

DOI:10.16356/j.1005-2615.2023.06.008

基于 UKDE 和 XGBoost 的航班过站时间动态预测

吴薇薇¹, 熊奥萍¹, 唐红武²

(1. 南京航空航天大学民航学院, 南京 211106; 2. 中航信移动科技有限公司, 北京 100029)

摘要: 为了提高机场运行高峰时航班过站时间预测的精度及可靠性, 研究了一种结合无偏核密度估计 (Unbiased kernel density estimation, UKDE) 和极端梯度提升决策树 (Extreme gradient boosting, XGBoost) 模型的航班过站时间动态预测方法。首先, 考虑模型输入变量航班密度的连续性和不确定性变化, 利用 UKDE 法估计机场航班密度, 将其作为动态指标输入模型。其次, 引入量子粒子群 (Quantum particle swarm optimization, QPSO) 法优化 XGBoost 模型。最后, 考虑前序航班延误发生前后输入特征的变化, 利用初始预测结果对航班密度进行修正, 得到二阶段预测结果。研究结果表明: 本文方法在高峰时段的预测平均绝对误差为 7.365 min, 效果优于随机森林 (Random forest, RF)、粒子群 (Particle swarm optimization, PSO)-XGBoost 和 XGBoost, 修正后的预测结果平均绝对误差减少了 3.373 min; 模型输入参数按敏感性程度由高到低依次为航班密度、前序航班提前到港时间和延误到港时间。

关键词: 航空运输; 时间预测; 极端梯度提升决策树; 航班过站保障; 核密度估计

中图分类号: V351

文献标志码: A

文章编号: 1005-2615(2023)06-1016-09

Dynamic Prediction of Flight Transit Time Based on UKDE and XGBoost

WU Weiwei¹, XIONG Aoping¹, TANG Hongwu²

(1. College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China;

2. Travelsky Mobile Technology Limited, Beijing 100029, China)

Abstract: To improve the accuracy and reliability of flight transit time prediction during airport peak periods, a dynamic flight transit time prediction method combining the unbiased kernel density estimation (UKDE) and the extreme gradient boosting (XGBoost) models is proposed. Firstly, considering the continuous and uncertain changes in the flight density of the model input variables, the UKDE method is used to estimate the flight density of the airport as a dynamic indicator. Secondly, the quantum particle swarm optimization (QPSO) method is introduced to optimize the XGBoost model. Finally, the changes in the input characteristics before and after the occurrence of the preceding flight delays are considered, and the initial prediction results are modified to obtain the two-stage prediction results. The final results verify that the average absolute error of the prediction in peak hours is 7.365 min, which is better than those of random forest (RF), particle swarm optimization (PSO)-XGBoost and XGBoost, and the average absolute error of the modified prediction results is reduced by 3.373 min. The model input parameters, in descending order of sensitivity, are flight density, early arrival time of preceding flights and delayed arrival time.

基金项目: 国家自然科学基金 (U2033205, U1933118); 南京航空航天大学校企协同育人平台工程实践计划项目 (2022QYGCJSJ59); 南京航空航天大学科研与实践创新计划 (xcxjh20220712)。

收稿日期: 2022-09-08; **修订日期:** 2023-02-06

作者简介: 吴薇薇, 女, 教授, 博士生导师。主要从事航空运输系统仿真与优化研究。在国内外权威刊物和会议上发表论文 50 余篇, 主持过多个重点研究项目; 参与编著《阻塞流理论及其应用》, 出版教材《航空运输经济学》、《运筹学实用教材习题集》。先后获南京航空航天大学教学优秀二等奖及教学成果一等奖, 江苏省研究生优秀教学案例等。

通信作者: 吴薇薇, E-mail: nhwei@nuaa.edu.cn。

引用格式: 吴薇薇, 熊奥萍, 唐红武. 基于 UKDE 和 XGBoost 的航班过站时间动态预测[J]. 南京航空航天大学学报, 2023, 55(6): 1016-1024. WU Weiwei, XIONG Aoping, TANG Hongwu. Dynamic prediction of flight transit time based on UKDE and XGBoost[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2023, 55(6): 1016-1024.

Key words: air transportation; time prediction; extreme gradient boosting(XGBoost); flight ground service; nuclear density estimation

航班过站时间,作为机场航班正常性管理的重要指标之一,是指前序航班到达本站的时间至后续航班离港时间之间的时段^[1]。近年来,在民航业快速发展的背景下,国内大型机场的进离港航班数日益增长,运行环境日趋复杂,特别是在高峰时段,航班过站保障管理面临着更大的压力。在机场实际运行管理中,由于机位分配、航班流量和天气等因素的影响,实际过站时间存在很大不确定性,使用计划或粗略估计的过站时间会产生较大误差,导致航班保障效率降低,甚至扩大航班延误情况。因此,准确把握航班过站时间,有助于机场部门预知过站延误进而制定应对措施,对机场实现航班推出控制精准化、减少航班延误传播具有重要意义。

