

http://bhxb.buaa.edu.cn jbuua@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2022.0625

进港航班滑入时间预测

唐小卫*, 丁叶, 张生润, 任思豫, 吴佳琦

(南京航空航天大学 民航学院, 南京 211106)

摘 要: 准确预测进港航班滑入时间对合理调配航班保障资源和提高机场场面运行效率具有重要意义, 可有效克服各大机场粗放式预测航班进港时刻的不足, 为此提出一种基于机器学习模型的滑入时间预测方法。以首都机场为具体研究对象, 分析进港航班滑入时间的影响因素并构建特征集; 将线性回归、K-最近邻、支持向量机、决策树、随机森林和梯度提升回归树 6 种在滑出时间预测方面得到广泛应用的机器学习模型用于进港航班滑入时间预测。研究结果表明: 在误差范围 ± 3 min 内 6 种机器学习模型的预测精度均超过 90%, 表明特征集的构建和模型的选择是有效的; 综合预测性能与模型拟合评估结果, 梯度提升回归树模型的预测效果最好; 在梯度提升回归树模型上场面流量特征的贡献度最大, 新引入的跨区特征对预测模型的贡献度超过了大部分传统特征。

关 键 词: 航空运输; 机场场面运行; 滑行时间预测; 机器学习; 梯度提升回归树

中图分类号: V351

文献标志码: A

文章编号: 1001-5965(2024)07-2218-07

航班滑行时间包括进港滑入时间和离港滑出时间。滑入时间是指航班从落地至入位上轮挡的时间, 滑出时间是指航班从机位撤轮挡推出至起飞离地的时间, 二者都是表征机场场面运行效率的重要指标。航班离港滑出时间的研究受到更多的重视, 这是因为离港航班在机位的目标推出时刻是由航班在跑道上的计算起飞时刻减去预测的滑出时间倒推得到的, 所以滑出时间的精准预测在缓解场面拥堵、提高跑道的利用率等方面具有重要作用。相比之下, 滑入时间的起点是航班在跑道落地, 而降落航班的次序在落地之前由空中管制部门确定, 所以对于机场几乎不存在优化空间。从作业操作来看, 离港滑出时飞行员无法直接面对机坪场景, 需使用牵引车将飞机从机位推出, 滑出时间持续较长; 而在滑入的整个过程中飞行员均直接面对运行场景且不需要牵引车, 所以滑入时间持续较短, 对

其研究也相应很少。在实际保障过程中, 各机场普遍以“航班实际落地时刻加上 10 min”近似代替“航班进港时刻”, 据此安排航班保障的人力与设备资源, 但这种粗糙的管理方法既未考虑到不同机场之间的规模差异, 也忽视了同一机场不同机位之间的位置差异, 缺乏科学性与合理性。因此准确预测进港航班的滑入时间并计算出较为准确的航班进港时刻, 对于机场保障资源的优化配置和航班过站的精细调度都具有重要的现实意义。

国内外对航班滑行时间预测的研究绝大多数集中在离港滑出时间预测上。早期的研究主要使用排队论模型^[1-2]与统计学中的多元线性回归 (linear regression, LR) 模型^[3-4]。随着适配开源软件的多种机器学习库的不断涌现, 机器学习模型在滑出时间预测研究中逐渐得到广泛应用。Lee 等^[5]发现属于集成学习的随机森林 (random forest, RF) 模

收稿日期: 2022-07-19; 录用日期: 2022-08-19; 网络出版时间: 2022-09-09 11:27

网络出版地址: link.cnki.net/urlid/11.2625.V.20220908.1849.004

基金项目: 国家自然科学基金 (61603178, U2333204, U2233208)

* 通信作者. E-mail: tangxiaowei@nuaa.edu.cn

引用格式: 唐小卫, 丁叶, 张生润, 等. 进港航班滑入时间预测 [J]. 北京航空航天大学学报, 2024, 50 (7): 2218-2224.

