



DOI:10.12404/j.issn.1671-1815.2401651

引用格式:郭姝娟,耿晓洁,李纲,等.基于机器学习的中欧集装箱货运方式选择预测[J].科学技术与工程,2025,25(1):357-364.

Guo Shujuan, Geng Xiaojie, Li Gang, et al. Prediction on Sino-Europe container transportation mode choice based on machine learning[J]. Science Technology and Engineering, 2025, 25(1): 357-364.

# 基于机器学习的中欧集装箱货运方式选择预测

郭姝娟<sup>1</sup>,耿晓洁<sup>1</sup>,李纲<sup>2\*</sup>,王翔<sup>1</sup>,魏梓峰<sup>1</sup>,李一义<sup>1</sup>

(1.大连海事大学交通运输工程学院,大连 116026;2.大连交通大学交通工程学院,大连 116028)

**摘要** 为高效准确预测重大突发事件下货运代理对于中欧间集装箱运输方式选择偏好,并揭示影响货运代理选择的相关因素,采用陈述性偏好方法对货运代理进行调查,同时考虑了运输属性和货物属性的影响,构建决策树、逻辑回归和随机森林预测模型,对货运代理的选择行为进行预测。通过准确率、精确率、召回率和  $F_1$  这4个评价指标,将3个机器学习模型与离散选择模型的预测结果进行了综合对比;并利用随机森林算法对疫情不同阶段下影响货运代理运输方式选择的属性重要性进行排序。研究表明:3个机器学习模型的预测精度均比离散选择模型高,其中随机森林模型相较于决策树模型和逻辑回归模型在中欧集装箱运输方式选择问题具有更高预测准确度,更加适用于该问题;影响因素方面:在平稳期,货物属性是最重要的影响因素,当重大突发事件发生时货运代理更加看重阈值延迟时间。此外,货物目的地和货物价值对中欧集装箱运输方式选择有着重要影响。该研究可为全球重大突发事件影响下更准确地分析货运代理的运输方式选择行为的决策机制,以及帮助航运公司和中欧班列经营者更好地理解货运代理偏好和决策因素,为应对类似的突发事件提供了有力依据。

**关键词** 交通工程;货物运输方式选择;机器学习;中欧班列与班轮运输

**中图分类号** U116.2; **文献标志码** A

## Prediction on Sino-Europe Container Transportation Mode Choice Based on Machine Learning

GUO Shu-juan<sup>1</sup>, GENG Xiao-jie<sup>1</sup>, LI Gang<sup>2\*</sup>, WANG Xiang<sup>1</sup>, WEI Zi-feng<sup>1</sup>, LI Yi-yi<sup>1</sup>

(1. College of Transportation Engineering, Dalian Maritime University, Dalian 116026, China;

2. School of Transportation Engineering, Dalian Jiaotong University, Dalian 116028, China)

**[Abstract]** In order to efficiently and accurately predict freight forwarders' transportation modes preferences between China and Europe during major emergencies, as well as to uncover the relevant factors influencing freight forwarders' choices, the stated preference method was employed to survey freight forwarders. Additionally, considering the influences of transportation and cargo attributes, decision trees, logistic regressions, and random forest prediction models were constructed to forecast the selection behavior of freight forwarders. The prediction results of the machine learning model and the discrete choice model were comprehensively compared through four evaluation metrics: accuracy, precision, recall, and  $F_1$  score. Furthermore, the random forest algorithm was utilized to rank the importance of attributes influencing freight forwarders' transportation mode choices during different stages of the pandemic. The study results demonstrate that the prediction accuracy of all three machine learning models is higher than that of the discrete choice model. Among them, the random forest model exhibits superior prediction accuracy compared to the decision tree and logistic regression models in addressing the choice of Sino-Europe container transport modes, making it more suitable for this problem. Regarding influencing factors, during stable periods, cargo attributes are identified as the most important factors. When major emergencies occur, freight forwarders place greater emphasis on the threshold delay time. Furthermore, the destination and value of the cargo are found to have significant impacts on the choice of Sino-Europe container transport modes. The study proposes an accurate analysis of the decision-making mechanisms guiding freight forwarders' mode choice behavior during major global emergencies. Furthermore, it is utilized by shipping companies and operators of the China Railway Express to gain a deeper understanding of the preferences and decision-making factors influencing freight forwarders. The insights derived from this study are considered a solid basis for effectively responding to similar emergency situations.

**[Keywords]** traffic engineering; freight transportation mode choice; machine learning; CR-Express and liner shipping

收稿日期:2024-03-09 修订日期:2024-10-11

基金项目:国家自然科学基金(72172023);辽宁省教育厅基本科研项目(LJKMR20220378);大连交通大学人文社科研究-支持人文社科融合发展专项研究项目(面上项目);辽宁省属本科高校基本科研业务费专项资金(LJ112410150021)