目前航班过站时间预测的主要方法有启发式算法、回归法、分布拟合、机器学习等。文献[2]基于分布拟合方法得到上客过程在各条件下持续时间分布,并利用最大似然法估计分布的未知参数,进而构建飞机上客时间分布模型。Luo等^[3]根据机场过站整体流程,建立了基于神经网络的地面过站保障 Petri网模型,为实现地面保障的协同决策提供决策依据。孙瑞山等^[4]根据机场实际保障流程建立作业网络图,并基于关键路径法得出该机场保障服务关键路径和保障总耗时。杨亚磊^[5]考虑机型、机场保障能力、天气等多参数的影响建立了过站作业时间模型,利用关键路径法求解得到各机型的最少过站时间,并建立航班过站地面过站流程,建立作业时间模型。邢志伟等^[6]基于机场运行数据拟合得到航班过站服务各理想化环节的分布,在此基础上提出了基于马尔科夫链蒙特卡洛法的航班过站时间估计方法,但所建立模型过于理想化,并未考虑机场保障资源变动的影响。Barratt等^[7]提出了基于飞机时空定位数据的航班保障流程数学模型,并基于高斯混合模型进行参数学习,但模型并未体现出航班保障过程的动态性。高强等^[8]通过构建波及延误树建立了航班过站松弛时间的优化模型。王立文等^[9]结合了初始样本空间生成算法与贝叶斯网络,提出了航班地面过站过程的动态预测方法。邢志伟等^[10]基于支撑向量机模型估计航班过站服务时间,并采用人工鱼群算法对模型超参数进行寻优,但在模型输入变量的选取上,将航班密度分为5类定性指标、将过站时间分为6个静态区间,忽略了其连续性变化的影响。丁建立等^[11]基于贝叶斯网预测了不同条件下的航班过站时间,在处理输入特征时对航班密度仅进行定

性分析,其估计精确性有待改善。冯霞等^[12]基于最邻近算法对进离港航班数量进行预测,将预测结果输入支撑向量回归模型进行航班滑出时间的预测,具有较好的精度,但没有考虑到港延误/提前对进离港航班数量和滑出时间的影响。

从研究现状来看,基于机器学习的预测方法在泛化能力及运算速度上具有显著优势,在预测精度上则较依赖于输入变量的选取与处理。目前大多数研究在预测过站时间时所选取的输入特征多为定性或静态指标(如将航班密度划分为若干类型等),对特征变量的连续性变化鲜有体现。航班实际运行中通常受地面保障、流量控制和天气等多重因素的影响,当发生航班延误后机场及航班的运行数据往往产生变动,事先考虑航班延误的影响建立预测模型,可以避免输入特征波动引起的预测精度下降,目前传统模型较少考虑航班延误对特征变量的影响,在航班实际运行时预测效果欠佳。因此,如何捕获输入特征变量随时间连续性变动的特征,进一步考虑航班延误的影响,实现更为精细的过站时间预测是本文的研究重点。

针对上述问题,本文创新性地提出一种融合无偏核密度估计(Unbiased kernel density estimation, UKDE)与XGBoost集成学习的航班过站时间二阶段预测方法。首先从影响航班过站时间的重要因素——航班密度的时间序列特性出发,利用UKDE生成机场航班密度概率分布函数,并进一步考虑航班延误发生前后输入特征的变化,构建基于量子粒子群算法(Quantum particle swarm optimization, QPSO)寻优的极端梯度提升决策树(Extreme gradient Boosting, XGBoost)过站时间二阶段预测模型,实现航班过站时间的动态预测,从而为机场相关部门过站保障资源的科学调度提供支撑。

1 基于UKDE估计机场航班密度

1.1 核密度估计

由于机场航班密度的不确定性,其不符合已知的某种分布规律,因此传统参数估计无法得到准确的概率密度。相较直方图,核密度估计(Kernel density estimation, KDE)避免了对组距大小和位置的依赖,能够反映变量连续的变化趋势。因此,本文采用无偏核密度估计机场航班密度概率分布,具体表达式为

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (1)$$

式中: $\hat{f}(x, h)$ 为航班概率密度函数, n 为一段时间内的航班数, $K(\cdot)$ 为核函数, x_i 为密度实际值, 带宽 h 一般根据所需精度进行设置。由于在设置带宽后, 核函数的形式对于核密度估计准确性的影响并不显著, 因此选择最常用的高斯函数作为核函数^[13]

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (2)$$

结合式(1~2)得到机场航班核密度估计函数为

$$\hat{f}(x, h) = \frac{1}{\sqrt{2\pi}nh} \sum_{i=1}^n \exp\left[-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2\right] \quad (3)$$

1.2 无偏交叉验证

带宽的选择问题是确保核密度估计准确性的必要条件^[14]。目前最常见的带宽选择方法是经验法、插入法和交叉验证法。经验法和插入法在估计多峰的数据时易出现过度平滑的现象。而无偏交叉验证法在估计波动性强的数据时具有更强的可变性, 因此采用无偏交叉验证方法进行最优带宽选择。

式(4)通过求解平均积分平方误差(Mean integrated squared error, MISE)的最小值来获得最佳带宽 h_0 , MISE的表达式为

$$h_0 = \arg \min \text{MISE}[\hat{f}(x, h)] \quad (4)$$

$$\text{MISE}(h) = E\left\{\int [\hat{f}(x, h) - f(x)]^2 dx\right\} \quad (5)$$

可以看出, MISE的表达式中含有未知函数 $f(x)$, 无法通过计算直接求得, 因此对其进行无偏估计得到表达式为

$$\text{UCV}(h) = \frac{1}{nh^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n K(x_i - x_j)} - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K\left(\frac{x_i - x_j}{h}\right) \quad (6)$$

