

# 基于 LASSO 变量选择的航空发动机相似性剩余寿命预测

于倩影<sup>1</sup>, 李娟<sup>1</sup>, 戴洪德<sup>2</sup>, 辛富禄<sup>1</sup>

- (1. 鲁东大学 数学与统计科学学院, 山东 烟台 264025;
2. 海军航空大学 航空基础学院, 山东 烟台 264001)

**摘 要:** 由于航空发动机监测变量众多, 传统方法直接选取性能退化趋势明显的变量进行寿命预测, 所以提出一种基于 LASSO(least absolute shrinkage and selection operator)的变量选取方法, 结合相似性寿命预测方法有效提高了预测精度。基于 K-means 聚类区分不同工况, 对航空发动机多个监测变量根据聚类结果进行变量转换。基于 LASSO 方法选取最优传感器变量。基于相似性方法进行航空发动机剩余寿命预测。将基于 LASSO 的变量选取方法与传统的根据退化趋势大小进行选择的方法进行剩余使用寿命预测的结果进行了对比研究。结果表明: 基于 LASSO 选取变量的相似性寿命预测误差的标准差在 3 种运行周期下分别减少了约 1.84、3.46、4.23。

**关 键 词:** 预测与健康管理; K-means 聚类; LASSO 方法; 相似性; 剩余寿命

中图分类号: V263.6

文献标志码: A

## LASSO based variable selection for similarity remaining useful life prediction of aero-engine

YU Qianying<sup>1</sup>, LI Juan<sup>1</sup>, DAI Hongde<sup>2</sup>, XIN Fulu<sup>1</sup>

- (1. School of Mathematics and Statistics, Ludong University, Yantai Shandong 264025, China;
2. School of Basic Sciences for Aviation,  
Naval Aviation University, Yantai Shandong 264001, China)

**Abstract:** Due to the large number of aero-engine monitoring variables, the variables with obvious performance degradation trend were directly selected by traditional method for the life prediction, so a variable selection method based on LASSO (least absolute shrinkage and selection operator) was proposed, which combined with the similarity life prediction method to effectively improve the prediction accuracy. Based on K-means clustering, different working conditions were distinguished, and multiple monitoring variables of aero-engine were transformed according to the clustering results. The optimal sensor variables were selected based on the LASSO method. The remaining useful life of aero-engine was predicted based on similarity method. The results of remaining useful life prediction based on the variable selection method by LASSO and the traditional selection method by the degradation trend were compared. The results showed that the standard deviation of the similarity life prediction error based on the variables selected by LASSO decreased by about 1.84, 3.46 and 4.23 under three operating cycles.

收稿日期: 2021-09-15

基金项目: 山东省自然科学基金面上项目(ZR2017MF036); 国防科技项目基金(F062102009)

作者简介: 于倩影(1998—), 女, 硕士生, 主要从事应用统计研究。

通信作者: 李娟(1981—), 女, 副教授、硕士生导师, 博士, 主要从事预测与健康管理研究。E-mail: daidaiquanquan123@126.com

引用格式: 于倩影, 李娟, 戴洪德, 等. 基于 LASSO 变量选择的航空发动机相似性剩余寿命预测[J]. 航空动力学报, 2023, 38(4): 931-938. YU Qianying, LI Juan, DAI Hongde, et al. LASSO based variable selection for similarity remaining useful life prediction of aero-engine[J]. Journal of Aerospace Power, 2023, 38(4): 931-938.

**Keywords:** prognostics health management; K-means clustering; LASSO method; similarity; remaining useful life

针对现代战争跨越式的武器装备建设,“十四五”规划明确了航空发动机的制造在我国航空领域发展中的重要地位。航空发动机作为一种为航空器提供飞行动力的高精度、高精度的热力机械装备,由成千上万个精密零件组成。作为飞机的“心脏”,航空发动机或直接或间接地影响了飞机的使用寿命、人民的经济利益乃至生命安全。一方面,航空发动机的制造水平衡量了一个国家的科技能力、工业生产能力和国防实力;另一方面,航空发动机的性能维护也带来了众多考验。因此,航空发动机的预测与健康管理(prognostics health management, PHM)具有重要意义<sup>[1]</sup>。

剩余寿命(remaining useful life, RUL)预测是预测与健康管理的核心内容。李琪等<sup>[2]</sup>将航空发动机的寿命预测方法分为基于模型驱动与基于数据驱动的方法。模型驱动采用以模型为主导的方法,但针对复杂的设备,构建较为完备的物理模型非常困难,而且不定期的停机检查会影响工作进程,甚至引起不必要的设备磨损<sup>[3]</sup>。数据驱动方法即有效利用历史或实时数据构建数学模型,从而进行剩余寿命预测。作为一种有效的数据驱动方法,Wang等<sup>[4]</sup>提出的基于相似性的剩余寿命预测方法,有效使用了航空发动机系统中大量从正常运行到故障的设备数据。在距离发生故障失效相对较远的情况下,基于相似度的剩余寿命预测具有长期预测的能力<sup>[5]</sup>,同时减轻了构建复杂设备退化模型的难度。

