴燃气涡轮发动机模型,并采用多变量牛顿-拉夫森方法进行性能自适应和气路实时分析对基本模型进行修正,实现压缩机变导叶等故障预测。文献[14]构建了发动机拟序火焰燃烧模型、气路系统模型、喷油器模型等模型,基于 AMESim 软件平台开展预测性研究。然而,物理模型的建立通常需要综合考虑发动机整体及部件所经历的物理、化学和气动热过程,对失效机理建立复杂的数学模型,其建模和分析的复杂性和难度较高,且可移植性较差,限制了这类方法的应用和推广。

在传感器监测技术及人工智能技术的不断发展下,在航空发动机的使用过程中,积累了大量与其健康状态相关的监测数据,充分利用监测数据对发动机健康状态进行预测成为提升其准确性的有效途径。例如,文献[15]引入 D-S 证据理论和模糊专家系统,能够有效应对对发动机故障诊断中的模糊信息。文献[16]和[17]均引入集成学习算法,结合多种机器学习算法,提升发动机退化模型的准确性,对发动机健康状态进行预测。文献[1]提出了一种基于双向门控循环单元和多门混合专家的双任务网络结构,可以同时发动机的健康状态与剩余寿命进行预测。文献[17]将 LSSVM 和 HMM 相结合,进行发动机状态预测,重点关注预测的准确性。相较于基于物理模型的方法,基于数据驱动模型的预测方法采

用统计学和机器学习相关技术,一方面不依赖于发动机运行与退化机制的物理知识,容易扩展至不同型号的发动机;另一方面,先进的机器学习技术支撑从复杂的监测数据中学习知识与信息,也使得建立的健康状态预测模型具备较高的准确性,因此在发动机健康状态预测中应用更为广泛。

随着研究的深入,研究者们逐渐意识到,先前工作往往局限于估计发动机的健康状态,没有充分挖掘健康状态和发动机部件之间的因果关系^[19]。文献[20]基于贝叶斯网络和可视化分析设计健康状态描述符模块,对健康状态不佳的原因进行解释,通过可解释性指数量化每个传感器对健康状态的影响。文献[21]与文献[20]类似,同样在预测的基础上增加了解释模块,通过计算空间注意力确定每个监测参数对预测结果的贡献。这类方法侧重于对某种故障类型的样本进行全局解释,且训练解释模块的成本较高。文献[22]在分析影响置信规则库可解释性因素的基础上,提出新的可解释置信规则库并开展健康状态评估,但本质上是将置信规则库默认为一类可解释的方法,并未对导致发动机健康状态恶化的因素展开深入分析。

1.2 置信规则库

作为一类专家系统,置信规则库的知识库由多条带信度结构的拓展“IF-THEN”规则组成,其中“IF”是前提项,在发动机健康状态预测中即为影响其健康状态的监测参数,“THEN”表示结论项,即为待预测的发动机健康状态。规则库中第 k 条规则如公式(1)所示:

$R_k: IF (x_1 \text{ is } A_1^k) \wedge (x_2 \text{ is } A_2^k) \wedge \dots \wedge (x_M \text{ is } A_M^k),$
 THEN $\{(D_1, \beta_{1,k}), \dots, (D_N, \beta_{N,k})\}$, with rule weight θ_k (1)
 and attribute weight $\delta_1, \delta_2, \dots, \delta_M$,

其中, $A_m^k (k=1, 2, \dots, K)$ 表示第 k 条规则中第 m 个监测参数的参考值。 M 表示监测参数的总数, K 表示规则总数, N 表示发动机健康状态的等级数。第 n 个等级的健康状态为 $D_n (n=1, 2, \dots, N)$, 信度为 $\beta_{n,k}$ 。

置信规则库可解释性的研究开始于 2020 年, 尚处于探索阶段。文献[23]将置信规则库可解释性划分为全局可解释性与局部可解释性, 即对模型决策过程的理解和对单个样本推理结果的理解。文献[22]总结了置信规则库建模中导致其可解释性降低的主要原因, 以此为依据设定了建模的约束条件。文献[24]从隶属函数、知识库和推理引擎等方面提出了置信规则库可解释性指标, 实现置信规则库可解释性的横向对比。然而, 这些研究并未深入挖掘影响置信规则库决策结果的关键因素, 难以满足置信规则库在发动机健康状态评估等实际工作中对可解释性的需求, 对后续视情维修等决策工作的支撑有限。

1.3 沙普利加性解释(SHapley Additive exPlanations, SHAP)

目前, 大多数常见的通用化机器学习模型解释方法都属于可加特征归因方法 (Additive Feature Attribution Methods), 即输出模型被定义为输入变量的线性相加, 例如 LIME (Local Interpretable Model-agnostic Explanations)^[12]、DeepLIFT (Deep Learning Important FeaTures)^[26]、分层相关传播 (Layer-wise Relevance Propagation)^[27] 和经典 Shapley 值估计 (包括 Shapley 回归值^[28], Shapley 采样值^[29]和定量输入影响^[30])。SHAP 是一种基于合作博弈论的有监督机器学习模型通用解释框架, 是

上述解释预测框架的统一, 通过 Shapley 值^[6]将输出模型定义为输入变量的线性相加^[9], 实现对预测模型与结果的解释。

SHAP 值可以通过多种方法进行近似, 例如 Kernel SHAP、Deep SHAP 和 Tree SHAP, 识别每个输入属性对预测结果贡献的程度与影响方向, 为模型提供较好的局部解释和全局解释^[31]。

2 方法与模型

本文在置信规则库基础上引入集成学习的思想, 通过多个子置信规则库的训练与集成, 开展发动机健康状态预测, 有效解决了多元监测数据导致的预测模型建模成本与难度增加的挑战, 同时提升了预测的准确性。在此基础上, 引入 SHAP 框架对基于集成置信规则库的发动机健康状态预测模型进行局部与全局的解释, 解决了因集成导致的预测可解释性降低的问题, 使预测模型与结果更好地为发动机健康管理等决策工作服务提供了有力支撑。

2.1 发动机健康状态预测集成置信规则库建模

传统的置信规则库在前提属性较多时会存在组合爆炸问题, 记第 m 个前提属性参考值的数量为 T_m , 则规则库中包含的规则总数为 $NOR = \prod_{i=1}^M T_i$, 显然, 规则总数随着前提属性数量的增加呈指数增长。如果, 将前提属性拆分为 L 组分别构建子置信规则库, 每个子置信规则库中包含 $m (m < M)$ 个前提属性, 则所需规则总数为 $NOR = L \prod_{i=1}^m T_i$, 将指数增加的规则总数变为线性增加, 大大降低了建模所需的成本与计算量, 这就是集成置信规则库的核心思想。集成置信规则库的实现需要重点解决如下三个问题: 一