TANG X W, DING Y, ZHANG S R, et al. Taxi-in time prediction of arrival flight [J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50 (7): 2218-2224 (in Chinese).

$$t = t_{\text{AIBT}} - t_{\text{ALDT}}$$

(1)

式中： t_{AIBT} 为航班实际上轮挡时刻； t_{ALDT} 为航班实际落地时刻。

有效样本进港滑入时间的统计数据如表 1 所示。由表 1 可知，进港滑入时间集中在 7~17 min，平均为 12.35 min。

表 1 进港滑入时间统计数据

Table 1 Statistical data of taxi-in time

统计指标	进港滑入时间/min	统计指标	进港滑入时间/min
最小值	3	最大值	59
25%位数	7	平均值	12.35
中位数	11	标准差	6.16
75%位数	17		

1.2 进港滑行时间预测的特征集构建

1.2.1 场面流量特征

场面流量对应了进港航班滑入过程中正在跑道、滑行道或停机坪上运动的航班数量,包括进港航班和离港航班,其数值直观反映了机场场面整体的拥堵情况。由首都机场跑滑系统构型和运行模式可知,进港航班滑行过程会受其他两条跑道上进离港航班的影响。为此,将第*i*个进港航班的滑行区间定义为 $[t_{\text{ALDT}}(i), t_{\text{AIBT}}(i)]$,在此区间运行的其他进出港航班共有 8 类,进港航班分别为 a1, a2, a3, a4 类,出港航班分别为 d1, d2, d3, d4 类。将这 8 类航班对应的场面流量定义为 8 个特征变量,其编号和定义如图 2 和表 2 所示。对于任意进港航班*j*,如果其落地时间早于航班*i*、入位时间位于 $[t_{\text{ALDT}}(i), t_{\text{AIBT}}(i)]$,则属于 a1 类进港航班, count(*j*)表示此类进港航班的数量,对应的场面流量特征标记为 A_1, A_2, A_3, A_4 ,定义与之类似。对于任意离港航班*k*,如果其推出时间早于航班*i*、离地时间位于 $[t_{\text{ALDT}}(i), t_{\text{AIBT}}(i)]$,则属于 d1 类离港航班, count(*k*)表示此类离港航班的数量,对应的场面流量特征标记为 D_1, D_2, D_3, D_4 ,定义与之类似。

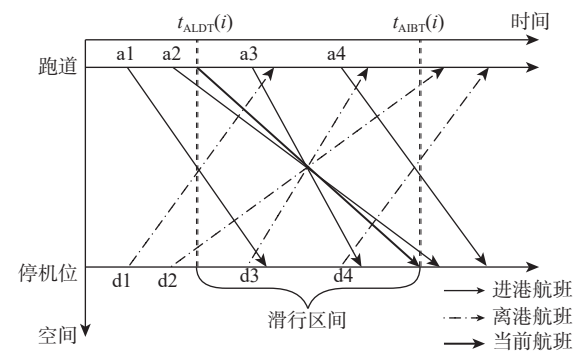


图 2 场面运行航空器类别

Fig. 2 Category of aircraft operation aircrafts on the surface

表 2 场面流量特征定义

Table 2 Definition of surface traffic flow features

特征变量	定义
A_1	$A_1(i) = \text{count}(j), t_{\text{ALDT}}(j) < t_{\text{ALDT}}(i) \& t_{\text{ALDT}}(i) < t_{\text{AIBT}}(j) < t_{\text{AIBT}}(i)$
A_2	$A_2(i) = \text{count}(j), t_{\text{ALDT}}(j) < t_{\text{ALDT}}(i) \& t_{\text{AIBT}}(j) > t_{\text{AIBT}}(i)$
A_3	$A_3(i) = \text{count}(j), t_{\text{ALDT}}(j) > t_{\text{ALDT}}(i) \& t_{\text{AIBT}}(j) < t_{\text{AIBT}}(i)$
A_4	$A_4(i) = \text{count}(j), t_{\text{ALDT}}(i) < t_{\text{ALDT}}(j) < t_{\text{AIBT}}(i) \& t_{\text{AIBT}}(j) > t_{\text{AIBT}}(i)$
D_1	$D_1(i) = \text{count}(k), t_{\text{AOBT}}(k) < t_{\text{ALDT}}(i) \& t_{\text{ALDT}}(i) < t_{\text{ATOT}}(k) < t_{\text{AIBT}}(i)$
D_2	$D_2(i) = \text{count}(k), t_{\text{AOBT}}(k) < t_{\text{ALDT}}(i) \& t_{\text{ATOT}}(k) > t_{\text{AIBT}}(i)$
D_3	$D_3(i) = \text{count}(k), t_{\text{AOBT}}(k) > t_{\text{ALDT}}(i) \& t_{\text{ATOT}}(k) < t_{\text{AIBT}}(i)$
D_4	$D_4(i) = \text{count}(k), t_{\text{ALDT}}(i) < t_{\text{AOBT}}(k) < t_{\text{AIBT}}(i) \& t_{\text{ATOT}}(k) > t_{\text{AIBT}}(i)$

