



DOI:10.12404/j.issn.1671-1815.2403356

引用格式:陈明强,郑文浩,孙雁君,等.基于回归树和随机森林的通航飞机燃油消耗预测[J].科学技术与工程,2025,25(7):3026-3034.

Chen Mingqiang, Zheng Wenhao, Sun Yanjun, et al. Fuel consumption prediction of general aviation aircraft based on regression tree and random forest[J]. Science Technology and Engineering, 2025, 25(7): 3026-3034.

航空、航天

基于回归树和随机森林的通航飞机燃油消耗预测

陈明强,郑文浩,孙雁君,林浩冬,段中航

(中国民用航空飞行学院飞行技术学院,广汉 618307)

摘要 现有研究中影响燃油消耗的特征值选择通常没有明确标准,研究成果难以与实际飞行相结合。基于塞斯纳 172 的飞行训练数据对通航教练机空中阶段的燃油消耗做出预测。首先,基于作者的飞行经验以及相关分析,从飞行员操作角度选择影响燃油流量的特征值。其次,使用回归树模型拟合不同飞行状态下的燃油流量,并将飞机实际飞行状态与燃油流量预测值相对应,便于后期从飞行技术层面研究具体的节油策略。最后,使用经过超参数寻优的随机森林模型对燃油流量做出预测。实验结果表明,本文所使用的模型精度优于现有研究成果,平均绝对误差为 0.286 gallon/h,均方根误差为 0.496 gallon/h,残差平方和为 0.968 4,平均绝对百分比误差为 4.00%。

关键词 回归树;随机森林;机器学习;燃油流量预测;塞斯纳 172;数据记录系统

中图分类号 V260.5;

文献标志码 A

Fuel Consumption Prediction of General Aviation Aircraft Based on Regression Tree and Random Forest

CHEN Ming-qiang, ZHENG Wen-hao, SUN Yan-jun, LIN Hao-dong, DUAN Zhong-hang

(School of Flight Technology, Civil Aviation Flight University of China, Guanghan 618307, China)

[Abstract] The selection of variables affecting fuel consumption in the existing studies usually has no clear criteria, and it is difficult to combine the research results with actual flight. The flight training data of a Cessna 172 was used to predict the fuel consumption during the airborne phase of general aviation trainer aircraft. Firstly, based on the authors' flight experience as well as correlation analysis, the features that influence fuel flow rate were selected from the pilot's operational perspective. Secondly, a regression tree model was used to predict fuel flow rate under different flight conditions, correlating the aircraft's actual flight status with the predicted fuel flow rate, in order to facilitate subsequent research on specific fuel-saving strategies from the flight technique perspective. Finally, a random forest model optimized with hyperparameter tuning was used to predict the fuel flow rate. The experimental results show that the accuracy of the model used is better than that of the existing research results, with a mean absolute error of 0.286 gallon/h, a root mean squared error of 0.496 gallon/h, a residual sum of squares of 0.968 4, and a mean absolute percentage error of 4.00%.

[Keywords] regression tree; random forest; machine learning; fuel flow rate prediction; Cessna 172; data recording system

目前,中国航空业正迅速发展。通用航空业在中国潜在需求大,涉及农业、林业、牧业、渔业、工业、建筑、科研、交通、娱乐等多个领域。早在“十三五”期间,中国就提出了通用航空和运输航空“双翼齐飞”的指导方针;近期印发的《“十四五”通用航空发展专项规划》进一步明确了未来通航产业发展的主要目标、重点任务和相关保障措施。

飞行员培训是中国通用航空的核心业务之一。现基于常用教练机塞斯纳 172 的飞行训练数据,利用机器学习的模型研究通航飞机的燃油消耗。燃

油消耗是航空运营中的主要成本之一,优化燃油消耗可大幅降低运营成本,提高经济效益。燃油消耗也与碳排放直接相关,降低燃油消耗可以减少碳排放和环境污染。此外,了解燃油消耗规律,可以帮助优化飞行计划和航线选择,同时培养飞行学员的节能意识和技巧,从而在未来的实际飞行中实现更高效、环保的飞行操作。目前,关于民用航空器燃油消耗有较多研究成果,大多基于神经网络算法^[1]、聚类算法^[2-3]以及它们的改进算法^[4-7]来研究不同飞行阶段的燃油消耗。陈聪等^[8]利用从飞机

收稿日期:2024-05-07 修订日期:2024-08-01

基金项目:民航飞行技术与飞行安全重点实验室项目(FZ2020KF09,FZ2021ZZ06)

第一作者:陈明强(1976—),男,汉族,重庆人,硕士,副教授。研究方向:视觉导航。E-mail:mq_chen@cafuc.edu.cn。

投稿网址:www.stae.com.cn

快速存储记录器中获取的大量数据设计了一种利用循环神经网络(recurrent neural network, RNN)及其改进网络门控循环单元(gate recurrent unit, GRU)进行飞机燃油流量预测的模型,解决了循环神经网络对历史信息利用能力不足,极易发生梯度消失与梯度爆炸的问题。陈聪等^[9]提出了一种基于偏最小二乘回归(partial least squares, PLS)和粒子群算法(particle swarm optimization, PSO)优化的Elman神经网络算法对客机燃油流量进行预测,该算法弥补了Elman算法训练不完善、学习精度差的缺陷。Sedat^[10]采用了一种新型卷积神经网络结合双向长短期记忆深度学习模型(convolutional neural network-bidirectional long short-term memory, CNN-BiLSTM)对民航客机燃油流量进行预测,该模型精度优于长短期记忆网络(LSTM)和双向长短期记忆网络(BiLSTM)模型;Seymour等^[11]提出了一种基于环境和起降机场的双组件燃油预测方法,预测精度高且耗时短,解决了预测精度和计算效率无法兼顾的问题;Oruc等^[12]将CSA(cuckoo search algorithm)算法首次应用于飞机下降阶段的燃油消耗预测研究中,构建了一个预测精度较高的新模型。

