

基于 XGBoost 的民航飞机发动机性能参数预测模型

樊智勇¹, 王振良², 刘哲旭²

(1. 中国民航大学 工程技术训练中心, 天津 300300;

2. 中国民航大学 电子信息与自动化学院, 天津 300300)

摘要: 为提高民航飞机发动机性能参数的预测精度, 提出一种基于模糊推理和 XGBoost 算法的发动机性能参数预测方法; 对发动机进行总体性能分析, 确定油门杆位置、气压高度、总温、全重、马赫数及飞行阶段为影响发动机性能参数的主要因素; 其次采用模糊推理对快速存取记录器 (QAR) 数据进行纵向飞行阶段划分, 消除人为划分训练数据对预测精度的主观影响; 最后, 建立各发动机性能参数的 XGBoost 预测模型, 并与多种预测模型进行对比实验; 实验结果表明: 对发动机 N1、燃油流量参数的预测, XGBoost 预测模型相比支持向量回归 (SVM)、线性回归模型和 BP 神经网络, 其精度更高且不需要对训练数据进行缩放。

关键词: 航空发动机; 模糊推理; XGBoost; QAR 数据; 性能参数预测

Prediction Model on Aircraft Engine Performance Parameters Based on XGBoost

FAN Zhiyong¹, WANG Zhenliang², LIU Zhexu²

(1. Engineering Technology Training Center, Civil Aviation University of China, Tianjin 300300, China;

2. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In order to improve the prediction accuracy of aircraft engine performance parameters, a new aeroengine performance prediction method based on fuzzy theory and XGBoost algorithm was proposed. Firstly, the overall performance of aeroengine was analyzed to determine main affecting aeroengine performance factors such as throttle angle, barometric altitude, total temperature, gross weight, Ma number in the flight phase; Secondly, the fuzzy theory was used to divide the QAR data into the vertical flight phase data, and the subjective influence on prediction accuracy was eliminated by artificially dividing the training data. Finally, the XGBoost prediction model of aeroengine parameters was established. Compared with various prediction models in the prediction of aeroengine N1 and fuel flow parameters, experimental results show that the XGBoost prediction model has higher accuracy than support vector regression (SVM), liner regression model and BP neural network, which does not require the scaling of the training data.

Keywords: aeroengine; fuzzy theory; XGBoost; QAR data; performance parameter prediction

0 引言

航空发动机是飞机的“心脏”, 是保障大型民用航空飞机飞行安全的核心部件, 目前航空发动机的健康监测方法大都采用航后诊断的方法, 主要对航空发动机的气路参数、振动参数和油液参数, 通过异常的参数找到飞机发动机潜在的风险状态数据, 进而判断发动机中与气路有关的单元体和子系统状态是否存在异常作为故障诊断的依据^[1]。对于航后飞机发动机的故障诊断存在延迟性, 若能根据飞机

发动机的性能参数变化趋势, 预判航空发动机的超限状态, 可防患于未然, 提高民航运行的安全性。航空发动机性能参数的预测在发动机健康管理、视情维修、设计研发、建模仿真等领域起到关键性作用^[2-5], 对航空发动机寿命预测有助于提高飞机发动机安全性, 降低航空公司的维护成本^[6]。

发动机性能参数预测是指通过一定的技术手段对表征发动机工况的主要性能参数, 如 N2 (高压转子转速)、N1 (风扇转速)、FF (燃油流量)、EGT (排气温度) 进行监

收稿日期: 2022-10-18; 修回日期: 2022-11-19。

基金项目: 天津市自然科学基金(21JCQNJC00710)。

作者简介: 樊智勇(1979-), 男, 山西大同人, 硕士, 教授, 硕士研究生导师, 主要从事航空电子设备自动化测试与维修方向的研究。

通讯作者: 刘哲旭(1987-), 男, 辽宁葫芦岛人, 博士, 讲师, 主要从事飞机系统仿真及测试方向的研究。

引用格式: 樊智勇, 王振良, 刘哲旭. 基于 XGBoost 的民航飞机发动机性能参数预测模型[J]. 计算机测量与控制, 2023, 31(6): 46-52.