注： t_{AOBT} 为航班实际撤轮挡时刻， t_{ATOT} 为航班实际起飞离地时刻。

分别计算上述 8 个场面流量特征与进港滑入时间的相关系数,如表 3 所示。由表 3 可知,在 1% 的显著性水平下,8 个特征与滑入时间的相关性绝对值均超过 0.45,说明本文建立的场面流量特征与进港滑入时间有较强的相关性。其中表示晚于航班*i*落地并早于航班*i*入位上轮挡的 a3 类航空器数量的场面流量特征 A_3 与滑入时间相关性最大。

表 3 场面流量与滑入时间的相关性

Table 3 The correlation coefficient of surface traffic flow and taxi-in time

特征变量	相关系数	特征变量	相关系数
A_1	0.573***	D_1	0.701***
A_2	-0.564***	D_2	-0.451***
A_3	0.878***	D_3	0.662***
A_4	0.662***	D_4	0.660***

注：***表示 1% 的显著性水平。

上述场面流量特征的设计思路在其他多跑道机场同样适用,但需要根据目标机场的构型和跑道运行模式确定跑道之间的运行是否相互影响。如果进港航班滑行时仅受同跑道运行航班的影响,则只需要考虑同跑道场面流量特征。

1.2.2 跨区特征

首都机场的 3 条远距平行跑道将机场划分成东、西两区,位于 01/19 跑道与 36R/18L 跑道之间的是东区,位于 36R/18L 跑道与 36L/18R 跑道之间的是西区。由中跑道 36R/18L 进港的航班可以从不同的脱离道进入东区或西区,不存在跨区运行,但由外侧 2 条跑道进港时,根据停靠机位所属区域,将出现跨区或不跨区两种情况。当航班进行跨区运行时,根据中跑道运行状况,一部分航班采取绕

滑方式,从东区进入西区的航空器需要从北面的 S6 滑行道绕行,从西区进入东区的航空器则需要从 S7 绕行,滑行距离较长;另一部分航班采取直接穿越中跑道的方式,但为了避免干扰中跑道的正常运行,可能需要在穿越前等待一段时间。上述 2 种跨区方式都会对进港滑入时长造成明显的影响。统计首都机场实际运行数据发现,近 30% 的进港航班需要跨区滑行,跨区滑行的平均滑入时间是不跨区滑行的 2 倍以上,如表 4 所示。

表 4 跨区与不跨区航班对应的平均滑入时间

Table 4 The average taxi-in time of cross-regional flights and non-cross-regional flights

是否跨区	样本量/%	平均滑行时长/min
不跨区	71	9.3
跨区	29	19.8

与首都机场不同,国内上海浦东机场、广州白云机场、杭州萧山机场等均为航站楼居中、两侧各一组窄距平行跑道的构型,此种情形下进港航班从一组窄距平行跑道的外侧跑道落地后需穿越内侧的起飞跑道,此时可将跨区特征调整为跨跑道特征,即进港航班在穿越起飞跑道时是否需要等待。此外,侧向跑道由于能够减少强侧风的不利影响,在我国大型机场的规划设计中不断出现,例如北京大兴机场、成都天府机场等^[17],此时跨区特征或者跨跑道特征的构建还需进一步研究。

1.3 特征集构建

借鉴滑行时间预测的已有研究,本文将运行时段、机型类别、航司属性、滑行距离和使用跑道等常见影响因素也纳入滑入时间预测模型的特征集,最终的特征集如表 5 所示。

表 5 用于滑入时间预测的特征集构建

Table 5 Construction of feature set for taxi-in time prediction

特征名称	特征代号	特征描述
场面流量	$A_1, A_2, A_3, A_4, D_1, D_2, D_3, D_4$	见表2
跨区运行	K	跨区为1,不跨区为0
滑行距离	d	跑道口至停机位的滑行距离
进港跑道	W	3条跑道共6端,跑道号为19、01、18L、36R、18R、36L的跑道端分别对应1~6编号
运行时段	T	全天聚类为[0:00—0:59,10:00—23:59]、[1:00—1:59,9:00—9:59]、[2:00—8:59]共3个时段,分别对应1~3编号
机型	M	机型共分为C、D、E、F四类,分别对应1~4编号
航司属性	I	国内航司为1,国外航司为0

