

文章编号: 1674-8190(2022)01-076-10

基于 XGBoost 的航空器动态滑行时间预测方法研究

赵征¹, 冯事成¹, 宋梅雯¹, 胡莉¹, 陆莎²

(1. 南京航空航天大学 民航学院, 南京 211106)

(2. 上海国际机场股份有限公司 运行指挥中心, 上海 201202)

摘要: 对航空器进港和离港滑行时间进行精确的动态预测, 可以有效提升机场的运行效率。首次提出基于 XGBoost 的航空器动态滑行时间预测方法, 该方法首先通过分析影响机场滑行时间的各类因素, 构建可变滑行时间预测的关键特征指标; 然后选取 XGBoost 算法建立可变滑行时间预测模型, 对模型的关键输入参数进行测试调整; 最后将 XGBoost 算法与随机森林和支持向量回归算法的预测效果进行对比。同时, 首次剖析样本数据量与滑行时间预测精度的关联, 并以广州白云国际机场为分析对象进行实验。结果表明: 采用 XGBoost 算法, 进/离港滑行时间的预测精度分别达到了 94.1% 和 96.6%, 优于主流算法随机森林和支持向量回归; 且实现白云机场动态滑行时间的精确和稳定预测所需样本量在 32 000 条(含)以上。

关键词: 航空运输; 动态滑行时间; 机器学习; XGBoost; 样本量

中图分类号: V355.1

DOI: 10.16615/j.cnki.1674-8190.2022.01.08

文献标识码: A

开放科学(资源服务)标识码(OSID):



Prediction Method of Aircraft Dynamic Taxi Time Based on XGBoost

ZHAO Zheng¹, FENG Shicheng¹, SONG Meiwen¹, HU Li¹, LU Sha²

(1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

(2. Operation Command Center, Shanghai International Airport Co., Ltd., Shanghai 201202, China)

Abstract: Accurate and dynamic prediction of aircraft arrival and departure taxi time can effectively improve the operation efficiency of the airport. An aircraft dynamic taxi time prediction method based on XGBoost is proposed for the first time. Firstly, the key characteristic index of variable taxi time prediction is constructed by analyzing various factors affecting the taxi time. Then, XGBoost algorithm is selected to establish the variable taxi time prediction model, and the key input parameters of the model are adjusted and tested. Finally, the prediction effect of XGBoost algorithm is compared with random forest and support vector regression. At the same time, the correlation between sample data size and prediction accuracy of taxi time is analyzed for the first time. Guangzhou Baiyun International Airport is taken as the analysis object. The results show that the prediction accuracy of the taxi-in and the taxi-out time reaches 94.1% and 96.6% by using the XGBoost algorithm, which are better than the mainstream algorithms of random forest and support vector regression. In addition, more than 32 000 samples are needed for accurate and stable prediction of dynamic taxi time at Baiyun Airport.

Key words: air transportation; dynamic taxi time; machine learning; XGBoost; sample size

收稿日期: 2021-05-07; 修回日期: 2021-09-05

基金项目: 国家自然科学基金(61903187)

通信作者: 赵征, zheng_zhao@nuaa.edu.cn

引用格式: 赵征, 冯事成, 宋梅雯, 等. 基于 XGBoost 的航空器动态滑行时间预测方法研究[J]. 航空工程进展, 2022, 13(1): 76-85.

ZHAO Zheng, FENG Shicheng, SONG Meiwen, et al. Prediction method of aircraft dynamic taxi time based on XGBoost[J].

Advances in Aeronautical Science and Engineering, 2022, 13(1): 76-85. (in Chinese)

0 引言

大型民用机场场面运行复杂,实际运行中,经常出现进港航空器的机位保障资源不能及时到位;离港航空器按照预计时间推出,但跑道头排队错位导致无法满足分配起飞时刻的情况。上述现象主要是由于机场协同决策(A-CDM)系统中滑行时间设置较为单一和静态,缺乏科学预测和动态修订而导致的。通过对航空器进港和离港滑行时间的精确和动态预测,可以更加准确地计算出预计上轮挡时间、预计推出时间等,从而提升机场的运行效率,提高航班准点率。

在滑行时间影响因素的分析方面,A-CDM系统实施文件^[1]中将影响滑行时间最常见的因素归为11种,包括机场布局和机场设施、使用的跑道、跑道交叉的次数、停机位、气象条件、航空器类型和飞行员、航空器重量、推出指令传达时间、除冰位、交通密度和机场当地的操作程序。R. A. Shumsky^[2]基于排队论,考虑航空公司、离港跑道和离港需求等影响因素,对离港滑行时间进行预测;H. Idris等^[3]分析了影响滑行时间的主要因素,如起飞队列长度、跑道配置、天气、飞机类型和停机位等,并使用排队论模型对离港滑行时间进行预测,结果显示起飞队列长度是最重要的影响因素,但是模型在 ± 5 min范围内的预测精度不到70%。

针对前人研究中关于进港航空器数量对离港滑行时间影响有限的观点,R. R. Clewlow等^[4]经过分析证明了进港航空器数量对离港滑行时间有着显著影响,且进/离港航空器流交互越多,这种影响就越显著;刘继新等^[5]分析了航空器滑行时间的影响因素,结果表明,场面流量对滑行时间的影响最为显著。

在滑行时间预测模型和方法的研究方面,M. S. Kistler等^[6]以达拉斯—沃思堡国际机场为对象,使用线性回归和对数线性回归模型对航空器滑行时间进行预测,结果表明,线性回归模型预测效果较好,在 ± 1 min误差范围内,预测精度达到71%,决定因子值达到0.95;P. Balakrishna等^[7-9]采用强化学习算法对美国的三个机场进行了离港滑行时间预测,算法在底特律国际机场和坦帕国际机场表现出较好的预测效果;S. Ravizza等^[10-11]提出滑行距离应分为推出段、转弯段和直线滑行段,并研