第一作者:郭姝娟(1983—),女,汉族,吉林辽东人,博士,副教授。研究方向:物流系统优化。E-mail:guoshujuan@dlmu.edu.cn。

\*通信作者:李纲(1982—),男,汉族,辽宁丹东人,博士,副教授。研究方向:交通规划。E-mail:lgpe2015@djtu.edu.cn。

中国对欧洲的出口货物主要通过两种运输方式实现,即班轮运输和中欧班列。班轮运输通常选择从中国东部沿海出发,途径新加坡港和苏伊士运河,最终抵达德国汉堡,这也是远东至西北欧的主要远洋货运路线<sup>[1]</sup>。班轮运输作为传统的集装箱运输方式,具有通用性好、运输成本低等优势。中欧班列通常从国内中部集散中心出发,通过铁路口岸出境,途径“一带一路”沿线各国,最终抵达欧洲。中欧班列作为按固定车次、线路、班期和全程运行时刻开行的集装箱国际铁路联运班列,具有连续性强、计划性强、安全可靠等优势<sup>[2]</sup>。随着地缘政治冲突等影响,中欧间的集装箱运输处于多变不确定的环境中。重大突发事件使得两种运输方式运输属性水平不断波动,进而影响货运代理的选择,班轮运输和中欧班列的市场分担率持续变化。因此,在面对重大突发事件,如何预测货运代理在中欧集装箱运输方式选择,以及分析影响货运代理选择运输方式的关键因素,是急需解决的问题。

已有学者对中欧间运输方式市场分担率及运输方式间的竞争关系进行了广泛研究。赵怡然等<sup>[3]</sup>基于 Logit 模型,对各类货物的时间价值进行计算,并针对货物的时间价值和中欧班列的发班频率属性,进行了中欧间贸易运输方式分担率的敏感性分析。Yang 等<sup>[4]</sup>探讨了贸易数据分类层次方面对商品模式选择的潜在影响,根据集计数据计算了 4 条贸易路线上各货物的货值与货重系数,通过细化贸易货物的种类,研究货物价值和重量对贸易运输分担率的影响,发现 4 条贸易路线的货运方式选择对商品重量比价值更敏感。孙昕等<sup>[5]</sup>构建竞争分析模型,从运输距离、运输时间和运价方面比较了中欧海铁联运和铁路运输的经济性。冯芬玲等<sup>[6]</sup>采用 Logit 模型,基于货物价值特性和数量化理论 III 对全货品进行市场分类,分析中欧班列、海运和空运三种运输方式在不同货物运输中的竞争力。但是以上研究均没有以货运代理为选择主体,分析货运代理的选择行为出发对中欧间运输方式的选择进行分析。

部分学者采用 Logit 模型进行个人选择行为分析,诸立超等<sup>[7]</sup>为探究不同时空的托运人选择行为差异,构建了多类线性或非线性效用函数的多项 Logit 模型 (multinomial logit, MNL) 及混合 Logit 模型 (mixed logit, ML)。刘浩等<sup>[8]</sup>采用多项 Logit 模型及混合 Logit 模型分析运输方式属性、空间特征、货物特征和选择惯性等对货运代理选择行为的影响。然而,Logit 模型有自己的模型假设,并且

需要因变量和解释变量之间的预定义潜在关系,且为了简化参数估计使得个体预测准确率较低。近年来,机器学习基于数据挖掘能够得到良好预测准确度的特性,已在诸多方面得到普遍应用,如船舶航速预测、交通流量预测、交通冲突预测等<sup>[9-11]</sup>。

随着机器学习在许多领域中的广泛应用,人们也越来越关注将其应用于个人选择行为建模<sup>[12-14]</sup>。与预先确定模型结构(通常是线性公式)的 Logit 模型不同,机器学习方法不是做出严格的假设,而是允许计算机探测数据结构,具有更灵活的建模架构,这可以减少模型与经验数据的不兼容性,这往往会导致更好的预测能力<sup>[15]</sup>。近年的实证研究已经证实,在预测能力方面机器学习普遍优于 Logit 模型。例如,Xie 等<sup>[16]</sup>应用分类与回归树 (classification and regression tree, CART) 和神经网络 (neural network, NN) 对旧金山湾区居民通勤出行模式选择进行建模,结果表明机器学习方法在预测方面变现出比 MNL 更好的性能。Zhang 等<sup>[17]</sup>根据同一地区的数据报告得到结论,支持向量机可以比神经网络和 MNL 更准确地预测通勤出行方式选择。Lhéritier 等<sup>[18]</sup>发现随机森林 (random forest, RF) 模型在精度和计算时间方面优于标准和潜在类 MNL 模型,建模工作量更少。Cheng 等<sup>[19]</sup>对比随机森林模型和多项 Logit 模型对居民出行方式选择的建模结果,评估个人属性和建成环境等各影响变量的不同重要性,并发现随机森林模型的预测精度相对最高。魏梓峰<sup>[20]</sup>为准确地预测在疫情背景下托运人对中欧集装箱运输方式的选择,引入机器学习模型并与传统 Logit 模型进行对比,结果表示机器学习的预测准确率要优于 Logit 模型。刘拥华等<sup>[21]</sup>为得到最准确预测货车司机出行路径选择的方法,将随机森林、迭代算法、梯度提升迭代决策树和传统 Logit 模型进行对比,结果表明,随机森林的分类准确率最优。目前,货物运输领域针对个体偏好的相关研究,其研究数据主要依托问卷方式进行行为调查和意向调查获取,研究方法则以建立离散选择模型分析托运人或货运代理对运输方式的偏好选择为主,鲜有应用机器学习模型在货运代理个体层面针对运输方式选择行为的探讨。