UCV(h)的最小值得到核密度估计的最佳带宽 h_0 。

2 QPSO-XGBoost模型建立

2.1 XGBoost

XGBoost是基于梯度提升决策树模型进行改进和优化的一种集成学习算法^[15]。具体模型为

$$\hat{y}_i^t = \hat{y}_i^{(t-1)} + \xi_i(x_i) \quad (7)$$

式中: \hat{y}_i^t 为模型第 t 次迭代的第 i 个航班的过站时间预测值, $\xi_i(x_i)$ 为第 t 棵回归树的预测值。

XGBoost模型的目标函数由损失函数

$\sum_i l(y_i, \hat{y}_i)$ 和避免过拟合的正则化项 $\Omega(f_k)$ 构成。

$$\text{Obj} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (8)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\tilde{w}\|^2 \quad (9)$$

式中: y_i 和 \hat{y}_i 分别为过站时间的实际值和预测值, T 和 \tilde{w} 分别为叶子节点的数量和每个叶子节点的预测值; γ 和 λ 为控制正则化的参数。

将第 t 次迭代所得预测值 \hat{y}_i^t 代入式(7), 并对目标函数式(8)进行二阶泰勒展开为

$$\text{Obj} = \sum_i l(y_i, \hat{y}_i^{(t-1)} + \xi_i(x_i)) + \Omega(f_i) \simeq \sum_i \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i \xi_i(x_i) + \frac{1}{2} h_i \xi_i^2(x_i) \right] + \Omega(f_i) \quad (10)$$

式中 g_i 和 h_i 为第 $t-1$ 次迭代 $l(y_i, \hat{y}_i^{(t-1)})$ 对 $\hat{y}_i^{(t-1)}$ 所求一阶导数和二阶导数。进一步移除其中的常数项得到最终目标函数

$$\text{Obj} = \sum_i \left[g_i \xi_i(x_i) + \frac{1}{2} h_i \xi_i^2(x_i) \right] + \Omega(f_i) \quad (11)$$

将样本集合改写为叶子节点集合, 即将 $\varphi_i(x_i)$ 转换为 ω_j , 并令 $G_j = \sum_i g_i$, $H_j = \sum_i h_i$, 令目标函数一阶导数值为0, 求得目标函数的最小值和叶子节点 j 对应的航班过站时间预测 ω_j

$$\omega_j = -\frac{G_j}{H_j + \lambda} \quad (12)$$

$$\text{Obj} = -\frac{1}{2} \sum_j \frac{(G_j)^2}{H_j + \lambda} + \gamma T \quad (13)$$

2.2 量子粒子群法

QPSO算法作为一种基于粒子群法(Particle swarm optimization, PSO)改进的启发式算法, 解决了标准PSO算法收敛速度较慢、易陷入局部最优解的问题^[16]。相较PSO算法, QPSO中引入了粒子平均最好位置 $C_j(t)$, 能够避免粒子陷入次优解的局面。

假设 R 维空间中存在由 N 个粒子组成的种群, 第 t 次迭代时, 依据个体历史最佳位置 $P_{i,j}(t)$ 、种群最佳位置 $G_j(t)$ 和粒子平均最佳位置 $C_j(t)$, 粒子 i 移动得到第 $t+1$ 次迭代的位置为

$$X_{i,j}(t+1) =$$

$$\alpha |C_j(t) - X_{i,j}(t+1)| \ln\left(\frac{1}{u_{i,j}(t)}\right) \pm s_{i,j}(t) \quad (14)$$

$$s_{i,j}(t) = \varphi_{i,j}(t) P_{i,j}(t) + G_j(t) [1 - \varphi_{i,j}(t)] \quad (15)$$

式中: $X_{i,j}(t+1)$ 为第 $t+1$ 次迭代第 j 维空间中第 i 个粒子的位置; α 为收缩扩张因子, 一般不大于1; $u_{i,j}(t)$ 和 $\varphi_{i,j}(t)$ 为第 t 次迭代第 j 维空间中第 i 个粒子的概率函数, 服从(0,1)上均匀分布。

2.3 参数优化步骤

采用 QPSO 对 XGBoost 模型的学习率(Learning_rate)、树的最大深度(Max_depth)以及叶子节点上的样本数(Min_child_weight)、最小损失函数下降值(Gamma) 4 个参数进行优化, QPSO-XGBoost 算法的流程如图 1 所示。

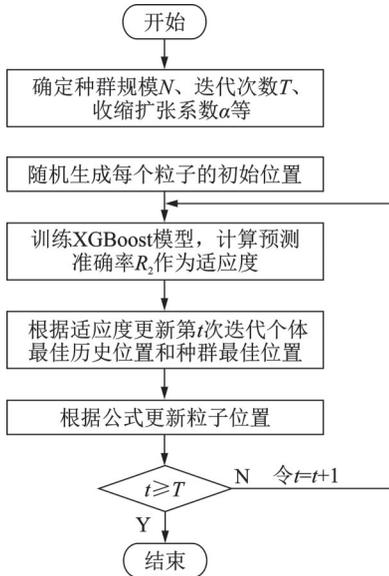


图1 QPSO-XGBoost算法流程图

Fig.1 Flow chart of QPSO-XGBoost algorithm

3 航班过站时间预测模型建立

3.1 模型输入参数的选取

某机场航班计划过站时间与实际过站时间散点分布如图 2 所示,可以看出航班实际过站时间围绕着计划过站时间上下浮动,其主要原因是航班实际运行过程中的一些内外界因素致使飞机实际过站时间偏离计划过站时间。本文结合经验和历史数据选取航班计划过站时间、机型、停靠机位、机场航班密度、前序航班延误时间作为模型输入参数。

