

# 基于两阶段模型的空铁联运客流量预测方法

——以沪蓉通道为例

郭九霞<sup>1</sup>, 田金玉<sup>1</sup>, 钟庆伟<sup>1</sup>, 陈 曲<sup>2</sup>

(1. 中国民用航空飞行学院空中交通管理学院, 四川 广汉 618307; 2. 民航局运行监控中心运行监控处, 北京 100710)

**摘要:** 随着国内综合交通立体枢纽的快速建设与发展,空铁联运模式给旅客出行带来了更多便利,准确地掌握空铁联运客流量对提升综合运输服务质量及保障运输安全至关重要。以城市群间空铁联运为背景,采用粒子群优化-随机森林(PSO-RF)模型和 Logit 模型相结合的两阶段模型,对出行路径客流量进行预测。以单向“空转铁”为实例。第 1 阶段,基于历史数据通过 PSO-RF 模型预测出行线路的日均民航客流量;第 2 阶段,对离港旅客进行行为调查,分析旅客换乘方式的特征及其选择行为,基于非集计理论构建旅客出行选择的二元 Logit 模型,并计算换乘高铁的分担率;最后,综合两阶段模型的结果,计算出行线路客流预测量。以上海至成都周边城市群构建算例,对所提出方法的有效性和可行性进行验证。结果表明,两阶段模型的准确率达到 80.40%。

**关键词:** 综合运输;空铁联运;客流预测;粒子群优化-随机森林(PSO-RF)模型;Logit 模型

**中图分类号:** U15 **文献标志码:** A **文章编号:** 1671-1807(2025)05-0082-06

随着国家立体交通规划纲要的提出,国内将建设 20 个左右国际性和 80 个左右全国性综合交通枢纽城市。预计到 2025 年,这些枢纽城市 80% 将实现空铁一体化。在距离较远的城市群之间,出行变得更加便捷,换乘耗时也越来越短,选择空铁联运作为出行方式的旅客将大幅度增加。空铁联运运行模式将成为城际旅客首选的出行方式,大型枢纽机场通常作为综合交通运输中心,可以实现多种交通的无缝衔接。

客流预测方法既有研究主要聚焦于旅客出行行为分析构建单个模型来预测客流。刘伟玲<sup>[1]</sup>从不同客流分析空铁联运客流影响因素,建立基于系统动力学的客流预测模型。吴薇薇等<sup>[2]</sup>构建出行效用函数,将客流分配到空铁联运出行路径上,得到了机场空铁联运需求。李一博<sup>[3]</sup>在研究空铁联运 OD (origin-destination,起讫点)间的协同优化时考虑了联运客流,从节点属性、直达情况、服务方案的数量和质量来分析客流的影响因素,并使用深度学习

建立了客流预测模型。刘翌洋等<sup>[4]</sup>研究了时空层面的空铁联运旅客出行情况,分析了对影响旅客选择的因素。李兴华等<sup>[5]</sup>基于问卷调查,分析了旅客对空铁联运的需求及偏好。胡晓伟等<sup>[6]</sup>对神经网络模型进行了优化和改进,考虑了时间和空间特征,针对地铁站的客流量进行了预测。随机森林回归预测(random forest regression, RFR)模型通过采用多个决策树有效预防训练数据集的过度拟合现象<sup>[7]</sup>。RFR 常被应用于多个领域的预测。Guo 等<sup>[8]</sup>等采用将随机森林回归模型与最大信息系数结合起来的方法预测航班离港延误。王宗博等<sup>[9]</sup>运用随机森林算法对雷暴气象下各气象细分因子进行重要性排序。罗凤娥等<sup>[10]</sup>采用包括随机森林算法在内的 5 类经典数据挖掘算法,考虑航班延误受多种因素交叉影响,对航班延误时间进行预测。陈滢<sup>[11]</sup>将随机森林用于风险预测,并与 K 近邻(KNN)和 Logistic 预测模型相对比,认为随机森林模型结果最优。