为此,现建立决策树、随机森林和逻辑回归 3 种机器学习模型,在重大突发事件的影响下,研究货运代理选择行为进行预测与分析。通过准确率、精确率、召回率、 $F_1$  得分 4 个指标,对机器学习模型预测性能进行评估,并与混合 Logit 离散选择模型预测结果进行对比。通过检查随机森林模型的属性重

要性输出,揭示驱动预测决策中起关键作用的影响因素。以期为在全球突发事件影响下更准确地分析货运代理的运输方式选择行为提供重要的参考依据,为航运公司和中欧班列经营者在应对市场动荡和提升运输行业竞争力方面提供理论和实践基础。

## 1 实验数据

### 1.1 数据来源

以全球突发事件新冠肺炎疫情为例,对中国货运代理在疫情不同发展阶段下集装箱运输方式选择进行陈述性偏好方法(stated preference, SP)调查。调查从2021年6月持续至2021年9月,以线上调查方式对中国货运代理进行了问卷发放,具体流程是通过电子邮箱发放问卷至在工商局注册的国际货运代理企业并通过电话联系确认回答情况。同时搜集了新冠肺炎疫情期间两种运输方式的运输服务水平数据和德国、波兰每日新增被感染人数。通过德国、波兰每日新增被感染人数可以发现欧盟疫情发展存在平稳期与严峻期的两个阶段,交通灯系统是欧盟推行的定义新冠肺炎疫情风险分级的“红绿灯”系统。其中绿色代表染病例增长相对平稳且欧盟国家取消过于严格的货物入境管控措施,即平稳期,对应的时间为2020年5月—10月。红色代表被感染病例激增且欧盟国家实行过于严格的货物入境管控措施,即严峻期,对应的时间为2020年12月—2021年3月。

研究的调查重点是中国对欧洲进行集装箱运输的主要区域,包括珠三角、长三角和东北地区等。调查问卷主要包含3个部分,公司基本业务信息、公司运输的最具代表性货物的运输特征信息,以及以中国上海到德国汉堡为起讫点的集装箱运输方式选择的SP情景设计。总计有171名货运代理参与并完成了调查问卷,每位代理提供其运输频率最高的货物种类作为本次调查中的代表性货物。要求货运代理根据其代表性货物回答第三部分SP调查,涵盖了货物属性(如货物价值、发货频率、单次发货量、中欧班列服务等)以及运输属性(如运输时间、运输费用、发班频率、延迟时间、延迟概率等)。由于部分受访者未完成意向行为调查并需经过数据清洗,最终在平稳期和严峻期分别收集了648个有效的调查样本数据。表1详细展示了所调查货运代理公司的属性。所有规模的货运代理公司均有样本纳入研究,尽管10人以下规模的公司占比相对较少,但在其他公司规模区间的比例分布相对均匀,确保了本次调查的代表性。

表1 货运代理公司属性

Table 1 Attribute of freight forwarding companies

货运代理公司属性	属性描述	水平设置	样本数量	占比/%
公司 规模	货运代理 公司员工 总人数/人	10 以下	7	4.09
		11 ~ 50	30	17.54
		51 ~ 100	19	11.11
		101 ~ 150	45	26.32
		151 ~ 200	32	18.71
		201 以上	38	22.22
总计		—	171	100.00

### 1.2 特征向量构建

集装箱运输方式选择是一个复杂的过程,受多方面因素影响,综合考虑了运输属性和货物属性,属性特征和描述性统计如表2所示。在平稳期和严峻期,中欧班列与班轮运输的运输时间、运输费用以及发班频率等运输服务水平具有显著差异,导致货运代理运输方式选择也会随之变化。收集了货运代理出口至欧洲最具代表性货物在平稳期与严峻期的货物属性值,包括两个时期的货物目的地、货物价值、发货频率、单次发货量等。不同的货物特征也会对运输方式选择产生重大影响。例如,在中欧集装箱运输方式选择过程中,货物价值种类属于低附加值类型货物,受运输成本限制,选择班轮运输可能性会更大。

## 2 模型构建

目前,机器学习已成功应用于许多领域,并取得了惊人的成果。同时,在过去的几年里,大数据彻底改变了运输行业。这两个热门话题启发了人们重新考虑传统的货运模式选择行为问题。故将各种因素影响下的货运模式选择视为二分类问题。选择常用的决策树、逻辑回归和随机森林3种机器学习算法进行模型构建。

### 2.1 决策树模型

决策树(decision tree)是一种常见的机器学习算法,用于分类和回归问题,且所用算法学习效率和精度较高,被广泛用于交通领域<sup>[22]</sup>。建立基于Gini指数和信息熵的决策树模型求解运输方式选择问题,并分析影响货运代理选择机理的关键影响因素。模型中所考虑的影响因素包括被选择方案及备选方案选择肢的属性,如1.2节所示。

为求得一棵使得在个人水平上预测准确率较高的决策树,采用遍历的方法以寻求最优的超参数。对于平稳期所涉及的模型而言,因特征空间上的特征数较少,故决策树的最大可用特征设置为全特征(6个被选择方案的方案特征、6个备选方案的