测，并根据实际监测数据建模分析，以获取各部件工作状态在不同外界环境下的变化趋势。航空发动机的性能参数变化趋势能够客观反映航空发动机的性能衰退状况^[7]。由于航空发动机工作过程复杂，工作环境多变，且核心部件长期处于高速旋转、高温状态下，其性能参数呈现非线性、非平稳性、波动性等特点，给发动机的性能趋势预测带来较大困难。传统的预测方法如多元线性回归^[8]、自回归滑动平均模型^[9]，其预测精度不高，且鲁棒性不强。随着机器学习的快速发展，人工神经网络和支持向量回归（SVM，support vector regression）广泛应用于民航飞机发动机性能参数预测。文献 [10] 采用改进量子粒子群优化支持向量回归对 V2500 发动机的排气温度进行预测；文献 [11] 建立了基于支持过程向量机的航空发动机排气温度预测模型；文献 [12] 利用模糊信息粒化方法对性能参数进行粒化处理，并采用优化的 SVM 对发动机 N2 参数进行了预测验证；文献 [13] 利用堆叠降噪自动编码器构建了燃油流量的预测模型，并使用真实飞行数据进行训练和验证；文献 [14] 从提高训练样本集的质量角度，提出一种基于样条函数拟合和相空间重构的训练集构造方法，并结合过程神经网络，对发动机的性能参数进行预测，取得了不错的效果。但是，支持向量回归存在核函数选取的困难，需要根据经验和反复试验确定，存在较大的随机性。使用神经网络预测存在收敛速度慢、训练时间长、需要大量训练数据样本、容易陷入局部最优等问题。XGBoost（eXtreme gradient boosting）算法是在 GBRT（gradient boosting regression tree）的基础上对 boosting 算法的一种改进，Chen^[15]等详细说明了 XGBoost 算法的原理并且证明了 XGBoost 计算速度优于传统的 GBRT 算法。除此之外，很多学者使用 XGBoost 与传统的 SVM、ANN 方法等进行比较皆得出 XGBoost 预测准确性更高，可以在相对较少的调参时间下，得到较高预测准确率^[16]，且对异常值的鲁棒性较强，不需要对原始数据进行标准化处理，是目前机器学习领域求解预测问题的常用模型之一。

对民航飞机发动机性能参数的预测，本文采用快速存取记录器（QAR，quick access recorder）记录的真实飞行数据作为训练样本集，QAR 数据记录的是整个飞行过程的各项飞行参数，包含各个飞行阶段。为提高预测模型的泛化能力，在构建发动机性能参数训练集时，采用模糊推理按照飞机的垂直飞行状态进行划分，进而构建训练集，消除人为划分数据集的主观影响。最后，基于 XGBoost 建立适用不同飞行阶段的发动机性能参数预测模型，对不同飞行状态下的发动机工作状况进行模拟。

1 建模基础

1.1 总体性能分析

民航飞机发动机是一个集热力学、机械、控制等多学科技术于一身的复杂系统，航空发动机是一个复杂的非线性系统，为了与之发动机的性能参数变化趋势采用总体法

对其进行性能分析即不对发动机内部的物理过程进行详细地数学描述，而是将整台发动机看作一个“黑盒”模型。模型的输入量包括发动机控制输入及飞机自身状态、外界环境等影响因素，输出量主要包括表征发动机推力和工作状况的性能参数。输入量与输出量之间的映射关系即为发动机的总体性能模型^[17]。通过对航空发动机工作原理的分析，选取油门杆位置（TLA）作为控制输入；根据民航飞机的性能手册，确定影响发动机性能的因素主要包括马赫数（Mach）、外界大气总温（TAT）、飞机总重（GW）、飞行高度（ALT）。另外，考虑到当前大型民航客机均装有自动推力系统，例如波音 737 系列、空客 A320 系列飞机，当飞机处于不同的飞行阶段时，自动推力系统采用不同的推力控制模式，其推力控制律会发生变化。因而，发动机的性能参数的变化规律会随着飞行阶段的改变有所不同，故将飞行阶段同样作为影响因素。图 1 为航空发动机总体性能模型结构图。

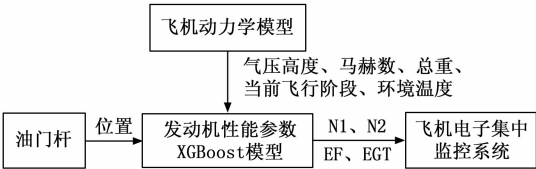


图 1 发动机总体性能模型结构

据上述分析，在构建 XGBoost 发动机性能参数预测模型时，选取油门杆位置、气压高度、大气总温、飞机总重以及马赫数作为输入特征变量，以所要预测的性能参数为状态变量，对模型进行训练。

1.2 飞行数据预处理

QAR 是飞行数据记录系统的设备之一，记录数据量大，可靠性高。相比“黑匣子”记录的飞行数据，QAR 数据存取方便可靠，读取速度快，具有更高的数据获取效率，更便于维护人员存取。本文的试验和验证数据来自某航空公司波音 737-800 客机，发动机型号为 CFM56-7B。

QAR 记录的飞行数据，其部分起飞爬升阶段的数据如表 1 所示。

表 1 737—800 型飞机部分起飞爬升阶段的数据

	D1	D2	D3	D4	...	Dn
ALT/ft	−223	−270	−197	−175	...	−485
TAT/℃	−2	1.5	4	7.25	...	−6
GW/lbs	143 200	142 160	148 400	153 840	...	159 760
TLA/deg	46	46	46	47	...	48
Mach	0.239	0.25	0.243	0.253	...	0.256
N1/%	91.75	91.63	92.38	93.75	...	92.88
N2/%	95.25	95.25	95.75	96.5	...	95.25
FF/(lb/h)	9 264	9 072	9 264	9 600	...	9 888
EGT/℃	742	738	751	774	...	742