民航客机常用的发动机(涡扇、涡桨)与本文所研究的塞斯纳172的发动机(活塞式)在原理、结构、性能方面有显著区别。燃油消耗是发动机的重要特征之一,不同类型的发动机在相同条件下可能有不同的油耗表现。目前,关于通航产业燃油消耗的研究相对较少,余芬等^[13]引入学习曲线构建了通用教练机燃油成本模型,通过蒙特卡洛法进行仿真分析,该方法与传统方法相比更为精确高效;Huang等^[14]利用分类与回归树模型(classification and regression tree, CART)以及SS-ANOVA(smoothing spline-analysis of variance)模型对通航飞机起降阶段

燃油流量进行预测;Huang等^[15]利用分类与回归树模型和神经网络模型对通航飞机的燃油消耗进行预测,使用FDR(flight data recorder)数据训练模型,利用ADS-B(automatic dependent surveillance-broadcast)数据验证模型,ADS-B数据可在飞行过程中实时获取。

现有研究对于影响燃油消耗的特征值选择通常没有明确标准,研究成果难以与实际飞行相结合。本文中从飞行员的操作角度选取特征值,基于通航教练机塞斯纳172的真实飞行训练数据,利用回归树和随机森林算法分析空中飞行阶段的燃油消耗,将飞机实际飞行状态与燃油流量预测值相对应,并对燃油流量进行预测。

1 数据

本文中整体工作流程如图1所示。首先,从塞斯纳172飞机上的飞行数据记录系统获得原始数据,通过分析相关飞行参数特点,提取出空中飞行阶段数据。基于相关性分析以及作者实际飞行经验选取发动机转速、垂直速度、真空速三个特征自变量。根据以往相关研究、机器学习模型特点以及数据特征,使用回归树和随机森林算法构建燃油流量预测模型。最后,计算燃油流量真实值和预测值之间的平均绝对误差(mean absolute error, MAE)、均方根误差(root mean squared error, RMSE)、平均绝对百分比误差(mean absolute percentage error, MAPE)以及 R^2 残差平方和值进行模型验证。

1.1 数据获取

本文中所采用的数据来自塞斯纳172飞机上的飞行数据记录系统。该机型配备的Garmin1000(以下简称“G1000”)航电系统^[16]以1 Hz的采样率记录着50多种有关飞机状态和外界环境的参数,这些

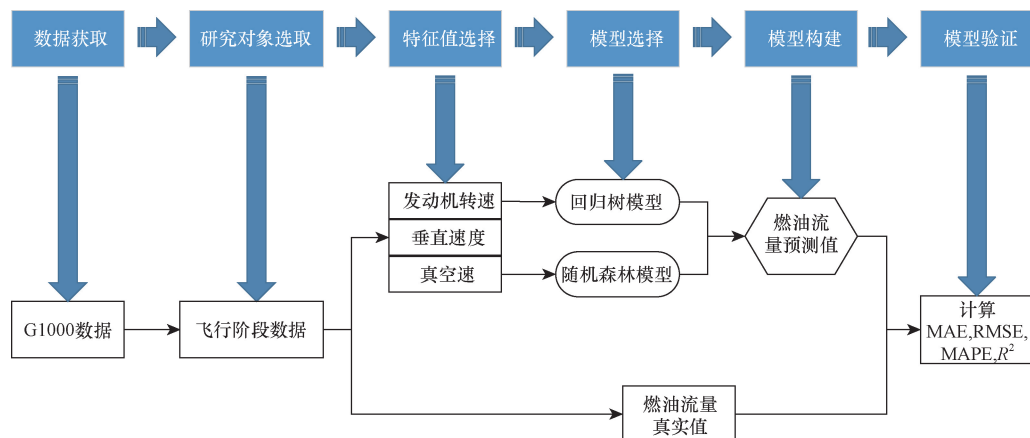


图1 整体工作流程

Fig. 1 Overall workflow

参数可大致分为:飞行参数、发动机参数、导航通信参数,表1列出了其中的一部分。

(1)飞行参数。飞行参数包括飞行环境参数、飞行操作参数以及飞行状态参数,反映了飞机在何种环境下通过何种操作处于何种状态之下。主要包括:风向、风速、外界大气温度、俯仰、横滚、地速、真空速、垂直速度、经纬度等。俯仰、横滚姿态可以反映飞机相对于自身横轴和纵轴的运动状态;地速、真空速、垂直速度可以反映飞机相对于地面和垂直方向上的运动状态。

(2)发动机参数。小型飞机上活塞发动机提供的推力是飞行的唯一动力来源。在飞行过程中,发动机参数是飞行员了解发动机工作状态的主要方式。发动机参数主要包括:汽缸头温度、排气温度、燃油流量、发动机转速、滑油流量、滑油温度等。燃油流量是本文关注的核心。

(3)通信导航参数。通信导航系统为飞机与地面之间的交流提供了保障,通过全球定位系统(global positioning system, GPS)系统为飞机提供实时位置并引导飞机飞往目的地,对保障飞行安全至关重要。通信导航参数主要包括通信频率、导航频率、磁航向等。

表1 G1000记录的主要飞行参数

Table 1 Main flight parameters recorded by G1000

类别	参数	来源	英文缩写	单位
飞行参数	平均海平面气压高度	直接测量	AltMSL	ft
	真空速	衍生参数	TAS	kt
	地速	衍生参数	GndSpd	kt
	垂直速度	衍生参数	VSpd	kt
	俯仰角	直接测量	Pitch	(°)
	横滚角	直接测量	Roll	(°)
	指示空速	衍生参数	IAS	kt
发动机参数	发动机转速	直接测量	E1 RPM	r/min
	气缸排气温度	直接测量	E1 EGT1	°F
	燃油流量	直接测量	E1 FFlow	gallon/h
	滑油压力	直接测量	E1 Oilp	psi
导航通信参数	航线角	直接测量	CRS	(°)
	磁航向	衍生参数	HDG	(°)
	航迹角	直接测量	TRK	(°)

