

基于无标签、不均衡、初值不确定数据的 设备健康评估方法

王村松, 陆宁云[†], 程月华, 姜 斌

(南京航空航天大学 自动化学院, 南京 211106)

摘 要: 缺少先验知识和完备信息的设备健康评估一直是预测与健康管理 (PHM) 领域的难点问题. 针对设备运行状态观测数据的无标签、不均衡、初值不确定性问题, 提出一种多变量深度森林的设备健康评估方法. 首先, 提出一种基于相关性指标和趋势性指标的特征选择方法以去除冗余特征; 然后, 利用三维数据标准化和量子模糊聚类方法, 动态设定设备健康状态并且解决数据初值的不确定问题; 最后, 采用一种多变量深度森林分类器实现设备健康状态的离线训练与在线评估. 案例分析结果验证了所提出的健康评估方法的有效性和可行性.

关键词: 健康评估; 深度森林; 量子模糊聚类; 特征提取; 特征选择

中图分类号: TP206.3

文献标志码: A

DOI: 10.13195/j.kzyjc.2018.1493

开放科学(资源服务)标识码(OSID):



引用格式: 王村松, 陆宁云, 程月华, 等. 基于无标签、不均衡、初值不确定数据的设备健康评估方法[J]. 控制与决策, 2020, 35(11): 2687-2695.

Equipment health risk assessment based on unlabeled, unbalanced data under uncertain initial condition

WANG Cun-song, LU Ning-yun[†], CHENG Yue-hua, JIANG Bin

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Equipment health risk assessment without prior knowledge and complete observation information is still a hard problem in the prognostic and health management (PHM) field. To solve the problems caused by unlabeled, unbalanced data under uncertain initial condition upon the health states of engineering equipment, a multivariate deep forest based health assessment method is proposed. Firstly, a correlation and trend metric based feature selection strategy is proposed to remove redundant features. Then, the three-dimensional data is unfolded to be a two-dimensional data, and a quantum fuzzy clustering is utilized to define the health states, which can overcome the problem of initial condition uncertainty. Finally, a multivariate deep forest classifier is adopted for offline training and online health assessment. Case study results show that the proposed health assessment method is effective and feasible.

Keywords: health assessment; deep forest; quantum fuzzy clustering; feature extraction; feature selection

0 引 言

航空航天装备结构精密复杂、对安全性和可靠性要求极为苛刻, 在极端使用环境下长时间连续工作过程中, 装备性能不可避免地产生退化, 极易引发各类故障, 威胁系统安全. 以健康评估和故障预测为核心的预测与健康管理的 (prognostics and health management, PHM) 技术, 是实现装备预测维护、保障系统可靠性的关键, 已成为航空航天领域的关注焦

点.

健康评估的主要目的是评估一个组件, 一个子系统或整个系统当前的健康状态^[1]. 现有的健康评估方法可分为 3 类: 基于模型的方法^[2-3]、数据驱动的方法^[4-5]和混合方法^[6-7]. 数据驱动的方法不需要退化过程的精确解析模型, 通过对设备运行数据直接处理与分析, 挖掘出数据背后隐藏的设备健康状态信息, 为故障预测与维护决策提供可靠信息. 这类方法适

收稿日期: 2018-10-31; 修回日期: 2019-04-11.

基金项目: 国家自然科学基金项目 (61873122, 61673206); 装备预研国防科技重点实验室基金项目 (61422080307); 航天器在轨故障诊断与维修重点实验室基金项目; 江苏省研究生科研与实践创新计划项目 (KYCX18_0300).

责任编辑: 方华京.

[†]通讯作者. E-mail: luningyun@nuaa.edu.cn.

合于系统极其复杂、难以建立退化机理模型、且拥有丰富监测数据的航空航天装备。

数据驱动的健康评估方法又可细分为3类^[8]:单变量方法^[9]、直接方法^[10]和多变量方法^[11]。单变量方法首先建立一个综合健康度指标(SHI),通过与事先设定的系统不同健康阶段的阈值相比,确定系统的实时健康状态,SHI构建和阈值确定这两个步骤极度依赖于应用对象的领域知识,是这类方法在实际应用中的掣肘问题;直接方法则是通过将当前样本与数据库中的历史样本进行相似度匹配,寻找出最为相似的样本作为参照进行健康评估,这类方法在样本充足完备的条件下,具有良好的健康评估性能和方法普适性;多变量方法不需要构建SHI,不需要预先设定健康度阈值,也不需要大量历史样本,直接利用设备运行数据,提取与系统健康状态退化相关的特征量,通过聚类方法对设备健康状态的特征量进行分类,进而获知设备所处的健康阶段信息,该方法可克服单变量方法和直接方法的不足或约束,近年来颇受PHM领域的关注。

现有的多变量方法需事先设定健康状态的类别个数,然后动态地评估设备所处的健康状态,然而很少涉及设备各个健康阶段及切换过程中存在的不确定性^[12]。多变量健康评估方法仍处于发展阶段,在逐步推向实际应用的过程中,仍存在一些挑战性问题^[8,13-14]。例如:在数据来源方面,如何优化配置传感器获取富含设备健康退化信息的高质量观测数据?如何有效融合多源异构传感信息?在数据处理方面,如何提取有效的健康退化特征?如何处理大量无标签、不均衡的观测数据?在数据应用方面,如何处理观测数据的多源不确定性?如何剔除系统工况或工作模式的影响?