表 1 包含飞机飞行高度、大气温度、飞行马赫数、发动机转速、排气管温度等基本参数。利用数据译码软件从机载记录器中获取 QAR 数据中的高度、总温、马赫数等所需参数。QAR 记录的飞行数据由于飞行环境的复杂多变,其中存在众多野值或频繁抖动,直接使用原始数据会包含野值这会影响发动机性能参数的预测精度,故在训练 XG-Boost 预测模型前先需要对飞行数据进行预处理,通过降噪滤波的方法将众多野值过滤进而排除野值对模型预测效果的影响。平稳小波变换和系数相关性原理相结合的方法处理飞行数据能较好地去除杂波^[18-19]。

小波阈值去噪是一种非线性的去噪方法,在最小均方误差意义下达到近似最有,可获得较好的去噪效果,小波阈值算法对飞行数据滤波平滑处理的基本原理为:QAR 数据经过小波变换分解后,得到一组小波分解系数,信号经过小波变换后的小波系数在各个尺度上有较强的相关性,而噪声数据经过小波变换后的小波系数相关性不强,得出信号的小波分解系数要大于噪声的小波分解系数,通过选取合适的阈值,将大于阈值的小波系数认为是信号产生的,予以保留,小于阈值的小波系数认为是噪声产生的,将其置零,最后将处理后的系数重构信号即可完成信号去噪。

1.3 飞行状态识别

民航飞机对其发动机的推力要求会根据当前飞行状态的不同发生改变。根据不同的飞行状态,民航涡扇发动机通常具备 5 种额定推力模式:起飞/复飞推力、最大连续推力、灵活起飞推力、最大爬升推力以及慢车推力。此外,现代大型民航客机均装有自动推力系统或者自动油门系统,在自动推力接通条件下,根据不同飞行状态,发动机推力的控制律也会发生相应的改变。因此,在不同的飞行状态下,民航发动机具有不同的运转状态和工作负荷,使得其主要的性能参数具有不同的变化规律或函数关系。故在根据真实飞行数据对发动机进行性能参数辨识过程时,需要对飞行数据进行飞行状态的识别,进而按照不同飞行状态进行建模分析。

民航飞机在爬升、下降飞行阶段中采用阶梯爬升和阶梯下降的方式,并不是连续地爬升及持续下降到目标高度。若仅从飞行阶段的定义上去简单地划分飞行数据,简单的将其分为起飞段、爬升段、巡航段、下降阶段以及处于地面滑行阶段,会造成辨识结果存在较大误差的问题。因此,需要将爬升和下降阶段中的阶梯平飞阶段单独进行考虑,同理若在飞机巡航阶段出现飞行高度改变时,需要将其中的爬升段和下降段单独考虑,若将该部分发动机性能数据视作巡航状态,会造成辨识结果具有较大误差。由于飞行阶段对飞机飞行状态的识别没有明确的界限,因此本文采用模糊推理^[20]的方式进行飞行阶段的识别。

1.3.1 输入、输出量的模糊化

如图 2 所示,模糊计算的过程可分为 4 个模块:模糊规

则库、模糊化、推理方法和去模糊化。模糊规则库是专家提供的模糊规则包括数据库和规则库。输入参数是根据选定的 QAR 参数设定论域进行模糊化,模糊化是根据隶属度函数从具体的输入得到对模糊集隶属度的过程,根据模糊规则进行模糊推理,经过模糊推理得到模糊结论,在进行去模糊化完成飞行状态的识别,去模糊化也称解模糊化是将模糊结论转化为具体的、精确的输出过程。

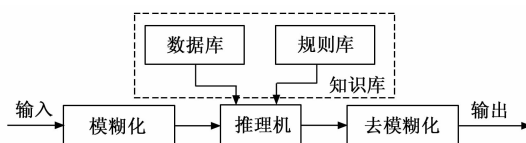


图 2 模糊计算的过程图

QAR 记录数据的频率为 1 Hz, 每条记录包含数百个飞行参数在某一时刻的值。从 QAR 数据中选取标准气压高度、垂直速度、空速为特征参量用于判断该条记录数据属于哪个飞行状态。飞行状态的模糊推理模型的输入主要有: 标准气压高度、垂直速度、空速, 将其进行模糊化, 隶属度取值为 $[0, 1]$ 。通过隶属度函数将垂直速度分为负(NV)、零(ZV)、正(PV)三个模糊集合, 其论域取值范围为 $[-5\,000, 5\,000]$ ft/min; 采用“Z”形、“S”形及三角形隶属度函数对气压高度、垂直速度、空速参量进行处理。其中, 垂直速度的隶属度函数如图 3 所示。