(1)机场航班密度。由于机场运行保障资源的有限性,单航班过站时间与机场航班密度存在相关关系。机场某日高峰与低峰时段航班过站时间分布如图 3 所示,高峰时段由于保障资源不足使飞机等待时间增加,航班过站时间相对更长。而低峰时段机场资源更充足,过站时间相对更短。

(2)机位。飞机停机位距离航站楼的远近会使车辆调度时间、飞机滑入/推出时间和上下客时间产生差异,进而影响航班过站时间。根据机场实际数据,将机位划分为近机位、远机位、远远机位 3 种类别。

(3)前序航班延误时间。当前序航班发生延误时,机场部门为尽可能利用过站时间吸收延误会要

求调度保障资源提前到位,进而促进过站时间缩短。而当航班提前到达时,由于保障资源到不及时可能导致航班过站时间延长。

(4)机型。单航班过站时间与其机型的最大座位数呈正相关。相较小型飞机,大型飞机过站时占用机场地面保障设备的时间更长,从而导致其他航班等待保障服务的时间延长。本文按座位数划分飞机类型如表 1 所示。

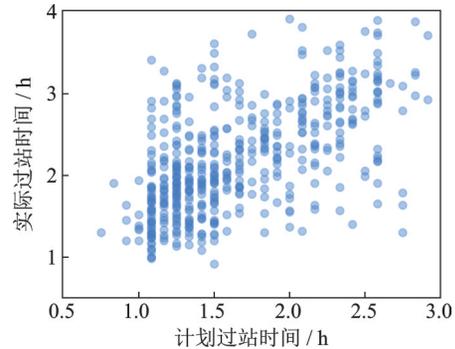


图 2 计划过站时间与实际过站时间散点分布图

Fig.2 Scatter diagram of planned time and actual time

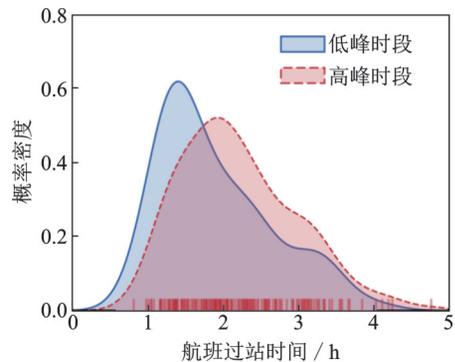


图 3 高低峰时段航班过站时间分布

Fig.3 Distribution of flight transit time during high and low peak periods

表 1 飞机类型划分结果

Table 1 Aircraft type classification results

| 飞机类型 | 座位数 |
|------|---------|
| A | <150 |
| B | 151~250 |
| C | 251~350 |
| D | >351 |

3.2 二阶段预测模型

单航班发生延误后,其后续保障服务所处时段也将随之产生变动,致使对应过站时间段内的机场航班密度发生改变。为了减小因航班延误所造成的特征变量的波动对预测精度的影响,提出了一种航班过站时间的二阶段预测法,具体如图 4 所示。

(1) 计划阶段:生成初始解

根据式(3)利用UKDE法估计一天内的航班密度的概率分布函数 $\hat{f}(x)$,在航班*i*的计划过站时段(s, e)内对 $\hat{f}(x)$ 进行积分得到航班密度。将所得的计划航班密度与机型、停机位、计划过站时间输入QPSO-XGBoost模型中,经过训练得到初步预测结果,数学表达式为

$$D = S_N \times P(s < t < e) = S_N \times \int_s^e \hat{f}(x) dx \quad (16)$$

$$T_i = \phi(T, D, P_k, P_l) \quad (17)$$

(2) 航班执行阶段:过站时段修正

当前序航班执行后更新航班概率分布函数得到 $\hat{f}^*(x)$,利用第1阶段的过站时间预测结果 T_i 和前序航班实际运行情况,在线计算更新航班*i*实际过站时间($s^*, s^* + T$)内的航班密度。根据式(19),将机型、停机位、前序航班延误时间、实际航班密度、计划过站时间作为QPSO-XGBoost模型的输入变量,得到最终预测结果。

$$D^* = S_N \times P(s^* < t < s^* + T) = S_N \times \int_{s^*}^{s^*+T} \hat{f}^*(x) dx \quad (18)$$

$$T_i^* = \phi(T, D^*, P_k, P_l, D_l) \quad (19)$$

式中: T_i, T_i^* 为航班*i*过站时间的初始预测值和最终预测值; $\phi(\cdot)$ 为QPSO-SVM模型; T 为航班*i*的机型; D 和 D^* 为在航班*i*过站时段的机场计划航班密度和估计航班密度; P_k 为航班*i*的停机位; D_l 为航班*i*的到达延误时间; P_l 为航班*i*的计划过站时间; s 和 s^* 分别为航班*i*计划过站开始时间和预计过站开始时间; e 为航班*i*过站计划结束时间。