**收稿日期:** 2024-09-24

**基金项目:** 四川省社科基金(SCJJ23ND186);中央高校基本科研业务费专项(PHD2023-041);民航教育人才类项目(MHJY2023010);中央高校教育教学改革专项(E2024024)

**作者简介:** 郭九霞(1981—),女,山西临汾人,博士,教授,研究方向为复杂系统韧性、空中交通管理、综合立体交通等;田金玉(1998—),女,四川遂宁人,硕士,研究方向为复杂系统韧性、空中交通管理、综合立体交通等;钟庆伟(1991—),男,四川什邡人,博士,副教授,研究方向为交通运输规划与管理、交通运输组织优化、机器学习、强化学习;陈曲(1987—),男,工程师,研究方向为空中交通管理、民航签派放行、飞机性能等。

Logit 模型被广泛应用于存在多种出行方式时,采用问卷调查讨论影响旅客出行方式选择的因素并得出各个出行方式的分担率。翁剑成等<sup>[12]</sup>为了研究旅客出行方式的选择,通过枢纽高铁站问卷调查数据,构建考虑心理倾向的 Logit 模型,从而定量分析不同时段下旅客选择接续方式的概率。徐明非等<sup>[13]</sup>结合了距离传输曲线模型和多变量 Logit 模型,并根据拟合结果分析了分担率对多个因素的敏感性。樊海博<sup>[14]</sup>通过对京沪通道的旅客进行了 SP(stated preference)和 RP(revealed preference)调查,研究了地区内多种交通方式的分担率。

综上,既有空铁联运客流预测研究一般仅采用单模型,且少见将两阶段模型运用于空铁联运。除此之外,还很少见将粒子群优化算法与随机森林结合起来运用在客流预测上。本文针对城市群间的空铁联运问题,通过在联运换乘衔接处设置问卷调查,构建两阶段模型。

两阶段法的运用,无论“空转铁”或是“铁转空”是一样的,所以下文将“空转铁”作为出行方式来分析。第 1 阶段采用粒子群优化-随机森林模型(particle swarm optimization-random forest, PSO-RF)模型来预测航空的客流量,该方法能有效提升客流预测的准确性;第 2 阶段采用 Logit 模型预测铁路旅客分担率。以沪蓉间空铁联运 OD 客流为例进行预测。准确预测空铁联运客流,不仅有助于保障旅客出行的畅通无阻,还对合理规划空铁联运交通运输路径有参考意义。

## 1 模型构建

### 1.1 问题描述

距离相隔较远的城市群间采取“空转铁”的方式出行,出行全过程分析如图 1 所示。出行全过程分为两个阶段。第 1 阶段,旅客选择从城市群外部城市 A 出发先乘坐飞机到达城市群内部城市 B;第 2 阶段,旅客从城市 B 乘坐高铁到达在同一城市群的目的城市 C。

### 1.2 客流影响因素分析及数据处理

将影响空铁联运客流的多种因素分为内在因

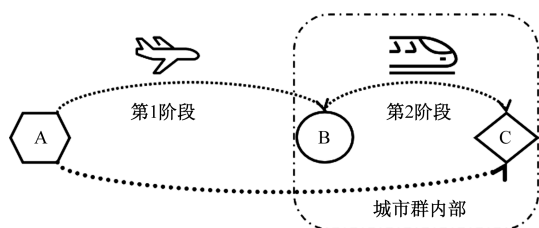


图 1 出行全过程分析

素及外在因素,并从这两方面进行讨论。内在因素主要为旅客的年龄、收入、性别和学历,外在因素为天气因素和节假日。其中,天气因素又包含了出发站天气和到达站天气。客流的影响因素见表 1。