2 预测结果分析与模型评估

2.1 模型预测结果分析

LR、KNN、SVR、DT、RF 和 GBRT 这 6 种机器学习模型在离港滑出时间预测方面得到了较为广泛的应用,其在现有滑出时间预测研究中可达到的预测性能^[11,13-14]如表 6 所示。本文将这 6 种模型用于首都机场进港航班的滑入时间预测,采用均方根误差 S_{RMSE} 、平均绝对误差 S_{MAE} 、决定系数 R^2 和预测精度作为评价指标,其中 R^2 可在一定程度上表征模型的拟合程度。预测精度为预测模型输出的时间与实际滑入时间的 S_{MAE} 差值在某一设定范围内的数量与总预测样本数之比,现有滑出时间预测研究主要关注 $\pm 3\text{ min}$ 、 $\pm 5\text{ min}$ 范围的预测精度,但在同一个机场平均滑入时间低于平均滑出时间,故在 $\pm 1\text{ min}$ 、 $\pm 2\text{ min}$ 、 $\pm 3\text{ min}$ 范围评估滑入时间预测精度。将处理好的数据集以 7:3 的比例^[11]划分为训练集与测试集,6 种模型的预测性能评估结果如表 7 所示。

表 6 滑出时间预测研究中 6 种模型的预测精度

Table 6 Prediction accuracy of six models in the research of taxi-out time prediction

模型	R^2	预测精度/%	
		$\pm 3\text{ min}$	$\pm 5\text{ min}$
LR ^[11]	0.81	73.74	88.40
SVR ^[13]	0.80	75.40	87.40
KNN ^[14]	0.79	59.69	81.46
DT ^[14]	0.81	63.07	85.10
RF ^[11]	0.84	77.79	90.53
GBRT ^[11]	0.83	77.36	90.19

表 7 6 种模型的滑入时间预测性能评估

Table 7 Evaluation of taxi-in time prediction performance by six models

模型	S_{RMSE}/min	S_{MAE}/min	R^2	预测精度/%		
				$\pm 1\text{ min}$	$\pm 2\text{ min}$	$\pm 3\text{ min}$
LR ^[11]	2.01	1.39	0.89	49.6	78.3	90.3
SVR ^[13]	1.63	1.12	0.93	58.4	85.7	94.3
KNN ^[14]	1.85	1.26	0.91	56.4	82.3	92.3
DT ^[14]	1.73	1.22	0.92	54.5	82.5	93.0
RF ^[11]	1.50	1.07	0.94	59.6	86.5	95.0
GBRT ^[11]	1.40	1.01	0.95	61.4	88.4	96.0

由表 6 与表 7 可知,6 种模型用于滑入时间预测时决定系数 R^2 较高,且高于用于滑出时间预测时的表现,表明这 6 种模型在测试集上的拟合优度较高,借鉴现有滑出时间研究构建的特征集可以有效用于滑入时间预测。

对比 6 种模型的预测精度,对首都机场进港航班滑入时间的预测值与实际值在 $\pm 2\text{ min}$ 的准确性平均达到 84%, $\pm 3\text{ min}$ 的准确性平均达到 93.48%,远高于用于离港滑出时间预测时在 $\pm 3\text{ min}$ 范围内的准确性,甚至超过了 $\pm 5\text{ min}$ 范围内的离港滑出时间预测精度。

综合模型误差、拟合优度和预测精度来看,GBRT 的预测性能最好: $S_{\text{RMSE}}=1.4\text{ min}$ 和 $S_{\text{MAE}}=1.01\text{ min}$ 均最低; $R^2=0.95$ 是 6 种模型中最高的; $\pm 1\text{ min}$ 、 $\pm 2\text{ min}$ 、 $\pm 3\text{ min}$ 的预测精度也最高,分别为 61.4%、88.4%、96%。