针对航空发动机的剩余寿命预测,监测变量众多,为降低工作的复杂程度,减少不必要的资源浪费,国内外的研究者主要从以下几方面进行研究。特征提取方面体现在:Wu等<sup>[6]</sup>采用动态差分技术分别从发动机21个传感器的原始数据中提取特征,建立性能退化的深度神经网络模型,得到了比传统神经网络模型更高的寿命预测精度。多传感器信息可以提高寿命预测精度,但过多传感器的安装会增加系统的体积、质量、功耗、复杂性,使得维护成本加大<sup>[7]</sup>。基于退化趋势选取传感器变量方面包括:Li等<sup>[8]</sup>对14个有明显变化趋势的传感器原始监测数据进行学习,提出了基于数据驱动的航空发动机卷积神经网络深度学习

的寿命估计方法。Wen等<sup>[9]</sup>考虑同一传感器在各训练单元上具有的相同递增或者递减趋势,识别出11个传感器作为模型的输入信号。Listou等<sup>[10]</sup>剔除整个引擎生命周期内恒定传感器信息测量,基于半监督的深度架构方法进行了剩余寿命预测。可见,基于退化趋势进行变量选取是常见的方法,但相应的理论支撑还需要进一步研究。

因此,基于统计理论的变量选取方法得到了广泛关注,Celestino等<sup>[11]</sup>考虑多重共线性,将相关性很强的变量剔除,最终提取了5个变量对航空发动机进行寿命预测。Lu等<sup>[12]</sup>通过评估剩余寿命的可预测性,选取了5个性能退化特征,对航空发动机进行寿命预测。张妍等<sup>[13]</sup>采用Relief算法最终选取8个退化特征,提出了一种基于退化特征相似性的寿命预测方法。Lasheras等<sup>[14]</sup>采用主成分分析的方法,对24个变量进行变量选择,最终在91.2%的贡献度下保留了17个主成分退化特征变量,并对航空发动机的剩余寿命进行了预测。然而,基于主成分的变量选择方法,本质是所有变量的线性组合,无法真正实现传感器降维。

而LASSO(least absolute shrinkage and selection operator)回归算法的基本思想是在线性回归模型的基础上加入L1范数(即向量中各个元素的绝对值之和)的约束项,通过控制惩罚参数 $\lambda$ 的值进行变量筛选和调整模型的复杂程度。LASSO是一种从大量数据筛选有效信息的方法,已经广泛应用于桥梁建筑<sup>[15]</sup>、轴承故障诊断<sup>[16]</sup>、基因选择<sup>[17]</sup>、药物研发、边坡稳定性分析等方面,但在航空发动机寿命预测领域,鲜有基于LASSO进行变量选取的理论与实践研究。

本文提出一种基于LASSO进行变量选择的相似性寿命预测方法,筛选重要的传感器变量,减少高维冗余信息,降低变量之间的共线性,提高航空发动机剩余寿命预测的精度,为航空发动机的预测与健康管理提供理论与实践参考。

## 1 基于LASSO的相似性寿命预测方法

该部分包括健康因子构建、基于LASSO的变量选取方法与基于相似性的剩余寿命预测。

### 1.1 健康因子构建

为定量描述航空发动机从正常状态到故障失效的退化参数,即航空发动机的健康状况,将不同传感器变量状态数据合成以映射到单一维度中,进行健康指标(health indicator, 记为  $I_h$ )的构建。健康指标构建的思想是将每组发动机的运转周期折合到区间 $[0, 1]$ 中( $I_h(t) \in [0, 1]$ ), 0 表示发动机处于完全失效状态, 1 表示发动机处于初始运转状态。

假设  $t_i$  表示发动机第  $i$  次循环次数,  $T_i$  表示当前发动机第  $i$  次循环次数与该发动机总循环次数的差值, 即  $T_i = \max(t_1, t_2, \dots, t_k) - t_i$ ,  $k$  表示每个发动机从初始运转到最终运转的总循环次数。

记  $I_h(t) \in [0, 1]$  为健康因子, 为当前循环次数的差值分别与该差值中的最大值的比值, 表现为每组发动机的健康状况从 1 退化到 0, 即

$$I_h(t) = \frac{T_i}{\max(T_1, T_2, \dots, T_k)}, \quad i = 1, 2, \dots, k \quad (1)$$