表2 属性特征和描述性统计  
Table 2 Attribute characteristics and descriptive statistics

属性 维度	属性 名称	属性 描述	平稳期		严峻期	
			均值	标准差	均值	标准差
运输属性	中欧班列运输时间	港到港/站到站的总运输时间/d	20.03	1.00	20.03	1.00
	班轮运输运输时间		36.26	3.49	36.26	3.49
	中欧班列运输费用	港到港/站到站的总运输费用/(美元·TEU <sup>-1</sup> )	2 015.20	242.85	4 926.85	1 052.33
	班轮运输运输费用		943.46	161.90	2 940.74	1 295.17
	中欧班列发班频率	每周发班次数/(班次·周 <sup>-1</sup> )	9.06	1.00	5.04	1.00
	班轮运输发班频率		4.99	1.00	4.00	1.00
	中欧班列延迟概率	货物遭延迟的比例/%	0.15	0.05	0.60	0.20
	班轮运输延迟概率		0.30	0.10	0.70	0.10
	中欧班列延迟时间	与约定交付时间的相差天数/d	1.50	0.05	7.42	2.50
	班轮运输延迟时间		2.05	1.00	9.07	3.00
中欧班列阈值延迟时间	将延迟时间值与货运代理最大可接受延迟时间进行对比,若延迟时间超过最大可接受值,则变量为两者差值,否则为0	0	0	1.06	1.92	
班轮运输阈值延迟时间		0.08	0.27	1.87	2.70	
货物属性	货物价值	货运代理货物是否为高附加值货物/%	0.44	0.50	0.44	0.50
	欧洲中部	货物目的地是否为中欧国家/%	0.57	0.49	0.58	0.49
	发货频率	货运代理关于代表性货物的每周发货次数	0.42	0.49	0.78	0.42
	中等发货量	每批集装箱货运量是否在20TEU~40TEU	0.37	0.49	0.26	0.44
	高发货量	每批集装箱货运量是否在40TEU~60TEU	0.31	0.50	0.44	0.50
	中欧班列服务	疫情前是否使用过中欧班列服务	0.24	0.43	0.06	0.24

注:TEU(twenty-feet equivalent unit)表示以长度为20英尺(20 ft)的集装箱为国际计量单位,也称国际标准箱单位,1 ft = 30.48 cm。

方案特征和6个人属性);模型的最小分支节点样本量和最小叶节点样本量均为20;为平衡样本中选择两种运输方式的比例,设置了样本平衡参数。相较于以上参数,树的最大深度和模型评判指标是影响决策树模型的最重要的超参数。选择信息熵和Gini指数作为备选模型评判标准,树的最大深度在1~30中遍历。在严峻期,最大特征数、最小分支节点样本量、最小叶节点样本量和平衡参数与平稳期一致,模型评判标准和树的最大深度也通过遍历确定,以确定最优参数。

调用Python中的sklearn库来求解最优模型,疫情不同阶段决策树最优参数如图1所示。在平稳期,信息熵和基尼指数对于模型无影响,两条准确率曲线完全重合,树的最大深度为9,最优预测准确率为69.73%;在严峻期,信息熵和基尼指数的最大值相同,均在深度为4的时候产生,最佳预测准确率为69.74%。

### 2.2 逻辑回归模型

逻辑回归(logistic regression)通常被用作二元分类问题的模型,被定义为一种用于学习输入和输出之间关系的监督学习算法,其主要思想是将输入特征通过一个线性变换,然后对其进行非线性映射得到预测结果,被广泛应用于离散选择模型领域<sup>[23]</sup>。构建基于损失函数最小的逻辑回归模型,并应用所有特征构建特征空间。将收集到的样本70%作为训练集,剩下的30%作为测试集。在平稳

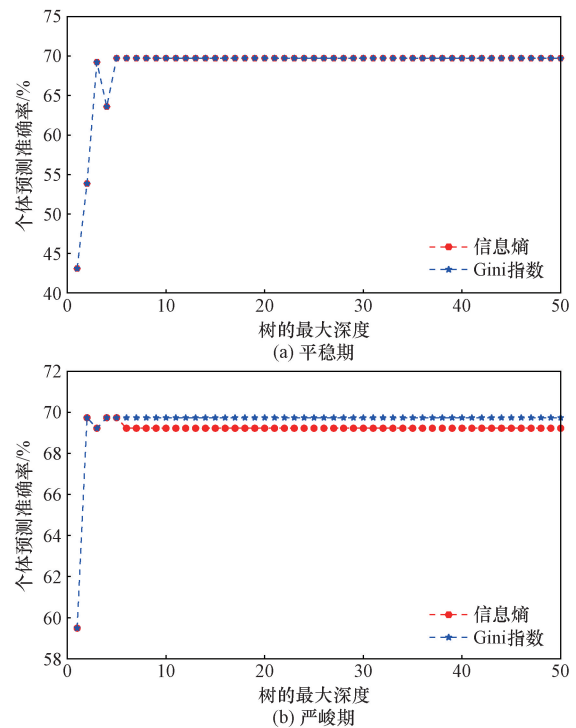


图1 疫情不同阶段决策树最优参数  
Fig. 1 Decision tree optimal parameters for different stages of the epidemic

期采用L<sub>2</sub>正则化防止过拟合问题,并采用随机梯度下降算法求解此问题,并在0~300遍历最大可迭代次数,发现在100代之后,增大可迭代次数,预测准确率基本保持不变,所以设置可迭代次数为100代。