2.2 模型拟合评估

过拟合和欠拟合是使用机器学习进行模型训练时的常见问题。当模型过度学习了训练集的非一般性特征时即导致过拟合,此时在测试集上远远达不到在训练集上的拟合效果,导致模型泛化能力较低;相反,模型的学习能力不够即导致欠拟

合,在训练集上也不能拥有较高的拟合优度,模型解释性较差。由表 7 可知,本文使用的 6 种模型都拥有较高的 R^2 值,可以判定其不存在欠拟合问题,为了确认预测模型在充分解释滑入时间的基础上是否也能具有较好的泛化能力,还需要进行过拟合检验。

以模型在训练集和测试集上的 R^2 作为得分指标,6 种模型在不同规模样本下的学习曲线如图 3 所示。由图 3(a)、3(b)、3(f)可知,LR、SVR 和 GBRT 这 3 个模型的训练集与测试集得分趋势逐渐收敛,即不存在过拟合问题,说明模型的拟合情况较好。由图 3(c)、3(d)、3(e)可知,KNN、DT 和 RF 模型在训练样本足够多的情况下,测试集得分仍远不及训练集,即存在过拟合问题。6 种模型中 GBRT 模型在训练集与测试集上的得分最高,在 0.95 左右,因此认为 GBRT 的拟合效果最好,同时也具有较高的解释能力与泛化性能。