本文主要针对观测数据的不均衡问题、无健康状态标签问题和健康状态初值不确定和阈值不确定问题,在现有多变量方法框架下提出一种数据驱动的健康评估新方法。主要的创新点在于:1)将多个样本的数据描述为不等长三维矩阵并统一标准化处理,解决健康状态初值不确定问题;2)利用量子模糊聚类,解决观测数据的无标签、不均衡问题;3)采用深度森林分类算法,实现设备所处不同健康阶段的在线概率型健康评估。

1 问题描述

1.1 观测数据的无标签、不均衡问题提出

用于系统健康评估的观测数据理论上可分成4类^[15]:有标签、均衡数据,有标签、不均衡数据,无标

签、均衡数据和无标签、不均衡数据。有标签,是指已知系统观测数据所对应的健康状态。一般假定一个系统或装备的健康状态分为4个阶段(即4个健康状态标签):完全健康、轻微损伤、严重损伤、失效阶段。轻微损伤一般定义为器件类失效并造成系统余度下降,而严重损伤可定义为部件失效并导致余度丧失。均衡数据,是指设备全寿命周期运行过程中,不同健康阶段中采集的观测数据数量比较接近,近似均匀分布。

通常,对于可进行加速寿命实验的一些简单部件,可以得到有标签、均衡的观测数据。但对于实际应用中的复杂工程设备,观测数据大都是无标签、不均衡数据。这是因为:一方面,人们很难事先得到某个观测时刻下准确的设备健康状态信息;另一方面,系统健康状态随时间的退化过程一般呈非线性,观测数据中处于完全健康或轻微损伤状态下的数据量远远高于严重损伤和临近失效状态下的数据,导致数据分布的不均衡性。

无标签问题导致一类有监督数据建模或机器学习算法很难直接适用于系统健康评估领域;而不均衡问题会导致数量多的数据类淹没掉数量少的数据类,使得传统分类或聚类算法效果不理想。目前,不均衡分类问题可以通过人为调整数据分布或采用数据重采样技术得以解决,本文中使用的随机森林算法是基于bagging的集成学习算法,而bagging本质上是一种数据自助重采样技术。

1.2 健康评估过程的多源不确定性

健康评估方法在实际应用中面临的最大问题是多源不确定性问题^[16-18],不确定性因素主要包括:

1) 系统工作参数的不确定性:不同工作任务或工作模式导致的系统物理参数值的不确定性。

2) 系统模型的不确定性:因系统机理不明、全局强非线性以及模型简化等原因导致的系统模型不确定性。

3) 退化过程的不确定性:实际中,即使是在相同条件下完成的加速寿命实验,所产生的观测数据往往也具有不同的退化趋势。

4) 观测数据的不确定性:观测量在测量过程中存在各种不确定性。

5) 初始状态和观测时间的不确定性:对于连续的健康评估过程,存在不同样本设备的初始健康状态不一致性(如图1);对于离散的健康评估过程,存在不同设备样本的各个观测时刻不一致性。

6) 健康状态不同阶段的阈值不确定性:实际中,

健康阶段阈值不是固定不变的,通常会随着时间而改变.

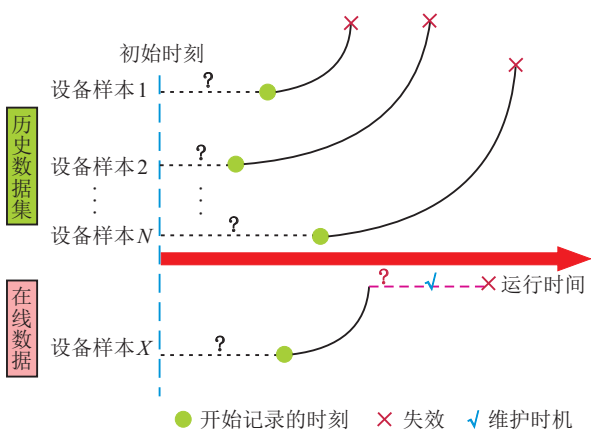


图 1 观测样本初始健康状态不确定性问题

上述罗列的多种不确定性中,2)和3)主要影响基于模型的健康评估方法,4)~6)主要影响基于数据的健康评估方法.由于本文提出的方法属于数据驱动中的多变量方法,主要考虑解决5)中的初始健康状态不确定性和6)中的阈值不确定性问题,对测量本身带来的不确定性暂不作考虑.

2 一种数据驱动的多变量健康评估方法

2.1 设备性能退化的特征选择

假定有 I 个设备样本的全寿命周期观测数据,每个设备样本有 J 个相同的观测变量(传感器输出的变量信号),每个设备样本的观测数据个数为 $K_i (i = 1, 2, \dots, I)$ (若为连续观测,则 K_i 等同于设备运行寿命).因此,每个设备样本的观测数据可用二维矩阵 $X_i (J \times K_i)$ 表示, I 个历史样本的数据可组成一个不等长的三维数据矩阵 $X (I \times J \times K)$,如图 2 所示.

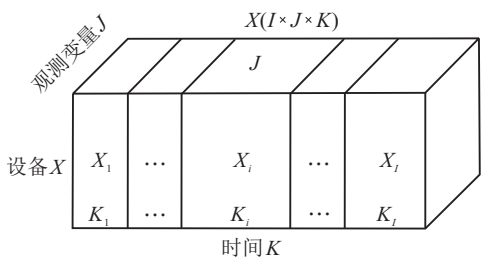


图 2 全寿命周期设备观测数据的三维表示

为了更好地描述设备性能退化,以开展健康评估研究,理想的性能退化特征需具有同类个体普适性、性能退化一致性等.因此,在对数据进行平滑处理的基础上,本文利用相关性和趋势性两种指标进行特征选择.相关性指标反映了特征与时间的线性相关程度,一定程度上体现了同类个体普适性.趋势性指标可以刻画出相似设备间退化特征的性能退化一致性.

1) 相关性指标.