在严峻期的超参数设置和平稳期相同。数据在训练集和测试集上的预测准确率如表3所示。

表3 不同阶段逻辑回归预测准确率  
Table 3 Logistic regression prediction accuracy at different stages

阶段	准确率/%	
	训练集	测试集
平稳期	94.70	93.85
严峻期	76.16	74.87

### 2.3 随机森林模型

随机森林是一种集成学习的机器学习算法,它可以用于分类和回归等任务,是基于 bagging 思想的算法,建立多个决策树构成一片森林以求得最优模型<sup>[24]</sup>。模型的评价指标为“基尼指数”,采用十折交叉验证来防止模型过拟合。不同阶段随机森林最优参数如图2所示。在平稳期,树的最大深度最优值是9,随机森林最重要的超参数是森林中树的数量,在0~300中遍历树的数量,发现树的数量等于6时,预测准确率达到最高为94.36%。在严峻期,树的最大深度与平稳期相同,树的数量在21时预测准确率达到最高为77.44%。

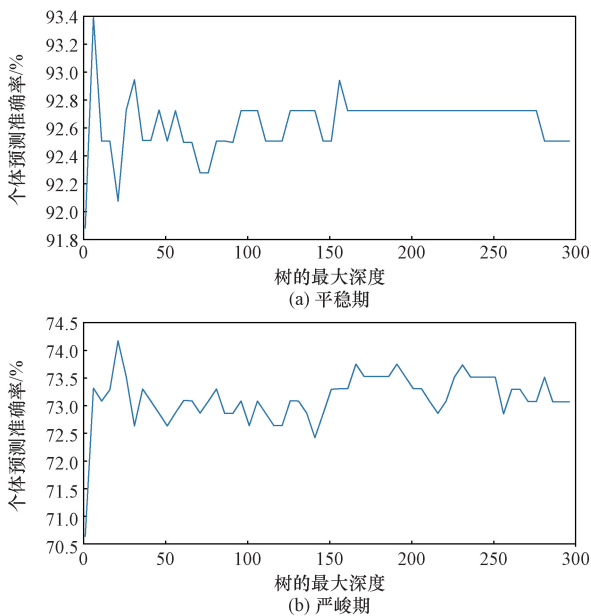


图2 不同阶段随机森林的最优参数

Fig. 2 Random forest optimization parameters for different stages

## 3 实验结果与分析

### 3.1 评价指标

为评估机器学习模型分类效果,选取准确率、精确率、召回率、 $F_1$ 得分4个指标对模型性能和精度进行综合评价。

$$A = (TP + TN) / (TP + TN + FP + FN) \times 100\% \quad (1)$$

$$P = TP / (TP + FP) \times 100\% \quad (2)$$

$$R = TP / (TP + FN) \times 100\% \quad (3)$$

$$F_1 = 2PR / (P + R) \times 100\% \quad (4)$$

式中:TP为模型将正类别样本正确预测为正类别的数量;TN为模型将负类别样本正确预测为负类别的数量;FP为模型将负类别样本错误预测为正类别的数量;FN为模型将正类别样本错误预测为负类别的数量; $A$ 为模型总体预测准确率,是衡量模型正确预测的样本数量与总样本数量的比率,它是针对整体数据集的评估,用于度量分类器正确分类样本的能力; $P$ 为精确率,是指在所有被模型预测为正类别的样本中,实际为正类别的比例,它衡量了模型的预测有多少是正确的; $R$ 为召回率,是指在所有实际正类别的样本中,模型成功预测为正类别的比例,它衡量了模型正确识别正类别的能力; $F_1$ 得分为精确率和召回率的调和平均值,它综合考虑了模型的准确性和完整性。由于精确度和召回率有时会出现冲突,即一个高一个低的情况,但理想的都比较高,因此可用 $F_1$ 得分进行综合评估模型性能, $F_1$ 得分越高,说明模型的性能越好。

### 3.2 模型预测结果对比

为了说明机器学习模型的预测准确率,与传统的Logit模型的预测性能进行了比较。李一义<sup>[25]</sup>针对不同阶段下货运代理中欧集装箱货运方式选择行为问题,提出了一种允许非补偿行为的软阈值混合Logit模型(mixed logit model with soft attribute cut-offs, ML-CO),并将其与基于补偿行为的二项Logit模型、基于非补偿行为的二项Logit模型进行对比,并证明了考虑货运代理对于货物延迟交付的忍耐程度异质性的ML-CO模型更能有效捕捉现实中中欧货运方式选择机制。因此,本节将三种机器学习模型与李一义提出的ML-CO模型的预测性能进行了比较。不同阶段的预测结果如表4和表5所示。

(1)针对不同模型而言,机器学习模型的预测性能普遍高于ML-CO模型,与前人研究结论一致,Logit模型的预测准确率偏低。在机器学习模型中,随机森林算法在分类的效果上表现最佳,决策树次之,而由于中欧贸易中,班轮运输始终是主要的运输线路,因此数据集中班轮运输的占比较大,而中欧班列的样本数量较少,导致数据不平衡,因此逻辑回归模型在预测过程中倾向于将所有的选择结果都预测为班轮运输。这导致在预测中欧班列时,逻辑回归模型的预测结果全部为班轮运输,无法进

表4 平稳期模型预测结果评价  
Table 4 Evaluation of model prediction results during stable period