本文采用线性回归方法构建健康指标, 以保留训练集中更多的退化数据信息。记  $\mathbf{Y} = (y_1, y_2, \dots, y_m)^T$  为  $m$  维健康因子变量矩阵,  $\mathbf{X} = (x_1, x_2, \dots, x_n)$  为  $n$  维传感器参数的协变量矩阵,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_n)^T$  为  $n+1$  维回归系数矩阵,  $\boldsymbol{\varepsilon}$  为噪声项矩阵, 假设航空发动机运行健康状况随时间从 1 线性下降到 0, 则健康因子与传感器拟合为线性回归模型表示为

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

将传感器测量值与其相关权重  $W_j^{(2)}$  相乘, 构造一个单一的健康指示器。基于权重, 平滑融合数据并偏移数据, 使得所有退化数据均从 1 开始。由此, 来自多个传感器的数据被融合到一个健康指示器中, 传感器运行状况指示器由移动平均值过滤器平滑。

$$\hat{T}_i = \sum_{j=-k}^k (\omega_j x_{t+j}) \quad (3)$$

### 1.2 基于 LASSO 的变量选取方法

LASSO 实际是一种保留子集收缩这一优点的压缩估计方法, 也是一种处理具有复共线性数据的有偏估计。LASSO 实际是 Bridge 回归的一种特殊的表达形式<sup>[18]</sup>。通过构造一个惩罚函数得到较为精练的模型, 设置惩罚参数  $\lambda$  的值, 从而压缩变量的系数。随着惩罚参数  $\lambda$  的不断增大, 变量的系数会持续趋向于零, 当  $\lambda$  足够大时, 一部分

变量的系数会被压缩至零, 从而达到减少原始变量的目的<sup>[19]</sup>。

区别于普通线性模型(式(2)), 基于 LASSO 方法构建的回归模型增加了惩罚函数。

如果传感器变量  $\mathbf{X}$  为列满秩设计矩阵, 可由普通最小二乘法(ordinary least squares, OLS)获得回归系数矩阵  $\boldsymbol{\beta}$ 。普通最小二乘估计是通过最小化残差平方和估计参数, 但普通最小二乘法的预测精度与稳定性较差, 不再适用于高维的数据分析, 因此需要加入惩罚函数, 以此对模型进行约束。

如果传感器变量  $\mathbf{X}$  为非列满秩设计矩阵, 则需通过增加惩罚项来约束模型的复杂度, 同时通过选择经验风险与模型复杂度相对较小的模型获得估计 LASSO 模型。普通最小二乘估计与基于 LASSO 估计表达式分别如下:

$$\hat{\boldsymbol{\beta}}_{OLS} = \operatorname{argmin} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\| = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (4)$$

$$\hat{\boldsymbol{\beta}}_{LASSO} = \operatorname{argmin} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad \text{s.t.} \sum_{j=1}^k |\beta_j| \leq \lambda \quad (5)$$

式中调节惩罚参数  $\lambda \geq 0$ , 在估计时压缩部分航空发动机传感器变量的系数, 使其缩减至 0, 最终完成对航空发动机传感器变量的降维目的。

最优模型的选择准则包括 AIC(Akaike information criterion)准则、BIC(Bayesian information criterion)准则、CP 准则(Mallow's CP 准则)等<sup>[18]</sup>。AIC 准则即赤池信息准则, 由日本统计学家赤池弘次于 1974 年提出, 通常情况下, 其为拟合精度和参数未知个数的加权函数。从可供选择的模型中选择最佳模型时, 通常选择使得 AIC 最小的模型。BIC 准则即贝叶斯信息准则, 由 Schwarz 于 1978 年提出, BIC 准则与 AIC 准则相似, 但 BIC 加入了样本容量  $N$ 。由马洛斯提出用以评估普通最小二乘法为假设的线性回归模型优良性的 CP 准则与前两者不同, CP 准则的评判标准为在所有子模型中使得  $C_s$  (即第  $s$  个传感器变量时的 CP 值) 最小时, 认为近似具有最小的预测误差, 即  $C_s$  值越小, 选择变量的精确度越高。本文在该部分选择运用 CP 准则为模型选取变量。

记  $R_{rss,s}$  为  $s$  个传感器变量时的残差平方和,  $\hat{\sigma}^2$  为全模型下普通最小二乘方差,  $L$  为似然函数, AIC 准则、BIC 准则、CP 准则表达式如下:

$$f_{AIC} = 2k - 2\ln L \quad (6)$$

$$f_{\text{BIC}} = k \ln N - 2 \ln L \quad (7)$$

$$C_s = \frac{R_{\text{FSS},s}}{\sigma^2} - n + 2s \quad (8)$$

### 1.3 基于相似性的剩余寿命预测

基于相似性剩余寿命预测的主要思想: 如果待服役航空发动机上各传感器的状态与已服役航空发动机上各传感器的状态参数在某时间内状态相似, 则它们可能存在相似的剩余寿命<sup>[20]</sup>。