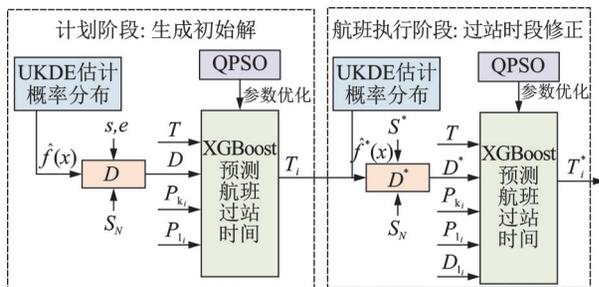


图4 航班过站时间预测模型示意图

Fig.4 Schematic diagram of flight transit time prediction model

4 实例分析

4.1 数据处理

将国内某机场2020年7月~2020年12月的航班过站相关10万余条数据按4:1的比例分为训练集和测试集。由于样本数据不可避免地存在小部分缺失或者异常情况,对样本数据进行预处理,剔

除异常和缺失数据。依据3.1节分析结果选取计划航班过站时间、机型、机位、前序航班延误时间、航班密度作为输入变量。为减小奇异样本数据的不良影响,按照式(20)对输入变量进行归一化处理。

$$w_i = \frac{v_i - v_{\min}}{v_{\max} - v_{\min}} \quad (20)$$

式中: v_i 为第*i*个特征样本值; w_i 为 v_i 归一化后的特征变量。

在采用UKDE法估计航班密度时,考虑实际应用中航班密度带宽 h_0 与当前时刻距离目标预测时刻时长 t_p 之间呈正相关关系,若在一天内采用同一带宽会产生较大偏差。本文假设06:00为当前时刻,并分别采用无偏交叉验证法得到距离当前时刻0~6 h、6~12 h、12~18 h时的最佳带宽 h_0 ,如表2所示。

表2 2020年10月1日~2020年10月5日航班密度分布函数各时段最佳带宽*h*取值

Table 2 Optimal bandwidth *h* values of flight density distribution function from 1 October 2020 to 5 October 2020

| t_p/h | 2020-10-01 | 2020-10-02 | 2020-10-03 | 2020-10-04 | 2020-10-05 |
|---------|------------|------------|------------|------------|------------|
| 0~6 | 0.208 | 0.208 | 0.205 | 0.223 | 0.245 |
| 6~12 | 0.232 | 0.266 | 0.269 | 0.274 | 0.262 |
| 12~18 | 0.231 | 0.293 | 0.250 | 0.314 | 0.283 |

其中6:00~12:00和12:00~18:00的航班密度概率分布曲线如图5、6所示,可以看出:在机场日间运行时,8:00~10:00和15:00~17:00左右整体航班密度较低,属于机场运行低峰时段;10:00~12:00和13:00~15:00为一天中的机场航班密度高峰时期。本文分别选取日间机场运行低峰时段(8:00~10:00)和机场运行高峰时段(13:00~15:00)作为样本划分依据,将样本数据按照飞机到港时间划分为低峰时段、高峰时段两类。

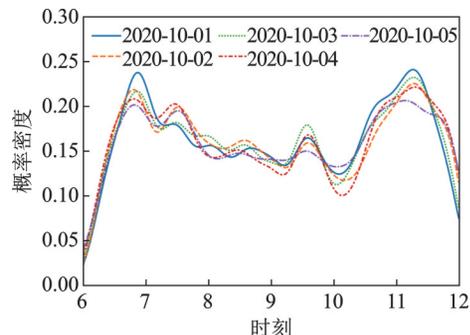


图5 6:00~12:00航班密度概率分布曲线

Fig.5 Flight density probability distribution curves from 6:00 to 12:00

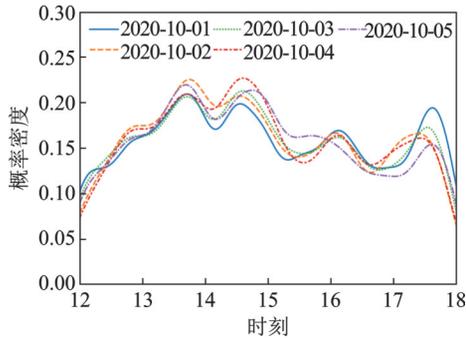


图 6 12:00~18:00 航班密度概率分布曲线

Fig.6 Flight density probability distribution curves from 12:00 to 18:00

4.2 运行结果分析

采用试值法设置 QPSO 的群体规模为 20, 创新参数 α 为 0.05, 优化目标选择为预测准确率 R2 最高。模型参数优化范围及结果如表 3 所示。

表 3 模型超参数优化范围及结果

Table 3 Range of model hyperparameters optimisation and results

| 待优化参数 | 取值范围 | 低峰阶段 | 高峰阶段 |
|------------------|------------|-------|-------|
| Max_depth | (1,20) | 12 | 5 |
| Min_child_weight | (1,10) | 1.537 | 8.129 |
| Learning_rate | (0.01,0.3) | 0.208 | 0.101 |
| Gamma | (0,0.2) | 0.107 | 0.085 |