模型	总体准确率/%	选择肢	精确率/%	召回率/%	$F_1$ /%
决策树	69.73	班轮运输	98.38	69.28	81.30
		中欧班列	13.36	80.56	22.92
逻辑回归	93.85	班轮运输	94.44	100	97.14
		中欧班列	0	0	0
随机森林	94.36	班轮运输	96.50	99.18	97.82
		中欧班列	73.68	38.89	50.91
ML-CO	52.16	班轮运输	66.37	52.74	58.78
		中欧班列	37.14	51.09	43.01

表5 严峻期模型预测结果评价  
Table 5 Evaluation of model prediction results during severe period

模型	总体准确率/%	选择肢	精确度/%	召回率/%	$F_1$ /%
决策树	69.74	班轮运输	84.81	70.64	77.01
		中欧班列	58.86	76.86	66.76
逻辑回归	74.87	班轮运输	64.67	100	78.56
		中欧班列	0	0	0
随机森林	77.44	班轮运输	91.71	89.74	90.71
		中欧班列	81.93	85.15	83.51
ML-CO	56.94	班轮运输	71.08	56.32	62.84
		中欧班列	42.09	58.08	48.87

行准确的分类。结果是中欧班列的精确率、召回率和  $F_1$  都为 0。因此逻辑回归模型的数据表现不佳。

(2) 不同阶段各个模型的预测准确率呈现出不同的结果。在严峻期, ML-CO 模型的预测准确率比在平稳期高, 但是机器学习模型的预测准确率比平稳期低。在实际情况下, 由于严峻期严格的管控政策, 导致运输属性波动较大, 所以严峻期预测的准确率要低于平稳期, 因此, 机器学习模型的预测准确率更符合实际。

(3) 对于不同的运输方式, 可以观察到所有模型对班轮运输的预测准确率普遍都高于中欧班列的预测准确率。造成这种结果的原因是班轮运输作为中欧贸易最主要的运输线路, 在数据集中的占比较大, 机器学习模型在学习过程中对班轮运输的特征和模式有更多的学习机会。这导致模型偏向于将更多的样本预测为班轮运输。而中欧班列在数据集中的样本数量较少, 模型对其特征和模式的学习能力相对不足。这使得模型很难准确预测中欧班列, 导致其预测结果较低。

### 3.3 影响因素分析

从模型的预测结果对比可知, 随机森林模型在

不同阶段都保持了相对较高的准确率, 且随机森林模型可基于基尼指数计算观测变量的特征重要度, 从而评估影响货运代理对于中欧集装箱货运方式选择的属性重要性差异。因此, 采用随机森林模型对不同阶段下两种运输方式的属性重要性进行排序, 其结果如表 6 所示。

在平稳期, 货物属性的权重值较高, 说明货运代理在线路划分时更加看重货物属性, 对运输属性的偏好较低。这是因为此时期的运输属性变化幅度不大, 货运代理更加看重重发货量、货物目的地是否是欧洲中部以及新冠肺炎疫情前是否使用过中欧班列服务。在严峻期, 由于各种运输限制和防疫措施的实施, 物流运输遭受了许多不可避免的延迟, 在此期间集装箱运输的平均延迟时间是平时的一倍, 货运代理对运输方式的阈值延迟时间最为关注。因此在重大突发事件到来时, 货主更加关注由于重大突发事件带来的延迟时间。

此外, 相对于平稳期, 班轮运输的运输费用、延迟时间以及延迟概率的权重值出现了上升趋势。对于中欧班列而言, 运输费用和延迟时间的权重值同样呈现增加的趋势。这表明在重大突发事件发生时, 运价波动将显著影响货运代理的选择, 同时这类事件也将导致运输时间的不确定性, 引发一定的延误。值得注意的是, 在这一时期内, 中欧班列的运输时间、发班频率以及延迟概率的权重值则呈下降趋势。这一变化主要归因于中欧班列本身具备运输时间短、发班频率稳定以及可靠性高的优势,

表6 随机森林模型属性重要性排序  
Table 6 Ranking of attribute importance of random forest model

属性维度	属性	平稳期	严峻期
运输属性	班轮运输运输费用	10	8
	中欧班列运输费用	16	5
	班轮运输运输时间	9	11
	中欧班列运输时间	8	16
	班轮运输发班频率	11	15
	中欧班列发班频率	6	14
	班轮运输延迟时间	13	10
	中欧班列延迟时间	14	13
	班轮运输延迟概率	15	13
	中欧班列延迟概率	7	17
	班轮运输阈值延迟时间	12	1
	中欧班列阈值延迟时间	17	3
货物属性	货物价值	4	2
	欧洲中部	2	4
	发货频率	5	6
	中等发货量	1	7
	高发货量	3	9
	中欧班列服务	2	12

使其在重大突发事件发生时,运输属性的波动相对较小。鉴于以上分析,针对航运公司和中欧班列的经营者,建议在突发事件期间更加关注和灵活应对运输费用和延迟时间的波动。应建立更为灵活的运价调整机制,以适应市场的不确定性。同时,加强危机管理和应急响应能力,以最大程度减少运输时间的延误,并提高运输的可靠性。中欧班列的经营者应继续维护中欧班列的短时效、稳定性等特点,有助于在动荡时期中保持相对竞争优势。