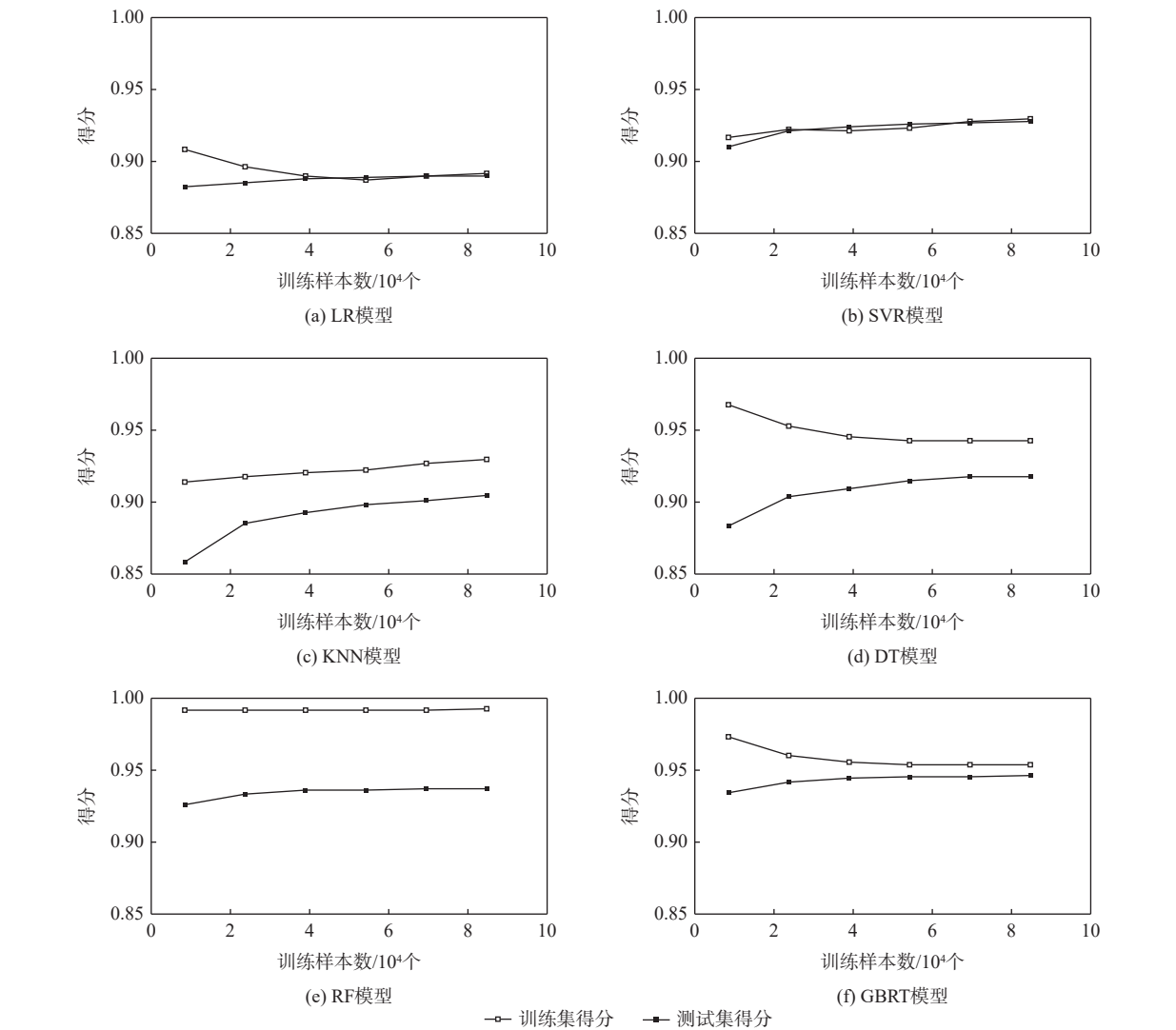


图 3 6 种模型的学习曲线

Fig. 3 Learning curves for six models

3 特征重要性分析

GBRT 模型拥有最好的预测和泛化性能, 故以 GBRT 模型为例进行进港滑入时间特征重要度的研究, 各特征变量对滑入时间预测的影响程度如表 8 所示。

表 8 GBRT 模型的特征重要度排序

Table 8 Feature importance order of GBRT model

排序	特征	特征重要度	排序	特征	特征重要度
1	A_3	0.636 5	8	D_4	0.013 6
2	d	0.121 0	9	D_2	0.008 1
3	K	0.055 0	10	T	0.003 6
4	D_3	0.050 9	11	W	0.003 3
5	A_4	0.048 6	12	A_1	0.001 4
6	D_1	0.040 3	13	M	0.000 6
7	A_2	0.017 0	14	I	0.000 1

由表 8 可知, 表征场面流量的特征变量对进港滑入时间预测的贡献较大, 第 1、4、5 位次均为场面流量特征, 其总贡献度达到 81.46%, 与文献 [11] 研究中滑出时间预测特征的重要度一致, 证明场面流量是影响滑行时间的最重要因素。首都机场三条跑道同时承担起降任务时, 根据管制运行规则, 进港航班比离港航班具有更高的优先级, 场面流量中的进港流量比离港流量更易对滑入时间产生影响, 因此反映进港流量的 4 个特征变量重要度远超离港流量特征, 总贡献度达到 70.35%。其中定义为 A_3 的进港流量特征对预测模型的贡献度达到 63.65%, 这是由于 a_3 类航班落地比预测航班晚但入位更早, 对场面滑行系统资源占用的优先级更高, 例如, 当在滑行交叉点相遇时, 预测航班有可能因为让行 a_3 类航班而发生等待, 导致滑入时间增长, 因此 a_3 类航班的数量对滑入时间预测的影响最大。

本文首次提出的跨区特征对首都机场滑入时间预测的重要度排在第 3 位, 达到 5.5%, 超过了 A_3 以外的所有流量特征和大部分传统特征, 说明引入的新特征是有效的。

4 结 论

为克服各大机场进港航班地面滑行粗放管理的不足, 将机器学习用于进港航班滑入时间的预测, 得出有如下结论:

1) 离港滑出时间预测中得到广泛应用的 LR、SVR、KNN、DT、RF 和 GBRT 这 6 种模型在进港滑入时间预测方面取得了更好的预测精度, $\pm 3\text{ min}$ 的

预测精度均超过 90%。

2) 从综合模型预测结果性能与模型拟合评估结果看, GBRT 模型在首都机场进港滑入时间预测问题上具有最好的表现, 拟合优度达到 0.95 且不存在过拟合和欠拟合, $\pm 3\text{ min}$ 的预测精度也最高, 达到 96%。

3) GBRT 模型中特征重要度的排序结果表明, 场面流量特征的贡献度最大, 这与离港滑出时间预测的研究结果一致。新引入的跨区特征对预测模型的贡献度达到 5.5%, 超过了大部分传统特征。

参考文献 (References)

[1] HEBERT J E, DIETZ D C. Modeling and analysis of an airport departure process[J]. Journal of Aircraft, 1997, 34(1): 43-47.

[2] IDRIS H, CLARKE J P, BHUVA R, et al. Queuing model for taxi-out time estimation[J]. Air Traffic Control Quarterly, 2002, 10(1): 1-22.