表 1 客流的影响因素

内在因素				外在因素		
旅客因素				天气因素		节假日
年龄	性别	收入	学历	出发站天气	到达站天气	

### 1.3 基于 PSO-RF 模型的客流预测建模

通过 PSO 算法优化 RF 模型来预测日均航班客流量。

**步骤 1:**划分训练集和测试集。从数据里面随机且有放回地抽取  $t$  个样本作为预测模型的训练集  $Q(x, y)$ 。定义航班客流预测训练集合  $x \rightarrow y$ 。其中,  $y \in (y_1, y_2, \dots, y_t)$  为模型中第  $t$  天内的历史航班客流量; $x$  为客流影响外在因素所建立的  $a$  个特征向量,  $x \in (x_1, x_2, \dots, x_a)$ 。剩余未被抽中的数据则作为样本测试集  $(x_i, y_i)$ 。

**步骤 2:**特征子集优化选择。通过不断迭代更新,利用 PSO 算法对特征子集进行优化,选择出最佳的特征子集。在 PSO 算法中,每个粒子代表一个特征子集,粒子的位置和速度是根据该特征子集的二进制表示随机生成的。粒子的位置表示特征的选择情况,速度表示特征的变化趋势。首先初始化粒子的位置和速度,然后更新粒子的速度和位置。Fitness 函数用于交叉验证评估每个粒子的质量,公式为

$$\text{Fitness} = \text{MAE}_i \quad (1)$$

式中:  $\text{MAE}_i$  为第  $i$  个验证集的预测平均绝对误差。

**步骤 3:**建立单棵回归决策树。根据训练集中的特征向量  $x$  和历史航班客流量  $y$ ,通过搜索和分裂,每个决策树的节点分裂为 2 个部分。当第  $n$  节点都分枝成  $c$  和  $d$  两个节点后,且  $c$  和  $d$  节点的基尼系数接近于最小时,将特征  $k$  对应的基尼系数作为在  $n$  节点处的基尼系数,即节点  $n$  的基尼指数为

$$\text{GI}_n = 1 - \sum_{k=1}^a [p(k/n)]^2 \quad (2)$$

$$p(k/n) = \frac{(X_k, Y_n)}{\sum_{k=1}^a (X_k, Y_n)} \quad (3)$$

式中:  $a$  为类别总数;  $p(k/n)$  为节点  $n$  处特征  $k$  的样本占比;  $(X_k, Y_n)$  为第  $n$  个样本对应的第  $k$  个特征值;  $\text{GI}_n$  为第  $n$  个节点的基尼指数节点在集合  $z$

中,则该单棵回归树的基尼系数为

$$\text{VIM}_{rk}^{\text{gini}} = \text{GI}_n - \text{GI}_c - \text{GI}_d, n \in z \quad (4)$$

式中: $\text{GI}_c$  和  $\text{GI}_d$  分别表示分枝后  $c$  和  $d$  两个新节点的基尼系数。

**步骤 4:** 构建多棵回归决策树。在步骤 2 已经建立好了单棵回归树,在此基础上,进一步随机抽样和特征选择,建立多个决策树,并集成它们的预测结果。

(1) 从原始样本中随机有放回地抽取  $t$  个样本集,作为循环深度学习训练的训练集,最终建立测试集  $(x_i, y_i)$  作为模型的预测集。

(2) 原始数据集变量个数为  $t$ , 在每个决策树模型的内部节点中随机选择  $i (i < t)$  个变量作为替代的分支变量,然后在构建单一决策树的过程中寻找最佳分支。

(3) 每个决策树自上而下分层。随机森林方法允许通过控制决策树的分支节点大小来调节决策树的生长。选择人为地控制决策树的分支数量,或让决策树自由生长,得到有约束的最优单一决策树。

(4)  $i$  棵决策树按照数据属性,遵循规则组合形成了随机森林模型,该模型为

$$e = \{h(X, \theta_i)\}, i = 1, 2, \dots, t \quad (5)$$

式中: $\theta_i$  为第  $i$  棵回归树; $X$  为客流影响因素向量; $h(X, \theta_i)$  为第  $i$  棵回归树的预测值。生成的随机森林是多元非线性回归分析模型,随机森林预测值是  $i$  棵树预测值的平均值。