注:1 ft = 0.304 8 m, 1 kt = 0.514 4 m/s, 1 gallon = 3.785 4 L, 1 psi = 6 894. 6 Pa。

1.2 研究对象选取

国际民航组织(International Civil Aviation Organization, ICAO)将民航客机飞行划分为五个阶段:推出滑行阶段、起飞离场阶段、巡航阶段、下降进近阶段以及落地和复飞阶段。民航客机爬升至巡航高度后就会接通自动驾驶并保持稳定平飞。由于客机的飞行阶段明确,识别、划分客机飞行阶段的方法也较为成熟,对于客机燃油消耗的研究多针对特定的飞行阶段。与民航客机不同,通航飞机在飞行训练中

会频繁的改变飞行高度,并且每一次飞行训练时所执行的科目不完全一致,加上手动操作和各种环境因素对小飞机的影响,最终导致SD卡中所记录的数据噪声较多,难以识别,划分明确的飞行阶段(图2展示了某次飞行训练任务中高度的剖面图)。此外,对于飞行训练的地面运行阶段来说,油耗与许多不确定因素有关,例如刹车的使用频率、使用轻重、地面粗糙程度、地面运行路线等等。地面操作的复杂多变会引起燃油流量的剧烈变化,因此本文中不考虑地面阶段的燃油消耗。综上所述,本文中除去地面阶段的所有飞行阶段整体作为研究对象。相较于空中飞行阶段,地面运行阶段高度、地速和俯仰角等参数有较明显的变化区间。本文通过分析飞行参数中较为重要的几个参数来确定地面运行阶段,处于两段地面运行阶段之间的就是空中飞行阶段。图2是单个csv文件里相关飞参(每个参数都是一个时间序列)的折线图,飞机处于地面运行阶段时,高度显示为机场标高。根据相关规定,飞机在地面运行时滑行速度不能超过10 kt,所以飞机的地速会在0~10 kt范围内变化。飞机停放在地面上时,飞机的重心位置和前三点式起落架决定了飞机会自带2°左右的仰角,所以飞机的姿态会在2°左右变化。

根据上文分析得到的地面运行阶段的参数特征,可定位每一段地面运行阶段数据的起点和终点,如图2虚线部分所示,前一部分地面运行段的终点就是相邻后一次空中飞行阶段的起点(绿色虚线),后一部分地面运行段的起点就是相邻前一次空中飞行阶段的终点(红色虚线)。在所绘制折线图中标记出的各起始点与原始数据表格中的数据一一对应,通过时间序列编号便可在原始数据中确定各空中飞行阶段的数据。

原始数据中共筛选出20次飞行训练任务,总飞行时间约17 h,即每个参数有61 869个观测结果。本文所使用数据未公开,表2列出了部分数据样例。

1.3 特征值选择

影响燃油流量的因素有很多,油耗预测模型的特征自变量如何选取并无明确标准。各参数与燃油流量之间的皮尔森相关系数的计算公式为

$$r = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y}) / (n - 1)}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2 / (n - 1)} \sqrt{\sum_{i=1}^n (y - \bar{y})^2 / (n - 1)}} \quad (1)$$

式(1)中: x 为自变量; y 为因变量; \bar{x} 为自变量均值; \bar{y} 为因变量均值; n 为样本总数。

计算结果按照由大到小的顺序整理在表3中。

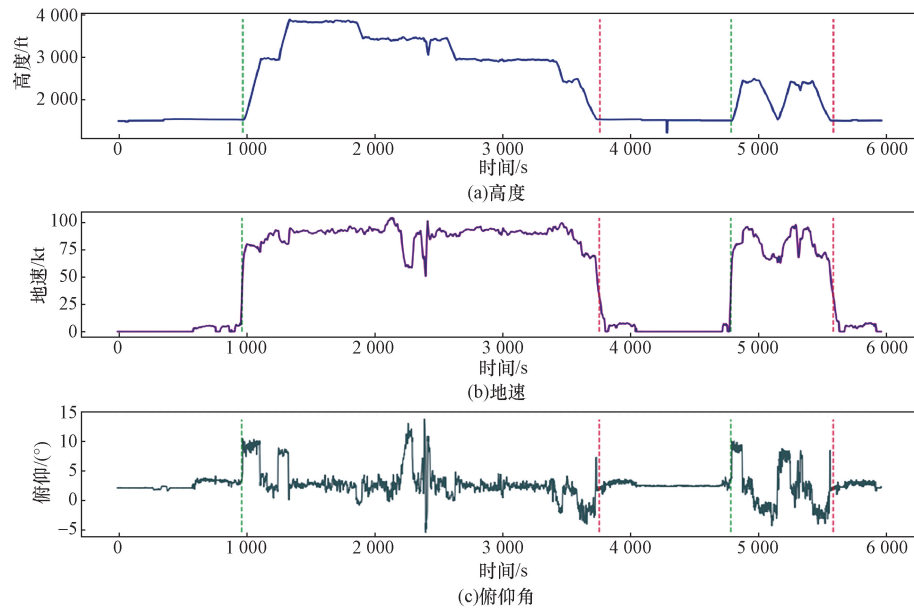


图2 某次飞行训练的高度、地速、俯仰变化

Fig. 2 Changes in altitude, ground speed and pitch during a flight training

表2 原始数据表

Table 2 Raw data table

当地时间	纬度/ (°)	经度/ (°)	高度/ ft	气压/ inHg	外界大 气温度/(°)	真空速/ kt	地速/ kt	垂直速度/ (ft·m ⁻¹)	俯仰/ (°)	横滚/ (°)	航向/ (°)	航迹/ (°)
10:09:07	30.469 49	105.614 0	894.4	30.36	11.8	30.24	30.93	-101.05	10.48	-0.52	307.9	306.4
10:09:08	30.469 59	105.613 9	893.4	30.36	11.8	32.21	33.8	-75.26	11.05	-0.62	308.1	306.6
10:09:09	30.469 68	105.613 7	893.4	30.36	11.8	33.77	36.54	-49.16	11.2	-0.67	307.8	306.8
10:09:10	30.469 79	105.613 5	893.4	30.36	11.8	35.31	39.09	-20.21	11.11	-0.56	307.7	306.6