本文采用斯皮尔曼等级相关系数.较其他两种相关分析方法(皮尔森和肯德尔相关性系数),斯皮尔曼等级相关系数法是一种无监督的方法,不在乎两个变量的总体分布形态和样本容量的大小.设备特征与时间之间的相关性的计算公式为

$$\text{Corr}(i, j) = \frac{6 \sum_{k=1}^{K_i} r^2(j(k))}{K_i^3 - K_i}. \quad (1)$$

其中: $\text{Corr}(i, j)$ 是特征序列 $j(K_i)$ 与其序列长度 $k = 1, 2, \dots, K_i$ 的斯皮尔曼相关系数; $j(K_i)$ 是设备 i 的第 j 个观测变量; $r^2(j(k))$ 是秩次的差值的平方.

相关性指标取值在 $[-1, 1]$ 内,正值代表与时间正相关,负值代表与时间负相关.相关性指标绝对值越大,其与时间的线性相关性越大.

2) 趋势性指标.

本文定义了观测变量 j 在不用样本间的趋势性指标 $\text{Tre}(j)$,即

$$\text{Tre}(j) = \frac{\sum_{i=1}^I \varepsilon(\text{Corr}(i, j))}{I}$$

$$\varepsilon(x) = \begin{cases} 1, & x > 0; \\ 0.5, & x = 0; \\ 0, & x < 0, \end{cases} \quad (2)$$

若 $\varepsilon(x)$ 取值为 0 或 1,则退化趋势一致;若其取值为 0.5,则代表变量恒定不变.

根据相关性指标和趋势性指标可以实现设备性能退化的特征选择.如果观测变量 j 满足式(3)和(4),则观测变量 j 被保留下来,否则被剔除.

$$|\text{Corr}(i, j)| \geq \theta, \quad (3)$$

$$\text{Tre}(j) = 0 \text{ or } 1, \quad (4)$$

其中 θ 是设定的阈值,本文取 $\theta = 0.5$.

2.2 健康状态退化特征的量子模糊聚类

针对退化数据存在的无标签、不均衡和不确定问题,本文提出一种退化特征的健康状态分类方法,主要包括三维数据的标准化和基于量子模糊聚类的健康状态标签定义.

2.2.1 三维数据标准化

经过 2.1 节的特征选择,假设设备的观测变量由 J 个减少到 F 个,此时可以重新定义三维数据矩阵为 $X (I \times F \times K)$.为了解决观测样本初始健康状态的不确定性问题,本文将 $X (I \times F \times K)$ 按 F 展开为二维数据 $X (Z \times F) (Z = K_1 + K_2 + \dots + K_I)$,再将数据统一标准化,得到数据集 $\tilde{X} (Z \times F)$.经过标准化处

理的数据再利用量子模糊聚类进行健康标签定义时,就可以完全消除初值不确定带来的影响.

$$\tilde{x}_{zf} = \frac{x_{zf} - \bar{x}_f}{s_f}. \quad (5)$$

其中: $s_f = \sqrt{\frac{1}{Z-1} \sum_{z=1}^Z (x_{zf} - \bar{x}_f)^2}$ 是展开的二维数据 $X(Z \times F)$ 的方差, \bar{x}_f 为展开的二维数据 $X(Z \times F)$ 的均值, $\bar{x}_f = \frac{1}{Z} \sum_{z=1}^Z x_{zf}$.

2.2.2 基于量子模糊聚类的健康标签定义

针对观测数据存在的无标签和不均衡问题,本文将量子聚类理论^[19]与模糊聚类理论^[20]相结合提出量子模糊聚类,对标准化后的特征数据集 $\tilde{X}(Z \times F) = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_Z\}^T$ ($\tilde{x}_z = \{\tilde{x}_{z1}, \tilde{x}_{z2}, \dots, \tilde{x}_{zF}\}$) 进行健康标签定义,用以解决无标签问题.同时,量子模糊聚类将观测数据的标签细分为多个,用以解决数据的不均衡问题.

相对于硬分类方法,量子模糊聚类是一种不需事先指定类别数的软分类方法,聚类算法本身也可以消除阈值不确定性给后续健康评估带来的影响.因此,基于量子模糊聚类的健康标签定义过程可以总结如下:

step 1: 通过输入的数据集 $\tilde{X}(Z \times F)$, 利用带有 Parzen 窗的高斯核函数估计波函数 ψ , 进而得到势能函数 $U(\tilde{x})$, 即

$$U(\tilde{x}) = E + \frac{(\sigma^2/2)\nabla^2\psi}{\psi} = E - \frac{d}{2} + \frac{1}{2\sigma^2\psi} \sum_{z=1}^Z \|\tilde{x} - \tilde{x}_z\|^2 \exp\left[-\frac{\|\tilde{x} - \tilde{x}_z\|^2}{2\sigma^2}\right]. \quad (6)$$

其中: $E = d/2$ 是 Hamilton 算子特征值, d 是 Hamilton 算子可能的最小特征值, 可以用样本的数据维数表示; σ 是波函数宽度调节参数. 之后通过判断势能函数的局部极小值点得到观测数据的类别数量 C .

step 2: 根据模糊理论将分类问题看作下式中目标函数 L 的极值求解问题:

$$y = \min_{\{m_c, \mu_c\}} \{L\},$$

$$L = \sum_{c=1}^C \sum_{z=1}^Z (\mu_c(\tilde{x}_z))^b \|\tilde{x}_z - m_c\|^2, 1 < b < \infty,$$

$$\sum_{c=1}^C \mu_c(\tilde{x}_z) = 1, z = 1, 2, \dots, Z. \quad (7)$$

其中: b 是模糊系数, $\mu_c(\tilde{x}_z)$ 是 \tilde{x}_z 对类别 c 的隶属程度, m_c 是类别的中心值, 且有

$$m_c = \frac{\sum_{z=1}^Z (\mu_c(\tilde{x}_z))^b \tilde{x}_z}{\sum_{z=1}^Z (\mu_c(\tilde{x}_z))^b},$$

$$\mu_c(\tilde{x}_z) = \frac{1}{\sum_{c=1}^C \left(\frac{\|\tilde{x}_z - m_c\|}{\|\tilde{x}_z - m_c\|} \right)^{\frac{2}{b-1}}}. \quad (8)$$

求解该极值问题, 先构造拉格朗日函数, 令 L 对 m_c 和 $\mu_c(\tilde{x}_z)$ 的偏导为 0, 可以得到 m_c 和 $\mu_c(\tilde{x}_z)$ 的迭代公式.

step 3: 利用 m_c 和 $\mu_c(\tilde{x}_z)$ 的迭代公式直到目标函数收敛.

step 4: 输出观测数据的健康标签 Tag.