不论是否有重大突发事件,货物目的地是否是欧洲中部和货物价值都对中欧集装箱运输方式选择有重要影响,侧面反映出欧洲中部国家使用中欧班列进行集装箱货物运输具有一定的地域优势。这一变量的重要性可能源于欧洲中部国家(如德国、波兰等)具备优越的地理位置、完善的铁路基础设施与运营网络,使其更具有抗击重大突发事件风险的能力,成为货运代理在特殊时期的偏好选择。货物价值占有较高的权重反映出货运代理更趋向于利用中欧班列运送高附加值货物。这是因为中欧班列相对较为稳定的运输服务更能满足高附加值货物对时效性的严格要求,而在面对重大突发事件时,确保货物安全、及时交付成为尤为重要的考虑因素。

## 4 结论

以班轮运输和中欧班列组成的中欧间集装箱货物运输为研究对象,基于决策树、逻辑回归和随机森林模型对货运代理对于中欧间集装箱运输方式选择行为进行模型构建与标定,得到主要结论如下。

(1)从准确率、精确率、召回率和  $F_1$  得分 4 个评价指标来看,机器学习模型预测准确度普遍高于 ML-CO 模型,证明了机器学习模型对此类问题的适应性且预测精度更高。且随机森林模型相较于决策树模型和逻辑回归模型在货运代理对于中欧间集装箱运输方式选择行为问题上具有更高的预测准确度,能有效处理不平衡样本,并具备描述解释变量与因变量间复杂非线性关系优势。

(2)根据随机森林模型的属性重要性排序可知,当重大突发事件发生时阈值延迟时间成为了决策的主要影响因素。这表明货运代理在重大事件发生时,选择运输方式会密切关注超过阈值的延迟时间。因此,当重大事件发生时班轮公司与中欧班列经营者应根据客户需求合理调整,尽量避免超出货运代理的延迟时间期望极限。建议他们通过提升运输服务质量、灵活调整运力和优化运输网络,以应对突发事件对物流行业带来的挑战。

(3)对中欧集装箱运输方式选择问题进行了初步探索,为运输服务优化和货运政策制定提供了指导。然而,在探讨中欧集装箱运输方式选择问题时,主要集中考虑了运输属性和货物属性的变化对运输方式选择的影响,但是在实际运输中,诸如地缘政治紧张、自然灾害、航线安全性等风险因素也在决策中扮演关键角色。未来研究可更深入地剖析这些风险因素如何影响运输方式的选择,以提供更为切实有效的应对策略。

## 参 考 文 献

- [1] 赵怡然. 中欧班列对中欧间海运的影响分析[D]. 大连: 大连海事大学, 2017.  
Zhao Yiran. Analysis of the impact of China-Europe trains on the shipping between China and Europe[D]. Dalian: Dalian Maritime University, 2017.
- [2] 张倩璐. 中欧班列运输模式研究[D]. 成都: 西南交通大学, 2020.  
Zhang Qianlu. Research on China-Europe railway express transportation mode[D]. Chengdu: Southwest Jiaotong University, 2020.
- [3] 赵怡然, 杨忠振, Theo Notteboom. 中欧铁路班列在中欧贸易运输中的作用[J]. 大连海事大学学报, 2018, 44(3): 33-40.  
Zhao Yiran, Yang Zhongzhen, Theo Notteboom. Roles of China Railway Express in Sino-Euro trade transportation system[J]. Journal of Dalian Maritime University, 2018, 44(3): 33-40.
- [4] Yang D, Ong G, Chin A T H. An exploratory study on the effect of trade data aggregation on international freight mode choice [J]. Maritime Policy & Management, 2014, 41: 212-223.
- [5] 孙昕, 殷文伟, 艾万政. 基于竞争分析模型的中欧海铁联运与铁路运输的经济性比较[J]. 水运管理, 2016(12): 4-6.  
Sun Xin, Yin Wenwei, Ai Wanzheng. The economic comparison of the Sino European sea rail transport and railway transportation based on the competitive analysis model[J]. Water Transport, 2016(12): 4-6.
- [6] 冯芬玲, 梁俊凯, 刘承光, 等. “一带一路”背景下中欧班列竞争力分析——基于货物细分和 Logit 模型[J]. 技术经济, 2020, 39(3): 119-128.  
Feng Fenling, Liang Junkai, Liu Chengguang, et al. Analysis of competitiveness of China and Europe in the context of “One Belt and One Road” based on goods segmentation and Logit model[J]. Journal of Technology Economics, 2020, 39(3): 119-128.
- [7] 诸立超, 刘昭然. 托运人在不同时空下的选择行为差异揭示[J]. 交通运输系统工程与信息, 2020, 20(1): 27-32.  
Chu Lichao, Liu Zhaoran. Revealing difference of shippers' choice behavior under different spatio-temporal conditions[J]. Journal of Transportation Systems Engineering and Information Technology, 2020, 20(1): 27-32.
- [8] 刘浩, 张戎, 诸立超. 考虑空间特征的货运方式选择行为模型[J]. 交通运输系统工程与信息, 2021, 21(1): 30-35.  
Liu Hao, Zhang Rong, Chu Lichao. Freight mode choice behavior model incorporating spatial characteristics[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(1): 30-35.