记  $y_j$  为第  $j$  个航空发动机真实的健康指标,  $\hat{y}_{ji}$  为拟合模型使用航空发动机  $i$  中确定的二阶多项式模型估计的航空发动机  $j$  的健康指标所得估计值, 则借助欧氏距离 (Euclidean distance, 见式 (9)) 与 1 范数, 定义距离函数 ( $d(i, j)$ ), 计算拟合模型的估计值与训练集中真实健康指标之间的距离得分情况。其中, 欧氏距离是指计算  $n$  维空间中两点间的真实距离。通过距离得分 ( $s(i, j)$ ) 来定量描述估计值与真实值两者之间的差别, 从而完成拟合模型估计值与真实健康指标之间相似度的测量。

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (9)$$

$$d(i, j) = \|y_j - \hat{y}_{ji}\|_1 \quad (10)$$

$$s(i, j) = e^{-[d(i, j)]^2} \quad (11)$$

由此可见, 拟合模型估计值与真实健康指标的距离越短时, 两者相似程度越高, 由此, 两者距离达到最短即  $d(i, j)$  最小时, 它们之间的相似性达到最高, 第  $k$  个参考样本的相似性最高时为

$$D_k = \min d_k(i, j) \quad (12)$$

在大样本情况下, 随着样本量的不断增大, 用样本的性质推断总体性质的精确度也不断增加。首先处理数据, 假设验证数据集中有一个集群, 其在训练数据集中获取最接近的 50 个样本。其次, 基于这 50 个样本构造统计量, 计算统计量, 拟合基于 50 个样本的概率分布, 判断落入拒绝域的情况, 并使用分布的中位数 (median) 作为航空发动机剩余寿命的估计值。

基于 LASSO 变量选择的航空发动机相似性剩余寿命预测流程图如图 1 所示。

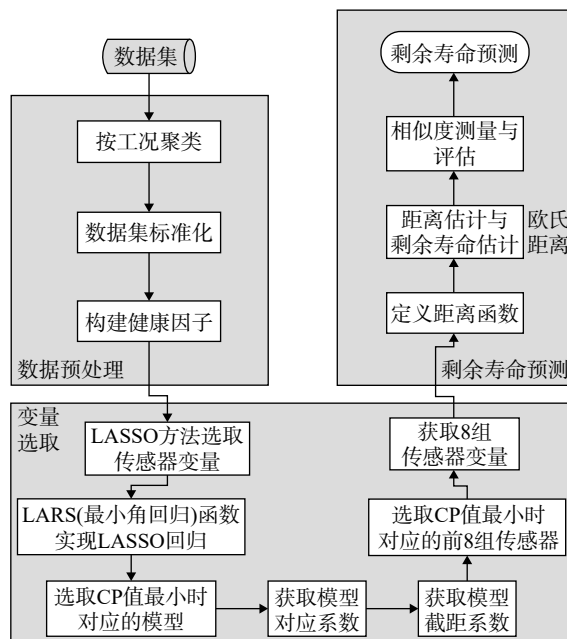


图 1 剩余寿命预测流程图

Fig. 1 Flow chart of remaining useful life prediction

## 2 实验验证

### 2.1 数据集简介

本文借助仿真工具 C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) 提供的一组涡扇发动机热力学仿真退化模型数据进行航空发动机剩余寿命预测的研究。该仿真工具由 NASA 埃姆斯研究中心开发, 通过改变燃油流量、空气流速、压力等输入参数模拟风扇、高压涡轮、低压涡轮、高压压气机和低压压气机等部件的性能退化过程, 使用传感器记录航空发动机的多个监测参数<sup>[21]</sup>。该数据由 218 组训练集与 218 组测试集的多维时间序列组成, 每个时间序列都来自同一个复杂工程系统的不同发动机, 即可以认为数据来自同一类型的发动机。每组发动机都被认为在每个时间序列开始时正常工作, 并在时间序列期间的某个点开始发生故障。本文采用的数据为该数据中的 218 组训练集。图 2 展示了航空发动机主要传感器的位置, 其中  $N_1$  为低压转子转速,  $N_2$  为高压转子转速, LPT (low pressure turbine) 为低压涡轮, HPT (high pressure turbine) 为高压涡轮, LPC (low pressure compressor) 为低压压缩机, HPC (high pressure compressor) 为高压压缩机。