2.3 基于深度森林的健康评估方法

2.3.1 主要思想

图3给出了多变量深度森林健康评估方法的框架, 可以分成离线训练和在线评估两部分. 在离线训练阶段, 利用具有全寿命周期并且带有健康标签(由2.2.2节得到)的数据集进行训练, 然后保存训练结果好的深度森林结构以便于直接用于在线评估阶段. 设备的在线评估阶段, 先将当前设备的运行数据(从开始记录到当前时刻, 非全寿命周期)用2.2.1节中的均值 \bar{x}_f 和方差 s_f 进行数据标准化, 然后输入离线训练阶段保存好的深度森林结构中, 得到当前设备的健康评估结果.

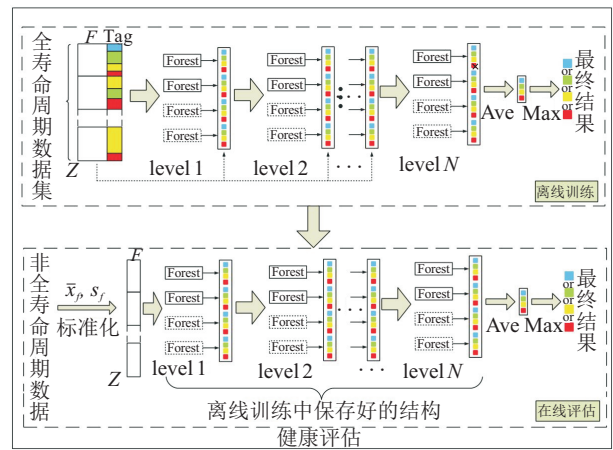


图3 基于深度森林的健康评估框架^[21]

2.3.2 基于深度森林的健康评估

类似于深度神经网络 (DNN) 层数的扩展, 深度森林^[22]也采用了 level 1、level 2 到 level N 这种森林的级联结构. 在每一层中包括了两个完全随机森林^[23](图3中实线部分)和两个随机森林^[24](图3中虚线部分), 完全随机森林和随机森林作为一种集成算法在分类过程中可以解决观测数据的不均衡问题.

本文采用深度森林算法是因为其相对于DNN方法具有超参数少、模型对超参数调节不敏感、根据结果精度自动扩展级联层数、可以用于小规模数据等优势。此外,深度森林算法的一个重要优势是能够给出分类结果的概率值,在健康评估中可以刻画不确定性。深度森林算法中,决策树可以在特征空间中不断划分子空间,并且为每个子空间打上标签,在叶节点能够得到训练样本中不同类别的概率分布。最后,在每个森林中通过对所有决策树的各类比例取平均,输出整个森林中各类的比例,图3中的最终结果是取各类中比例最大的值。

基于深度森林的健康评估方法训练过程如图4所示。

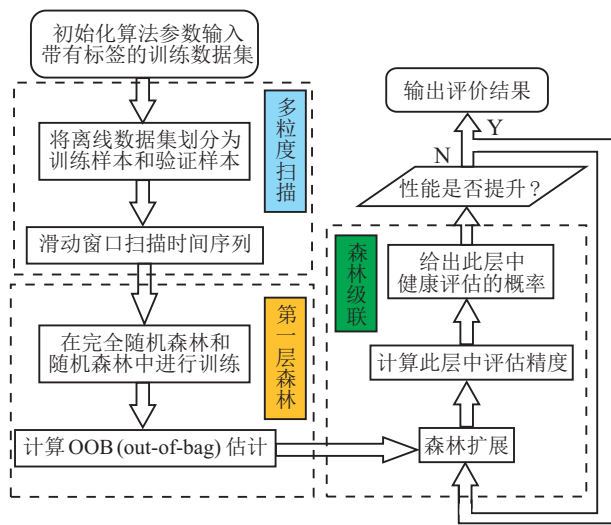


图 4 基于深度森林的健康评估算法训练过程

首先,将输入的训练集划分成训练样本和验证样本,并且类似于卷积神经网络(CNN),利用滑动窗口对数据进行扫描切割;然后,在第1层中利用完全随机森林和随机森林进行训练,并且计算OOB(out-of-bag)估计值;最后,进行森林级联层次的扩展,计算健康评估精度并且给出相对应的类别概率信息,如果评估精度有所提高则自动继续进行层次扩展,否则终止层次扩展输出最终结果。本文将常用的分类准确率作为算法的评价指标。此外,由于深度森林模型对超参数调节不敏感,参照文献[22]一些主要的参数可以设置如下:

$$n_cascade\ RF = 2, n_cascade\ RF\ tree = 101,$$

$$cascade_test_size = 0.2, tolerance = 0.$$

其中: $n_cascade\ RF$ 代表每个级联层完全随机森林或随机森林中的数量, $n_cascade\ RF\ tree$ 代表每个级联层中单个完全随机森林或随机森林包含的完全随机决策树或随机决策树的数量, $cascade_test_size$ 代表

验证样本所占训练样本的比例大小, $tolance$ 代表级联层扩展的精度差。

3 案例分析

3.1 数据描述

本文以航空发动机为例,验证所提出的设备健康评估算法。航空发动机数据源自NASA故障预测研究中心公布的民用航空发动机全寿命周期数据集^[25]。此发动机数据集是在推进系统仿真平台C-MAPSS上大量模拟某涡扇发动机机性主要部件(风扇、低压压气机、高压压气机、高压涡轮、低压涡轮)退化过程而产生,发动机主要部件简图如图5所示。