**步骤 5:** 对 PSO-RF 模型进行验证。可以使用精度  $P$ 、平均绝对百分比误差 MAPE、平均绝对误差 MAE 和拟合优度  $R^2$ 、均方根误差 RMSE。

精度  $P$  计算公式为

$$P = \frac{1}{a} \sum_{i=1}^a \left(1 - \left| \frac{y_i - y_i^*}{y_i^*} \right| \right) \times 100\% \quad (6)$$

式中: $y_i$  为实际客流量; $y_i^*$  为模型的预测客流量; $t$  为样本数量。

平均绝对百分比误差 MAPE 计算公式为

$$\text{MAPE} = \frac{100\%}{a} \sum_{i=1}^a \left| \frac{y_i^* - y_i}{y_i} \right| \quad (7)$$

平均绝对误差 MAE 计算公式为

$$\text{MAE} = \frac{1}{a} \sum_{i=1}^a |y_i^* - y_i| \quad (8)$$

拟合优度  $R^2$  计算公式为

$$R^2 = \frac{\sum_{i=1}^a (y_i^* - \bar{y})^2}{\sum_{i=1}^a (y_i - \bar{y})^2} \quad (9)$$

式中: $\bar{y}$  为实际客流量的平均值。

均方根误差 RMSE 计算公式为

$$\text{RMSE} = \sqrt{\frac{1}{a} \sum_{i=1}^a (y_i^* - y_i)^2} \quad (10)$$

**步骤 6:** 分析特征重要性。将步骤 3 中得到的单棵回归树基尼系数进行加权平均,得到总体的特征重要性为

$$\text{VIM}_k^{\text{gini}} = \sum \text{VIM}_{rk}^{\text{gini}} \quad (11)$$

可以得到每个特征在整个随机森林中的相对重要性,特征的重要性越高,表示该特征对于随机森林的建模和预测有着更大的影响。

#### 1.4 Logit 模型构建

在空铁联运衔接处所考虑的换乘涉及的出行方式仅有选择高铁与不选择高铁两种模式,因此采用二项 Logit 模型构建客流分担率模型。选取了包括 4 个旅客因素在内的 8 个影响因素来分析客流,假设效用函数  $V_x$  与  $\partial_{xmi}$  呈线性关系<sup>[15]</sup>。则旅客选择第  $n$  种运输方式的效用函数形式为

$$V_x = V_{xm} + a = \sum_{i=1}^8 Q_i \partial_{xmi} + a \quad (12)$$

式中: $V_x$  为旅客选取第  $x$  种运输方式出行的效用函数值; $\partial_{xmi}$  为旅客  $x$  的选择第  $n$  种运输方式的第  $i$  个影响因素; $Q_i$  为第  $i$  个影响因素对应的参数; $a$  为常数。

则旅客换乘选择高铁出行的概率为

$$P = \frac{e^{V_1}}{e^{V_1} + e^{V_2}} \quad (13)$$

式中: $V_1$  为选择高铁; $V_2$  为不选择高铁。

## 2 实例分析

成都作为拥有双机场的西南枢纽城市,上海作为全国经济第一的城市,机场的繁忙程度可想而知,因此本文将沪蓉通道作为典型的实例进行具体分析。研究分为以下两个阶段。

第 1 阶段以沪蓉通道(上海虹桥机场-成都双流机场、上海浦东机场-成都双流机场、上海虹桥机场-成都天府机场、上海浦东机场-成都天府机场 4 条出行链)为研究背景,选择 2023 年 8 月 1 日至 2024 年 1 月 31 日上海到成都的航班客流数据作为实验数据,采用 PSO-RF 模型从日均客流角度对上海-成都的航班客流进行预测。

第 2 阶段,随着成都天府机场高铁站的开通运行,两个机场成为交通运输综合枢纽中心,实现空地地铁一体化下的“空转铁”联运变得十分方便。以

成都“一市两场”为 RP 问卷的主体,对机场出站换乘的旅客进行出行调查,设计了行为意向的调查问卷。采用线上和线下相结合的方式进行调查,在 2024 年 1 月 1—14 日,共计回收问卷 1 000 份,有效问卷 950 份,有效率为 95%。其中包含了个人属性 8 项,将采集到的数据用于分析与旅客选择行为的关系,获取高铁出行的分担率。