究了飞机滑行通过这三段时的滑行角度和速度,然后基于多元线性回归、最小中位数平方线性回归、支持向量机和M5模型树等对离港滑行时间进行了预测;H. Lee等^[12]以夏洛特机场为例,基于神经网络、线性回归模型、支持向量机、 k 近邻和随机森林等算法对滑行时间进行预测,并比较了这几种算法的预测效能,结果显示,支持向量机预测效果优于其他模型;T. Diana^[13]比较了集成机器学习、普通最小二乘法 and 惩罚算法在预测滑行时间方面的性能,以探究机器学习是否比传统的线性回归具有更好的预测效能,结果表明,在对预测结果进行偏差和方差的评估比较方面,两者各有优劣,故在模型选择时应当以偏差和方差的均衡作为选择依据;Wang X W等^[14]比较了线性回归、梯度提升决策树和随机森林等算法对滑行时间的预测效果,并研究了各个特征对模型的影响程度,得出使用一小部分特征(包括所有机场的一般性特征起飞/到达、滑行距离、总转弯数、平均速度和航空器数量),以及针对特定目标机场的少数特征,就可以实现较高的预测准确性;李楠等^[15]基于多元回归模型,分别采用传统统计学和机器学习预测航空器离港滑行时间,结果显示,机器学习的预测准确度较高, ± 5 min误差范围内预测精度达到87%;冯霞等^[16]将离港滑行时间预测划分为两个阶段,先利用已知特征预测离港滑行过程中使用相同跑道起降的航班数量,再将其作为变量输入滑行时间预测模型,结果显示,该方法的预测效果优于直接预测。

现有研究成果大都只针对离港滑行时间进行分析预测,且未分析样本数据量对预测模型的影响。考虑到进/离港滑行在影响因素和滑行时间上均存在差异,进港滑行时间的准确预测对计算预计上轮挡时间和确定地面保障资源的到位时间至关重要,而样本数据量与预测效果的关联分析可为机场的系统开发和运行提供参考。因此,本文采用XGBoost算法针对进港滑行时间和离港滑行时间分别进行预测,并剖析机场滑行时间预测所需样本数据量与滑行时间预测精度的关联。

1 航空器滑行时间影响因素分析

虽然滑行时间预测的算法具有一定程度的普适性,但因机场机位和机坪布局、跑滑构型、滑行规则、气象环境等的不同,导致不同机场的航空器

滑行时间影响因素有其定制性。因此,本文首先介绍滑行预测方法的实证机场信息,然后针对实证机场进行影响因素的分析提取。

1.1 机场运行环境分析

广州白云国际机场(以下简称“白云机场”)是中国三大门户复合枢纽机场之一,也是中国最繁

忙的五大机场之一,年旅客吞吐量超 5 000 万人次。白云机场场面布局图如图 1 所示,机场建有三条跑道,跑道运行方向为南北运行,01/19 号跑道进、离场混合运行,02L/20R 号跑道主要用于离场,02R/20L 号跑道主要用于进场。机场共有 2 个航站楼,分别为 T1 和 T2,停机位 275 个,主要分布在东机坪、西机坪、北机坪以及其他区域。

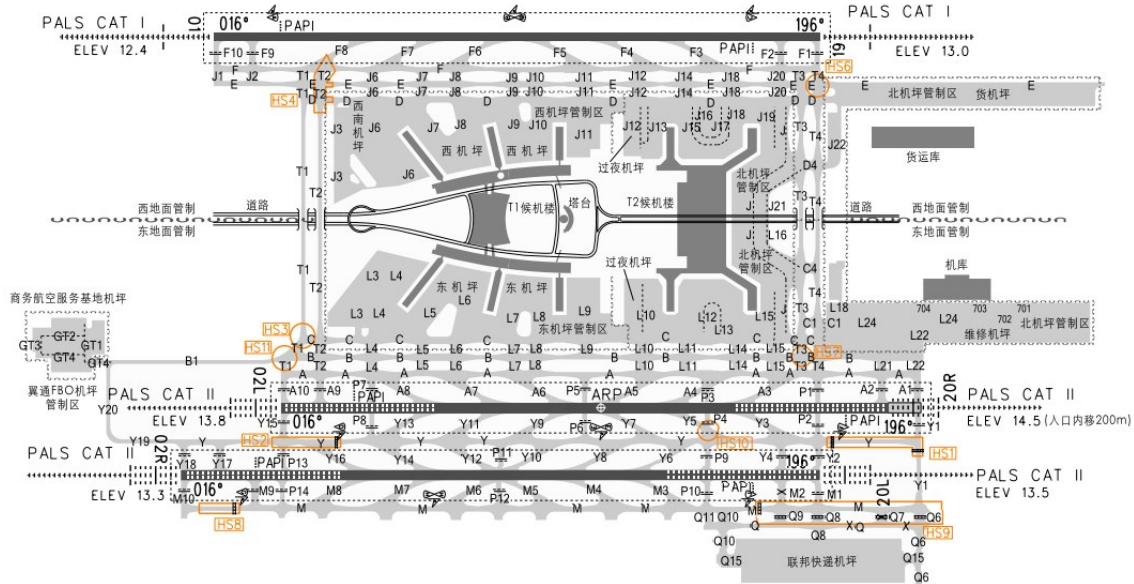


图 1 广州白云国际机场场面布局

Fig. 1 Layout of Guangzhou Baiyun International Airport

1.1.1 航空器滑行规则

进港航空器滑行方面,由 01 号和 19 号跑道进港的航空器主要滑入北机坪和东机坪,由 02R 号和 20L 号跑道进港的航空器主要滑入北机坪和西机坪。离港航空器滑行方面,东机坪、西机坪和北机坪的航空器经由 01 号、02L 号和 20R 号跑道离港,而 19 号跑道只有东机坪和北机坪的航空器经此跑道离港。