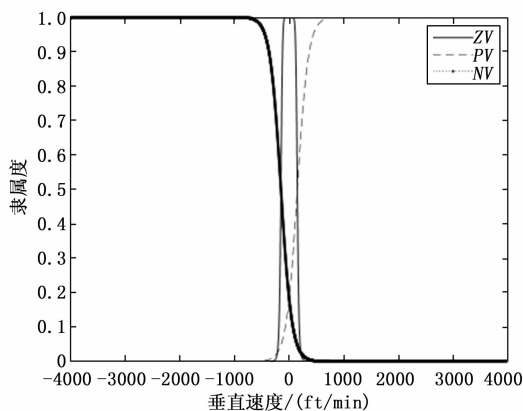


图 3 垂直速度的模糊隶属度函数

同理, 将标准气压高度分为低(LH)、中(MH)、高(HH)三个模糊集合, 其论域取值范围为 $[-2\,000, 40\,000]$ ft; 将空速分为低(LV)、中(MV)、高(HV)三个模糊集合, 其论域取值范围为 $[0, 400]$ knots。以飞行阶段为输出, 将其划分为滑行/停止(T/S)状态、起飞爬升(CLB)状态、阶梯平飞(LEV)状态、巡航(CRU)状态以及下降(DES)状态, 其论域取值范围为 $[0, 5]$ 。

1.3.2 模糊推理

根据每个飞行阶段飞机高度、垂直速度、空速参量的特点, 采用“if-then”形式制定模糊规则, 用于推导每条数

据属于那个飞行阶段。具体的模糊规则见表 2。根据模糊规则库, 采用 Mamdani 模糊推理法计算每条记录数据的总输出隶属度函数。

1.3.3 去模糊化

去模糊化就是将模糊结论转化为具体的、精确的输出过程。采用最大隶属度法进行去模糊化, 以得到该记录数据所处的飞行阶段。由于模糊集合通常是由多个模糊子集的并集构成的, 它的隶属度曲线中可能有多处的隶属度都取最大值, 故在最大隶属度的基础上, 采取最大值法选择模糊集合代表点。

表 2 模糊推理规则表

序号	气压高度	空速	垂直速度	飞行状态
1	LH	LV	ZV	T/S
2	MH	MV	PV	CLB
3	HH	HV	ZV	CRU
4	MH	MV	NV	DES
5	MH	MV	ZV	LEV
6	HH	HV	PV	CLB
7	HH	MV	NV	DES

通过模糊推理的方式来进行飞行阶段的划分, 可以很好地解决真实飞行数据中飞行阶段界限不明确的问题, 并且排除以往人工划分飞行阶段对预测结果的主观影响。

2 基于 XGBoost 的民航发动机性能参数建模

2.1 XGBoost 原理

XGBoost 是一种 Boosting 集成学习算法, 它通过不断迭代, 生成新的树来拟合前一棵树的残差, 与其他算法相比 XGBoost 不仅使用一阶导数, 而且还使用二阶导数, 损失更精确^[21]。从本质上来看, XGBoost 是由多个弱分类器组合而成的学习器:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \tag{1}$$

其中: $\hat{y}_i$ 为第 i 个样本的模型预测值, K 为树的数量, f_k 为第 k 棵树, x_i 为输入的第 i 个样本。XGBoost 通过最小化损失函数 $l(y_i, \hat{y}_i)$ 和正则函数 $\Omega(f_k)$ 来寻找最佳的函数集, 其目标函数为式 (2) 和 (3):

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \tag{2}$$

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T \omega_j^2 \tag{3}$$

其中: 损失函数为预测值 $\hat{y}_i$ 与真实值 y_i 的误差, 正则函数代表每棵树的复杂度, T 为叶子节点的数量, ω 为叶子节点权重, γ 为新叶子节点复杂度代价, λ 可以控制叶子节点权重不会过大, 防止过拟合。由式 (1) 可知, 第 t 轮预测结果为第 $t-1$ 轮预测结果与第 t 次迭代产生的决策树函数之和, 故式 (2) 变为:

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{4}$$

对损失函数进行二阶泰勒展开得:

$$Obj^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{5}$$

其中: $g_i = l'(y_i, \hat{y}_i^{(t-1)})$ 为损失函数的一阶导数, $h_i = l''(y_i, \hat{y}_i^{(t-1)})$ 为损失函数的二阶导数。又因 $\sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)})$ 为常数项, 不影响求最优解, 所以目标函数可简化为:

$$Obj^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{6}$$

将决策树参数代入目标函数, 定义第 j 叶的样本集如下: $I_j = \{i | q(x_i) = j\}$, 目标函数变为:

$$Obj^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \tag{7}$$

定义 $G_j = \sum_{i \in I_j} g_i$ 与 $H_j = \sum_{i \in I_j} h_i$, 则目标函数变为:

$$Obj^{(t)} = \sum_{j=1}^T \left[G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma T \tag{8}$$