是如何确定每个子置信规则库包含的前提属性；二是如何保证子置信规则库的建模精度；三是如何集成子置信规则库的预测结果。

为此，集成置信规则库采用集成学习中的 Bagging 框架应对上述问题，其应用于发动机健康状态预测的建模流程如下：

步骤 1：健康状态预测子置信规则库建模

步骤 1.1：无放回地随机抽取 m 个监测参数作为子置信规则库 BRB_l 的前提属性；

步骤 1.2：从发动机监测数据集中有放回地随机抽取 q 个样本，且仅保留步骤 1.1 中确定的 m 个监测参数，作为子置信规则库 BRB_l 的训练集；

步骤 1.3：建立优化目标及约束条件，对子置信规则库 BRB_l 中的参数进行优化。

$$\min \quad MSE = \frac{1}{Q} \sum_{q=1}^Q (HI_q - \hat{HI}_q)^2 \quad (2.1)$$

$$s.t. \quad lb_m \leq A_{m,j} \leq ub_m, A_{m,1} = lb_m, A_{m,T_m} = ub_m, \quad (2.2)$$

$$0 \leq \delta_m \leq 1, 0 \leq \theta_k \leq 1, 0 \leq \beta_{n,k} \leq 1, \sum_{n=1}^N \beta_{n,k} \leq 1 \quad (2.3)$$

其中，公式 (2.1) 中 MSE 为优化目标，即发动机健康状态预测值与真实值的均方误差。公式 (2.2) 限制了前提属性 A_m 的取值要在最大值 ub_m 和最小值 lb_m 之间，公式 (2.3) 作为约束条件，分别约束了前提属性权重 δ_m 、规则权重 θ_k 和结论项信度 $\beta_{n,k}$ 的取值。

步骤 2：确定子置信规则库的权重

对于子置信规则库 BRB_l ，将发动机监测数据集未选中的样本作为测试集，计算其在发动机健康状态预测中的准确性，如公式 (3) 所示。

$$w_l = \sum_{i=1}^{Q-q} (HS_i - \hat{HS}_i)^2 \quad (3)$$

步骤 3：子置信规则库集成

步骤 3.1：令 L 个子置信规则库独立地对发动机健康状态进行预测，第 l 个子置信规则库 BRB_l 的预测结果记为 HS_l 。

步骤 3.1.1：根据公式 (4) 计算输入数据与规则前提项参考值的匹配度，输入数据记为 $(x_1, \varepsilon_1) \wedge (x_2, \varepsilon_2) \wedge \dots \wedge (x_M, \varepsilon_M)$ ， x_m 表示第 m 个监测参数的输入， ε_m 描述监测数据 x_m 中的不确定性。

$$\alpha_{i,j}^k = \frac{\varphi(x_i, A_{i,j}^k) \varepsilon_i}{\sum_{l=1}^{T_k} \varphi(x_i, A_{i,l}^k)}, \quad \varphi(x_i, A_{i,j}^k) = \begin{cases} \frac{A_{i,l+1}^k - x_i}{A_{i,l+1}^k - A_{i,l}^k} & j = l (A_{i,l}^k \leq x_i \leq A_{i,l+1}^k) \\ \frac{x_i - A_{i,l}^k}{A_{i,l+1}^k - A_{i,l}^k} & j = l+1 (A_{i,l}^k \leq x_i \leq A_{i,l+1}^k), \\ 0 & else \end{cases} \quad (4)$$

其中， $\alpha_{i,j}^k$ 表示在第 k 条规则中第 i 个前提属性的输入值 x_i 与其第 j 个参考值的匹配程度， $A_{i,j}^k$ 表示第 k 条规则中第 i 个前提属性的第 j 个参考值， T_k 表示第 k 条规则中第 i 个前提属性参考值的数量。

步骤 3.1.2：根据公式 (5) 计算输入数据对规则的激活度。

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_{i,j}^k)^{\bar{\delta}_i}}{\sum_{l=1}^K \left[\theta_l \prod_{i=1}^{T_l} (\alpha_{i,j}^l)^{\bar{\delta}_i} \right]} \quad \text{and} \quad \bar{\delta}_i = \frac{\delta_i}{\max_{j=1,2,\dots,T_k} \{\delta_j\}}, \quad (5)$$

其中， ω_k 表示第 k 条信度规则的激活权重。

步骤 3.1.3：根据公式 (6) 聚合被激活的规则。

$$\beta_n = \frac{\mu \left[\prod_{k=1}^K (\omega_k \beta_{n,k} + 1 - \omega_k \sum_{n=1}^N \beta_{n,k}) - \prod_{k=1}^K (1 - \omega_k \sum_{n=1}^N \beta_{n,k}) \right]}{1 - \mu \left[\prod_{k=1}^K (1 - \omega_k) \right]} \quad (6)$$

$$\mu = \left[\sum_{n=1}^N \prod_{k=1}^K \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{n=1}^N \beta_{n,k} \right) - (N-1) \prod_{k=1}^K \left(1 - \omega_k \sum_{n=1}^N \beta_{n,k} \right) \right]^{-1}$$

步骤 3.1.4: 根据公式 (7) 计算健康状态。

$$HS_l = \sum_{n=1}^N u_n \beta_n \quad (7)$$

其中, 信度结构 $B = (\beta_1, \beta_2, \dots, \beta_N)$ 表示预测得到的发动机健康状态 (Health state, HS) 处于各等级的信度, u_n 为第 n 个等级的健康状态效用值。

步骤 3.2: 根据子置信规则库权重对其进行集成, 得到最终的预测结果, 如公式 (8) 所示。

$$HS = \sum_{l=1}^L \bar{w}_l HS_l, \quad (8)$$

其中, \bar{w}_l 为归一化后子置信规则库 BRB_l 的权重,

$$\text{即 } \bar{w}_l = \frac{\min(w_1, w_2, \dots, w_L)}{w_l}。$$

集成置信规则库建模有三个特点, 一是前提属性的随机抽样, 保证了子置信规则库的多样性; 二是训练集的随机抽样, 既可以保证监测数据集规模较小时相同前提属性的子置信规则库的差异性, 同时能够控制监测数据规模较大时的计算开销, 并且无需预留额外样本估计子置信规则库的权重; 三是子置信规则库的集成, 相当于从多个角度对数据进行分析, 在降低计算成本的同时保证了应用于发动机健康状态预测中的准确性。