1.1.2 机场天气概况

白云机场位于广州市北部花都区,处于雷雨多发地带,据统计白云机场近 50 年来,年平均出现雷暴日数为 75.4 天,最多为 109 天,最少为 49 天。此外,因受台风或台风外围因素影响,经常出现短时强降雨。雷暴和强降雨是影响白云机场航空器滑行两个最主要的天气因素。

1.2 影响因素分析方法

航空器滑行时间受多种因素影响,不同机场

存在共同的影响因素,也有其特定的影响因素。基于对机场运行环境的分析,白云机场场面存在较多交叉路口,容易造成滑行冲突,因此存在较多冲突热点区域(Hot Spot,简称 HS);且由于机位分配的原因,使得航空器存在穿越跑道滑行和交叉滑行的可能。针对白云机场场面运行环境选取三个特定的影响因素,即经过 HS 的数量、是否交叉滑行、是否穿越跑道;两个一般性影响因素,即所属航空公司、停机位,通过统计单个影响因素下的平均滑行时间,直观展示各个因素对滑行时间的影响情况,为最终影响因素集的确定提供分析方法。

1.2.1 经过 HS 的数量

经过 HS 区域时,航空器可能需要减速或者停车,待其他航空器通过后才能继续滑行。随着经过 HS 的数量增多,航空器需要减速或者停车的几率会增加,滑行时间的受影响程度也会增大。由

HS 数量统计的平均滑行时间如图 2 所示,可以看出:当经过 HS 的数量从 0 增加到 2 时,进港平均滑行时间依次增加约 5% 和 58%,离港平均滑行时间依次增加约 2% 和 13%。

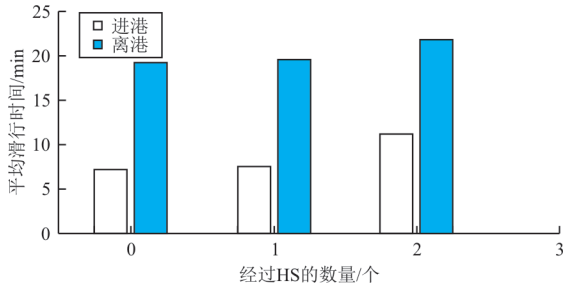


图 2 由经过的 HS 数量统计的平均滑行时间
Fig. 2 Average taxi time measured by HS number

1.2.2 交叉滑行

交叉滑行是反映航空器滑行距离的重要因素,若从东机坪离港的航空器由 01/19 号跑道起飞,或从西机坪离港的航空器由 02L/20R 号跑道起飞,则视为存在交叉滑行;同理,若从 01/19 号跑道进港的航空器滑入东机坪,或从 02R/20L 号跑道进港的航空器滑入西机坪,则视为存在交叉滑行。由交叉滑行统计的平均滑行时间如图 3 所示,可以看出:当存在交叉滑行时,进港平均滑行时间增加约 70%,离港平均滑行时间增加约 10%。

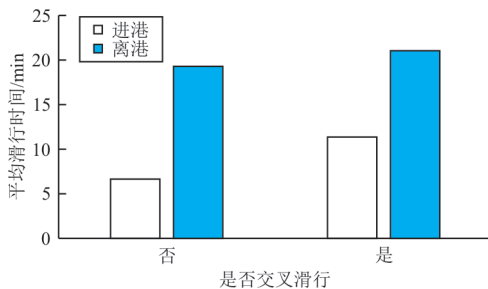


图 3 由交叉滑行统计的平均滑行时间
Fig. 3 Average taxi time calculated by cross taxi

1.2.3 穿越跑道

对于白云机场由 02R/20L 号跑道进港的航空器存在穿越 02L/20R 号跑道的情况,若此时 02L/20R 号跑道上有机场离港起飞,则由 02R/20L 号跑道进港的航空器必须在穿越点外等待,进港滑行时间会受到影响。由穿越跑道统计的平均进港滑行时间如图 4 所示,可以看出:当航空器存在穿越跑道的情况时,进港平均滑行时间增加约

54%。

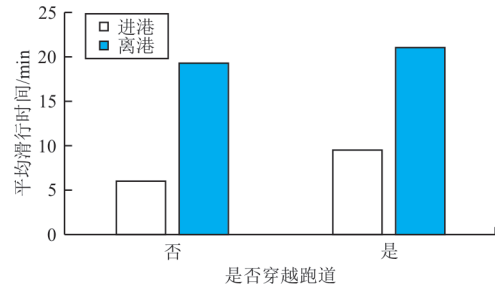


图 4 由穿越跑道统计的平均滑行时间

Fig. 4 Average taxi time calculated by crossing runway

1.2.4 所属航空公司

不同航空公司对航空器推出模式和航空器滑行速度有不同的规定,会对航空器滑行时间造成影响。由航空公司统计的平均进港滑行时间和平均离港滑行时间分别如图 5~图 6 所示,可以看出:不同航空公司航空器平均进/离港滑行时间均存在较大差异。