对于固定结构 $q(x)$, 可通过顶点式 (9) 计算第 j 叶子最优权重 ω_j^* :

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \tag{9}$$

将式 (9) 代入式 (8) 即可得到目标函数最小值, 其值越小, 树的结构越好, 此时即为目标函数的最优解:

$$\arg \min_x (Obj) = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{10}$$

2.2 建模流程

采用 XGBoost 建立发动机主要性能参数的流程如图 4 所示, 具体步骤如下:

- 1) 通过总体性能分析, 从真实飞行数据中选取油门杆的位置、标准气压高度、空速、大气总温及飞机总重作为输入;
- 2) 采用小波变换和系数相关性原理相结合的方法对原始数据进行滤波处理以剔除其中的野值, 并实现数据平滑;
- 3) 采用模糊推理对预处理后的数据进行状态划分, 并对其进行标准化, 然后分别构建爬升阶段数据集、巡航阶段数据集、阶梯平飞阶段数据集、下降阶段数据集;
- 4) 对每一飞行阶段的数据集, 按照 7: 3: 2 的比例来随机构建训练集、验证集和测试集;
- 5) 使用训练集数据训练模型得到一个 XGBoost 模型, 并采用交叉验证的方式, 使用验证集对该 XGBoost 模型进行评估, 最后得到一个近似最优的 XGBoost 模型;
- 6) 最后, 将测试集数据输入到步骤 5) 得到的模型, 检验其拟合精度。若模型精度不满足误差要求, 则重返第 4) 步, 重新划分数据集进行训练。

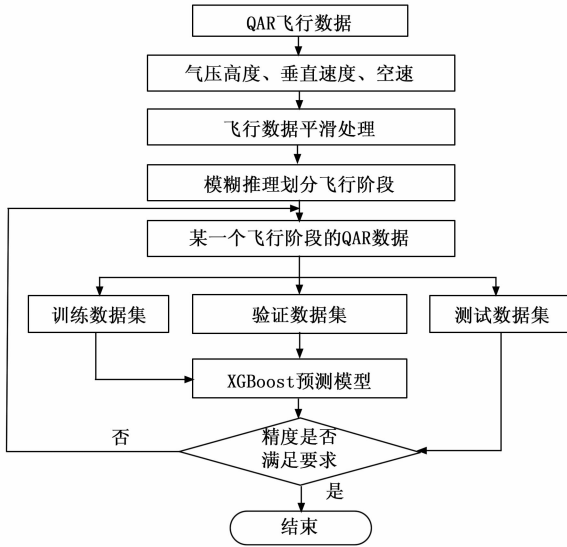


图 4 XGBoost 发动机性能参数建模流程

3 实验与分析

3.1 模型建立

由于构建不同飞行阶段的发动机性能参数模型时的区别在于数据集的不同,其所采用的建模方法是一致的。因此,以爬升阶段、下降阶段为例详细阐述建模流程。试验数据选取 120 个航班的飞行数据,用于构建发动机主要性能参数的 XGBoost 模型,其中 70 个航班数据用作模型的训练集,30 个航班数据用作模型的验证集,20 个航班数据用作模型的测试集。飞机通常装有两套相同的发动机引气系统,本文选用的数据均是在两套引气系统正常工作情况下记录的。在确定的引气状态下,对发动机工作状态的影响因素可以归结为如下几个特征变量,如式 (11) 所示。采用 XGBoost 建模的目标就是构建一个拟合函数关系,如式 (12) 所示,用于仿真飞机在爬升阶段,随着外界条件改变其发动机性能参数的变化规律。

$$X = \{TLA, MACH, ALT, GW, TAT\} \quad (11)$$

$$N1 = f_1(TLA, MACH, ALT, GW, TAT)$$

$$FF = f_2(TLA, MACH, ALT, GW, TAT)$$

$$N2 = f_3(TLA, MACH, ALT, GW, TAT) \quad (12)$$

$$EGT = f_4(TLA, MACH, ALT, GW, TAT)$$

对于不同的数据集样本,各类回归预测算法的表现不同。为验证 XGBoost 算法在发动机性能参数预测方面具有的优势,本文分别构建爬升和下降阶段发动机性能参数的数据集,并进行对比试验,以说明该算法的有效性。分别采取线性回归 (LR, liner regression)、BP 神经网络以及支持向量回归对该样本集进行建模,并在验证集上进行测试,得到了各自最优的训练模型。

采用决定系数 R^2 、平均绝对误差 (MAE, mean absolute error,) 及均方误差 (MSE, mean square error) 为衡

量标准,来评估模型输出的预测值与真实值之间的偏差,如公式 (13) 所示:

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (\bar{y}_i - y_i)^2}$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (13)$$