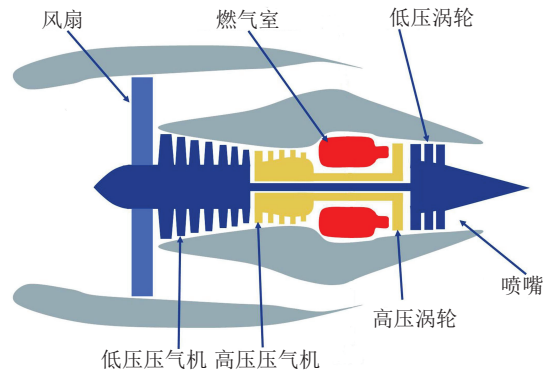


图 5 发动机主要部件^[26]

本文采用的“FD001.txt”中的数据集合包含了3个操作状态和21个传感器观测变量(见表1),模拟了100个相似发动机高压压气机性能退化故障。其中数据中包含有噪声并且观测样本的初始健康状态是不确定的。

表 1 传感器数据集描述

序号	符号	描述	单位
1	T_2	风扇入口总温	$^{\circ}R$
2	T_{24}	低压压气机出口总温	$^{\circ}R$
3	T_{30}	高压压气机出口总温	$^{\circ}R$
4	T_{50}	低压涡轮出口总温	$^{\circ}R$
5	P_2	风扇入口压力	psia
6	P_{15}	外涵总压	psia
7	P_{30}	高压压气机出口总压	psia
8	N_f	风扇物理转速	rpm
9	N_c	核心机物理转速	rpm
10	epr	发动机压比 (P_{50}/P_2)	—
11	P_{s30}	高压压气机出口静压	psia
12	ϕ	燃油流量与 P_{30} 比值	pps / psi
13	NR_f	风扇换算转速	rpm
14	NR_c	核心机换算转速	rpm
15	BPR	涵道比	—
16	$farB$	燃烧室燃气比	—
17	$htBleed$	引气焓值	—
18	N_f_dmd	设定风扇转速	rpm
19	$PCNfR_dmd$	设定核心机换算转速	rpm
20	W_{31}	高压涡轮冷却引气流量	lbm / s
21	W_{32}	低压涡轮冷却引气流量	lbm / s

3.2 数据预处理

根据3.1节描述,航空发动机退化数据集表示为 $X(100 \times 21 \times K)$,下面给出发动机1中部分传感器采集的原始数据和数据经平滑处理后的效果图.此退化数据集大致可分为3类:恒定不变(图6(a))、具有上升趋势(图6(b))和具有下降趋势(图6(c)).

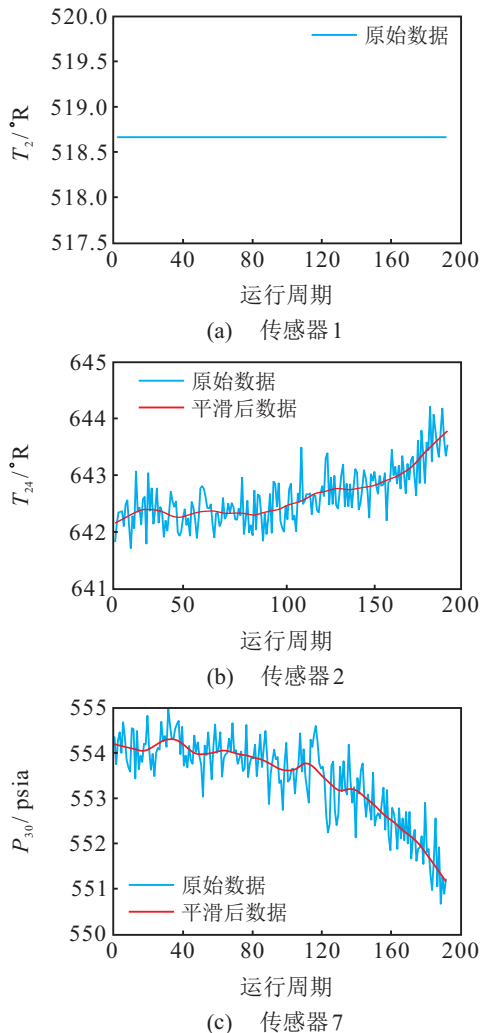


图6 发动机1数据平滑示意图

首先,利用式(1)分别计算各传感器变量与运行周期的相关性指标;然后,利用式(2)计算各变量的单调性指标;最后,满足式(3)和(4)的传感器观测变量4、7、11、12、15、20和21被保留下来,数据可表示为 $X(100 \times 7 \times K)$.

图7给出了部分具有代表性的传感器变量的相关性分析示意图.图7(a)传感器1采集的数据一直保持恒定,所以相关性指标一直为0.5,被剔除;图7(b)传感器4相关性指标一直为正且都大于0.5,代表其各发动机中传感器4采集的数据与运行周期呈正相关,满足单调性,故最终被保留下来;图7(c)传感器9相关性指标在不同发动机中正负不定,不满足单调性,被剔除;图7(d)传感器13相关性指标在有的发动机值

小于0.5,不满足设定的阈值,被剔除.