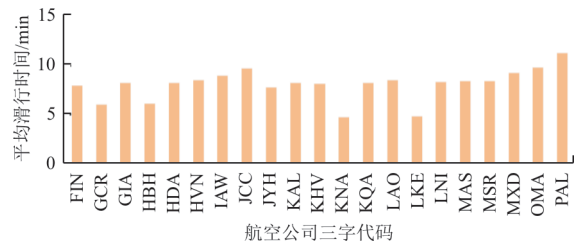


图 5 由航空公司统计的平均进港滑行时间

Fig. 5 Average taxi-in time calculated by airline

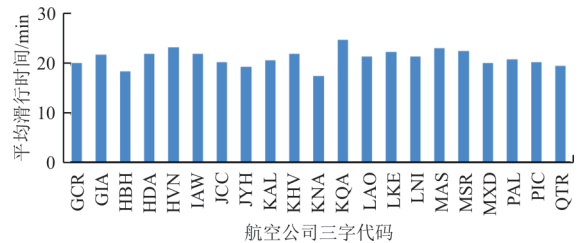


图 6 由航空公司统计的平均离港滑行时间

Fig. 6 Average taxi-out time calculated by airline

1.2.5 停机位

从同一跑道起飞/落地的航空器,由于停放区域不同,会对航空器滑行造成两方面的影响,一是滑行距离不同,二是滑行过程中受其他航空器干扰的程度不同,每一方面都会对滑行时间造成影响。由停机位统计的平均进港滑行时间和平均离港滑行时间分别如图 7~图 8 所示,可以看出:不同停机位上的航空器平均进/离港滑行时间均存在较大差异。

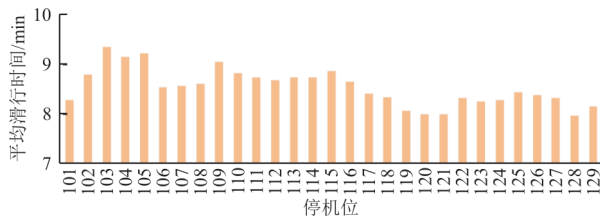


图 7 由停机位统计的平均进港滑行时间

Fig. 7 Average taxi-in time calculated by stand

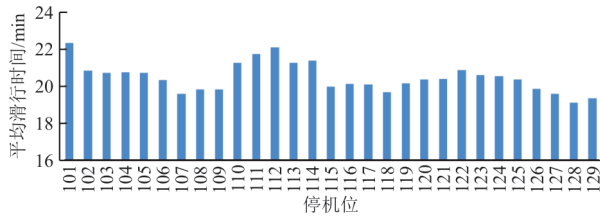


图 8 由停机位统计的平均离港滑行时间

Fig. 8 Average taxi-out time calculated by stand

1.3 最终影响因素集

基于现有研究选取影响滑行时间的其他因素,如航空器类型、起飞/到达跑道、时段、起飞队列长度、场面同时滑行的航空器数量和天气。针对场面滑行的航空器流量统计方法,本文采用 Yin Jianan 等^[17]基于宏观网络拓扑图对流量的计算方法,将其分为离港瞬时流量、离港累计流量、进港瞬时流量和进港累计流量。在对天气的处理上,结合机场运行的天气统计数据 and 管制一线经验,将对航空器滑行会造成较严重影响的雷暴和强阵雨天气视为恶劣天气,其他对航空器滑行几乎没有影响的天气视为正常天气,而后根据机场记录的历史天气情况给每一个航班标定对应的天气类型。结合对特定影响因素的分析提取,最终确定的进/离港滑行时间影响因素如表 1 所示。

表 1 进/离港滑行时间影响因素

Table 1 Influencing factors of taxi-in and taxi-out time

序号	进港滑行影响因素	离港滑行影响因素
1	航空公司	航空公司
2	航空器类型	航空器类型
3	跑道号	跑道号
4	停机位	停机位
5	时段	时段
6	场面同时滑行的航空器数量	场面同时滑行的航空器数量
8	恶劣天气	恶劣天气
9	是否穿越跑道	起飞队列长度
10	是否交叉滑行	是否交叉滑行
11	经过 HS 的数量	经过 HS 的数量

2 预测模型

2.1 模型及评估标准

XGBoost 的英文全称为 Extreme Gradient Boosting, 是 Chen Tianqi 等提出的一种梯度提升模型,也是一种基于决策树的集成机器学习算法^[18]。相较于前几代基于决策树的算法,XGBoost 算法可在最短的时间内,使用较少的计算资源,得到较为出色的结果。

为了更加直观地比较 XGBoost 算法在航空器滑行时间预测方面的性能,本文选择随机森林和支持向量回归两种常用的滑行时间预测算法用以对比分析。

在对模型预测性能的评估上,选择决定系数 (R^2)、均方根误差 (RMSE) 和平均绝对误差 (MAE) 三个指标。通常情况下, R^2 值越大,自变量可以解释的滑行时间变化的比例越大,然而, R^2 值过大也可能意味着一些过拟合的问题;RMSE 衡量观测值与真实值之间的偏差;MAE 反映预测值误差的实际情况,它们是机器学习中评价模型的两把重要标尺。这两个指标的计算公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - t_i)^2} \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - t_i| \quad (2)$$

式中: N 为所有进港航班或者所有离港航班; p_i 为航班 i 滑行时间的预测值; t_i 为航班 i 滑行时间的真实值。

2.2 模型输入与输出

2.2.1 模型输入

本文先进行训练集和测试集的划分,然后进行数据预处理,包括异常值和空缺值的处理,以避免模型提前学习到测试集数据信息。在测试不同划分比例的预测效果后,最终选定以 80% 的数据作为训练集,20% 的数据作为测试集。以进港为例,经过预处理后的进港航班数据如表 2 所示。

基于前文对滑行时间影响因素分析的结果,从预处理后的数据中提取、计算用于输入模型各个特征,而后通过编码将计算机无法识别的字符信息,如航空公司代码、机型和跑道号等转化为