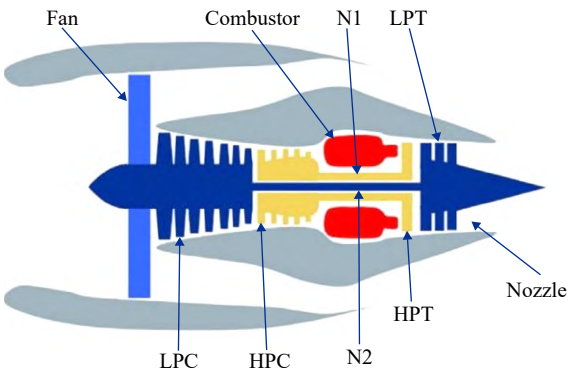


图 2 航空发动机主要传感器示意图  
Fig. 2 Schematic diagram of main sensors of aero-engine

2.2 工况识别与数据转换

K-means 聚类是一种基于两点间距离归类的于非层次聚类法。本文借助 K-means 方法基于训练集的 3 种操作设置聚类, 最终形成 6 种工作状态。该聚类方法执行过程主要包括: 初始化过程, 即在训练集中随机选择 6 个数据点作为质心; 循环过程, 即基于就近原则借助欧氏距离计算训练集中的数据点与每个质心的距离, 每个数据点与距离最短的质心划分为一类, 不同初始条件下重复运算, 最终每组发动机训练数据相应地自动划分到 6 个集群, 即认为分别对应 6 种不同工况, Opsetting\_1、Opsetting\_2、Opsetting\_3 分别表示 3 种操作设置, 6 个集群的聚类中心在 3 种操作设置下的位置如图 3 所示。

图 3 体现了 6 个聚类中心基于 3 种操作设置的位置情况, 求得的 6 个集群的聚类中心在 3 种操作设置下的数值, 即具体的 6 个聚类中心 3 种操作设置的参数如表 1 所示。

记  $x'_{i,a}$  表示标准化的传感器测量值,  $x_{i,a}$  表示航空发动机传感器变量标准化之前传感器变量  $i$  中参考样本  $a$  的传感器测量值,  $\bar{x}_j$  和  $\sigma_j$  分别表示第  $j$

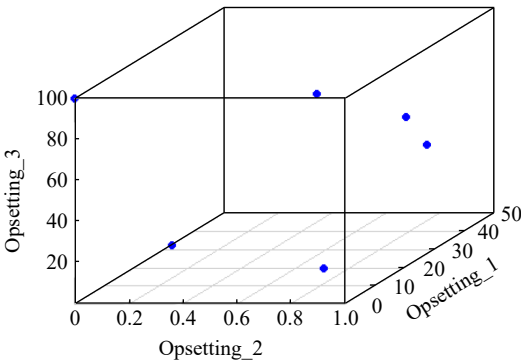


图 3 6 个聚类中心示意图  
Fig. 3 Schematic diagram of 6 cluster centers

表 1 6 个聚类中心参数表  
Table 1 Parameter table of 6 cluster centers

聚类中心	Opsetting_1	Opsetting_2	Opsetting_3
1	10.003 0	0.250 5	20
2	25.003 0	0.620 5	80
3	20.002 9	0.700 5	0
4	35.003 0	0.840 5	60
5	0.001 5	0.000 5	100
6	42.003 0	0.840 5	40

个簇心的平均值与标准差。将训练集基于不同工作状态下分类后进行类标准化, 即  $x'_{i,a} = \frac{x_{i,a} - \bar{x}_j}{\sigma_j}$ 。

2.3 基于 LASSO 的变量选取

借助 LARS 算法, 可以通过选定算法的步数或者选定  $\beta$  的饱和度 (即  $|\beta|/|\beta_{\max}|$ ), 以确定 LASSO 模型的参数。随着惩罚参数  $\lambda$  的不断增大, 基于 LASSO 获得航空发动机传感器变量的选取结果如图 4。交叉验证的步数达到最小时, 对应的 CP 值同时达到最小, 此时可视为准确选取变量, 得到最小误差对应的系数。航空发动机传感器变量系数的大小可以定量表示航空发动机传感器变量的重要程度。

为了与竞赛冠军的剩余寿命预测相比较, 本文基于 LASSO 方法也选用 8 个传感器, 即  $\lambda$  不断增大时, 最先被 LASSO 回归模型选入的 8 个变量作为具有代表性的传感器组合, 从而控制变量个数, 消除因参数变量个数的不同, 影响与根据退化趋势大小选取变量的方法进行剩余寿命预测的分析结果。本文采用的 LARS 算法中的 LASSO 模型。

图 4 上方横坐标的数字表示 LASSO 算法的步数, 即迭代次数, 右侧纵坐标的数字表示先后

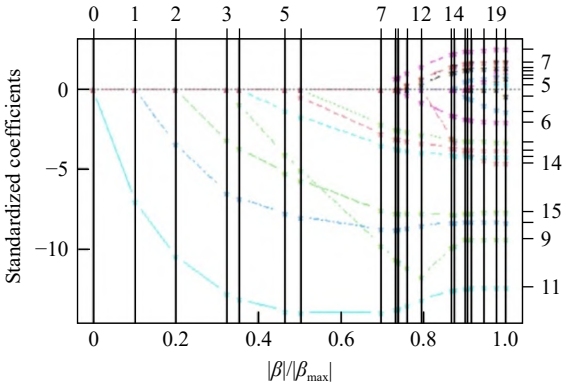


图 4 LASSO 拟合系数示踪图  
Fig. 4 LASSO fitting coefficient tracer diagram

被选入的传感器变量。由图4可以看出,最先选入的8组航空发动机的传感器为第11组传感器,其次为第9组传感器,依次有第4组传感器、第15组传感器、第14组传感器、第17组传感器、第2组传感器、第3组传感器。