注:1 inHg = 3 386.39 Pa; 1 ft = 0.304 8 m。

表3 Pearson 相关系数

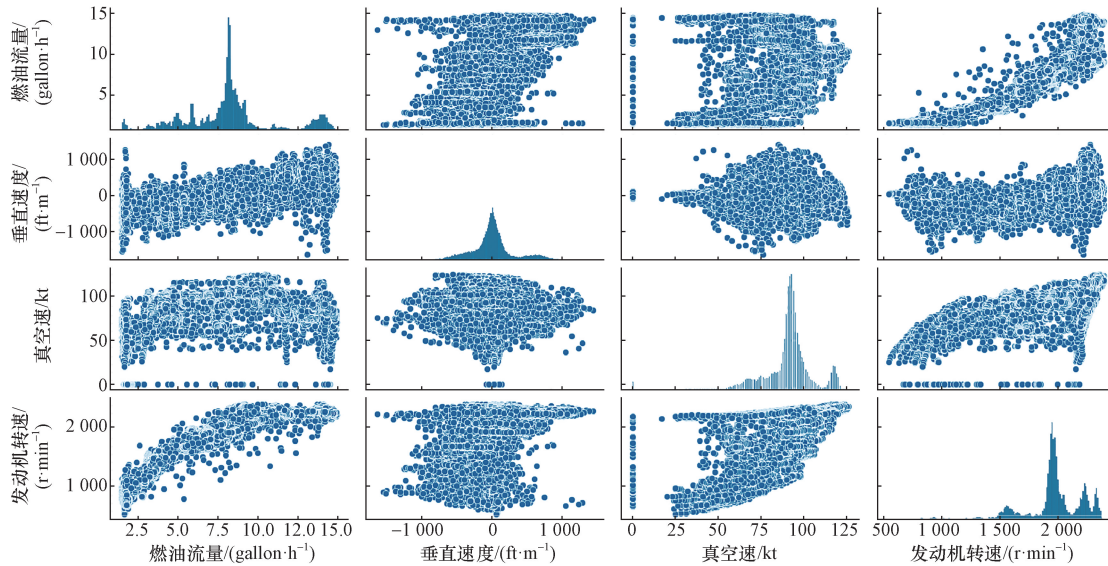
Table 3 Pearson correlation coefficient

参数	发动机转速	垂直速度	俯仰	真空速	表速	地速	横滚	高度
Pearson 相关性系数	0.819 3	0.778 6	0.634 9	0.137 3	0.131 8	0.103 4	0.082 0	0.065 7

对飞机油耗影响较大的因素可大致分为三个方面:飞机设计、操作方法、空管控制环境。飞机设计是飞机制造厂商关注研究的重点,空管控制环境是空管、签派、运行人员关注的重点,操作层面是飞行员关注的重点。在实际飞行训练中,发动机转速、垂直速度、真空速是飞行员操作层面主要去控制、观察的变量,飞行员可以从 PFD(primary flight display)和 MFD(multi-function display)面板上直接读出并且监控这些参数。发动机转速直接反映了飞机推力大小,垂直速度反映了飞机在垂直方向上的变化趋势,真空速反映了飞机沿自身纵轴运动的快慢。通过这些参数飞行员可以知道飞机当前所处飞行状态,通过油门、方向舵脚蹬以及操纵杆改变或保持这些参数使飞机机动并动态稳定在预选的航道、高度、速度,让飞机以飞行员预想的飞行状

态飞行。俯仰与垂直速度之间存在强线性相关,特征变量之间的强线性关系会直接影响模型的预测精度。综上所述,选择发动机转速、垂直速度、真空速作为本文的特征变量。

图3展示了燃油流量和发动机转速、垂直速度、真空速之间的两两相互关系。对角线上为各参数的直方图,而非对角线上是两个不同参数之间的关系图。一次函数状或团簇状图像代表两参数线性相关或非线性相关,根据图像的变化趋势可以判断两参数之间的正负相关性。通过观察不难发现:①燃油流量与各自变量之间都是比较复杂的非线性关系;②发动机转速和燃油流量之间呈现较为明显的正相关性,垂直速度、真空速与燃油流量之间存在一定程度的正相关性;③各自变量之间无明显相关性。



1 gallon = 3.785 4 L; 1 ft = 0.304 8 m

图3 各自变量与因变量之间的两两关系

Fig. 3 A pairplot between the respective variable and the dependent variable

2 模型选择

本文中利用回归树和随机森林模型,对通航教练机的燃油流量进行预测。两种模型均有较强的鲁棒性且适用于非线性关系的变量,能更好的应用于数据噪声较多的通航教练机燃油流量研究。回归树模型更易解释,算法较为简单,通过数据可视化能清晰的展示不同飞行状态下飞机的燃油流量。随机森林模型算法更加复杂,解决了回归树易过拟合的问题,能进一步提高模型的预测精度。

2.1 回归树原理

回归树模型^[17]旨在将输入的特征空间分割成多个子空间,使最终得到的子空间内所包含的样本尽可能属于同一类别,即节点的纯度达到最高。信息熵 $E(D)$ 常被用于度量样本集合的纯度,计算公式为

$$E(D) = - \sum_{k=1}^{|Y|} P_k \log_2 P_k, \quad k = 1, 2, \dots, |Y| \quad (2)$$

式(2)中: P_k 为样本集合 D 中第 k 类样本所占比例; $E(D)$ 越小,代表样本集合 D 的纯度越高。在分割节点时,可以通过计算信息增益、信息增益比、基尼系数等衡量纯度的参数作为分割节点时的依据。在构造回归树时,通常会存在三种节点:根节点、内部节点、叶子节点。根节点位于回归树顶端,是起始节点。内部节点是回归树中间部分的节点。叶子节点位于回归树最底端,也就是决策结果。