运行模型及算法, 随机选取 50 个测试样本, 得到高低峰时段预测结果如图 7、8 所示。

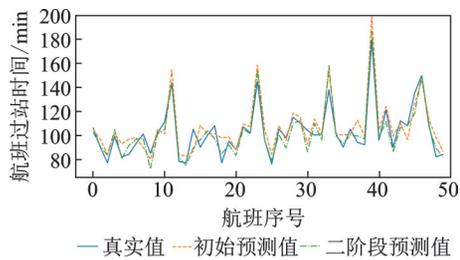


图 7 低峰时段预测结果

Fig.7 Prediction results in low-peak period

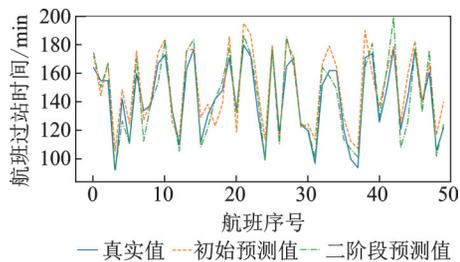


图 8 高峰时段预测结果

Fig.8 Prediction results in high-peak period

采用平均绝对误差 (Mean absolute error, MAE) 和均方根误差 (Root mean squared error, RMSE) 对预测模型进行误差分析和模型评价, 如

表 4 所示, 其他 3 个对比模型的预测误差为经过修正后的结果。本文模型在机场各运行时段内的平均绝对误差都在 10 min 以内, 体现了模型的可靠性和有效性; 模型预测误差能力在高峰时段内稍次于低峰时段, 但修正后的高峰时段的 MAE 和 RMSE 误差分别减少了 3.373 min 和 3.474 min, 体现了本文模型通过对初始解进行过站时段修正产生了较好的效果。

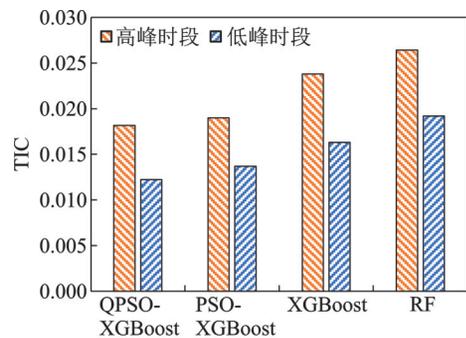
表 4 预测误差分析结果

Table 4 Results of analysis of forecast errors

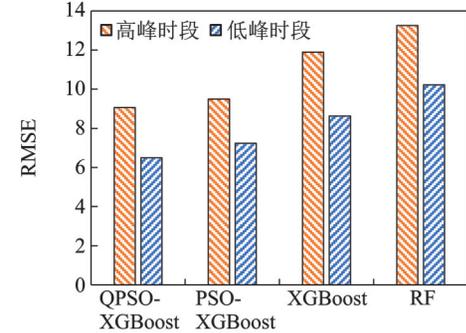
| 方法 | RMSE | | MAE | |
|--------------------------|--------|--------|--------|-------|
| | 高峰时段 | 低峰时段 | 高峰时段 | 低峰时段 |
| 随机森林 (Random forest, RF) | 13.248 | 10.225 | 11.092 | 8.332 |
| XGBoost | 11.895 | 8.633 | 9.189 | 6.615 |
| PSO-XGBoost | 9.497 | 7.233 | 7.608 | 5.574 |
| QPSO-XG 初始误差 | 12.531 | 9.315 | 10.738 | 7.725 |
| Boost 修正后误差 | 9.057 | 6.490 | 7.365 | 5.127 |

4.3 对比分析

为体现 QPSO-XGBoost 模型在机场高峰时段的预测准确度, 采用基于粒子群法寻优的 XGBoost 模型 (PSO-XGBoost)、标准 XGBoost 模型、RF 与其对比。随机抽取高峰时段测试样本, 计算 4 种方法的均方根误差和希尔不等系数 (Theil inequality coefficient, TIC), 结果如图 9 所示。



(a) Comparison of TIC



(b) Comparison of RMSE

图 9 各方法预测误差对比

Fig.9 Comparison of prediction errors of various methods

在预测误差上, QPSO-XGBoost模型的RMSE和TIC值均小于PSO-XGBoost、XGBoost、随机森林3种模型。相较其他模型, 本文模型在高峰阶段运行航班的预测表现明显更佳。在预测速率上, 本文模型的运行时间为113 s, 远远低于PSO-XGBoost模型190 s的运行时间, 较XGBoost和RF模型96 s和64 s运行时间长, 但预测精度大幅提升。虽然QPSO-XGBoost模型并不是对每个航班过站服务时间的估计都非常准确, 但QPSO-XGBoost模型的总体估计更加稳健。

4.4 敏感性分析

相较机型、机位等参数, 机场航班密度及前序航班到港时间在实际运行中往往会产生波动。因此, 为分析航班密度和前序航班到港时间两类易变动参数对航班过站时间的影响程度, 分析敏感性。

4.4.1 航班密度敏感性分析

以原样本的航班密度为基准, 设置步长20%, 将航班密度分别以-40%、-20%、+20%和+40%输入模型中, 同时固定其他变量值, 得到预测结果中各航班过站时间变化率核概率分布曲线如图10所示。