- [9] 朱晓晨, 尹奇志, 赵福芹, 等. 基于 LightGBM 的船舶航速预测模型[J]. 大连海事大学学报, 2023, 49(1): 56-65.  
Zhu Xiaochen, Yin Qizhi, Zhao Fuqin, et al. Ship speed prediction model based on LightGBM[J]. Journal of Dalian Maritime University, 2023, 49(1): 56-65.
- [10] 王金水, 欧雪雯, 陈俊岩, 等. 基于组合模型的城市轨道站点短时客流分类预测[J]. 铁道科学与工程学报, 2023, 20(6): 2004-2012.  
Wang Jinshui, Ou Xuwen, Chen Junyan. Short-term passenger flow classification prediction of urban railway stations based on combined model[J]. Journal of Railway Science and Engineering, 2023, 20(6): 2004-2012.
- [11] 任丽丽, 吴江玲, 郭旭亮, 等. 无信号环形交叉口机非冲突机器学习预测方法[J]. 科学技术与工程, 2023, 23(31): 13592-13600.  
Ren Lili, Wu Jiangling, Guo Xuliang, et al. Traffic conflict prediction of motorized and non-motorized vehicles at unsignalized roundabout using machine learning[J]. Science Technology and Engineering, 2023, 23(31): 13592-13600.
- [12] Xia Y T, Chen H F, Zimmermann R. A random effect Bayesian neural network (RE-BNN) for travel mode choice analysis across multiple regions[J]. Travel Behaviour and Society, 2023, 30: 118-134.
- [13] Tamim K M, Jamal A, Samim K M, et al. Predicting the travel mode choice with interpretable machine learning techniques: a comparative study[J]. Travel Behaviour and Society, 2022, 29: 279-296.
- [14] 周艳聪, 郝园媛. 基于机器学习的运营商客户行为分析[J]. 科学技术与工程, 2022, 22(2): 585-592.  
Zhou Yancong, Hao Yuanyuan. Research of operator customer behavior analysis based on machine learning algorithm[J]. Science Technology and Engineering, 2022, 22(2): 585-592.
- [15] Zhao X L, Yan X, Yu A L, et al. Prediction and behavioral analysis of travel mode choice: a comparison of machine learning and Logit models [J]. Travel Behaviour and Society, 2020, 20: 22-35.
- [16] Xie C, Lu J Y, Parkany E. Work travel mode choice modeling with data mining-decision trees and neural networks[J]. Transportation Research Record Journal of the Transportation Research Board, 2003(10): 50-61.
- [17] Zhang Y L, Xe Y C. Travel mode choice modeling with support vector machines [J]. Transportation Research Record, 2008, 2076(1): 141-150.
- [18] Lheritier A, Bocamazo M, Delahaye T, et al. Airline itinerary choice modeling using machine learning [J]. Journal of Choice Modelling, 2019, 31: 198-209.
- [19] Cheng L, Chen X W, De Vos J, et al. Applying a random forest method approach to model travel mode choice behavior[J]. Travel Behaviour and Society, 2019, 14: 1-10.
- [20] 魏梓峰. 疫情下西部陆海新通道市场分担率与托运人选择行为预测[D]. 大连: 大连海事大学, 2023.  
Wei Zifeng. Prediction of market share rate and shipper choice behavior of the new western land-sea corridor under the epidemic [D]. Dalian: Dalian Maritime University, 2023.
- [21] 刘拥华, 段瑞坤, 段莉珍, 等. 分路段差异化收费条件下货车司机出行路径选择意愿模型[J]. 科学技术与工程, 2023, 23(35): 15239-15245.  
Liu Yonghua, Duan Ruikun, Duan Lizhen, et al. Truck driver's travel route choice willingness model under the condition of differentiated tolls by road section [J]. Science Technology and Engineering, 2023, 23(35): 15239-15245.
- [22] 温慧敏, 朱珊, 孙建平, 等. 轨道新线对公共交通乘客吸引力影响因素研究[J]. 交通运输系统工程与信息, 2023, 23(4): 282-289.  
Wen Huimin, Zhu Shan, Sun Jianping, et al. Assessment of impact factors on passenger attraction of new metro line [J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(4): 282-289.
- [23] 许佳佳, 陈慧敏. 基于机器学习的城市居民低碳出行行为研究[J]. 青岛大学学报(工程技术版), 2023, 38(3): 30-38.  
Xu Jiajia, Chen Huimin. Research on low-carbon travel behavior of urban residents based on machine learning [J]. Journal of Qingdao University (Engineering & Technology Edition), 2023, 38(3): 30-38.
- [24] 杨敏, 任怡凤, 盛强, 等. 基于随机森林算法的旅客空铁联运中转城市选择模型[J]. 东南大学学报(自然科学版), 2022, 52(1): 162-171.  
Yang Min, Ren Yifeng, Sheng Qiang, et al. Passenger air-rail intermodal transit city choice model based on random forest algorithm [J]. Journal of Southeast University (Natural Science Edition), 2022, 52(1): 162-171.
- [25] 李一义. 疫情背景下货运代理中欧集装箱运输方式选择行为研究[D]. 大连: 大连海事大学, 2022.  
Li Yiyi. The mode choice of Sino-Europe container transportation in the context of COVID-19 for freight forwarder [D]. Dalian: Dalian Maritime University, 2022.