3.2 爬升阶段预测结果分析

选取真实飞机从 15 000 ft 高度爬升到 30 000 ft 高度的一段真实爬升数据,该数据未用作模型训练,对所建立的 XGBoost 发动机性能参数预测模型进行检验。真实飞机影响发动机性能参数的状态变量变化,如图 5 所示。将其作为输入,对所建立的模型进行检验。

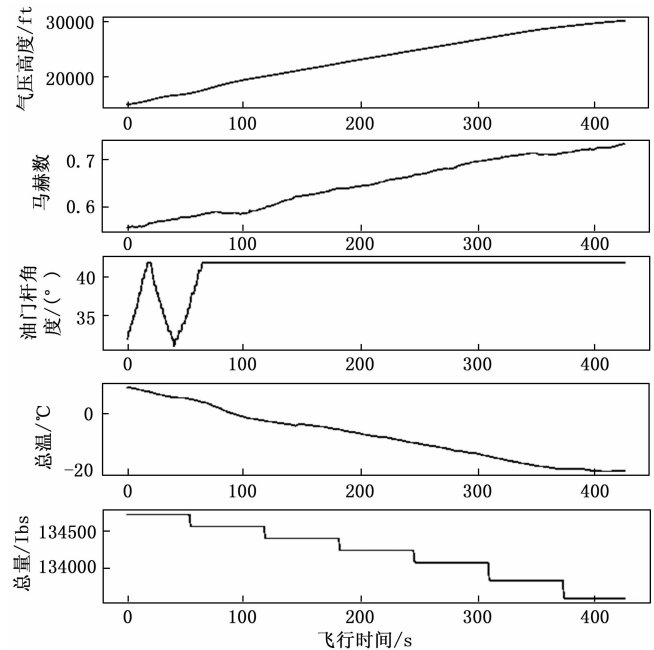


图 5 爬升阶段飞机状态参数变化趋势

将如图 5 所示数据,输入到所建立的性能参数辨识模型中,在民航飞机从 15 000 ft 爬升到 30 000 ft 高度过程中,发动机 N1、燃油流量参数的预测结果与真实记录的数据对比,如图 6~7 所示,观测值即为真实数据,图中用不带标记的实线表示,支持向量回归 (SVM) 预测模型曲线在图中用附带实心小黑圆的实线表示、XGBoost 算法的预测曲线在图中用附带实心的黑五角星的实线表示、线性回归模型 (LR) 预测曲线在图中用附带实心黑三角的实线表示、BP 神经网络预测曲线在图中用附带实心小黑矩形的实线表示。

通过对比分析图 6 和图 7 可知飞机爬升阶段对发动机 N1、燃油流量参数预测结果,对比四种预测算法和观测值

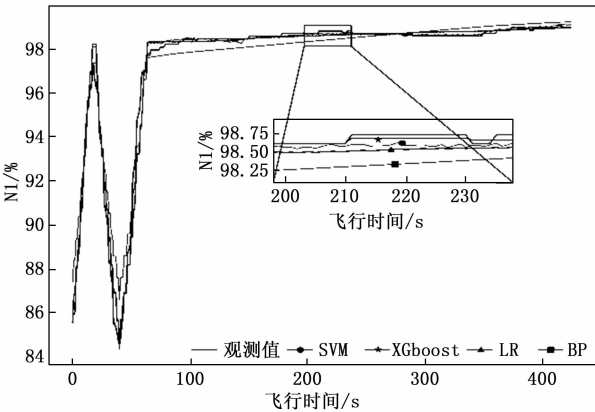


图 6 爬升阶段发动机 N1 参数预测结果对比

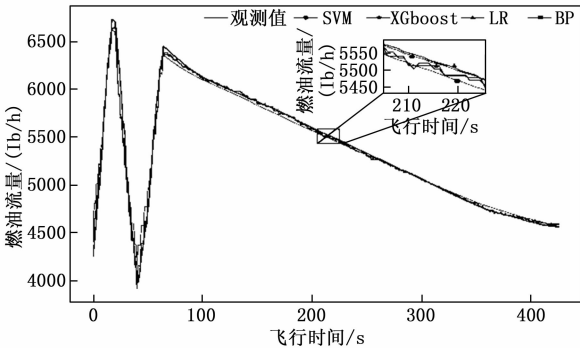


图 7 爬升阶段发动机燃油流量参数预测结果对比

曲线的跟随性可得出: XGBoost 模型预测曲线、SVM 模型预测曲线、BP 模型预测曲线、LR 模型预测曲线中 XGBoost 模型预测曲线更接近观测值, 预测结果更好。

表 3 爬升阶段各模型对 N1 的预测结果评估

模型	R^2	MAE	MSE
线性回归	0.981 2	0.221 6	0.172 7
SVM	0.982 0	0.193 8	0.165 7
BP 神经网络	0.953 0	0.450 9	0.433 6
XGBoost	0.997 3	0.024 7	0.002 5