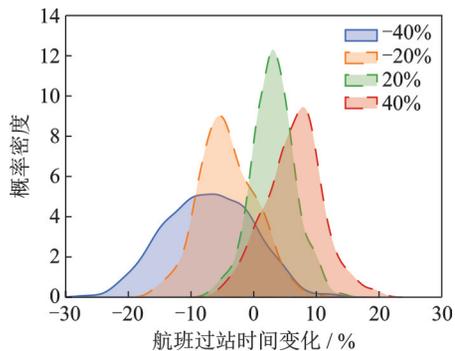


图10 不同航班密度对航班过站时间的影响

Fig.10 Effect of different flight densities on flight transit time

令置信度为90%, 取条件分位数0.95和0.05所对应值作为航班过站时间波动区间的上下限, 计算过站时间变化均值。如表5所示, 当航班密度增加时, 航班平均过站时间增加; 航班密度的变化在±20%时, 大部分航班过站时间的变化率在

表5 航班过站时间变化波动区间及均值

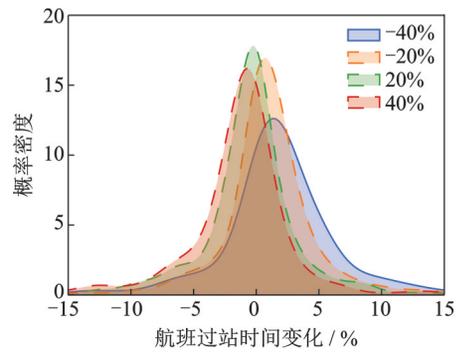
Table 5 Fluctuation interval and mean value of flight transit time variation %

| 航班密度 | 航班过站时间变化率 | |
|------|----------------|-------|
| | 波动区间 | 均值 |
| -40 | (-18.11, 4.00) | -7.13 |
| -20 | (-11.56, 3.63) | -4.18 |
| 20 | (-2.42, 9.18) | 3.15 |
| 40 | (-1.45, 13.83) | 6.42 |

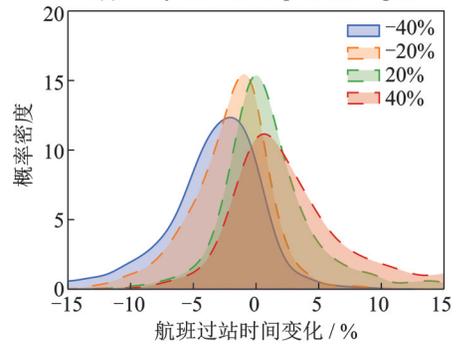
-11.56%~9.18%内波动; 当航班密度的变化为±40%时, 过站时间在-18.11%~13.83%波动, 说明过站时间对航班密度的变化表现出较高敏感性。

4.4.2 前序航班到港时间敏感性分析

将前序航班到港时间分为提前到港和延误到港两类进行分析。固定其他变量值, 将前序航班延误/提前到港时间变化以-40%、-20%、+20%和+40%分别输入模型中, 得到各航班过站时间变化率分布曲线如图11所示。



(a) Delayed arrival of previous flights



(b) Early arrival of previous flights

图11 不同前序航班到港时间对航班过站时间的影响

Fig.11 Effect of different preceding flight arrival time on flight transit time

航班过站时间变化波动区间及均值由表6所示, 可以看出: 当航班延误到港时, 延误时间越长, 则平均过站时间越短; 当航班提前到港时, 提前时间越长, 则平均过站时间越长。航班延误和提前到港时间变化±20%时, 航班过站时间变化率分别

表6 航班过站时间变化波动区间及均值

Table 6 Fluctuation interval and mean value of flight transit time variation %

| 航班到港时间 | 航班过站时间变化率 | | | |
|--------|----------------|-------|----------------|-------|
| | 延误到港 | | 提前到港 | |
| | 波动区间 | 均值 | 波动区间 | 均值 |
| -40 | (-5.83, 10.65) | 2.13 | (-10.73, 1.42) | -3.43 |
| -20 | (-4.94, 6.79) | 1.00 | (-7.76, 2.18) | -2.07 |
| 20 | (-6.69, 5.45) | -0.48 | (-3.93, 8.16) | 0.97 |
| 40 | (-8.61, 4.04) | -1.41 | (-2.93, 14.81) | 2.96 |

在 $-6.69\% \sim 6.79\%$ 和 $-7.76\% \sim 8.16\%$ 内波动;当航班延误和提前到港时间变化 $\pm 40\%$ 时,航班过站时间变化率分别在 $-8.61\% \sim 10.65\%$ 和 $-10.73\% \sim 14.81\%$ 内波动。由上述分析可知,相较于延误到港,过站时间对前序航班提前到港时间的变化表现出更高的敏感性。

5 结 论

本文构建一种结合UKDE与QPSO-XGBoost集成学习模型的两阶段预测法,实现了机场高低峰时段内的各航班过站时间的精确预测,并利用机场实际运行数据进行验证,得到主要结论有:

(1)构建了融合QPSO的XGBoost集成学习模型,相较PSO-XGBoost、XGBoost、RF法具有更好的预测精度和泛化能力,能够有效挖掘过站相关特征和航班过站时间之间的内在关联,实现按不同时段细化的航班过站时间预测;

(2)基于无偏交叉验证改进的核密度法估计航班密度的概率分布,揭示了机场过站航班数量的波动特性随时间而呈现的非线性变化规律,突破了传统过站时间预测方法缺乏考虑航班密度连续性变化的问题;