[3] RAVIZZA S, ATKIN J A D, MAATHUIS M H, 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.

[4] JORDAN R, ISHUTKINA M A, REYNOLDS T G. A statistical learning approach to the modeling of aircraft taxi time[C]// 29th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2010.

[5] LEE H, MALIK W, JUNG Y C. Taxi-out time prediction for departures at charlotte airport using machine learning techniques[C]// Proceedings of the 16th AIAA Aviation Technology, Integration, and Operations Conference. Reston, Virginia: AIAA, 2016.

[6] HERREMA F, CURRAN R, VISSER H, et al. Taxi-out time prediction model at Charles de Gaulle Airport[J]. Journal of Aerospace Information Systems, 2018, 15(3): 120-130.

[7] 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 X, MENG J S. Flight taxi-out time prediction based on KNN and SVR[J]. Journal of Southwest Jiaotong University, 2017, 52(5): 1008-1014(in Chinese).

[8] 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.

[9] LIAN G, ZHANG Y P, XING Z W, et al. RETRACTED: a new dynamic pushback control method for reducing fuel-burn costs: using predicted taxi-out time[J]. Chinese Journal of Aeronautics, 2019, 32(3): 660-673.

[10] 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(自然科学版), 2021, 40(3): 1-6.
LI N, JIAO Q Y, ZHANG L D, et al. Taxi time prediction of departure aircraft[J]. Journal of Chongqing Jiaotong University (Natural Science), 2021, 40(3): 1-6(in Chinese).

[11] WANG X W, BROWNEE 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.

[12] RAVIZZA S, CHEN J, ATKIN J A D, et al. Aircraft taxi time prediction: comparisons and insights[J]. *Applied Soft Computing*, 2014, 14: 397-406.

[13] LI N, JIAO Q Y, ZHU X H, et al. Prediction of departure aircraft taxi time based on deep learning[J]. *Transactions of Nanjing University of Aeronautics and Astronautics*, 2020, 37(2): 232-241.

[14] 陈祯, 唐小卫, 林有超, 等. 基于决策树的航空器离港滑出时间预测方法和模型[J]. *武汉理工大学学报 (交通科学与工程版)*, 2021, 45(3): 448-453.
CHEN Z, TANG X W, LIN Y C, et al. Prediction method and model of aircraft taxi-out time based on decision tree[J]. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2021, 45(3): 448-453(in Chinese).

[15] 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 Z, FENG S C, SONG M W, 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).

[16] BROWNLEE A E I, WEISZER M, CHEN J, et al. A fuzzy approach to addressing uncertainty in Airport Ground Movement optimisation[J]. *Transportation Research Part C: Emerging Technologies*, 2018, 92: 150-175.

[17] 沈志远, 胡莹莹. 考虑尾流影响的侧向双跑道机场的跑道容量研究[J]. *南京航空航天大学学报*, 2020, 52(1): 161-170.

SHEN Z Y, HU Y Y. Runway capacity of lateral double-runway airport considering wake effect[J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2020, 52(1): 161-170(in Chinese).

Taxi-in time prediction of arrival flight

TANG Xiaowei*, DING Ye, ZHANG Shengrun, REN Siyu, WU Jiaqi

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

Abstract: Accurate prediction of flight taxi-in time has a significant meaning in allocating aircraft support resources reasonably and improving airport surface movement efficiency. Therefore, a method of taxi-in time prediction based on machine learning model is proposed. It can effectively overcome the deficiency of extensive aircraft arrival time prediction in major airports currently. Using Beijing Capital International Airport as the research object, we firstly analyzed the factors that influence the taxi-in time and created the feature set. Next, we applied various techniques that are commonly used to predict taxi-out times, such as linear regression, K-nearest neighbor, support vector regression, decision tree, random forest, and gradient boosting regression tree, to predict the taxi-in time. The results show that the prediction accuracy of the six machine learning models is over 90% within ± 3 min, which means that the construction of the feature set and the selection of models are effective. The gradient boosting regression tree model has the best performance based on the prediction results and model fitting evaluation results. The prediction results of gradient boosting regression tree show that the surface traffic flow features contribute most to the prediction model, and the newly proposed cross-regional feature contributes more than most traditional features.

Keywords: air transportation; airport surface movement; taxi time prediction; machine learning; gradient boosting regression tree