其中: R^2 取值范围为 $[0, 1]$, 其值越靠近于 1, 则说明模型的解释性越强; MAE、MSE 的值越小, 表示模型预测效果越好。从表 3 结果中可以看出, XGBoost 对该数据集的预测效果, 相比线性回归、SVM、BP 神经网络, 其 R^2 最接近于 1, MAE 和 MSE 的值最小, 则说明 XGBoost 预测模型对爬升阶段 N1 参数的预测精度较高。

3.3 下降阶段预测结果分析

选取飞机从 30 000 ft 高度下降到 3 000 ft 高度的一段真实下降数据, 对所建立的下阶段 XGBoost 模型进行检验。该过程中飞机的马赫数、油门杆输入、外界大气总温以及飞机自身重量的变化, 如图 8 所示。各种预测模型对该过程中发动机 N1、燃油流量的预测效果如图 9、图 10

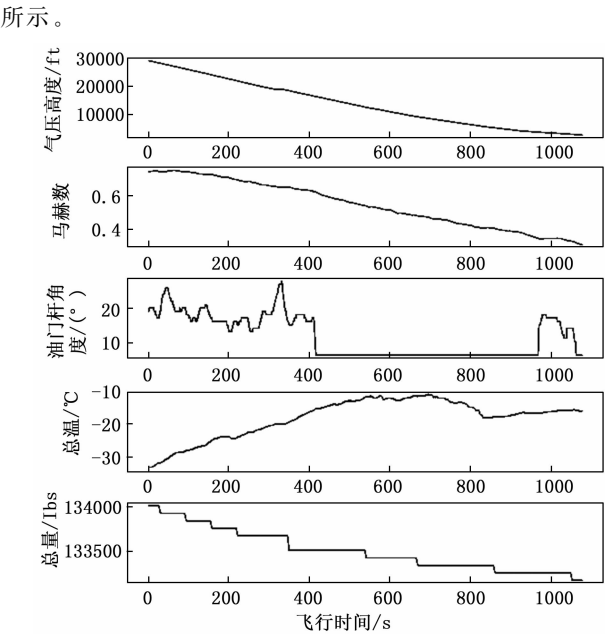


图 8 下降阶段飞机状态参数变化趋势

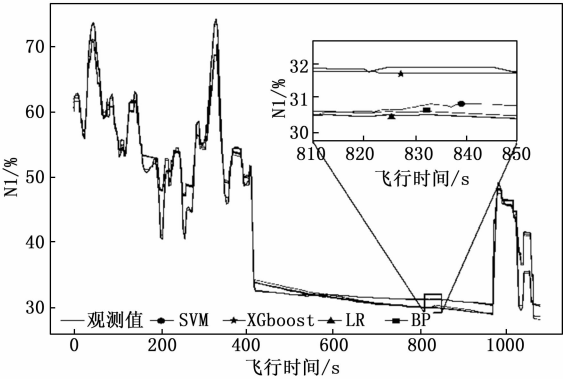


图 9 下降阶段发动机 N1 参数预测结果对比

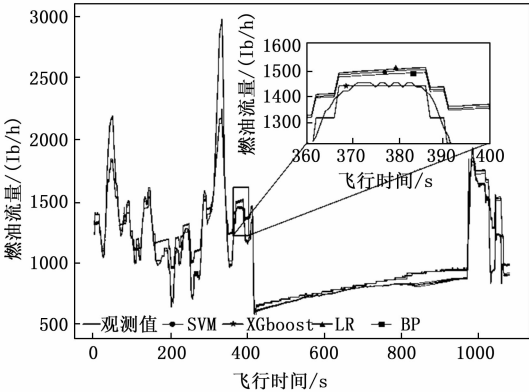


图 10 下降阶段发动机燃油流量参数预测结果对比

观测值即为真实数据, 图中用不带标记的实线表示, 支持向量回归 (SVM) 预测模型曲线在图中用附带实心小黑圆的实线表示、XGBoost 算法的预测曲线在图中用附带实心的黑五角星的实线表示、线性回归模型 (LR) 预测曲

线在图中用附带实心黑三角的实线表示、BP 神经网络预测曲线在图中用附带实心小黑矩形的实线表示。

通过对比分析图 9 和图 10 可知飞机下降阶段对发动机 N1、燃油流量参数预测结果，比较四种预测算法和观测值曲线的跟随性可得出：XGBoost 模型预测曲线、SVM 模型预测曲线、BP 模型预测曲线、LR 模型预测曲线中 XGBoost 模型预测曲线更接近观测值，预测结果更好。

表 4 下降阶段各模型对 N1 的预测结果评估

模型	R^2	MAE	MSE
线性回归	0.960 9	1.497 1	4.468 7
SVM	0.962 1	1.469 2	4.3277
BP 神经网络	0.961 3	1.489 7	4.451 9
XGBoost	0.988 8	0.254 4	0.184 9

