



DOI:10.12404/j.issn.1671-1815.2308202

引用格式:邹杰,曹宏禄,李平安,等.基于随机森林的高速公路变路径偷逃费行为识别[J].科学技术与工程,2024,24(36):15694-15700.

Zou Jie, Cao Honglu, Li Pingan, et al. Identification of fee evasion behavior in expressway changing path based on random forest[J]. Science Technology and Engineering, 2024, 24(36): 15694-15700.

基于随机森林的高速公路变路径偷逃费行为识别

邹杰¹, 曹宏禄², 李平安¹, 黄诗音¹, 赵建东^{3*}

(1. 中公华通(北京)科技发展有限公司, 北京 100088; 2. 北京交通大学交通运输学院, 北京 100044;

3. 北京交通大学系统科学学院, 北京 100044)

摘要 为提高高速公路变路径偷逃费行为识别效率,针对改变路径偷逃费行为进行研究,建立一种基于随机森林的高速公路变路径偷逃费行为识别模型,能够有效识别该类偷逃费行为,协助高速公路相关管理部门追缴偷逃费用。首先,分析原始收费数据,筛选出与本次研究相关的字段,经过运算得到 12 个模型可输入的初始特征;然后,通过计算各个特征的方差膨胀因子(variance inflation factor, VIF)和容忍度(tolerance, TOL)值来剔除存在共线性的特征,并利用 Boruta 算法筛选高重要性特征(“行驶方向是否一致”“入出站是否一致”“通行时间”和“最小费额里程”);其次,使用 SMOTETomek 综合采样技术来平衡数据集;再其次,利用网格搜索法对随机森林进行超参数调优;最后,利用所建立模型进行训练和识别,并与基准模型的识别效果进行对比。结果表明:所建立模型能够更好地对高速公路变路径偷逃费行为进行识别,Macro-F1 分数达到了 0.966,优于极限梯度提升(extreme gradient boost, XGBoost)(0.943 1)、决策树(decision tree, DT)(0.956 3)和梯度提升决策树(gradient boosting decision trees, GBDT)(0.938 2),能够为运营管理部门稽查该类偷逃费车辆提供参考。

关键词 随机森林(RF);改变路径偷逃费;Boruta 算法;数据不平衡处理

中图分类号 U495;

文献标志码 A

Identification of Fee Evasion Behavior in Expressway Changing Path Based on Random Forest

ZOU Jie¹, CAO Hong-lu², LI Ping-an¹, HUANG Shi-yin¹, ZHAO Jian-dong^{3*}

(1. TransChina(Beijing) Technology Co., Ltd., Beijing 100088, China;

2. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China;

3. School of Systems Science, Beijing Jiaotong University, Beijing 100044, China)

[Abstract] In order to improve the efficiency of identifying toll evasion behavior by changing paths on highways, toll evasion behavior was studied by changing paths. A model for identifying toll evasion behavior by changing paths on highways based on random forests was established, which can effectively identify such behavior of toll evasion and assist relevant management departments of highways in recovering evaded fees. Firstly, the original toll data were analyzed to filter out the fields related to this study, and the 12 initial features that can be inputted into the model were obtained after arithmetic. Secondly, the features with covariance were eliminated by calculating the variance inflation factor (VIF) and tolerance (TOL) values of each feature, and the Boruta algorithm was used to filter out the high-importance features (“whether the driving direction is consistent” “whether the entry and exit stations are consistent” “travel time” and “minimum fare mileage”). Thirdly, the data set was balanced using the SMOTETomek integrated sampling technique. Then, the grid search method was used to tune the hyperparameters of the random forest. Finally, the model built was utilized for training and recognition, and the recognition effect was compared with that of the benchmark model. The results show that the model developed can better recognize the toll evasion behavior by changing paths on highways, and the Macro-F1 score reaches 0.966, which is better than the extreme gradient boost (XGBoost) (0.943 1), decision tree(DT) (0.956 3) and gradient boosting decision trees (GBDT) (0.938 2), and it can provide reference for operation management departments to inspect such toll evasion vehicles.