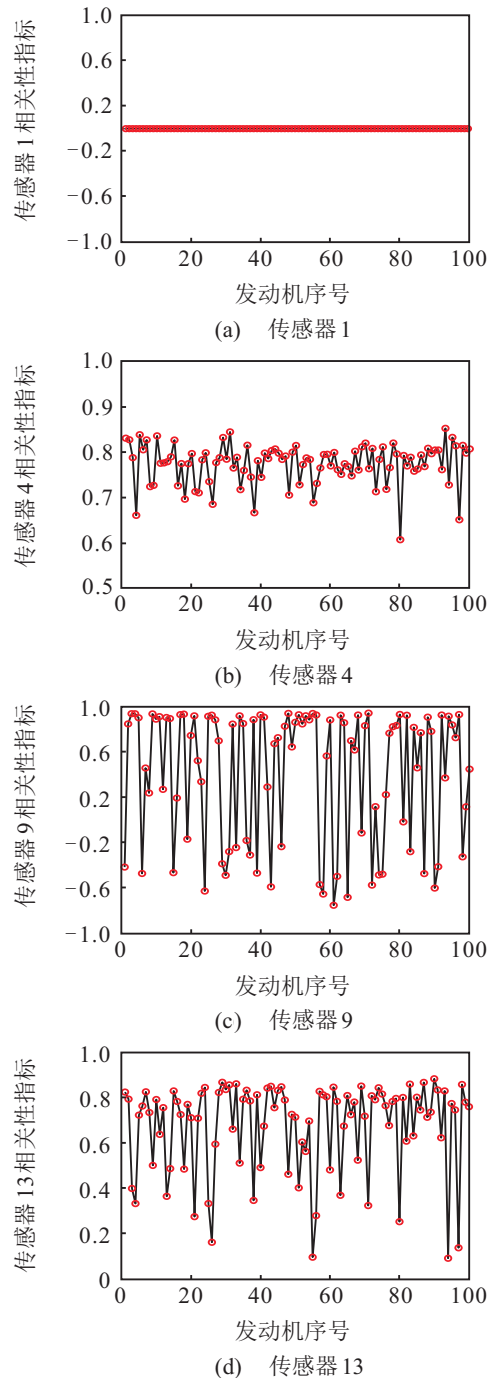


图7 不同传感器变量相关性分析

3.3 健康状态分类与评估

退化特征数据集 $X(100 \times 7 \times K)$ 符合图1所示的初始状态不确定特点,并且具有不平衡、无标签特性.然而,现有部分健康评估方法^[27]通常忽略了观测样本初始健康状态存在的不确定性,认为所有观测样本一开始都处于健康阶段,往往会降低不同健康阶段的评估结果精度.

本节首先对退化特征数据集按变量展开,进行统一标准化,解决初始状态不确定问题,经过多次测试选择量子模糊聚类的波函数宽度调节参数 $\sigma = 0.6$,

根据式(6)计算得到类别数 $C = 4$, 得到退化数据集的健康阶段标签, 解决不均衡、无标签问题; 然后, 通过随机选取前 80 个样本退化特征数据集来训练多变量深度森林健康评估器, 采用编号 81~90 的样本测试训练的结果, 当分类准确率均大于 90% 时, 保存训练结果较好的深度森林结构; 最后, 采用编号 91~100 的样本非全寿命周期的数据进行在线评估。

图 8 和图 9(a) 分别给出发动机样本 81 和 83 的离线训练结果, 其他样本结果图与之类似。图 8 和图 9 中横坐标代表发动机的运行周期, 纵坐标的“1”、“2”、“3”和“4”分别代表发动机处于“正常”、“轻微损伤”、“严重损伤”和“失效”4 个不同的健康阶段。图 8 和图 9 中黑色的圈划线是将数据健康分类结果作为实际结果, 红色的点划线是利用深度森林算法的健康评估训练结果。

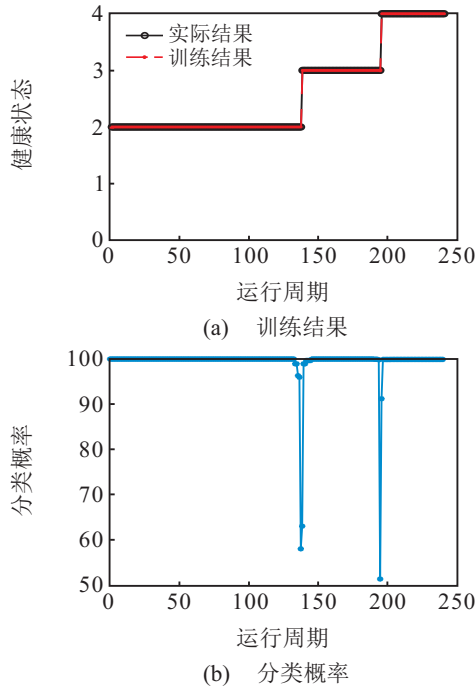


图 8 发动机 81 健康状态训练结果

从图 9(a) 可以看出, 发动机不同健康阶段的运行周期长短不一, 本文提出的健康状态分类方法有效地将不均衡、无标签退化特征集分成 4 类来刻画发动机的健康状态。从图 9(a) 中局部放大图可以看出, 在健康状态由“1”切换到“2”和由“3”切换到“4”的阶段存在误分类, 并且给出了该阶段健康类别的概率, 在误分类阶段概率较低, 可以将其视为过渡阶段。从图 8(a) 可以看出, 发动机初始阶段处于“2”阶段, 本文提出的健康状态分类方法有效地解决了发动机初始状态不确定问题。