通常来讲,对于回归问题,一般通过计算均方误差(mean squared error, MSE)值作为节点划分依

据。而对于分类问题,则是通过计算熵值作为其分类依据。每一个结点 MSE 的计算方法为

$$\text{MEAN} = \frac{\sum_{i=1}^n c_i}{n} \quad (3)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (c_i - \text{MEAN})^2 \quad (4)$$

式中: c_i 为集合中每一个样本的目标值; n 为集合内样本总数。

在原始数据中,遍历每一个特征值所对应的因变量值,用尽每一个因变量值分别将原始数据分割成 2 个集合:左集合(\leq 此次选定因变量值的样本)和右集合($>$ 此次选定因变量值的样本)。分别计算两个集合的 MSE,当(左-MSE + 右-MSE)取得最小值时,此时所对应的特征值和因变量值,就是最佳分割特征以及最佳分割值。

找到最佳分割特征和最佳分割目标值之后,原始集合分裂成新的两个集合,每一个集合也叫作一个节点,每分割一次,回归树的深度(depth)就会 + 1,直到分割结束。为了防止过拟合,通常会设置最大深度 max-depth,当 $\text{depth} \geq \text{max-depth}$,树就停止分裂了,最后生成的不再分裂的集合就叫作叶子节点。在回归树中,落在相同集合内的样本以该集合内因变量的平均值作为该区间内所有样本的预测值。式(3)中计算出的 mean 值就是该节点的值,也就是落在该节点内样本的预测值。

2.2 随机森林原理

随机森林^[18]将多颗回归树进行集成,不仅提高了预测精度,还解决了单颗回归树易过拟合的问

题。随机森林内所包含的回归树越多,其算法鲁棒性越强,精确度越高。

随机森林算法的工作步骤如下(图4)。

(1) 样本随机:从原始数据中有放回的随机抽取若干样本(该样本量小于总样本),每一次取出的数据都不完全一样,分别作为各回归树的训练数据集。

(2) 特征随机:从每个样本集合的所有特征中随机挑选若干个特征(该特征数量小于总特征数量),以挑选出的特征作为结点构建单颗回归树。

(3) 重复第一步和第二步,建立 n 颗回归树,且这些树都需完全成长不被剪枝,最终自由森林由 n 颗未被剪枝的回归树组成。

(4) 根据这些树的预测结果进行投票,决定样本的最后预测值。

3 模型拟合

数据集被随机分成两组:70%作为训练集,用于训练模型,30%作为测试集,用于验证模型的可行性。为了模型的预测精度更高,在构建随机森林模型之前,本文对随机森林模型的超参数进行寻优,寻优结果展示在图5、图6中。主要考虑的超参数为:① $n_estimators$:随机森林算法中回归树的数量;

② max_depth :回归树的最大深度;③ $min_samples_leaf$:回归树中叶节点上的最小样本数;④ $min_samples_split$:回归树分裂所需的最小样本数。两图中均以 MSE 值来衡量不同超参数组合下模型的优劣, MSE 越小,模型精度越高。超参数热图(图5)用颜色表示 MSE 大小,黑白两级分别代表 MSE 的极大值与极小值。热图只能进行粗略的划分,确定最优超参数区间,无法得到精确值。图6用折线表示 MSE 随超参数改变的变化,固定某两个超参数,可以分析另一个超参数改变对模型精度带来的影响。例如:当 $max_depth = None$ 、 $n_estimators = 50$ 时, $min_samples_leaf$ 越大,模型精度越高。通过寻优分析,得到本文随机森林算法最佳超参数组合: $max_depth = None$ 、 $n_estimators = 200$ 、 $min_samples_leaf = 4$ 、 $min_samples_split = 2$ 。

利用训练集对模型进行训练分类,最终得到回归树,如图7所示。该图展示了如何根据三个特征变量的状态确定出燃油流量预测值,得到最终的燃油流量预测值需要经过多个内部节点的条件判断,若符合节点条件则进入左集合继续进行判断,否则进入右集合继续进行判断,直到到达叶子节点,最终获得该状态下的燃油流量预测值。随机森林算法