数值型信息,作为模型的最终输入数据。以进港 为例,模型最终输入的进港数据如表 3 所示。

表 2 预处理后的进港航班数据
Table 2 Preprocessed incoming flight data

航空公司	航班号	实际进港时间	上轮挡时间	机型	跑道号	停机位
CSH	9317	2019/01/15 23:48:58	2019/01/15 23:54:00	73H	02R	131
CHH	7526	2019/01/16 00:00:20	2019/01/16 00:04:00	73H	1	408

表 3 输入模型的进港数据
Table 3 Incoming flight data input to model

月份	小时	航空公司	机型	跑道号	停机位	进港瞬时流量	离港瞬时流量	进港累计流量	离港累计流量	是否穿越跑道	天气	是否交叉滑行	经过 HS 的数量
1	23	0	1	0	5	2	0	5	0	0	0	0	0
1	0	1	2	1	12	3	0	7	0	0	0	1	0

2.2.2 模型输出

模型输出即是预测的滑行时间。本文参照中国民用航空局 A-CDM 系统的相关定义,进港航空器滑行时间($T_{\text{taxi-in}}$)指航空器从落地(T_{arrive})时刻到滑入停机位上轮挡($T_{\text{in-block}}$)时刻之间的时间。

$$T_{\text{taxi-in}} = T_{\text{in-block}} - T_{\text{arrive}} \quad (3)$$

离港航空器滑行时间($T_{\text{taxi-out}}$)指航空器从撤轮挡时刻($T_{\text{off-block}}$)到起飞离地时刻(T_{off})之间的时间。

$$T_{\text{taxi-out}} = T_{\text{off}} - T_{\text{off-block}} \quad (4)$$

2.3 模型调参

机器学习模型的预测性能主要取决于两方面因素,一是数据本身的质量,二是模型参数的调整。为了进一步提高模型预测效果,需对基于 XGBoost 的滑行时间预测模型进行调参。

首先,介绍对 XGBoost 算法预测性能影响较大的关键参数。 $n_estimators$ 为集成算法中弱评估器的数量,此参数值越大,模型的学习能力越强,但模型也越容易过拟合; min_child_weight 参数控制叶子上所需的最小样本量,它的调整对于样本量较大的数据比较有效; max_depth 控制模型中树的最大深度,这个参数值越大模型越复杂,且模型容易过拟合; $Subsample$ 参数控制随机抽取的用于训练的数据比例; $learning_rate$ 参数控制迭代速率,两者都可以防止模型过拟合; reg_lambda 和 reg_alpha 参数分别是权重的 L1 和 L2 正则化项; $Gamma$ 控制节点分裂所需的最小损失函数下降值,它们可以降低模型的复杂度,加快算法收敛速度。然后,本文使用网格搜索对 XGBoost 算法的关键参数进行调整,并以交叉验证结果作为参数选择的标准。

最后,经过多轮测试调参,确定的部分关键参数最佳值如表 4 所示。

表 4 模型关键参数最佳值
Table 4 The optimal value of key parameters of the model

参数	最佳值	参数	最佳值
$n_estimators$	1 200	$learning_rate$	0.1
max_depth	5	$Gamma$	0.4
$Subsample$	1		

2.4 滑行时间预测步骤

在对滑行时间影响因素分析、模型及模型预测性能评估标准和模型输入输出介绍的基础上,总结出滑行时间预测的步骤,如图 9 所示。

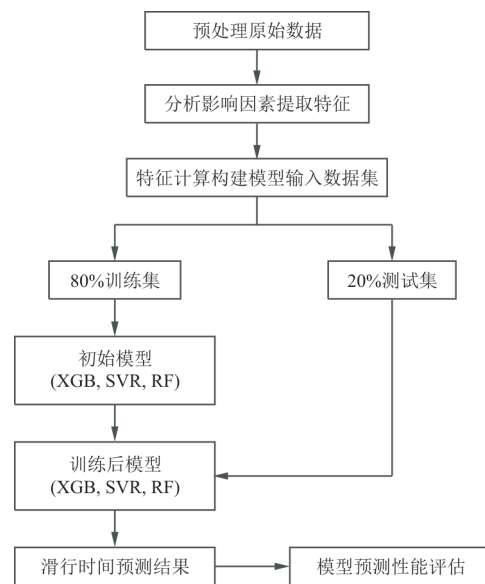


图 9 滑行时间预测步骤

Fig. 9 The steps of taxi time prediction

3 预测结果分析

基于随机森林、支持向量回归和 XGBoost 三种模型对进港和离港滑行时间分别进行预测。首先对白云机场进/离港滑行时间进行统计,为预测结果分析中误差范围的确定提供参考;然后比较三种模型的预测性能,分析特征的重要性;最后分析样本数据量对模型预测精度的影响。

3.1 误差范围确定

本文所用航班数据来源于白云机场塔台运行管理系统(TOMS)2019年全年数据,数据内容包括航空公司、航班号、机型、进港跑道、离港跑道、进港机位、离港机位、进港时间、出港时间、上轮挡时间和撤轮挡时间。去除数据中的缺失值和时间逻辑错误值后得到进港航班数据约 19 万条,离港航班量数据约 20 万条。

分别对进港滑行时间和离港滑行时间进行统计,统计结果如表 5 所示,可以看出:进港滑行时间集中在 5~10 min,离港滑行时间集中在 16~23 min,进港滑行时间远小于离港滑行时间。

表 5 进港和离港滑行时间统计
Table 5 Statistics of taxi-in and taxi-out time

进/离港时间	最小值	第 25 百分位数	平均数	第 75 百分位数	最大值
进港滑行时间	1	5	7	10	30
离港滑行时间	4	16	20	23	59

结合白云机场进/离港滑行时间统计结果和现有研究成果,选择进港滑行时间以 ± 3 min 误差范围,离港滑行时间以 ± 5 min 误差范围为评估标准。

3.2 预测模型性能对比

支持向量回归、随机森林和 XGBoost 模型的拟合度和预测误差的评估结果如表 6~表 7 所示。

表 6 进港滑行时间预测模型评估结果
Table 6 Evaluation results of the prediction model for taxi-in time