应用集成置信规则库开展发动机健康状态预测的时间复杂度和空间复杂度均为 $O(N)$ 。如监测参数有 13 个, 假设每个监测参数有 4 个参考值, 健康状态等级为 5, 那么采用传统的置信规则库, 共需优化 402,653,200 $((5+1) \times 4^{13} + 4 \times 3 + 4)$ 个参数,

而采用集成置信规则库 ($L=20$), 仅需优化 2,160 $(20 \times ((5+1) \times 4^2 + 4 \times 2 + 4))$ 个参数, 大大降低了预测模型构建的时间和空间成本, 能够较好地应对发动机健康状态预测中监测参数多元的特点。

2.2 发动机健康状态预测模型 SHAP 框架分析

第 2.1 节所述发动机健康状态预测集成置信规则库建模方法使建模过程有效地利用了发动机运行中收集的多元监测参数数据。然而, 其健康状态预测需集成多个子置信规则库独立的预测结果, 在反向追溯影响预测结果的因素时需综合考虑多个子置信规则库中结构与参数均存在差异的多条规则, 导致分析与解释的难度大大增加, 预测结果难以为发动机健康管理相关的决策工作提供支撑。

为此, 本文在发动机健康状态预测模型的基础上, 使用 SHAP 框架开展预测模型与结果的可解释性分析, 通过局部特征重要性分析、全局特征重要性分析, 确定监测参数对发动机健康状态的影响, 并为后续预测模型的调整优化提供意见与建议。

假设发动机健康状态预测模型的输入数据样本为 $\mathbf{x} = (x_1, x_2, \dots, x_p)$, 其中 p 为样本总数, 则原始模型 $f(\mathbf{x})$ 的简化输入 \mathbf{x}' 解释模型 $g(\mathbf{x}')$ 表示为:

$$f(\mathbf{x}) = g(\mathbf{x}') = \phi_0 + \sum_{m=1}^M \phi_m x'_m, \quad (9)$$

其中, M 表示前提属性的个数, ϕ_0 表示所有输入缺少时的恒定值。输入 \mathbf{x} 和 \mathbf{x}' 通过映射函数 $\mathbf{x} = h_x(\mathbf{x}')$ 相关联。文献[3]指出, 公式 (9) 存在唯一解, 并将模型的条件期望与博弈论中经典 Shapley 值组合到每个属性的归因值 ϕ_m 中, 可根据公式 (10) 计算:

$$\phi_m(f, \mathbf{x}) = \sum_{\mathbf{z}' \subseteq \mathbf{x}'} \frac{|\mathbf{z}'|! (M - |\mathbf{z}'| - 1)!}{M!} [f_x(\mathbf{z}') - f_x(\mathbf{z}' \setminus m)] \quad (10)$$

其中, $|\mathbf{z}'|$ 表示 \mathbf{z}' 中非零实例的数量且 $\mathbf{z}' \subseteq \mathbf{x}'$, ϕ_m 为 Shapely 值。文献[3]提出了公式 (9) 的解, 其中 $f_x(\mathbf{z}') = f(h_x(\mathbf{z}')) = E[f(\mathbf{z})|\mathbf{z}_s]$, S 是 \mathbf{z}' 中非零指标的集合, 称为 SHAP 值。

基于 SHAP 值, SHAP 框架提供了多种可视化分析手段, 能够较好地辅助决策者对影响发动机健康状态的关键因素进行分析。在局部可解释性上, 可采用力导向图 (Force Plot) 和瀑布图 (Waterfall Plot), 在全局可解释性上, 可采用概括图 (Summary Plot)、依赖图 (Dependence Plot)、热力图 (Heatmap Plot)、队列条形图 (Queue Bar Plot) 和特征聚类图 (Feature Cluster Plot) 等。

2.3 基于 EnsembleBRB-SHAP 的发动机健康状态预测及其可解释性分析流程

综合第 2.1 和 2.2 小节研究内容, 本小节提出了基于 EnsembleBRB-SHAP 的发动机健康状态预测及其可解释性分析流程, 该流程既能有效利用发动机状态监测多元不确定数据, 保证预测的准确性, 又借助 SHAP 框架对预测结果进行全局和局部解释, 增强预测结果的可信度, 并更好地辅助后续发动机健康管理等决策工作。具体流程如图 1 所示。

如图 1 所示, 基于 EnsembleBRB-SHAP 的发动机健康状态预测及其可解释性分析流程主要分为 3 步。首先是监测参数数据集的构建, 需要在收集监测数据的基础上, 采用数据预处理手段剔除异常值与缺项样本、降低数据噪声等, 将所有监测参数作为前提属性集, 预处理后的监测数据作为样本集, 为后续发动机健康状态预测建模奠定数据基础。

其次, 基于 EnsembleBRB 构建发动机健康状态

预测模型, 在抽取子置信规则库的前提属性及训练数据集的基础上对其参数进行优化, 进而对子置信规则库的结果进行集成, 预测每个样本的健康状态其详细步骤见第 2.1 小节。其中, 每个子置信规则库的前提属性从前提属性集中采用无放回的随机抽样获得, 保证每个子置信规则库的前提属性不重复; 每个子置信规则库的训练数据集从样本集中采用有放回的随机抽样抽取 (仅包含该子置信规则库的前提属性), 既可以保证每个子置信规则库之间的差异性, 同时无需额外预留样本组成测试集用于计算子置信规则库的误差, 便于确定其权重。

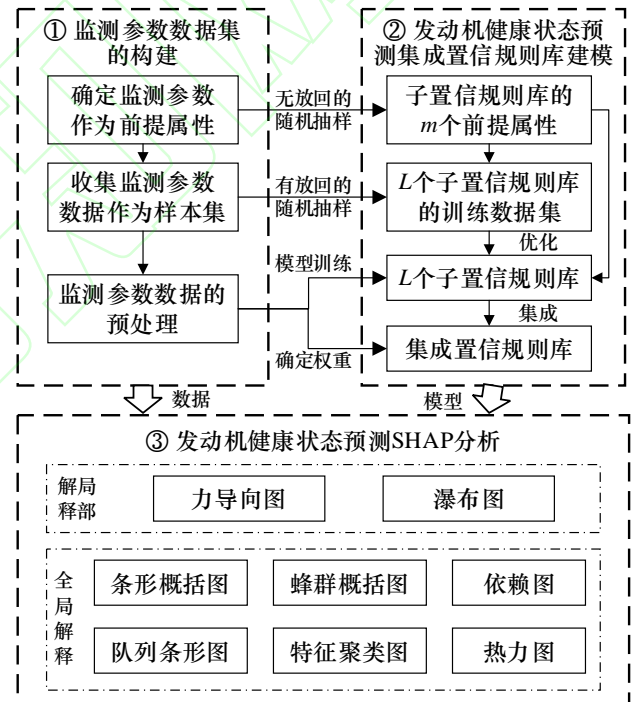


图1 基于EnsembleBRB-SHAP的发动机健康状态预测及其可解释性分析流程

Fig.1 EnsembleBRB-SHAP-based interpretable prediction and analysis process of engine health state