[Keywords] random forest(RF); toll evasion by changing path; Boruta algorithm; data imbalance processing

随着高速公路省界收费站的取消,全国高速公路形成了“收费一张网”,收费模式开始转为利用电

子收费(electronic toll collection, ETC)门架系统进行不停车分段式收费,极大效率地提高了高速公路路

收稿日期:2023-10-21; 修订日期:2024-10-01

基金项目:河北高速公路集团有限公司科技创新计划(03032111KY0229)

第一作者:邹杰(1976—),男,汉族,山东泰安人,博士,副研究员。研究方向:交通工程、交通信息化。E-mail:297632999@qq.com。

*通信作者:赵建东(1975—),男,汉族,山西忻州人,博士,教授。研究方向:智能交通系统。E-mail:zhaojd@bjtu.edu.cn。

投稿网址:www.stae.com.cn

网的通行效率,但由于新模式下单车单次通行费额的增加,随之出现了大量的偷逃费行为^[1]。如何利用大量的高速公路收费数据对偷逃费车辆进行稽核,以减少高速公路管理部门的损失,成为了当下研究的热点。

随着科技的进步,可获取的数据不断增多,机器学习(machine learning, ML)技术被广泛应用于各大领域的研究。目前,中外学者已经开始使用机器学习方法对高速公路偷逃费车辆进行识别稽核。张凡等^[2]针对绿通车数据集不平衡的问题,通过分析非绿通车与绿通车之间的特征差异,构建一种基于SMOTE-Logistic的绿通车分类模型。Zhao等^[3]从减轻车辆载重、缩短车辆行驶距离和其他方法这3个方面对偷逃费行为进行分析,提出了一种基于高斯混合模型(Gaussian mixture model, GMM)的聚类方法来识别偷逃费车辆。李松江等^[4]为提高高速公路偷逃费车辆的预测精度,利用ReliefF算法选取最优特征,建立一种基于改进遗传算法-改进BP神经网络(improved genetic algorithm-improved back propagation, IGA-IBP)的偷逃费车辆预测模型。刘昱岗等^[5]首先对假冒绿通车行为特性进行分析,采用改进的合成少数类过采样技术(Borderline synthetic minority over-sampling technique, Borderline-SMOTE)解决绿通车数据集不平衡问题,通过属性关联性检验和共线性检验来剔除无关属性,最后构建分类与回归树(classification and regression tree, CART)对正常/假冒绿通车进行分类。向红艳等^[6]利用随机森林(random rorest, RF)算法对与逃费车辆预测相关的特征进行重要度排序,基于此建立逻辑回归(logistic regression, LR)模型对逃费车辆进行预测。

综上所述,目前关于高速公路偷逃费车辆识别的研究在数据和算法层面的应用较为单薄,且对某一具体偷逃费行为的识别鲜有研究。鉴于此,针对改变路径偷逃费行为(U/J型行驶)进行研究,分析真实高速公路收费数据,进行一系列数据预处理工作,采用网格搜索法选取随机森林的最优超参数组合,最终建立基于随机森林的高速公路变路径偷逃费行为识别模型。并比较本文模型与其他典型机器学习算法的识别性能。研究成果可辅助管理者稽查偷逃费车辆并追缴实际通行费用。

1 数据来源与研究过程

1.1 数据来源

以河北省青银高速作为研究对象,数据表包括入口通行流水表、出口通行流水表、ETC门架读卡

数据表和ETC门架牌识数据表,数据的时间跨度为2022年8月1日—2022年8月31日,筛选出本文所需的关键字段如表1所示。由于需要从多个数据表中获取同一辆车的信息进行分析,首先根据各数据表之间的关联匹配规则,合并得到单车单次的通行记录;然后将数据集中存在的缺失数据(如关键字段数值缺失)、冗余数据(如多条重复的车辆通行信息)和噪声数据(如关键字段数值异常)进行清洗,剩余共计360 999条数据信息。

由于原始收费数据部分字段无法直接用于模型的建立,根据改变路径偷逃费行为的行驶特性制定相关的计算规则,得到研究所需特征如表2所示。

表1 原始收费数据字段

Table 1 Original charge data field

字段名称	字段含义
PASS_ID	通行标识
EN_TIME	进站时间
EN_TOLL_STATION_NAME	入口收费站名称
EX_TIME	出站时间
EX_TOLL_STATION_NAME	出口收费站名称
FEE_MILEAGE	计费里程
VEHICLE_PLATE	车牌号
SHORT_FEE_MILEAGE	最小费额里程