## 2.1 第 1 阶段: PSO-RF 模型

### 2.1.1 数据收集、清洗

查找相关机场所在城市的历史天气情况,以及城市、日期、天气、机场、节假日等信息,基于飞常准上的历史航班信息,关联处理形成数据信息。对这些数据进行清洗和规范化。将所有特征值处理成为数值类型。

### 2.1.2 模型精度验证

利用式(6)~式(10),验证基于 PSO-RF 模型的航班客流量预测模型的精度  $P$ 、平均绝对百分比误差 MAPE、平均绝对误差 MAE 和拟合优度  $R^2$ 、均方根误差 RMSE。

### 2.1.3 影响因素重要性评估

根据式(11)可对影响因素重要性进行评估。影响因素重要性排序如图 2 所示。由图 2 可知,影响因素的重要性排序的前 5 依次为航线、日期、班次、出发站天气、节假日。航线的不同对航班客流量的影响较大,可见对不同航线的客流都进行预测是很有必要的。

### 2.1.4 用重要性排序前 5 的影响因素预测客流及精度验证

在 2.1.3 中对影响因素重要性进行了评估,得到了重要性排序。使用重要性排序前 5 的影响因素重新训练模型再次预测客流。以 2024 年 1 月 1—

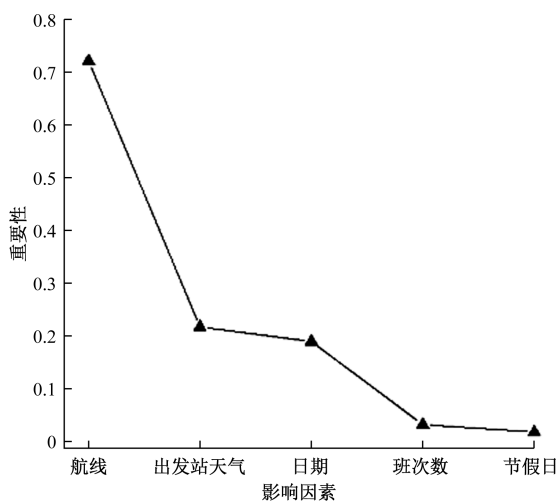


图 2 影响因素重要性排序

14 日为例,日均客流预测结果见表 2。

预测客流和真实客流之比为 PSO-RF 模型的客流准确率。由表 2 可知,2024 年 1 月 1—14 日,上海飞成都的日均航班预测客流为 7 208 人,而日均真实航班客流为 6 477 人。从预测客流和真实客流可计算得到预测客流准确率为 89.9%。

将使用全部影响因素预测客流结果的评价,与使用重要性排序前 5 的影响因素预测客流结果的评价进行对比,PSO-RF 模型的评价见表 3。

表 2 日均客流预测结果

起点	终点	预测客流/人
上海虹桥机场	成都双流机场	2 157
上海浦东机场	成都双流机场	1 971
上海虹桥机场	成都天府机场	1 604
上海浦东机场	成都天府机场	1 476

表 3 PSO-RF 模型的评价

分类	$R^2$	MAE	MAPE/%	$P$ /%	RMSE
前 5	0.85	152.34	10.63	90	224.52
全部	0.84	154.09	10.74	89	224.52

并利用式(6)~式(10),再一次验证 PSO-RF 预测模型的精度  $P$ 、平均绝对百分比误差 MAPE、平均绝对误差 MAE 和拟合优度  $R^2$ 、均方根误差 RMSE。将模型的评价结果对比可知,使用重要度排序前 5 的影响因素来训练模型,得到的结果更精确。使用影响程度更大的特征预测得到的客流量与真实值更接近。