最后, 基于 SHAP 框架对构建的发动机健康状态预测集成置信规则库及预测结果进行可解释性分析, 首先计算监测参数的 SHAP 值, 在此基础上以可视化形式对待分析样本及样本集进行分析, 包括

局部解释和全局解释两部分，为后续发动机健康管理等决策工作提供更多依据。

3 案例研究

本节以涡轮发动机健康状态为例，验证本文所提基于 EnsembleBRB-SHAP 的航空发动机健康状态可解释预测方法的可行性与有效性。

3.1 案例背景与数据来源

本节使用的数据来源于美国埃姆斯研究中心的卓越预测中心 (Prognostics Center of Excellence, PCoE) 提供的涡轮发动机退化数据集，该数据集由商用模块化航空推进系统仿真平台 (C-MAPSS) 生成，共包含 100 个带有不同程度初始磨损的发动机在寿命周期内记录的 21 个监测参数，各传感器参数的物理含义如表 1 所示。

表 1 各传感器参数的物理含义
Table 1 Physical meaning of sensors

英文缩写	描述	单位
TFI	风扇进口温度	°R
T_LPC_O	低压压缩机出口温度	°R
T_HPC_O	高压压缩机出口温度	°R
T_LPT_O	低压涡轮机出口温度	°R
P_FI	风扇出口压力	psia
P_BD	涵道压力	psia
P_HPC_O	高压压缩机出口压力	psia
PFS	风扇转速	r/min
ECS	核心转速	r/min
EPR	压力比率 (P50/P2)	--
SP_HPC_O	高压压缩机出口静态压力	psia
R_FL_P	高压压缩机出口燃料流量与静压比	pps/psi
CFS	校正后的风扇转速	r/min
CCS	校正后的核心转速	r/min
BR	涵道比	--
BFAR	燃烧器的燃烧空气比	--
BE	排气焓	--
DFS	要求的风扇转速	r/min
CDFS	修正后要求的风扇速度	r/min
HPT_CB	高压涡轮机冷却液泄漏量	lbm/s
LPT_CB	低压涡轮机冷却液泄漏量	lbm/s

通过对样本集的简单分析可以发现，在上述 21 个监测参数中，TFI、P_EI、P_BD、EPR、BFAR、DFS、CDFS 等七个参数未发生变化，为常值，因此

仅取剩余的 14 个监测参数作为前提属性，构建样本集，并剔除异常样本。

3.2 结论与分析

采用第 2.1 节所述步骤，取 $L=20$ 构建发动机健康状态预测集成置信规则库，测试集上的均方误差为 $MSE=0.0122$ 。由于该数据集中发动机的故障模式相同，因此仅以第 1 号发动机为例，对比其真实值与基于该 EnsembleBRB 模型的预测值，如图 2 所示。

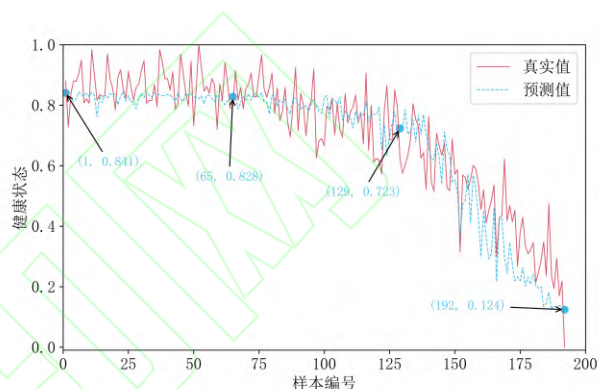


图 2 #1 号发动机健康状态真实值与预测值对比

Fig. 2 Comparison of real and predicted values of #1 engine health state

为了验证集成置信规则库在发动机健康状态预测中的有效性，本文引入了支持向量机、决策树、梯度提升回归、Lasso 算法、多层感知机等 5 种经典回归模型，以预测值与真实值的 MSE 作为衡量指标，对模型的准确性进行对比，结果如表 2 所示。可以看出，集成置信规则库具有最优的预测准确性。

表 2 六种预测方法间准确性对比
Table 2 Comparison of MSE among six methods

预测方法	MSE	预测方法	MSE
支持向量机	0.0293	Lasso 算法	0.0227
决策树	0.0136	多层感知机	0.4704
梯度提升回归	0.0244	集成置信规则库	0.0122

采用 Python 的 SHAP 库^[5]，使用 Kernel SHAP 对上述发动机健康状态预测集成置信规则库进行解释，并从样本集中选取了部分发动机的退化实例进行研究，分析影响发动机健康状态的因素。

3.2.1 局部可解释性分析

局部可解释性分析主要针对单个样本开展,挖掘影响样本健康状态预测结论的重要因素。由于本例中涉及的样本较多,本小节中仅选取部分具有代表性的样本展开局部可解释性分析,采用的分析手段包括力导向图和瀑布图。

(1) 力导向图

力导向图通过展示样本的每个监测参数对预测结果走势的影响(即导致预测结果增加或减少),来解释单个样本的预测结果是如何产生的。本文选取了在不同健康状态下的4个样本(在图2中标出)绘制了力导向图,如图3所示。