### 3 对比分析

传统方法使用最具趋势的传感器变量构建健康指标,这些最具趋势的传感器变量对退化表现为较高的敏感性。传统方法是建立一个数组并将其进行初始化处理,以储存各传感器变量的斜率值。基于所有传感器变量拟合关于健康状况的线性回归模型,在训练模型过程中获得各传感器的退化数据斜率。基于估计线性退化模型获得各传感器变量斜率后,将所得传感器变量的斜率进行从大到小排序,选取斜率最大的8组传感器基于相似性预测航空发动机的剩余寿命。选取的8组传感器变量分别为第2组传感器、第3组传感器、第4组传感器、第7组传感器、第11组传感器、第12组传感器、第15组传感器、第17组传感器。

基于LASSO方法选取的传感器分别为第2组传感器、第3组传感器、第4组传感器、第9组传感器、第11组传感器、第14组传感器、第15组传感器、第17组传感器。两种方法选取传感器区别在于传统方法选取了第7组传感器、第12组传感器,本文基于LASSO方法选取了第9组传感器、第14组传感器。

本文为了评估相似性剩余寿命模型,使用航空发动机传感器测量值的50%(即航空发动机运行周期的50%)、70%(即航空发动机运行周期的70%)和90%(即航空发动机运行周期的90%)的验证数据来预测其剩余寿命。同时,为了控制单一变量以探究基于LASSO方法选取变量的有效性,本文在相似性模型构建和剩余寿命的加权平均方法等方面均使用传统相似性方法,计算航空发动机的剩余寿命真实值与基于相似性方法预测的剩余寿命预测值两者之间的误差。对数据集重复相同的相似度评估的过程,并计算每个断点的估计剩余寿命和真实剩余寿命之间的误差。计算预测误差的平均值和标准偏差,继而比较基于LASSO方法与基于传统方法选择变量预测航空发动机剩余寿命的有效性。借助航空发动机50%运行周期的退化数据时,基于LASSO方法选择的传感器进行寿命预测的误差均值为-4.6671,而基于传统方法的估算误差均值为-5.8944。由

此可见,数据退化信息较少的情形下,基于LASSO方法选择的传感器进行的寿命预测估计较传统方法更好。这一点在较难获取全程退化数据时,预测航空发动机的剩余寿命的过程中尤为重要。

图5将90%运行周期的验证数据及其最接近的样本数据进行可视化,纵坐标表示为拟合后的健康因子退化参数,横坐标表示每个样本运行的循环次数,由此可以看出,90%运行周期的退化数据拟合结果与最接近的样本数据呈现的退化趋势的吻合度较好。

在90%置信度下,与图6所示的基于传统方法的航空发动机剩余寿命预测情况相比,图7所示的基于LASSO方法选取变量后的剩余寿命预测结果更接近剩余寿命的真实值。可见基于LASSO方法选取变量有效提高了航空发动机的剩余寿命预测。

表2体现了基于两种方法的预测误差的标准

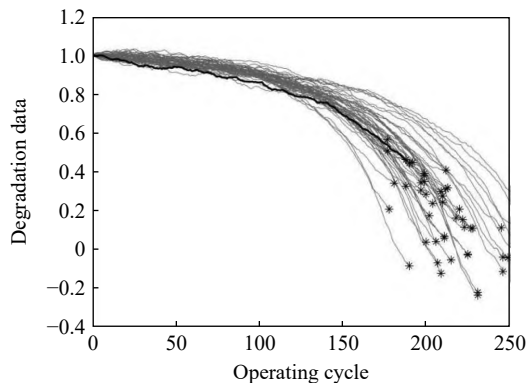


图5 90%运行周期的退化数据拟合

Fig. 5 Degradation data fitting for 90% operating cycle

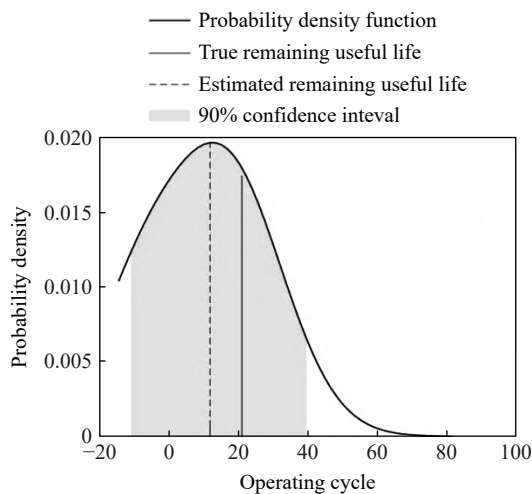


图6 90%运行周期的数据拟合概率分布(传统方法)