## 2.2 第 2 阶段: Logit 模型

二项 Logit 模型的选择分别为选择高铁和不选择高铁。影响换乘过程中出行者的属性变量定义见表 4。

对属性进行定义后,使用 SPSS26.0 对 950 条数据进行参数值标定,其结果见表 5。显著性小于 0.05 代表变量对结果有显著的影响,应该纳入模型中;而大于 0.05 的属性则不应该纳入模型中。因此,是否愿意换乘、是否有儿童与老人、大件行李数目、天气情况、性别,以上 5 个属性应纳入模型中;相反,受教育程度、年龄、月收入,这 3 个属性不应纳入。

将参数标定结果代入式(13)中,再根据问卷可计算得到换乘高铁的客流分担率为 26.3%。

根据 RP 问卷调查结果可知,在成都天府机场和双流机场换乘高铁的客流分担率,也就是采用“空转铁”出行方式的客流占比为 36.4%。本文认为预测分担率与调查分担率之比为客流准确率,可知该模型计算得到的客流准确率为 72.25%。

表 4 属性变量定义

属性	定义
是否愿意换乘	令愿意换乘=1;不愿意换乘=2
是否有儿童和老人	令有儿童和老人=1;没有儿童和老人=2
大件行李数目	令行李数分别定义为 0,1,2,3
天气情况	令小雨雪;中雨雪;大雨雪分别定义为 1,2,3
受教育程度	令初中以下;高中或中专;本科或专科;研究生及以上分别定义为 1,2,3,4
月收入	令 1 500 元以下;1 500~3 000 元;3 000~5 000 元;5 000~8 000 元;8 000~10 000 元;10 000~15 000 元;15 000 元以上分别定义为 1,2,3,4,5,6,7
年龄	令 18 岁以下;18~29 岁;30~39 岁;40~49 岁;50~59 岁;60 岁以上分别定义为 1,2,3,4,5,6
性别	令女性=1;男性=2

表 5 参数标定结果

属性	B	标准误差	瓦尔德	自由度	显著性	Exp(B)
是否愿意换乘	1.166	0.282	17.103	1	0.000	3.208
是否有儿童和老人	0.931	0.347	7.215	1	0.007	2.537
大件行李数目	-0.327	0.148	4.899	1	0.027	0.721
天气情况	-0.306	0.133	5.302	1	0.021	0.737
受教育程度	0.135	0.093	2.099	1	0.147	1.144
月收入	-0.081	0.073	1.239	1	0.266	0.922
年龄	-0.075	0.080	0.871	1	0.351	0.928
性别	-0.659	0.249	6.993	1	0.008	1.409

2.3 结果分析

(1)以历史航班客流数据为训练集,得到一周内的日均航班客流预测值,将预测值和真实值之比作为 PSO-RF 模型预测的客流准确率,其结果为 89.90%。真实值与预测值对比如图 3 所示。由图 3 可知,预测客流量与实际客流量较接近,能够清晰地捕捉到真实流动的变动趋势。

(2)采取问卷调查的方式获取调查分担率,并根据建立的 Logit 模型求得预测分担率,预测分担率与调查分担率之比作为客流准确率,其结果为 72.25%。由第 1 阶段 PSO-RF 模型预测的日均客流和第 2 阶段 Logit 模型求得的分担率,即求得日均“空转铁”预测客流量。如此,将上海飞成都的日均航班预测客流结果分别乘以预测分担率即为日均“空转铁”预测客流量。将上海飞成都的日均真实客流量乘以调查分担率即为日均“空转铁”真实客流量。

(3)由日均“空转铁”预测客流量与日均真实客流量之比得到两阶段预测模型的客流准确率,其结果为 80.40%。模型预测结果评价见表 6。

此外,通过建立离散选择行为模型分析发现,旅客同行是否有儿童和老人、是否愿意换乘、天气情况、大件行李数目与是否选择联运有显著影响。旅客选择出行关系如图 4 所示。图中很直观地反映出了旅客是否有儿童和老人、是否愿意换乘这两个出行属性与是否选择“空转铁”出行的联系。