从表 4 结果中可以看出，XGBoost 对下降阶段数据集同样具有较好的预测效果，相比线性回归、SVM、BP 神经网络，其 R^2 最接近于 1，MAE 和 MSE 值最小，说明 XGBoost 预测模型对爬升阶段 N1 参数的预测精度较高。

4 结束语

航空发动机性能参数预测是故障预判和视情维修的决策前提，本文将集成学习算法中的 XGBoost 用于发动机性能参数预测模型的构建，在构建模型训练数据集时，考虑了飞行阶段对预测精度的影响，提出一种基于模糊推理和 XGBoost 算法的发动机性能参数预测方法。

利用真实民航飞行数据对预测模型进行验证，并将其与传统的 BP 神经网络算法、支持向量回归、线性回归模型在预测精度上进行了比较。实验结果表明，基于 XGBoost 的发动机性能参数模型的精度较高，同时不需要对数据进行缩放。本文对航空大飞机航空发动机性能参数的研究对通用航空机载信息网络数据共享技术中数据处理和飞行阶段划分有借鉴意义。

参考文献：

[1] 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.

[2] KIAKOJOORI S, KHORASANI K. Dynamic neural networks for gas turbine engine degradation prediction, health monitoring and prognosis [J]. Neural Computing and Applications, 2016, 27 (8): 2157 - 2192.

[3] GIORGI M G D, CAMPILONGO S, FICARELLA A. A diagnostics tool for aero-engines health monitoring using machine learning technique [J]. Energy Procedia, 2018 (148): 860 - 867.

[4] WANG C, LI Y G, YANG B Y. Transient performance simulation of aircraft engine integrated with fuel and control systems

[J]. Applied Thermal Engineering, 2017 (114): 1029 - 1037.

[5] ZHANG C X, WANG N. Aero-engine condition monitoring based on support vector machine [J]. Physics Procedia, 2012, 24: 1546 - 1552.

[6] 刘 康, 肖 娜. 基于堆叠稀疏自编码神经网络的航空发动机剩余寿命预测方法研究 [J]. 计算机测量与控制, 2019, 27 (12): 29 - 33, 38.

[7] 孙绍辉, 王华伟, 陈福立. 多元退化信息的航空发动机可靠性预测 [J]. 火力与指挥控制, 2013, 38 (11): 32 - 35.

[8] 宋云雪, 张科星, 史永胜. 基于多元线性回归的发动机性能参数预测 [J]. 航空动力学报, 2009, 24 (2): 427 - 431.

[9] 常博博, 毛 宁, 刘 骁, 等. 基于 ARIMA 模型的航空发动机性能预测 [J]. 自动化应用, 2013 (5): 45 - 46, 53.

[10] 皮 骏, 黄江博, 黄 磊, 等. 基于改进 QPSO-SVR 的航空发动机排气温度预测 [J]. 振动 · 测试与诊断, 2019, 39 (2): 267 - 272, 440.

[11] 于广滨, 丁 刚, 姚 威, 等. 基于支持过程向量机的航空发动机排气温度预测 [J]. 电机与控制学报, 2013, 17 (8): 30 - 36.

[12] 李艳军, 张 建, 曹愈远, 等. 基于模糊信息粒化和优化 SVM 的航空发动机性能趋势预测 [J]. 航空动力学报, 2017, 32 (12): 3022 - 3030.

[13] 黄 曦, 卿新林, 王奕首, 等. 基于 SDAE 的航空发动机燃油流量基线模型构建 [J]. 航空发动机, 2019, 45 (2): 29 - 33.

[14] 丁 刚, 付旭云, 钟诗胜. 基于过程神经网络的航空发动机性能参数预测 [J]. 计算机集成制造系统, 2011, 17 (1): 198 - 207.

[15] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system [C] //Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco: ACM, 2016: 785 - 794.

[16] 徐文英, 王大军, 卢朝阳, 等. 基于 XGBoost 算法的终端区进场航空器飞行时间预测 [J]. 北京交通大学学报, 2022, 46 (6): 72 - 79.

[17] 熊海国. 飞行模拟器发动机系统建模与面向对象的仿真研究 [D]. 哈尔滨: 哈尔滨工业大学, 2010.

[18] 李正欣, 张凤鸣, 张晓丰, 等. 基于平稳小波变换的飞行数据去噪方法 [J]. 计算机应用, 2011, 31 (10): 2790 - 2792, 2810.

[19] 邸亚洲, 李富荣, 于建立, 等. 小波分析在飞参数据降噪中的应用 [J]. 计算机仿真, 2010, 27 (10): 1 - 4.

[20] 赵振峰, 白晓虎, 陈 强, 等. 基于模糊集合理论的体积压裂水平井产量预测方法 [J]. 石油钻采工艺, 2019, 41 (4): 521 - 528.

[21] CHEN T Q, GUESTIN C. XGBoost a scalable tree boosting system [C] //Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 16) . Association for Computing Machinery, New York, USA, 2016: 785 - 794.