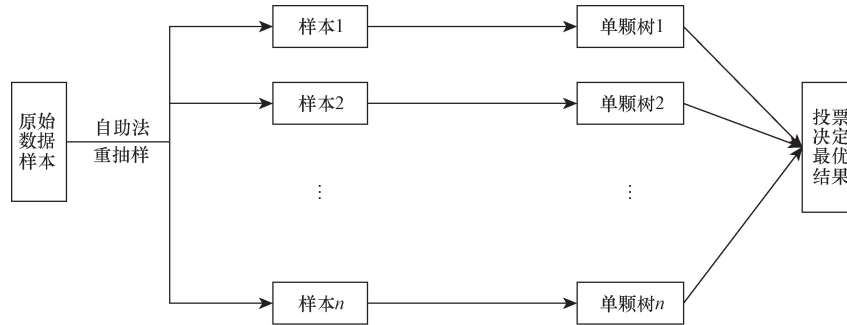


图4 随机森林原理示意图

Fig. 4 Schematic diagram of random forest principle

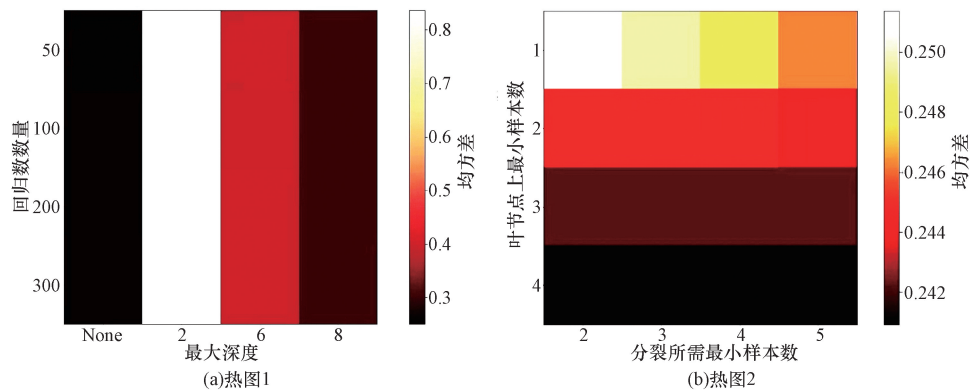


图5 超参数热图

Fig. 5 Hyperparameter heat map

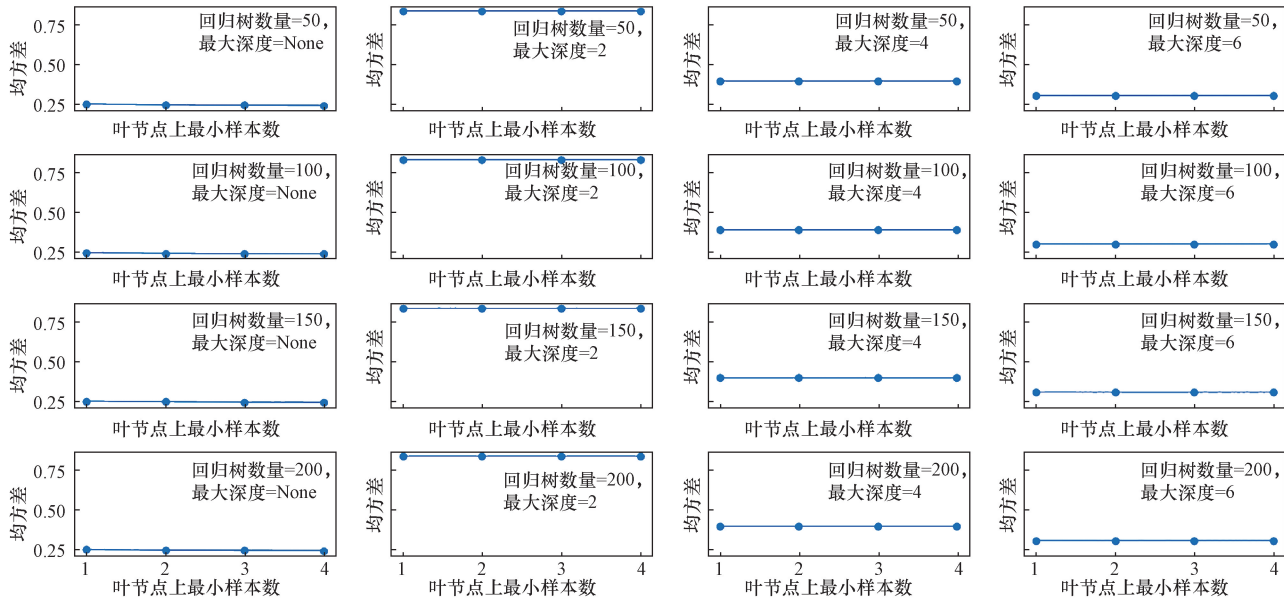


图6 不同超参数所对应模型 MSE 值

Fig. 6 MSE values of the model corresponding to different hyperparameter

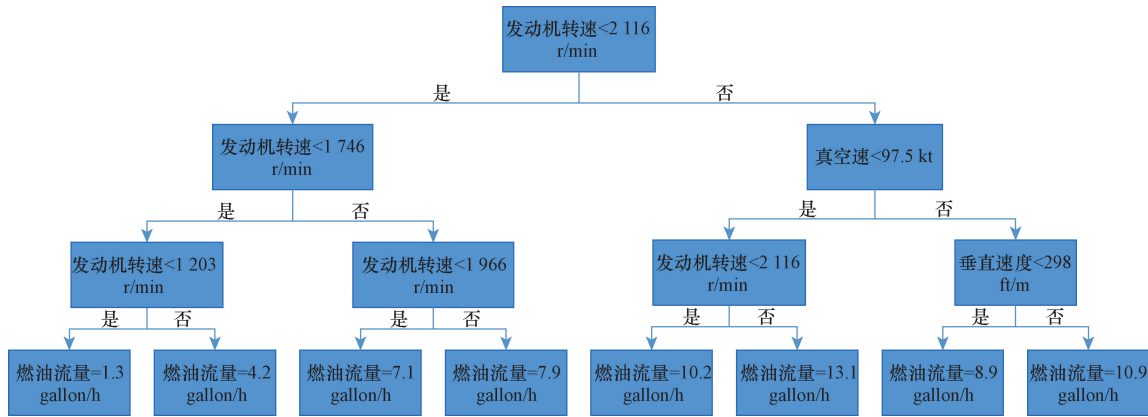


图7 回归树模型结果

Fig. 7 Regression tree model results

结果与回归树算法相似,它通过构建的多颗回归树投票,最终获得某时刻的燃油流量预测值。图8分别展示了回归树和随机森林模型中各自变量对燃油流量预测的贡献度大小,由图8可知,发动机转速是影响燃油流量变化的关键性变量,与之前图3中分析所得结论一致。

塞斯纳 172R 的平飞功率一般维持在 1 950 ~ 2 050 r/min,图7所示树状图的根节点转速以 2 116 r/min分界,左半区飞机的状态应该处于平飞或者下降着陆状态,而右半区飞机应处于爬升或者平飞增速状态。图7中所示最小燃油流量值为 1.3 gallon/h,此时飞机的发动机转速小于 1 203 r/min,可以基本判定此时飞机处于着陆状态,需要持续收油门,减小飞机功率,进而减小飞机的整体能量,最终保证飞机能够安全着陆。当飞机转速处于

1 950 ~ 2 050 时,飞机的燃油流量维持在 7 ~ 8,由此可以推断飞机平飞时的耗油量为 7 ~ 8 gallon/h。塞斯纳 172R 平飞真空速约为 90 kt,当飞机的真空速大于 90 kt,转速在 2 150 r/min 左右时,此时飞机可能处于平飞增速初期,油门缓慢增加,速度也缓慢增加;当飞机真空速小于 97.5 kt,但转速较大时,此时飞机可能处于爬升阶段初期,此时飞行员前推油门杆加满油门,但是由于飞机状态的改变需要一定的反应时间,所以刚加满油门时飞机的状态还没有改变,表现出的只有转速快速增加,此时的瞬间耗油率达到最大值;当飞机的真空速大于 97.5 kt 但是垂直速度较小时,此时飞机已经增速成功,飞机飞行速度增加到飞行员预期值,飞机速度越快,维持飞机状态所需能量越多,所以此时燃油流量略大于正常飞行;当飞机真空速大于 97.5 kt 且垂直速度也很大时,