Model	R^2		MAE		RMSE	
	Train	Test	Train	Test	Train	Test
SVR	0.72	0.70	1.40	1.35	1.95	1.91
RF	0.96	0.77	0.46	1.27	0.64	1.74
XGBoost	0.84	0.81	1.08	1.16	1.45	1.60

表 7 离港滑行时间预测模型评估结果
Table 7 Evaluation results of the prediction model for taxi-out time

Model	R^2		MAE		RMSE	
	Train	Test	Train	Test	Train	Test
SVR	0.65	0.72	2.15	2.09	3.10	2.8
RF	0.96	0.76	0.71	1.91	0.99	2.65
XGBoost	0.82	0.79	1.64	1.78	2.26	2.48

从表 6~表 7 可以看出:无论是训练集还是测试集,三个模型均得到较高的 R^2 值和较小的误差值,证明模型在进/离港滑行时间的预测上均有良好的性能。其中,虽然进港滑行时间预测结果的误差值小于离港,但考虑到进港滑行时间远小于离港滑行时间,故从误差比例的角度来看并不能认为模型对进港滑行时间的预测效果优于离港滑行时间。支持向量回归和 XGBoost 在训练集和测试集上的表现较为稳定,而随机森林则有较大的波动。在训练集上,随机森林 R^2 值过大,达到了 0.96,表明模型存在过拟合的情况。在测试集上,XGBoost 的预测效能优于随机森林和支持向量回归,其 R^2 值比两者都高,而误差值比两者都低。

进港和离港滑行时间预测结果与实际值误差在 ± 1 、 ± 2 、 ± 3 、 ± 4 和 ± 5 min 范围内的模型预测精度如表 8~表 9 所示。

表 8 进港滑行时间预测精度
Table 8 Prediction accuracy of taxi-in time

Model	误差/%				
	± 1 min	± 2 min	± 3 min	± 4 min	± 5 min
SVR	49.9	78.3	91.2	95.5	97.8
RF	50.8	80.7	92.8	96.9	98.5
XGBoost	54.3	84.0	94.1	97.7	98.9

表 9 离港滑行时间预测精度
Table 9 Prediction accuracy of taxi-out time

Model	误差/%				
	± 1 min	± 2 min	± 3 min	± 4 min	± 5 min
SVR	32.8	57.3	78.0	88.3	93.1
RF	34.3	62.3	80.9	91.0	95.5
XGBoost	36.7	65.7	83.8	92.8	96.6

从表 8~表 9 可以看出:进港滑行时间预测方面, ± 3 min 误差范围内,三个模型精度均在 91% 以上。其中,支持向量回归预测精度最低,仅有

91.2%, XGBoost 预测精度最高, 达到了 94.1%。离港滑行时间预测方面, ± 5 min 误差范围内, 三个模型精度均在 93% 以上。其中, 支持向量回归预测精度最低, 仅有 93.1%, XGBoost 精度最高, 达到了 96.6%。

三个模型进港和离港滑行时间预测值与实际值之差的对比如图 10~图 11 所示, 可以看出: 支持向量回归算法预测结果的误差值范围最大, XGBoost 算法最小。

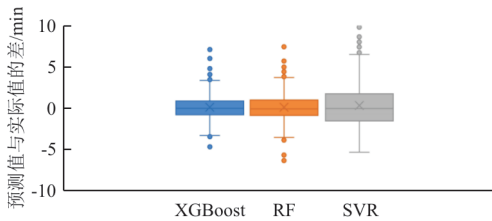


图 10 三种算法进港滑行时间预测值与实际值之差
Fig. 10 Difference between the predicted and actual taxi-in time of the three algorithms

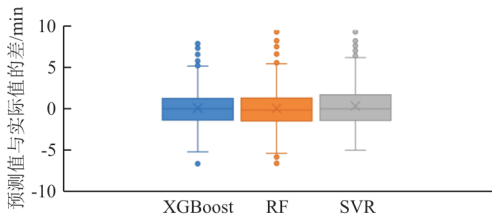


图 11 三种算法离港滑行时间预测值与实际值之差
Fig. 11 Difference between the predicted and actual taxi-out time of the three algorithms

3.3 特征重要性分析

根据 XGBoost 和随机森林得出的特征重要度排名, 探究针对白云机场提取的特定影响因素对于滑行时间预测的重要性。随机森林和 XGBoost 的进港特征重要度排名如图 12~图 13 所示, 可以看出: 在两者的结果中, 交叉滑行都具有最高的特征重要度; 在 XGBoost 的结果中, 穿越跑道和经过 HS 数量的特征重要度也排在前 4。

随机森林和 XGBoost 的离港特征重要度排名如图 14~图 15 所示, 可以看出: 交叉滑行和经过 HS 的数量只在 XGBoost 的结果中具有约 0.05 的特征重要度, 而在随机森林的结果中, 这两个特征的重要度只有约 0.01。此外, 所属航空公司在进港和离港中的重要度都不超过 0.02。