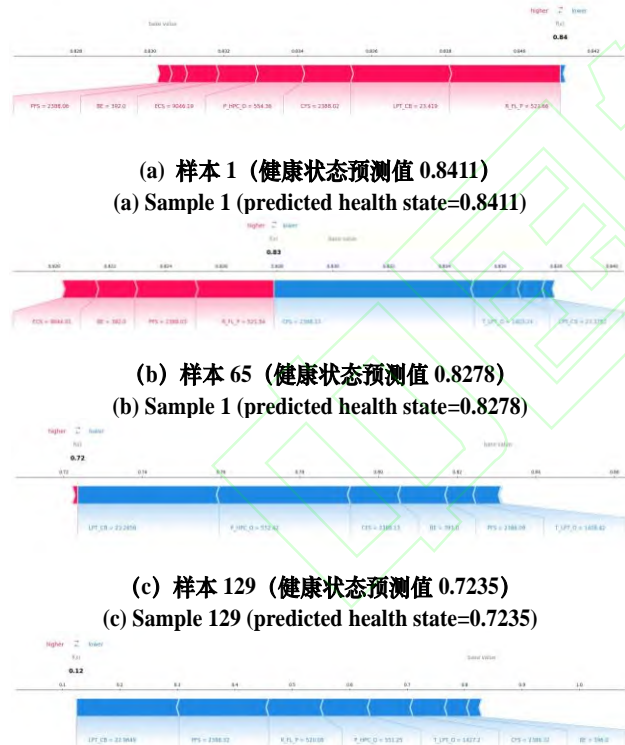


图3 #1号发动机部分样本 SHAP 力导向图

Fig. 3 SHAP force plots for #1 engine

在图3中,红色(右向箭头)表示正贡献,蓝色(左向箭头)表示负贡献,可以看出,在发动机

健康状态较好的时候(如图3(a)),大多监测参数均对健康状态预测值起到正贡献,尤其是高压压缩机出口燃料流量与静压比(R_{FL_P})和低压涡轮机冷却液泄漏量(LPT_CB)。在发动机退化初期,风扇转速(CFS)对其健康状态影响较大,如图3(b)所示。随着发动机健康状态的退化,对健康状态预测值起到负贡献的监测参数越来越多,其中 LPT_CB 的是推动发动机健康状态退化的主要因素之一。

(2) 瀑布图

瀑布图和力导向图相同,都是对单个预测的解释,其底部以模型输出的期望值开始,每行显示单个监测参数如何影响发动机健康状态的预测结果,红色(右向箭头)表示监测参数使发动机健康状态更好,而蓝色(左向箭头)表示其导致发动机健康状态更差。

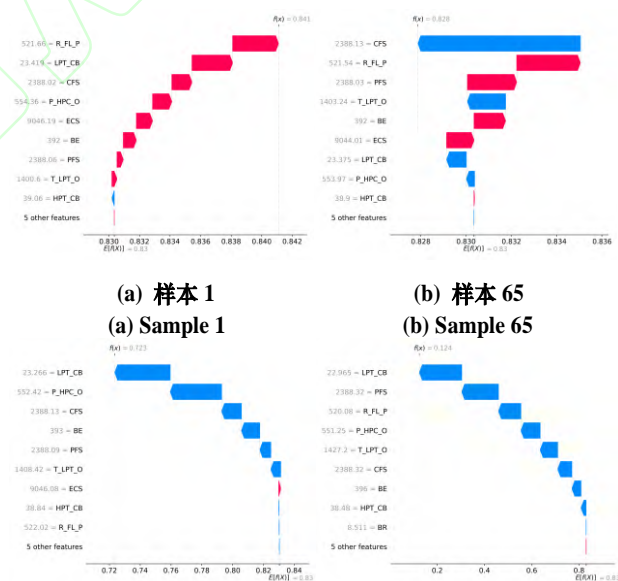


图4 #1号发动机部分样本 SHAP 瀑布图

Fig. 4 SHAP waterfall plots for #1 engine

3.2.2 全局可解释性分析

(1) 蜂群概括图

蜂群概括图是对所有样本预测情况的整体分析,通过展示监测参数分布及其与预测结果之间的关系,解释监测变量对发动机健康状态预测结果的影响。数据集中每个样本以散点形式绘制于图 5 中,纵轴表示监测参数,该监测参数的值越大,则点的颜色越红,反之越蓝。对健康状态影响力(即平均绝对 SHAP 值)越大的监测参数在纵轴的排序越靠上。沿着横轴水平检查每个监测参数的颜色分布,可以初步分析监测参数与其 SHAP 值之间的关系(具体分析需结合依赖图开展)。在本例中,低压涡轮冷却液泄漏量(LPT_CB)对发动机健康状态的影响最大,可以看到,其值越小,越对发动机健康状态产生负面影响。高压压缩机出口压力(P_HPC_O)等参数与发动机健康状态的关系也是如此。与之相反,风扇转速(PFS)等监测参数在 SHAP 值基准线左侧红色点较多,说明这些监测参数的值越大,越对发动机健康状态产生负面影响。高压压缩机出口静态压力(SP_HPC_O)等 5 个监测参数对发动机健康状态的影响较小,在后续建模中可考虑移除这些参数。

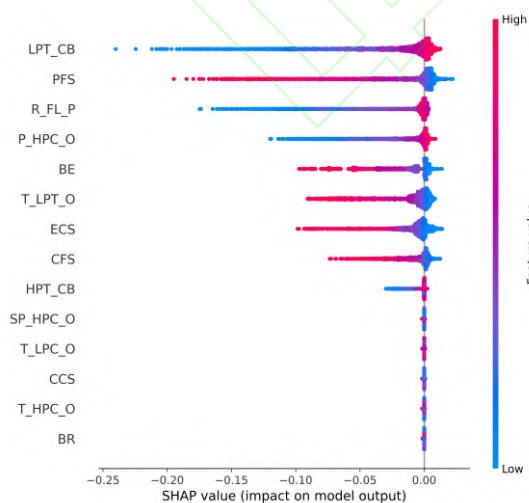


图 5 SHAP 蜂群概括图

Fig. 5 SHAP beeswarm summary plot

(2) 依赖图

为了更好地理解每个监测参数与发动机健康状态之间的关系。本文依据图 5 选取了对发动机健康状态影响程度最大的两个监测参数 LPT_CB 和 PFS、影响程度最小的监测参数 BRB 以及监测参数 BE,绘制了依赖图,如图 6 所示。