此时飞机应该在大速度平飞时受到了不稳定气流的影响,从而导致飞机同时拥有较大的飞行速度和垂直速度,这种情况是少见的,因此落在该区间的样本点是最少的。表4总结了由回归树模型得到的不同飞行状态下,飞机的燃油流量和发动机转速值。

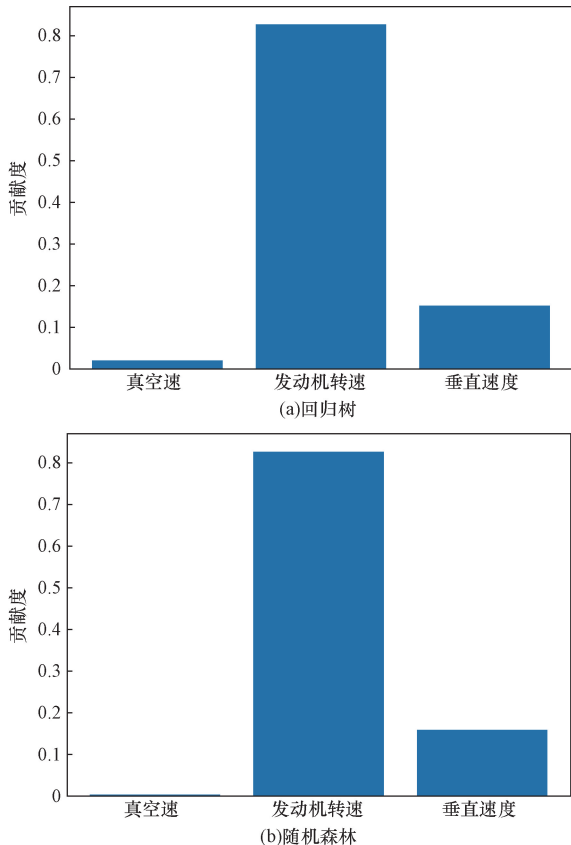


图8 自变量贡献度

Fig. 8 Independent variable contribution

表4 不同飞行状态下飞行参数值

Table 4 Flight parameter values under different flight conditions

	flight conditions				
飞行状态	爬升	平飞	平飞增速	下降	进近着陆
燃油流量/ (gallon·h ⁻¹)	13.1	7~8	10.2	4.2	1.3
发动机转速/ (r·min ⁻¹)	2 350	1 950~2 050	2150	1 250~1 700	<1 200

注:1 gallon=3.785 4 L。

4 模型评估

利用测试集数据在已建立的回归树模型以及随机森林模型中进行数据预测,同样将测试集中的发动机转速、真空速和垂直速度作为输入变量,在训练好的模型中得到燃油流量的预测值,再计算每一点所对应的燃油流量真实值与预测值之间误差大小来评估模型的准确性。本文中通过以下几种常见算法来进行模型评估。

(1)平均绝对误差(mean absolute error, MAE):它代表着真实值 x_i 与预测值 \hat{x}_i 之间绝对误差的平均值,MAE 范围 $[0, +\infty)$,当真实值与预测值越接近时,误差越小,该值越小。计算公式为

$$MAE = \frac{SAE}{N} = \frac{\sum_{i=1}^N |\hat{x}_i - x_i|}{N} \quad (5)$$

(2)均方根误差(root mean squared error, RMSE):它代表着真实值 x_i 与预测值 \hat{x}_i 之间误差平方和取平均之后再求平方根, RMSE 范围 $[0, +\infty)$,当真实值与预测值越接近时,误差越小,该值越小。计算公式为

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2} \quad (6)$$

(3)残差平方和 R^2 : R^2 用于度量因变量中的变化可由自变量解释部分所占比例,取值范围为 $0 \sim 1$, R^2 越接近 1,表明能用 x 的变化来解释 y 的变化部分越多,回归的拟合程度就越好。一般情况下,如果 $R^2 > 0.8$,则认为该模型较好。计算公式为

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{x}_i - x_i)^2}{\sum_{i=1}^N (\bar{x}_i - x_i)^2} \quad (7)$$

(4)平均绝对百分比误差(mean absolute percentage error, MAPE):MAPE 使用百分率来衡量偏离的相对大小,更易理解和解读,MAPE 范围为 $[0, +\infty)$ 。计算公式为

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{x}_i - x_i|}{|x_i|} \times 100\% \quad (8)$$

式中: \hat{x}_i 为燃油流量预测值; \bar{x}_i 为燃油流量真实值的平均值; x_i 为燃油流量真实值; N 为观测值总数;SAE 为绝对误差的总和。

表5列出了空中飞行阶段燃油流量预测模型计算所得验证参数,其中回归树模型 MAE 值为 0.786 gallon/h, RMSE 值为 1.030 gallon/h, MAPE 为 11.42%, $R^2 = 0.865 1$;随机森林模型的 MAE 值、RMSE 值和 MAPE 均小于回归树模型,其中 MAE 为 0.286 gallon/h, RMSE 为 0.496 gallon/h, MAPE 为 4.0%, $R^2 = 0.968 4$ 。通常来说,MAE 会略小于 RMSE,因为 RMSE 受异常值的影响更大,其值随误差呈次方形式增长。通过观察 MAE 值可以判断系统的预测精度如何,通过比较 MAE 和 RMSE 的值可以发现预测数据中是否存在较大的偏差。对于教练机空中飞行阶段而言,随机森林的预测精度远高于回归树。本文所得模型与相关文献的研究结果整理在表5中。