图 9(b) 和图 9(c) 展示了分别利用人工神经网络 (ANN) 和支持向量机 (SVM) 两种方法进行的发动机

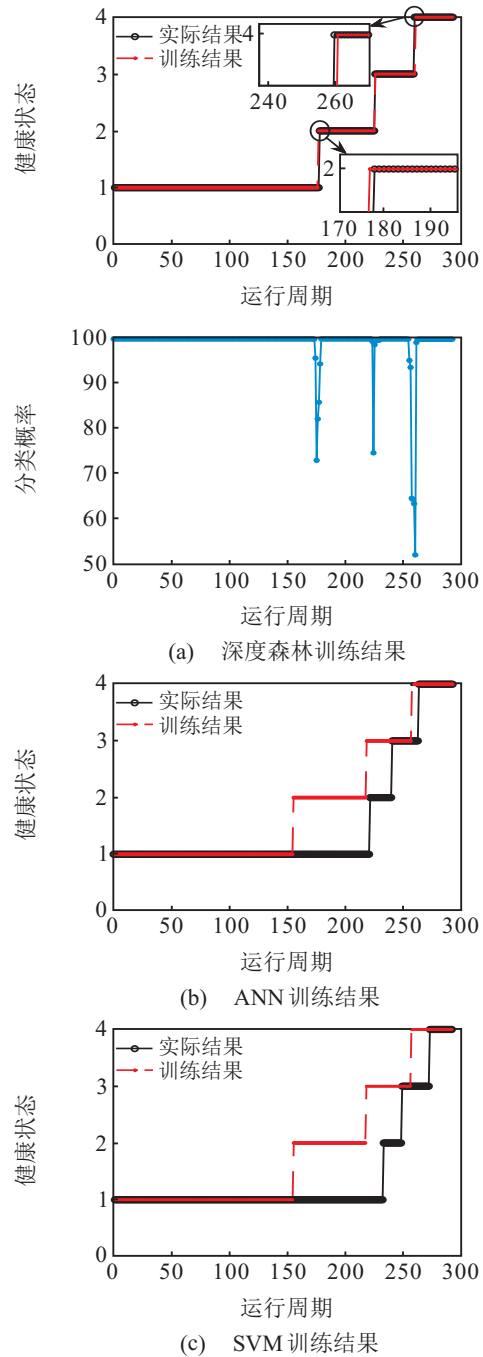


图 9 发动机 83 健康状态训练结果

健康状态评估结果(假设所有观测样本一开始都处于健康阶段), 可以看出这两种方法在健康状态“4”评估误差较小, 对后续剩余使用寿命的预测结果不会有太大的影响。然而, 如果涉及到其他健康状态, 这两种方法评估误差较大, 则会直接影响设备各个不同阶段的维护策略。更重要的是, 相比深度森林算法, 这两种方法不能提供相对应的分类概率, 整体分类的准确性较差(如表 2)。从表 2 可以看出, 本文提出的深度森林健康评估方法的离线训练结果准确率均大于 90%。然而, ANN 和 SVM 的离线训练结果在个别样本中分类准确率只能到达 60% 和 50% 左右, 较低的健康评估准确率不能为设备不同阶段的维护策略

提供可信的理论支撑,不能够用于设备的在线健康评估.

表2 离线训练准确率

序号	深度森林	ANN/%	SVM/%
81	100	90.42	86.25
82	98.60	81.76	80.37
83	99.32	68.94	62.46
84	97.75	60.30	82.02
85	100	92.55	78.72
86	99.64	89.93	91.73
87	98.31	85.96	50.99
88	97.65	75.59	98.12
89	99.54	86.64	86.64
90	99.35	74.68	63.64

图10展示了发动机93的在线评估结果,与图8和图9(a)比较可以看出,在线评估结果虽然在健康状态由“1”切换到“2”和由“2”切换到“3”存在的误分类略高于离线训练结果,但是在线评估准确率也大于90%.其他样本的在线评估结果与发动机93在线评估类似,评估准确率也均大于90%.

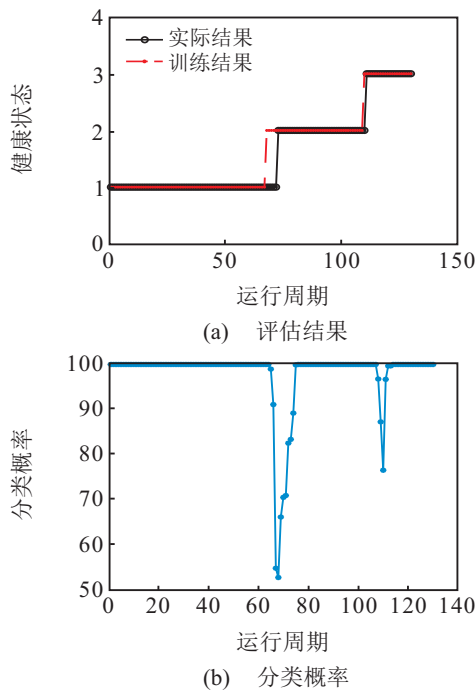


图10 发动机93健康状态在线评估结果

4 结论

本文针对设备观测数据的无标签、不均衡、初始状态不确定等问题,提出了一种多变量深度森林的设备健康评估方法,能够动态设定设备不同健康状态阈值,给出上述问题的解决方法.案例结果表明,本文提出的健康评估方法优于一些常用的健康评估方法,如ANN和SVM等.更重要的是,本文能够准确评

估设备各阶段的健康状态并且给出不同阶段的分类概率结果.

本文方法应用的两个前提是:1) 有比较充足的全寿命周期退化数据样本集;2) 假定历史样本处于相同的工作工况.然而,忽略了在缺少全寿命周期观测数据和处于不同工况下的设备健康评估,在后续的研究中将重点考虑.此外,PHM技术的重要核心是对设备的故障预测,它的主要目的是估计设备的故障趋势及剩余使用寿命,未来将在本文的基础上对此进行重点研究.