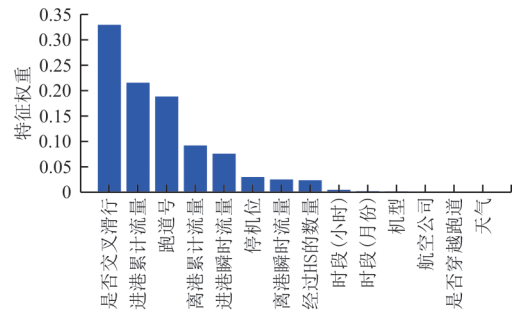


图 12 随机森林进港特征重要度排名
Fig. 12 Feature importance of taxi-in time of random forest

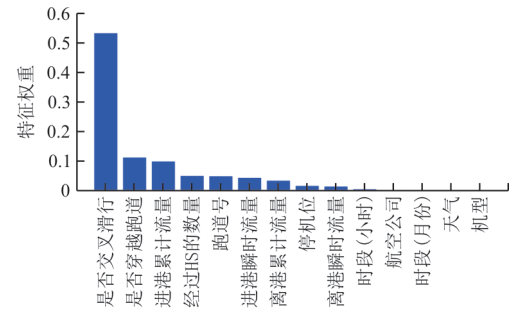


图 13 XGBoost 进港特征重要度排名
Fig. 13 Feature importance of taxi-in time of XGBoost

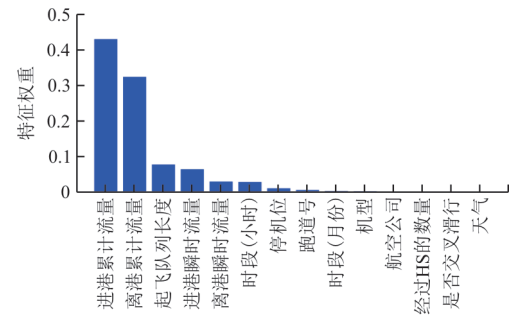


图 14 随机森林离港特征重要度排名
Fig. 14 Feature importance of taxi-out time of random forest

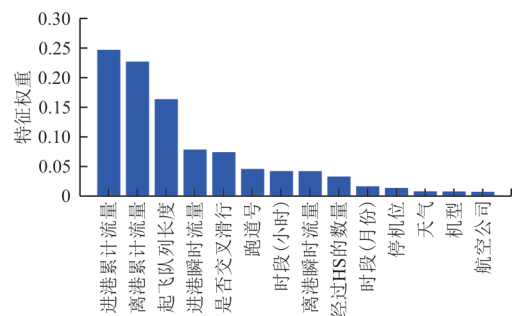


图 15 XGBoost 离港特征重要度排名
Fig. 15 Feature importance of taxi-out time of XGBoost

对特征重要度的分析表明, 针对白云机场提取的特征对于进港滑行时间预测的重要性高于离港滑行时间预测, 其中交叉滑行、穿越跑道和经过

HS 的数量这三个特征的重要性高于所属航空公司。

3.4 样本数据量对模型预测精度的影响

滑行时间预测使用的进港航班数据量约 19 万条,离港航班数据量约 20 万条,过大的数据量会造成数据采集、处理和计算负担。为了研究 XGBoost 模型预测精度随样本数据量的变化情况,在保留所有特征信息的基础上,对进港和离港数据分别进行 3 次随机采样,数据量从 2 000 条开始,并以 2 000 作为梯度,一直增加到 40 000 条,之后用 3 次样本预测的均值作为最终结果。

进港和离港滑行时间预测精度随数据量的变化情况如图 16~图 17 所示,可以看出:进港和离港滑行时间预测精度随数据量的增加呈逐步提高的趋势,但随着数据量的增加,这种趋势逐渐变缓。当数据量在 26 000 条以下时,预测精度的变化趋势波动较大,预测精度的提高速度较快;当数据量超过 26 000 条,预测精度的变化趋势变得平缓;当数据量超过 32 000 条,进港 ± 3 min 误差范围的预测精度稳定在 93.0%~93.5%,离港 ± 5 min 误差范围的预测精度稳定在 95.0%~95.5%。

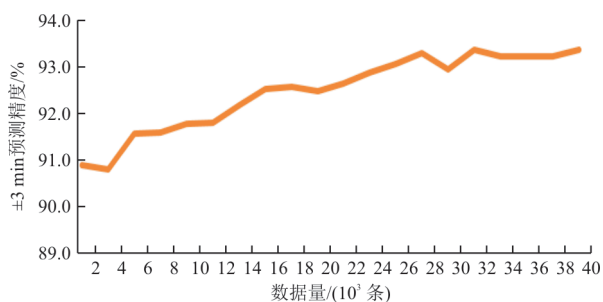


图 16 进港滑行时间预测精度随数据量的变化情况

Fig. 16 The change of the prediction accuracy of the taxi-in time with the data quantity

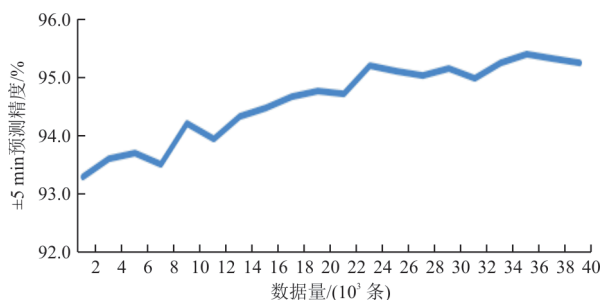


图 17 离港滑行时间预测精度随数据量的变化情况

Fig. 17 The change of the prediction accuracy of the taxi-out time with the data quantity

对模型预测精度随样本数据量变化的分析表明,当数据量超过 32 000 条时,模型就能获得较高的预测精度和较稳定的预测表现。

4 结 论

(1) 本文提出了基于 XGBoost 的航空器动态滑行时间预测方法,并以白云机场为实例,验证了方法的有效性。

(2) 模型预测结果表明,XGBoost 算法在预测误差和预测精度方面都优于随机森林和支持向量回归;且使用 XGBoost 算法,进/离港滑行时间的预测精度分别达到了 94.1% 和 96.6%。