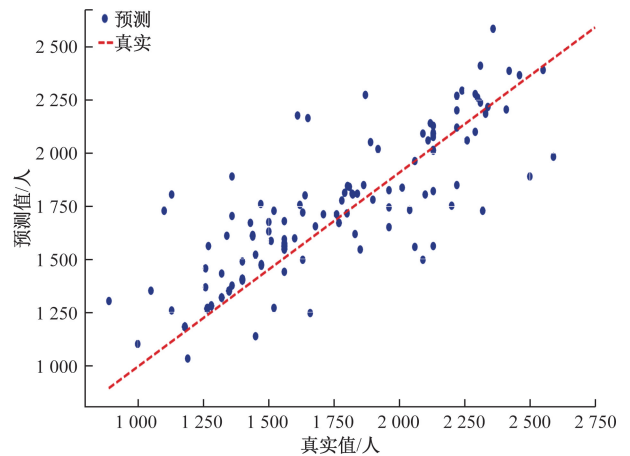


图 3 真实值与预测值对比

表 6 模型预测结果评价

模型	客流量准确率/%
PSO-RF 模型	89.90
Logit 模型	72.25
两阶段预测模型	80.40

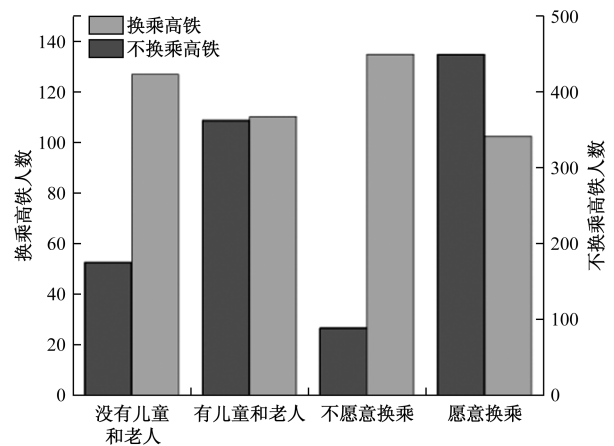


图 4 旅客选择出行关系图

3 结论

第 1 阶段利用 PSO 效率高、优化性能强的特点,结合 RF 模型的优点,使用了一种将 PSO 与 RF 相结合的混合方法。该混合方法有效地避免了优化过程中的过拟合或欠拟合问题,提高了预测模型的鲁棒性和泛化性。后续融合了 PSO-RF 模型及 Logit 模型的优点,提出了一个两阶段法来预测“空转铁”联运客

流。该方法不再局限于用一个方法或一个阶段计算出客流,在数据缺乏情况下同样适用。由两阶段模型预测的客流量准确率可知,这个模型具有较好的准确性,可用于有效地预测“空转铁”的客流。

PSO-RF 模型对客流影响因素进行了重要性测评。结果表明,航线、日期、班次、出发站天气、节假日对空铁联运客流影响较大。将这 5 个影响因素重新训练模型再次预测客流,发现 5 个评价指标都有明显的提高,这些指标可以量化预测值与真实值之间的差异,从而判断模型的拟合程度。可知,采用重要性排序前 5 的影响因素重新训练模型,得到的预测结果比采用全部影响因素训练模型得到的预测结果,准确度更高,更接近实际客流数据。

旅客行为选择模型验证了旅客是否有儿童和老人、是否愿意换乘这两个出行属性与是否选择“空转铁”有显著联系。对建立的旅客离散选择行为模型进行分析,可为运营商吸引旅客使用“空转铁”联运、政府推广空地铁一体化以及为旅客的出行提供更好的服务提供参考。

### 参考文献

- [1] 刘伟玲. 基于系统动力学的空铁联运客流预测[J]. 科技创新与应用, 2015(13): 29.
- [2] 吴薇薇, 徐幼婷, 刘硕, 等. 考虑出行效用的空铁联运需求研究[J]. 华东交通大学学报, 2021, 38(1): 136-141.
- [3] 李一博. 考虑联运客流的空铁时刻表协同优化研究[D]. 北京: 北京交通大学, 2023.