参考文献(References)

- [1] 姜斌, 吴云凯, 陆宁云, 等. 高速列车牵引系统故障诊断与预测技术综述[J]. 控制与决策, 2018, 33(5): 841-855.
- [2] Moghaddass R, Zuo M J. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process[J]. Reliability Engineering & System Safety, 2014, 124: 92-104.
- [3] 祁海铭, 程月华, 姜斌, 等. 基于多状态故障的卫星姿态控制系统剩余寿命预测方法[J]. 南京航空航天大学学报, 2015, 47(1): 29-36.
- [4] Wang X L, Jiang B, Lu N Y, et al. Dynamic fault prognosis for multivariate degradation process[J]. Neurocomputing, 2018; 275: 1112-1120.
- [5] 顾胜, 魏蛟龙, 皮德常. 一种粒子群模糊支持向量机的航天器参量预测方法[J]. 宇航学报, 2014, 35(11): 1270-1276.
- [6] Gu S, Wei J L, Pi D C. Particle swarm optimization-fuzzy support vector machine based prediction of spacecraft parameter[J]. Journal of Astronautics, 2014, 35(11): 1270-1276.
- [7] Liao L, Köttig F. A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction[J]. Applied Soft Computing, 2016, 44(C): 191-199.
- [8] Hong M, Wang Q, Su Z, et al. In situ, health monitoring for bogie systems of CRH380 train on Beijing-Shanghai high-speed railway[J]. Mechanical Systems & Signal Processing, 2014, 45(2): 378-395.
- [9] Javed K, Gouriveau R, Zerhouni N. State of the art and taxonomy of prognostics approaches, trends of

- prognostics applications and open issues towards maturity at different technology readiness levels[J]. *Mechanical Systems & Signal Processing*, 2017, 94: 214-236.
- [9] Tamilselvan P, Wang P. Failure diagnosis using deep belief learning based health state classification[J]. *Reliability Engineering & System Safety*, 2013, 115(7): 124-135.
- [10] Ramasso E, Rombaut M, Zerhouni N. Prognostic by classification of predictions combining similarity-based estimation and belief functions[C]. *Proceedings of the 2nd International Conference on Belief Functions*. Berlin: Springer, 2012: 61-68.
- [11] Javed K, Gouriveau R, Zerhouni N. A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering[J]. *IEEE Transactions on Cybernetics*, 2015, 45(12): 2626-2639.
- [12] Zhou H, Huang J, Lu F. Reduced kernel recursive least squares algorithm for aero-engine degradation prediction[J]. *Mechanical Systems & Signal Processing*, 2017, 95: 446-467.
- [13] Atamuradov V, Medjaher K, Dersin P, et al. Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation[J]. *International Journal of Prognostics & Health Management*, 2017, 8: 1-31.
- [14] 胡昌华, 施权, 司小胜, 等. 数据驱动的寿命预测和健康管理技术研究进展[J]. *信息与控制*, 2017, 46(1): 72-82.
(Hu C H, Shi Q, Si X S, et al. Data-driven life prediction and health management: State of the art[J]. *Information and Control*, 2017, 46(1): 72-82.)
- [15] Gouriveau R, Ramasso E, Zerhouni N. Strategies to face imbalanced and unlabelled data in PHM applications[J]. *Chemical Engineering Transactions*, 2013, 33: 115-120.
- [16] Sankararaman S. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction[J]. *Mechanical Systems & Signal Processing*, 2015, 52/53(1): 228-247.
- [17] Atamuradov V, Medjaher K, Dersin P, et al. Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation[J]. *International Journal of Prognostics & Health Management*, 2017, 8: 1-31.
- [18] 孙强, 岳继光. 基于不确定性的故障预测方法综述[J]. *控制与决策*, 2014, 29(5): 769-778.
(Sun Q, Yue J G. Review on fault prognostic methods based on uncertainty[J]. *Control and Decision*, 2014, 29(5): 769-778.)
- [19] Horn D, Axel I. Novel clustering algorithm for microarray expression data in a truncated SVD space[J]. *Bioinformatics*, 2003, 19(9): 1110-1115.
- [20] Bezdek J C. Pattern recognition with fuzzy objective function algorithms[J]. *Advanced Applications in Pattern Recognition*, 1981, 1171(22): 203-239.
- [21] Wang C S, Lu N Y, Cheng Y H, et al. Deep forest based multivariate classification for diagnostic health monitoring[C]. *Proceedings of the 30th Chinese Control and Decision Conference*. Shenyang: IEEE, 2018: 6233-6238.
- [22] Zhou Z H, Feng J. Deep forest[J]. *National Science Review*, 2018, DOI: 10.1093/nsr/nwy108.
- [23] Fan W, Wang H, Yu P S, et al. Is random model better? on its accuracy and efficiency[C]. *Proceedings of the 3rd IEEE International Conference on Data Mining*. Melbourne: IEEE, 2003: 51-58.
- [24] Breiman L. Random forest[J]. *Machine Learning, Machine Learning*, 2001, 45(1): 5-32.
- [25] Saxena A, Goebel K, Simon D, et al. Damage propagation modeling for aircraft engine run-to-failure simulation[C]. *Proceedings of the 2008 International Prognostics and Health Management*. Denver: IEEE, 2008: 1-9.
- [26] Frederick D K, DeCastro J A, Litt J S. User's guide for the commercial modular aero-propulsion system simulation (C-MAPSS)[Z]. *NASA/ARL Technical Manual TM2007-215026*, 2007.
- [27] Ramasso E, Rombaut M, Zerhouni N. Joint prediction of continuous and discrete states in time-series based on belief functions.[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2013, 43(1): 37-50.

作者简介

王村松(1990—), 男, 博士生, 从事数据驱动的故障预测与健康管理的研 究, E-mail: wangcunsong@nuaa.edu.cn;

陆宁云(1977—), 女, 教授, 博士生导师, 从事数据驱动的故障诊断、故障预测与健康管理等研究, E-mail: luningyun@nuaa.edu.cn;

程月华(1977—), 女, 教授, 博士, 从事航天器故障预测、故障诊断与容错控制等研究, E-mail: chengyuehua@nuaa.edu.cn;

姜斌(1966—), 男, 教授, 博士, 从事故障诊断和容错控制、航空航天应用等研究, E-mail: binjiang@nuaa.edu.cn.

(责任编辑: 闫 妍)