表5 模型验证指标
Table 5 Model validation metrics

使用方法	MAE/ (gallon·h ⁻¹)	RMSE/ (gallon·h ⁻¹)	MAPE/%	R ²
本文回归树	0.786	1.030	11.42	0.865 1
本文随机森林	0.286	0.496	4.00	0.968 4
回归树 ^[14]	1.244	—	—	—
SS-ANOVA ^[14]	1.086	—	—	—
回归树 ^[15]	1.359	1.881	6.30	—
神经网络 ^[15]	1.210	1.651	14.14	—

5 结论

本文中基于塞斯纳 172 飞机机载飞行数据,选择了 3 个与飞行员实际操作相关联的特征变量,采用了基于回归树和随机森林算法来预测通航教练机飞行阶段的燃油流量,通过计算各类误差评价参数,再对比以往相关研究所构建的燃油流量回归模型,表明了本文所提出模型能有效提高教练机燃油流量的预测精度,为通航领域燃油流量预测提供了一种新的可靠的方法。

未来研究中,将以每一次飞行训练任务整体的燃油消耗量作为研究对象,尝试不同方法进行建模,进一步提高实验的现实可行性以及燃油的预测精度,进而为通航培训提供基于数据的节油建议,提高燃油经济性,降低航校燃油成本。

参 考 文 献

- [1] Baumann S, Klingauf U. Modeling of aircraft fuel consumption using machine learning algorithms[J]. The Council of European Aerospace Societies Aeronautical Journal, 2020, 11(1): 277-287.
- [2] Jefry Y, Rhea P L. Cluster-based aircraft fuel estimation model for effective and efficient fuel budgeting on new routes[J]. Aerospace, 2022, 9(10): 624.
- [3] Zhu Q, Pei J, Liu X, et al. Analyzing commercial aircraft fuel consumption during descent: a case study using an improved K-means clustering algorithm[J]. Journal of Cleaner Production, 2019, 223: 869-882.
- [4] 赵元棣, 王中义. 基于卷积神经网络的飞机燃油消耗预测方法[J]. 重庆交通大学学报(自然科学版), 2023, 42(8): 125-131.
Zhao Yuandi, Wang Zhongyi. Prediction method of aircraft fuel consumption based on convolutional neural network[J]. Journal of Chongqing Jiaotong University(Natural Science), 2023, 42(8): 125-131.
- [5] Luo W, Wu Z, Chen C. An aircraft fuel flow model of cruise phase based on LSTM and QAR data[C]// International Symposium on Computational Intelligence and Design (ISCID). Piscataway: IEEE, 2020: 118-121.
- [6] Lin Y, Guo D, Wu Y, et al. Fuel consumption prediction for pre-departure flights using attention-based multi-modal fusion[J]. Information Fusion, 2024, 101: 101983.
- [7] Mevlut U, Umud M D, Gokhan I. Physics guided deep learning for data-driven aircraft fuel consumption modeling[J]. Aerospace, 2021, 8(2): 44.
- [8] 陈聪, 候磊, 李乐乐, 等. 基于 GRU 改进 RNN 神经网络的飞机燃油流量预测[J]. 科学技术与工程, 2021, 21(27): 11663-11673.
Chen Cong, Hou Lei, Li Lele, et al. Prediction of aircraft fuel flow based on recurrent neural network[J]. Science Technology and Engineering, 2021, 21(27): 11663-11673.
- [9] 陈聪, 李乐乐, 陈灏, 等. 基于 PLS-PSO-Elman 的飞机燃油流量预测[J]. 科学技术与工程, 2021, 21(5): 1963-1973.
Chen Cong, Li Lele, Chen Hao, et al. Prediction about fuel flow of airplanes based on PLS-PSO-Elman[J]. Science Technology and Engineering, 2021, 21(5): 1963-1973.
- [10] Sedat M. A new proposal for the prediction of an aircraft engine fuel consumption: a novel CNN-BiLSTM deep neural network model[J]. Aircraft Engineering and Aerospace Technology, 2023, 95(5): 838-848.
- [11] Seymour K, Held M, Georges G, et al. Fuel estimation in air transportation: modeling global fuel consumption for commercial aviation[J]. Transportation Research Part D: Transport and Environment, 2020, 88: 102528.
- [12] Oruc R, Sahin O, Baklacioglu T. Fuel flow rate modeling for descent using cuckoo search algorithm: a case study for point merge system procedure at Istanbul airport[J]. Aircraft Engineering and Aerospace Technology, 2022, 94(5): 824-833.
- [13] 余芬, 刘博, 张军. 通用教练机燃油成本仿真[J]. 科学技术与工程, 2019, 19(24): 392-397.
Yu Fen, Liu Bo, Zhang Jun. Simulation analysis of general trainer fuel cost[J]. Science Technology and Engineering, 2019, 19(24): 392-397.
- [14] Huang C, Xu Y, Johnson M E. Statistical modeling of the fuel flow rate of GA piston engine aircraft using flight operational data[J]. Transportation Research Part D, 2017, 53: 50-62.
- [15] Huang C, Cheng X. Estimation of aircraft fuel consumption by modeling flight data from avionics systems[J]. Journal of Air Transport Management, 2022, 99: 102181.
- [16] 游磊, 孙宏, 姜迪, 等. Garmin 航电系统飞参数据的特性分析[J]. 科技和产业, 2021, 21(9): 187-191.
You Lei, Sun Hong, Jiang Di, et al. Analysis of flight data characteristics of garmin avionics system[J]. Science Technology and Industry, 2021, 21(9): 187-191.
- [17] Breiman L, Friedman J, Olshen R, et al. Cart[J]. Classification and regression trees, 1984.
Breiman L, Friedman J H, Olshen R A, et al. Classification and regression trees[M]. Belmont California: Wadsworth, Inc., 1984.
- [18] Liu Y, Wang Y, Zhang J. New machine learning algorithm; Random forest[C]//Information Computing and Applications: Third International Conference(ICICA). Chengde; Springer Berlin Heidelberg, 2012: 246-252.