- [4] 刘翌洋, 聂磊, 佟璐, 等. 基于旅客感知的空铁联运路径选择研究[J]. 铁道运输与经济, 2022, 44(12): 36-42.
- [5] 李兴华, 李思雨, 成诚, 等. 空铁一体枢纽联运服务需求及偏好研究[J]. 综合运输, 2020, 42(6): 8-12.
- [6] 胡晓伟, 吴则洋, 卢泓博, 等. 基于 GA 优化深度神经网络的地铁站点客流预测[J/OL]. 交通运输工程与信息学报, 1-16 [2024-08-23]. <https://doi.org/10.19961/j.cnki.1672-4747.2024.07.023>.
- [7] BREIMAN L. Random forests [J]. Mach Learning, 2001, 45: 5-32.
- [8] GUO Z, YU B, HAO M Y, et al. A novel hybrid method for flight departure delay prediction using random forest regression and maximal information coefficient[J]. Aerospace Science and Technology, 2021, 116: 1-10.
- [9] 王宗博, 包杰, 单小轩, 等. 考虑雷暴气象细分因子的航班延误集成学习预测[J]. 航空计算技术, 2022, 52(3): 52-56.
- [10] 罗凤娥, 王波, 李娜, 等. 基于数据挖掘技术的航班延误预测综述[J]. 科技和产业, 2020, 20(11): 75-80.
- [11] 陈滢. 基于网格搜索和随机森林的汽车信贷违约预测研究[J]. 科技和产业, 2023, 23(9): 116-121.
- [12] 翁剑成, 马靖媛, 潘晓芳, 等. 考虑心理偏好的枢纽抵站旅客接续出行方式选择模型[J]. 北京交通大学学报, 2024, 48(4): 32-42.
- [13] 徐明非, 李昌铃, 王元庆, 等. 公铁复合城际走廊多模式客流分担特征[J]. 交通运输工程学报, 2020, 20(5): 176-186.
- [14] 樊海博. 基于多项 Logit 模型的客运通道交通分担率研究[J]. 铁道建筑技术, 2023(12): 174-176, 193.
- [15] 张睿, 马瑜, 赵冰茹, 等. 京沪线高铁与民航旅客出行选择行为研究[J]. 交通运输系统工程与信息, 2016, 16(1): 223-228.

## Prediction Method of Air-rail Transportation Passenger Flow Based on Two-stage Model: Taking Shanghai-Chengdu Corridor as an Example

GUO Jiuxia<sup>1</sup>, TIAN Jinyu<sup>1</sup>, ZHONG Qingwei<sup>1</sup>, CHEN Qu<sup>2</sup>

(1. School of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, Sichuan, China;

2. Flight Planning Department, Civil Aviation Administration Operation Monitoring Center, Beijing 100710, China.)

**Abstract:** With the rapid construction and development of comprehensive transportation hubs, the air-rail intermodal transportation model has brought more convenience to passengers' travel. Accurately understanding the air-rail intermodal passenger flow is crucial for improving the overall transportation service quality and ensuring transportation safety. A two-stage model combining particle swarm optimization-random forest model (PSO-RF) and Logit model is used to predict the passenger flow of travel paths in the context of air-rail intermodal transportation between urban agglomerations. Take one-way "Shifting from Railways to Aviation" as an example. In the first stage, the average daily civil aviation passenger flow of travel routes was predicted based on historical data by PSO-RF model. In the second stage, a behavioral survey was conducted through the airport outbound passengers to analyze the characteristics of passengers' transfer mode and choice behavior. Then, a Binary Logit model of passenger travel choices was constructed based on the disaggregate theory, and the high-speed rail transfer sharing rate was calculated. Finally, the results of the two-stage model were combined to calculate the travel route passenger flow forecasts. The effectiveness and feasibility of the proposed method by constructing a case study from Shanghai to Chengdu neighboring urban agglomeration were validated. The results indicate that the accuracy of the two-stage model reaches 80.40%.

**Keywords:** integrated transportation; air-rail intermodal transport; passenger flow forecast; particle swarm optimization-random forest (PSO-RF) model; Logit model