Fig. 6 Data fitting probability distribution of 90% operation cycle (traditional method)

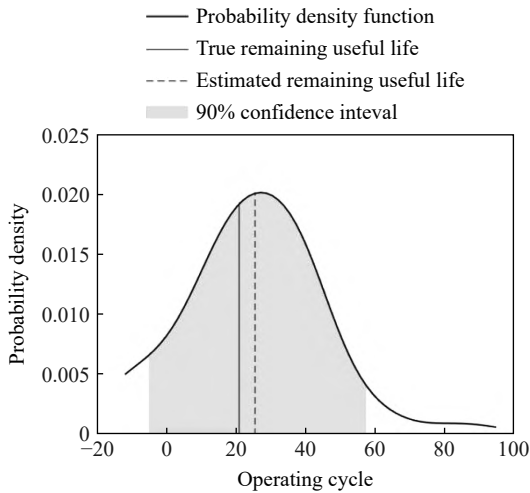


图 7 90% 运行周期的数据拟合概率分布(LASSO 方法)  
Fig. 7 Data fitting probability distribution of 90% operation cycle (LASSO method)

表 2 两种方法预测误差的标准差结果  
Table 2 Standard deviation results of prediction errors of the two methods

运行时间	方法	预测误差的标准差
50% 周期	LASSO 选取方法	24.660 5
	传统方法	26.496 1
70% 周期	LASSO 选取方法	16.614 8
	传统方法	20.072 0
90% 周期	LASSO 选取方法	13.804 8
	传统方法	18.031 3

误差。当机器处于中间健康阶段(即生命运行周期的 50%)时,两种方法估计的航空发动机剩余寿命与真实剩余寿命值之间均存在较大的误差;当观测到更多的数据时,剩余寿命的估计能力增强。分别将 50%、70%、90% 三种运行周期下基于 LASSO 方法与传统方法预测误差的标准差做差,分别得到约为 1.84、3.46、4.23 的结果,说明在 3 种运行周期下,基于 LASSO 选取变量的相似性寿命预测误差的标准差更小。基于 LASSO 方法在 3 种运行周期情况下的预测误差的标准差均较传统方法更小,生命运行周期越接近于最终寿命。由此可见,基于 LASSO 方法结合相似性寿命预测方法有效提高了剩余寿命的预测精度。

4 结论与展望

综上所述,本文结合 2008 PHM Data Challenge 挑战赛数据集,提出了基于 LASSO 变量选择与相似性融合的寿命预测方法。与传统的方法相比,

在 3 种运行周期内,预测误差的标准差都有所减小。在筛选变量的过程中提供了有力的理论支撑,说明 LASSO 方法不仅可以应用于桥梁建筑、轴承故障诊断等方面,也可以应用于航空领域中,在真正意义上实现变量的降维。

然而,本文方法仍有改进的空间,可根据 LASSO 方法增加一些条件构建模型,使得越重要的变量,其惩罚函数变得越小,使重要变量更容易被挑选出来。同时,可基于自适应 LASSO、随机 LASSO 等一系列的改进后的 LASSO 方法重新针对该问题进行研究与分析,讨论是否更有效地提高航空发动机的剩余寿命预测精度。

参考文献：

[ 1 ] LEI Y, LI N, GUO L, et al. Machinery health prognostics: a systematic review from data acquisition to RUL prediction[J]. Mechanical Systems and Signal Processing, 2018, 104(5): 799-834.

[ 2 ] 李琪, 高占宝, 李善营, 等. 变工况条件下基于相似性的剩余使用寿命预测方法[J]. 北京航空航天大学学报, 2016, 42(6): 1236-1243.

LI Qi, GAO Zhanbao, LI Shanying, et al. Similarity-based remaining useful life prediction method under varying operational conditions[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(6): 1236-1243. (in Chinese)

[ 3 ] 孟光, 尤明懿. 基于状态监测的设备寿命预测与预防维护规划研究进展[J]. 振动与冲击, 2011, 30(8): 1-11.

MENG Guang, YOU Mingyi. Review on condition-based equipment residual life prediction and preventive maintenance scheduling[J]. Journal of Vibration and Shock, 2011, 30(8): 1-11. (in Chinese)

[ 4 ] WANG Tianyi, YU Jianbo, SIEGEL D, et al. A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems[C]//2008 International Conference on Prognostics and Health Management. Denver, US: IEEE, 2008: 1-6.