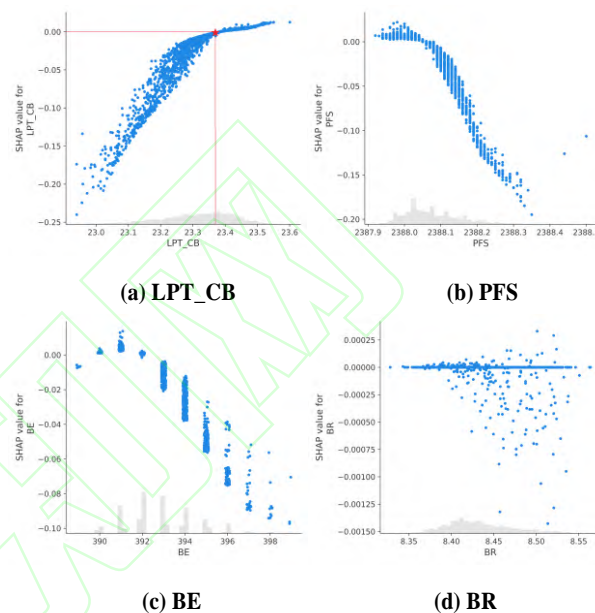


图 6 单监测参数 SHAP 依赖图

Fig. 6 SHAP dependence plot for single parameter

在图 6 (a) 和图 6 (b) 中, LPT_CB 和 PFS 很明显呈现出单调递增和单调递减的趋势, 分别表示 LPT_CB 越大, SHAP 值也越大, 对健康状态起到正向作用, PFS 与此相反, 与蜂群图中的结论是一致的。注意到图 6 (a) 中, 低于红线 $y=0$ 的点表示这些样本与健康状态下降的预测有关。图 6 (c) 中 BE 呈现出先升后降的趋势, 说明 BE 在发动机健康状态预测中存在最优取值, 高于或低于该值都将导致发动机健康状态的恶化。如图 6 (b) 和图 6 (c) 所示, 在固定监测参数值的情况下, 样本呈现垂直分布是由该参数与其他监测参数的交互关系导致的, 需要结合其他监测参数进一步分析。图 6 (d) 中大

部分样本的 SHAP 值均为 0，表示对健康状态预测结果几乎不产生影响，少量样本的分布不均匀，可能是噪声数据。

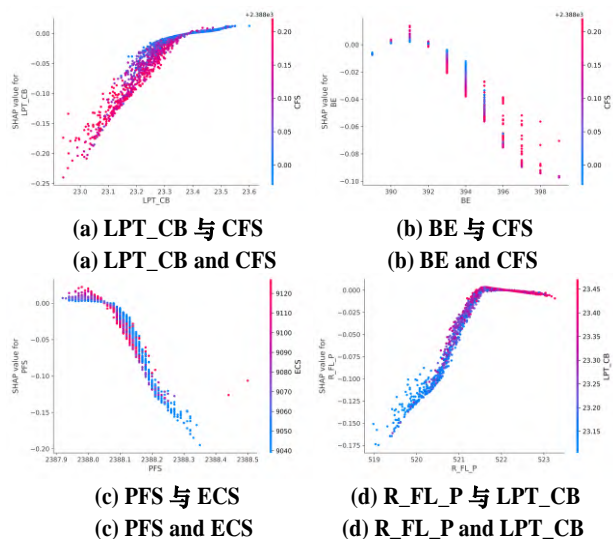


图 7 双监测参数 SHAP 依赖图
Fig. 7 SHAP dependence plot for a pair of parameter

图 7 通过双监测参数依赖图分析不同监测参数对健康状态预测结果的共同影响。图 7 中点的颜色由纵轴变量决定。从图 7 (a) 中可以看出，LPT_CB 与 CFS 呈负相关关系，当 LPT_CB 值越大、CFS 值越小时，预测得到的健康状态越好。在图 7 (b) 中，以 BE=394 为例，可以明显看出 CFS 的取值越大，对应的 SHAP 值越小，说明在固定 BE 时，较高的 CFS 与发动机健康状态的恶化有关。在图 7 (c) 中，以 PFS 的 SHAP=0 为界，当其 SHAP>0 且固定 PFS 时，ECS 的取值越大对发动机健康状态的正影响程度越高；当其 SHAP<0 且固定 PFS 时，ECS 的取值越大对发动机健康状态的负影响程度越高。图 7 (d) 可知 R_FL_P 在发动机健康状态预测时存在拐点，即健康状态与 R_FL_P 先呈现正相关性后呈现负相关性，而 LPT_CB 随着 R_FL_P 的增加而增加，即二者间呈现正相关关系。

(3) 热力图

热力图展示了全部样本在各监测参数下 SHAP 值的分布情况。在图 8 所示的热力图中，左侧纵轴为监测参数影响力排名，按从大到小排序，右侧柱状图展示了每个监测参数对模型的影响程度，颜色深浅表示 SHAP 值的大小，红色表示正影响，蓝色表示负影响，颜色越深 SHAP 绝对值越大，对模型的影响越大。由于模型预测的基准值较高，因此监测参数在多数情况下对发动机健康状态产生负影响，即在图 8 中表示为蓝色。顶部黑色曲线展示了模型的预测结果。可以看出，对模型预测负影响最大的三个参数分别为低压涡轮机冷却液泄漏量 (LPT_CB)、风扇转速 (PFS) 和高压压缩机出口燃料流量与静压比 (R_FL_P)，主要与发动机健康状态恶化有关。与图 3 和图 4 中的结论一致，一旦上述三个监测参数对应零部件发生潜在故障，发动机健康状态将明显恶化。

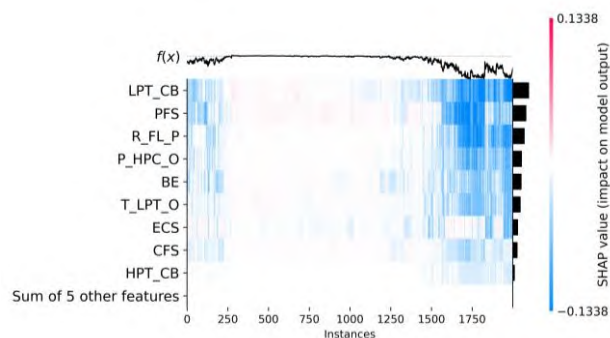


图 8 SHAP 热力图
Fig. 8 SHAP heatmap plot

(4) 队列条形图

本文依据 SHAP 库中提供的自动群组划分功能，采用决策树模型自动地选取监测参数及其取值，以对样本进行划分，即以高压压缩机出口燃料流量与静压比 (R_FL_P) 为参照，将样本分为两组，并