(3) 对特征重要度的分析表明,提取的特征能够较好地反映白云机场滑行态势,且提取的特征对于进港滑行时间预测的重要性要高于离港滑行时间预测,其中交叉滑行、穿越跑道和经过 HS 的数量这三个特征的重要性高于所属航空公司。

(4) 对模型预测精度随样本数据量变化的分析表明,当数据量超过 32 000 条时,模型就能获得较高的预测精度和较稳定的预测表现,可为白云机场 A-CDM 系统中滑行时间动态预测所需样本量提供参考。

参考文献

- [1] EUROCONTROL Airport CDM Team. Airport CDM implementation manual[S]. Brussels: The European Organization for the Safety of Air Navigation, 2017.
- [2] SHUMSKY R A. Real-time forecasts of aircraft departure queues[J]. Air Traffic Control Quarterly, 1997, 5(4): 281-308.
- [3] IDRIS H, CLARKE J P, BHUVA R, et al. Queuing model for taxi-out time estimation[J]. Air Traffic Control Quarterly, 2007, 10(1): 1-22.
- [4] CLEWLOW R R, SIMAIKIS I, BALAKRISHNAN H. Impact of arrivals on departure taxi operations at airports [C]// Proceedings of the AIAA Guidance, Navigation, and Control Conference. Washington, D. C.: AIAA, 2010: 15-37.
- [5] 刘继新, 尹旻嘉, 朱学华, 等. 基于航空器滑出时间的离场影响因素研究[J]. 武汉理工大学学报, 2018, 42(2): 195-200.
- [6] LIU Jixin, YIN Minjia, ZHU Xuehua, et al. Research on departure influencing factors based on aircraft taxi-out time [J]. Journal of Wuhan University of Technology, 2018, 42(2): 195-200. (in Chinese)
- [6] KISTLER M S, GUPTA G. Relationship between airport

- efficiency and surface traffic[C]// 9th AIAA Aviation Technology, Integration, and Operations Conference. Washington, D. C. : AIAA, 2009: 1-13.
- [7] BALAKRISHNA P, GANESAN R, SHERRY L, et al. Estimating taxi-out times with a reinforcement learning algorithm [C]// Proceedings of the IEEE/AIAA 27th Digital Avionics Systems Conference. Washington, D. C. : AIAA, 2008: 3-12.
- [8] BALAKRISHNA P, GANESAN R, SHERRY L. Application of reinforcement learning algorithms for predicting taxi-out times[C]// Eighth USA/Europe Air Traffic Management Research and Development Seminar. [S. l. : s. n.], 2009: 1-15.
- [9] BALAKRISHNA P, GANESAN R, SHERRY L. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: a case-study of Tampa Bay departures[J]. Transportation Research Part C: Emerging Technologies, 2010, 18(6): 950-962.
- [10] RAVIZZA S, ATKIN J, MAATHUIS M, et al. A combined statistical approach and ground movement model for improving taxi time estimations at airports[J]. Journal of the Operational Research Society, 2013, 64(9): 1347-1360.
- [11] RAVIZZA S, CHEN J, JASON A D, et al. Aircraft taxi time prediction: comparisons and insights[J]. Applied Soft Computing Journal, 2014, 14(1): 397-406.
- [12] LEE H, MALIK W, ZHANG B, et al. Taxi time prediction at charlotte airport using fast-time simulation and machine learning techniques [C]// The 15th AIAA Aviation Technology, Integration, and Operations Conference. Washington, D. C. : AIAA, 2015: 1-21.
- [13] DIANA T. Can machines learn how to forecast taxi-out time? a comparison of predictive models applied to the case of Seattle/Tacoma International Airport[J]. Transportation Research Part E: Logistics and Transportation Review, 2018, 119: 149-164.
- [14] WANG X W, BROWNLEE A E I, WOODWARD J R, et al. Aircraft taxi time prediction: feature importance and their implications [J]. Transportation Research Part C: Emerging Technologies, 2021, 124: 102892.
- [15] 李楠, 焦庆宇, 张连东, 等. 离场航空器滑行时间预测研究 [J]. 重庆交通大学学报, 2021, 40(3): 1-6.
- LI Nan, JIAO Qingyu, ZHANG Liandong, et al. Taxi time prediction of departure aircraft [J]. Journal of Chongqing Jiaotong University, 2021, 40(3): 1-6. (in Chinese)
- [16] 冯霞, 孟金双. 基于 KNN 和 SVR 的航班滑出时间预测 [J]. 西南交通大学学报, 2017, 52(5): 1008-1014.
- FENG Xia, MENG Jinshuang. Flight taxi-out time prediction based on KNN and SVR[J]. Journal of Southwest Jiaotong University, 2017, 52(5): 1008-1014. (in Chinese)
- [17] YIN Jianan, HU Yuxin, MA Yuanyuan, et al. Machine learning techniques for taxi-out time prediction with a macroscopic network topology[C]// 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). London, UK: IEEE, 2018: 18308349.
- [18] CHEN Tianqi, GUESTRIN C. XGBoost: a scalable tree boosting system[C]// The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: [s. n.], 2016: 785-794.

作者简介:

赵征(1978—),男,博士,讲师。主要研究方向:空域与机场运行评估与优化。

冯事成(1994—),男,硕士研究生。主要研究方向:交通运输规划与管理。

宋梅雯(1997—),女,硕士研究生。主要研究方向:交通运输规划与管理。

胡莉(1996—),女,硕士研究生。主要研究方向:交通运输规划与管理。

陆莎(1987—),女,学士,中级工程师。主要研究方向:交通运输规划与管理。

(编辑:马文静)