[ 5 ] YOU M Y, MENG G. A generalized similarity measure for similarity-based residual life prediction[J]. Proceedings of the Institution of Mechanical Engineers Part E: Journal of Process Mechanical Engineering, 2011, 225(3): 151-160.

[ 6 ] WU Y, YUAN M, DONG S, et al. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks[J]. Neurocomputing, 2018, 275(1): 167-179.

[ 7 ] 任子强, 司小胜, 胡昌华, 等. 融合多传感器数据的发动机剩余寿命预测方法[J]. 航空学报, 2019, 40(12): 134-145.

REN Ziqiang, SI Xiaosheng, HU Changhua, et al. Remaining useful life prediction method for engine combining multi-sensors data [J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(12): 134-145. (in Chinese)

[ 8 ] LI X, DING Q, SUN J. Remaining useful life estimation in prognostics using deep convolution neural networks[J]. Reliability Engineering and System Safety, 2018, 172(4): 1-11.

[ 9 ] WEN Pengfei, ZHAO Shuai, CHEN Shaowei, et al. A generalized remaining useful life prediction method for complex systems based on composite health indicator[J]. Reliability Engineering and System Safety, 2021, 205(1): 107241.1-107241.15.

[10] LISTOU E A, EMIL B, VILMAR A, et al. Remaining useful life

- predictions for turbofan engine degradation using semi-supervised deep architecture[J]. *Reliability Engineering and System Safety*, 2019, 183(3): 240-251.
- [11] CELESTINO O, FERNANDO S L, JAVIER R P, et al. A hybrid ARIMA-SVM model for the study of the remaining useful life of aircraft engines[J]. *Journal of Computational and Applied Mathematics*, 2019, 346(1): 184-191.
- [12] LU F, WU J, HUANG J, et al. Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm[J]. *Aerospace Science and Technology*, 2019, 84: 661-671.
- [13] 张妍, 王村松, 陆宁云, 等. 基于退化特征相似性的航空发动机寿命预测[J]. *系统工程与电子技术*, 2019, 41(6): 1414-1421.
- ZHANG Yan, WANG Cunsong, LU Ningyun, et al. Remaining useful life prediction for aero-engine based on the similarity of degradation characteristics[J]. *Journal of Systems Engineering and Electronics*, 2019, 41(6): 1414-1421. (in Chinese)
- [14] LASHERAS F, NIETO P, DE COS JUEZ F, et al. A hybrid PCA-CART-MARS-based prognostic approach of the remaining useful life for aircraft engines[J]. *Sensors*, 2015, 15(3): 7062-7083.
- [15] MANGALATHU S, JEON J S, DESROCHES R. Critical uncertainty parameters influencing seismic performance of bridges using LASSO regression[J]. *Earthquake Engineering and Structural Dynamics*, 2018, 47(3): 784-801.
- [16] ZHAO Z, WU S, QIAO B, et al. Enhanced sparse period-group LASSO for bearing fault diagnosis[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66(3): 2143-2153.
- [17] 张靖, 胡学钢, 李培培, 等. 基于迭代LASSO的肿瘤分类信息基因选择方法研究[J]. *模式识别与人工智能*, 2014, 27(1): 49-59.
- ZHANG Jing, HU Xuegang, LI Peipei, et al. Informative gene selection for tumor classification based on iterative lasso[J]. *Pattern Recognition & Artificial Intelligence*, 2014, 27(1): 49-59. (in Chinese)
- [18] 曾津, 周建军. 高维数据变量选择方法综述[J]. *数理统计与管理*, 2017, 36(4): 678-692.
- ZENG Jin, ZHOU Jianjun. Variable selection for high-dimensional data model: a survey[J]. *Journal of Applied Statistics and Management*, 2017, 36(4): 678-692. (in Chinese)
- [19] 宋瑞琪, 朱永忠, 王新军. 高维数据中变量选择研究[J]. *统计与决策*, 2019, 35(2): 13-16.
- SONG Ruiqi, ZHU Yongzhong, WANG Xinjun. Research on variable selection in high-dimensional data[J]. *Statistics and Decision*, 2019, 35(2): 13-16. (in Chinese)
- [20] 雷从英, 夏良华, 林智崧. 基于相似性的装备部件剩余寿命预测研究[J]. *火力与指挥控制*, 2014, 39(4): 91-94.
- LEI Congying, XIA Lianghua, LIN Zhisong. Research on similarity-based remaining life prediction of equipment components[J]. *Fire Control and Command Control*, 2014, 39(4): 91-94. (in Chinese)
- [21] SAXENA A, GOEBEL K. PHM08 challenge data set[EB/OL]. [2022-05-22]. <https://ti.arc.nasa.gov/tech/dash/groups/pcoc/prognostic-data-repository/>.

(编辑: 陈 越)