(3)为减小因航班延误造成的航班密度的波动对预测精度的影响,提出了一种预测航班过站时间的二阶段方法,并通过实例验证得到经过站时段修正后的预测结果在高峰时段具有更高的精度,能够为运营管理者提供更可靠的航班过站信息预警;

(4)通过敏感性分析呈现了在不同航班密度和前序航班到达时间下过站时间预测值的变化规律,得到了模型中易变动参数按敏感性程度由高到低依次为航班密度、前序航班提前到港时间和前序航班延误到港时间。

参考文献:

- [1] 戚彦龙. 航班延误分级和过站时间研究[D]. 南京:南京航空航天大学,2014.
QI Yanlong. Research on flight connection time and delays classification[D]. Nanjing: Nanjing University of Aeronautics and Astronautics,2014.
- [2] 冯霞,张鑫,陈锋. 飞机过站上客过程持续时间分布[J]. 交通运输工程学报,2017,17(2): 98-105.
FENG Xia, ZHANG Xin, CHEN Feng. Boarding duration distribution of aircraft turnaround[J]. Journal of Traffic and Transportation Engineering,2017,17(2): 98-105.
- [3] LUO Q, CHEN Y, CHEN L, et al. Research on situation awareness of airport operation based on Petri nets[J]. IEEE Access, 2019(99): 25438-25457.
- [4] 孙瑞山,张子全. 基于CPM的停机坪航班保障工作方法研究[J]. 中国民航大学学报,2011,29(5): 23-26.
SUN Ruishan, ZHANG Zitong. Study on apron flight service work method based on CPM[J]. Journal of Civil Aviation University of China,2011,29(5): 23-26.
- [5] 杨亚磊. 机场航班过站地面保障建模与时间优化[D]. 天津:中国民航大学,2020.
YANG Yalei. Modeling and time optimization of airport flight passing ground assurance[D]. Tianjing: Civil Aviation University of China, 2020.
- [6] 邢志伟,唐云霄. 枢纽机场航班保障服务时间估计[J]. 系统仿真学报, 2017, 29(11): 2856-2864, 2874.
XING Zhiwei, TANG Yunxiao. Flight support service time estimation of hub airport[J]. Journal of System Simulation,2017,29(11): 2856-2864,2874.
- [7] BARRATT S T, KOCHENDERFER M J, BOYD S P. Learning probabilistic trajectory models of aircraft in terminal airspace from position data[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(9): 3536-3545.
- [8] 高强,周覃,陈欣. 基于波及延误的航班过站松弛时间重分配[J]. 华南理工大学学报(自然科学版), 2019,47(10): 151-156.
GAO Qiang, ZHOU Tan, CHEN Xin. Redistribution of flight delay slack time based on spread delay[J]. Journal of South China University of Technology(Natural Science Edition),2019,47(10): 151-156.
- [9] 王立文,李彪,邢志伟,等. 过站航班地面保障过程动态预测[J]. 北京航空航天大学学报,2021,47(6): 1095-1104.
WANG Liwen, LI Biao, XING Zhiwei, et al. Dynamic prediction of ground support process for transit flight[J]. Journal of Beijing University of Aeronautics and Astronautics,2021,47(6): 1095-1104.
- [10] 邢志伟,韩大浩,罗谦. 基于改进GA的神经网络航班保障时间估计[J]. 计算机工程与设计, 2020,41(1): 107-114.
XING Zhiwei, HAN Dahao, LUO Qian. Estimation of flight support time based on improved GA neural network[J]. Computer Engineering and Design,2020,41(1): 107-114.
- [11] 丁建立,赵键涛,曹卫东. 基于贝叶斯网的航班过站时间动态估计[J]. 南京航空航天大学学报,2015,47(4): 517-524.
DING Jianli, ZHAO Jiantao, CAO Weidong. Dynamic estimation about turnaround time of flight based on Bayesian network[J]. Journal of Nanjing University of Aeronautics & Astronautics,2015,47(4): 517-524.
- [12] 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.
- [13] LI M Z, RYERSON M S, BALAKRISHNAN H. Topological data analysis for aviation applications[J]. Transportation Research Part E Logistics and Transportation Review, 2019, 128: 149-174.
- [14] 刘玉梅,陈云,赵聪聪,等.基于KDE及Markov的高速列车传动系振动评价及可靠性分析[J].西南交通大学学报,2022,57(4):783-790,796.
- LIU Yumei, CHEN Yun, ZHAO Congcong, et al. Vibration evaluation and reliability analysis of high-speed train transmission system based on kernel density estimator and Markov model[J]. Journal of Southwest Jiaotong University, 2022, 57(4): 783-790, 796.
- [15] 张薇薇,刘盾,贾修一.基于XGBoost的三分类优惠券预测方法[J].南京航空航天大学学报,2019,51(5):643-651.
- ZHANG Weiwei, LIU Dun, JIA Xiuyi. Three classified coupon prediction based on xgboost algorithm[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 643-651.
- [16] OMKAR S N, KHANDELWAL R, ANANTH T V S, et al. Quantum behaved particle swarm optimization (QPSO) for multi-objective design optimization of composite structures[J]. Expert Systems with Applications, 2009, 36(8): 11312-11322.

(编辑:陈珺)