基于 SHAP 的平均绝对值分别对各组样本的监测参数对预测结果影响程度进行分析, 绘制 SHAP 队列条形图如图 9 所示。

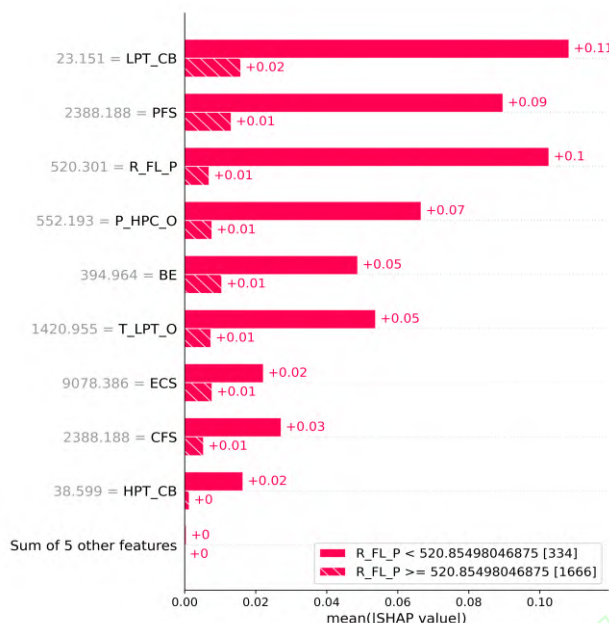


图 9 SHAP 队列条形图

Fig. 9 SHAP queue bar plot

在图 9 中, 左侧纵轴为监测参数影响力排名, 按参数影响力从大到小进行排序, 与图 5 蜂群概括图相同。在 $R_FL_P < 520.85$ 的情况下, 各监测参数对发动机健康状态的影响程度明显增加。结合样本数据可以看出, 绝大多数 $R_FL_P < 520.85$ 的样本健康状态均较差 (远低于基准值 0.83), 因此将与其他监测参数共同作用影响预测结果。

(5) 聚类图

特征聚类图通过对每组监测参数进行层次聚类, 挖掘在高度相似的监测参数。这些监测参数在预测中存在一定的冗余, 后续预测模型优化时可考虑调整监测参数避免冗余。如图 10 所示, 当聚类临界点为 0.25 时, 风扇转速 (PFS) 和校正后的风扇转速 (CFS)、核心转速 (ECS) 和校正后的核心转速 (CCS) 呈现出明显的相关性, 说明在实际预测

发动机健康状态时, 风扇转速 (PFS) 和校正后的风扇转速 (CFS)、核心转速 (ECS) 和校正后的核心转速 (CCS) 具有高度相似的数据特征, 存在冗余。结合图 9 中的队列条形图, 校正后的风扇转速 (CFS) 和校正后的核心转速 (CCS) 对模型预测准确性的影响力较小, 在构建发动机健康状态预测模型时, 为了提高建模与预测效率, 可移除这两个监测参数。

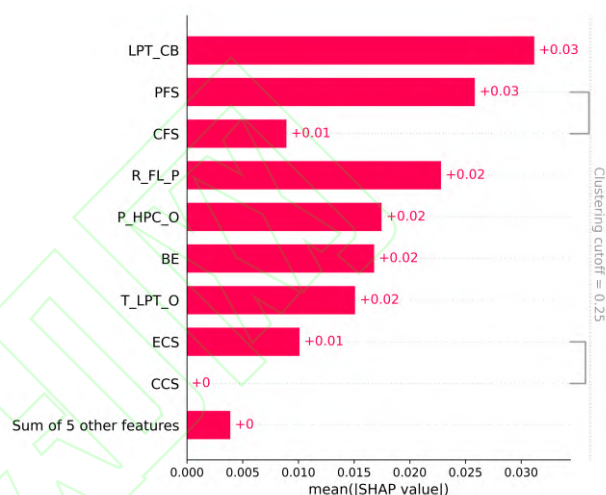


图 10 SHAP 特征聚类图

Fig. 10 SHAP feature cluster plot

3.2.3 总结

3.2.1 节中基于 SHAP 力导向图和瀑布图的单样本局部可解释性分析可以得到影响每个样本健康状态预测结果的原因, 可以看出, 在发动机即将失效时, 低压涡轮机冷却液泄漏量 (LPT_CB) 和风扇转速 (PFS) 两个监测参数起到决定性作用。

通过 3.2.2 节的全局可解释性分析可得:

1) 低压涡轮机冷却液泄漏量 (LPT_CB)、风扇转速 (PFS) 和高压压缩机出口燃料流量与静压比 (R_FL_P) 是决定发动机健康状态最关键的 3 个监测参数, 发动机日常管理应多关注其对应零部件的工作状态;

2) 高压压缩机出口静态压力 (SP_HPC_O)、

低压压缩机出口温度 (T_LPC_O)、校正后的核心转速 (CCS)、高压压缩机出口温度 (T_HPC_O) 和涵道比 (BR) 等 5 个监测参数对发动机健康状态的影响较小, 在后续预测建模中可考虑移除上述参数, 以提升建模效率。

3) 发动机健康状态与低压涡轮机冷却液泄漏量 (LPT_CB) 和高压压缩机出口压力 (P_HPC_O) 呈现出明显的正相关关系, 与风扇转速 (PFS)、低压压缩机出口温度 (T_LPT_O)、核心转速 (ECS)、校正后的风扇转速 (CFS) 呈现明显的负相关关系, 高压压缩机出口燃料流量与静压比 (R_FL_P) 和排气焓 (BE) 存在明显的最优值, 偏离最优值都将导致发动机健康状态的恶化。

4) 风扇转速 (PFS) 和校正后的风扇转速 (CFS)、核心转速 (ECS) 和校正后的核心转速 (CCS) 具有高度相似的数据特征, 可结合参数影响力考虑仅保留其中一种参数, 以提升建模效率。

4 结束语

本文提出了一种具备可解释性的航空发动机健康状态预测方法, 该方法融合了集成置信规则库和 SHAP 解释框架, 能够有效利用航空发动机运行过程中积累的多元监测数据构建预测模型, 通过信度结构描述数据不确定性, 并基于多个子置信规则库的集成保证了预测的准确性。在此基础上, 通过 SHAP 框架对发动机健康状态预测开展局部解释与全局解释, 能够有效挖掘影响发动机健康状态的关键因素, 指导后续发动机健康管理及预测模型调整优化等决策工作。

应用本文所提方法在 C-MAPSS 仿真数据集上

开展了发动机健康状态预测的案例研究, 结果表明, 该方法既能通过数据驱动的 EnsembleBRB 预测建模实现多元发动机监测参数的有效利用, 保证预测的客观性与准确性, 又能通过基于 SHAP 框架的局部与全局可解释性分析, 确定影响发动机健康状态的关键监测参数以及建模中存在的冗余参数、掌握监测参数与发动机健康状态的关联关系等, 为该型号发动机健康管理提供决策支撑。

良好的可解释性使 EnsembleBRB-SHAP 方法同样适用于装备健康管理领域的其他问题中, 后续将在更多复杂非线性系统的分析与决策实践中进一步验证所提方法的有效性与可拓展性。

参考文献

- [1] ZHANG Y, XIN Y, LIU Z, et al. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE[J]. Reliability Engineering & System Safety, 2022, 220: 108263.
- [2] CUBILLO A, PERINPANAYAGAM S, ESPERON-MIGUEZ M. A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery[J]. Advances in Mechanical Engineering, 2016, 8(8).
- [3] VICHARE N M, PECHT M G. Prognostics and health management of electronics[J]. IEEE Transactions on components and packaging technologies, 2006, 29(1): 222-229.
- [4] YOU Y, SUN J, CHEN Y, et al. Ensemble belief rule-based model for complex system classification and prediction[J]. Expert Systems with Applications, 2021, 164: 113952.
- [5] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[J]. Advances in neural information processing systems, 2017, 30: 4765-4774.
- [6] YANG J B, LIU J, WANG J, et al. Belief rule-base inference methodology using the evidential reasoning approach-RIMER[J]. IEEE

- Transactions on systems, Man, and Cybernetics-part A: Systems and Humans, 2006, 36(2): 266-285.
- [7] 韩文策, 康潇, 李红宇, 等. 基于人在回路策略的在线置信规则库疾病诊断方法[J]. 计算机工程, doi: 10.19678/j.issn.1000-3428.0068838.
HAN W C, KANG X, LI H Y, et al. A novel online belief rule base method with human-in-the-loop strategy for disease diagnosis[J]. Computer Engineering, doi: 10.19678/j.issn.1000-3428.0068838.
- [8] SHAPLEY L. A value for n-person games. Contributions to the theory of games II[M]. New York, USA: Cambridge university press, 1953.
- [9] MANGALATHU S, HWANG S H, JEON J S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach[J]. Engineering Structures, 2020, 219: 110927.
- [10] 武宇翔, 韩肖清, 牛哲文, 等. 基于变权重随机森林的暂态稳定评估方法及其可解释性分析[J]. 电力系统自动化, 2023, 47(14): 93-104.
WU Y X, HAN X Q, NIU Z W, et al. Transient stability assessment method based on variable weight random forest and its interpretability analysis[J]. Automation of Electric Power Systems, 2023, 47(14): 93-104. (in Chinese)
- [11] SAXENA A, GOEBEL K, SIMON D, et al. Damage propagation modeling for aircraft engine run-to-failure simulation[C]//2008 international conference on prognostics and health management. IEEE, 2008: 1-9.
- [12] HANACHI H, LIU J, BANERJEE A, et al. A physics-based modeling approach for performance monitoring in gas turbine engines[J]. IEEE Transactions on Reliability, 2014, 64(1): 197-205.
- [13] KIM S, IM J H, KIM M, et al. Diagnostics using a physics-based engine model in aero gas turbine engine verification tests[J]. Aerospace Science and Technology, 2023, 133: 108102.
- [14] 于拓舟, 王宇, 张建锐, 等. 基于 AMESim 发动机物理模型的预测性研究[C]// 仿真与试验技术大会论文集, 2016:1-5.
YU T Z, WANG Y, ZHANG J R, et al. Predictive research of engine physical model based on AMESim[C]//2016 Siemens PLM Software, 2016:1-5. (in Chinese)
- [15] SUN X, TAN J, WEN Y, et al. Rolling bearing fault diagnosis method based on data-driven random fuzzy evidence acquisition and Dempster-Shafer evidence theory[J]. Advances in Mechanical Engineering, 2016, 8(1).
- [16] LI Z, GOEBEL K, WU D. Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning[J]. Journal of Engineering for Gas Turbines and Power, 2019, 141(4): 041008.
- [17] CHENG Y, ZENG J, WANG Z, et al. A health state-related ensemble deep learning method for aircraft engine remaining useful life prediction[J]. Applied Soft Computing, 2023, 135: 110041.
- [18] 崔建国, 高波, 蒋丽英, 等. LSSVM 与 HMM 在航空发动机状态预测中的应用研究[J]. 计算机工程, 2017, 43(10): 310-315.
CUI J G, GAO B, JIANG Y L, et al. Application Research of LSSVM and HMM in Aeroengine Condition Prediction[J]. Computer Engineering, 2017, 43(10): 310-315.
- [19] SKORDILIS, E., MOGHADDASS, R. A deep reinforcement learning approach for real-time sensor-driven decision making and predictive analytics[J]. Computers & Industrial Engineering, 2020, 147: 106600.
- [20] VISHNU, T. V., NARENDHAR GUGULOTHU, PANKAJ MALHOTRA, et al. Bayesian networks for interpretable health monitoring of complex systems[C]//Workshop on AI for Internet of Things at IJCAI. 2017.
- [21] GAO JIAHAO, YOUREN WANG, and ZEJIN SUN. An interpretable RUL prediction method of aircraft engines under complex operating conditions using spatio-temporal features[J]. Measurement Science and Technology, 2024, 35(7): 076003.
- [22] ZHOU ZHIJIE, YOU CAO, GUANYU HU, et al. New health-state assessment model based on belief rule base with interpretability[J]. Science China Information Sciences, 2021, 64(7): 172214.
- [23] SACHAN S, YANG J-B, XU D-L. Global and

- Local Interpretability of Belief Rule Base[C]// Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020). 2020:68–75.
- [24] CAO Y, ZHOU Z, HU C, et al. On the Interpretability of Belief Rule-Based Expert Systems[J]. IEEE Transactions on Fuzzy Systems. 2020, 29 (11): 3489–3503.
- [25] RIBEIRO M T, SINGH S, GUESTRIN C. Why should I trust you? Explaining the predictions of any classifier[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016: 1135-1144.
- [26] SHRIKUMAR A, GREENSIDE P, KUNDAJE A. Learning important features through propagating activation differences[C]//International conference on machine learning. PMIR, 2017: 3145-3153.
- [27] BACH S, BINDER A, MONTAVON G, et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation[J]. PloS one, 2015, 10(7): e0130140.
- [28] LIPOVETSKY S, CONKLIN M. Analysis of regression in game theory approach[J]. Applied stochastic models in business and industry, 2001, 17(4): 319-330.
- [29] STRUMBELJ E, KONONENKO I. Explaining prediction models and individual predictions with feature contributions[J]. Knowledge and information systems, 2014, 41: 647-665.
- [30] DATTA A, SEN S, ZICK Y. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems[C]// IEEE symposium on security and privacy (SP). IEEE, 2016: 598-617.
- [31] MANGALATHU S, HWANG S H, JEON J S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach[J]. Engineering Structures, 2020, 219: 110927.