表2 初始提取特征

Table 2 Initial extraction feature

序号	特征名称	说明
1	入站时段	取值为[0,23]
2	出站时段	取值为[0,23]
3	入站日期	取值为[0,30]
4	出站日期	取值为[0,30]
5	计费里程	单位:m
6	最小费额里程	单位:m
7	通行时间	单位:h
8	行驶均速	单位:km/h
9	是否为省内车牌	0为否,1为是
10	入出站是否一致	0为否,1为是
11	是否超时行驶	0为否,1为是
12	行驶方向是否一致	0为否,1为是

1.2 改变路径逃费行为分析

主要针对改变路径相关逃费方式进行研究,改变路径逃费是指利用高速公路设计上的漏洞和管理的盲区,通过改变行驶方向和路线,缩短实际缴费里程,通过两辆车在中途换客换货等方式,从而达到少缴或逃缴通行费的目的,该逃费方式主要包括U/J型行驶,其逃费原理如图1所示。其中,U型行驶为车辆行驶中途在某处掉头,如中央分隔带缺口、互通式匝道掉头,然后从入口收费站出站,实现“跑长买短”。J型行驶指车辆驶入高速公路后,中

途行驶到相邻高速并从某处掉头返回到靠近入口的收费站下高速,从而达到“跑长买短”的目的。

根据 U/J 型行驶特性从原始收费数据中筛查出偷逃费车辆,以供后续研究。表 3 为改变路径偷逃费行为的数据特征及数量。

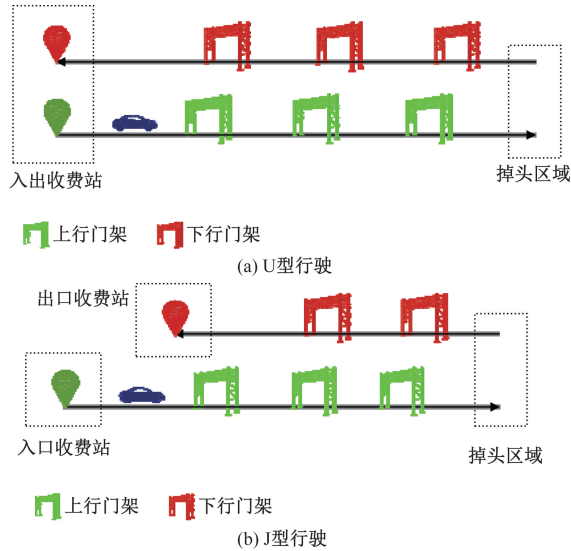


图 1 改变路径偷逃费行为示意图

Fig. 1 Schematic diagram of changing the path to evade fees

表 3 改变路径偷逃费行为统计

Table 3 Statistics on behavior of changing paths to evade fees

偷逃费类型	数据特征	车辆数/辆
正常行驶	无	74 849
U 型行驶	车辆进出口行驶方向相反 车辆进出口站名相同	129
J 型行驶	车辆进出口行驶方向相反 车辆进出口站相邻	4 567

1.3 研究过程

为得到较优的高速公路变路径偷逃费行为识别模型,首先对收费数据进行初始分析,然后对数据进行一系列数据预处理工作。在确定模型最优输入特征后,将数据集按照 7:3 比例划分为训练集和测试集,对随机森林算法的超参数进行调优,基于此来识别高速公路变路径偷逃费行为。同时,以决策树 (decision tree, DT)^[7]、极限梯度提升 (extreme gradient boost, XGBoost)^[8] 和梯度提升决策树 (gradient boosting decision trees, GBDT)^[9] 作为基准模型来验证本文模型的识别效果。研究过程如图 2 所示。

2 数据预处理

2.1 数据离散化

在机器学习任务中,将连续型特征离散化可以增强算法对特征的理解和区分能力,有效降低模型计算的时间复杂度。

Chi-merge 算法是一种自下而上的有监督离散化算法,在离散化过程中考虑了决策属性的信息,数据按照连续变量的取值从小到大排序,然后根据卡方检验来合并相邻的取值,直到满足停止准则为止^[10],其计算公式为

$$X^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

式(1)中: m 为每次比较相邻区间的数量,取 2,即 2 个区间进行比较; k 为类别的数量; A_{ij} 为第 i 区间中第 j 类实例的数量; E_{ij} 为 A_{ij} 的期望,其计算公式为

$$E_{ij} = \frac{\left(\sum_j A_{ij}\right) \left(\sum_i A_{ij}\right)}{\sum_i \sum_j A_{ij}} \quad (2)$$

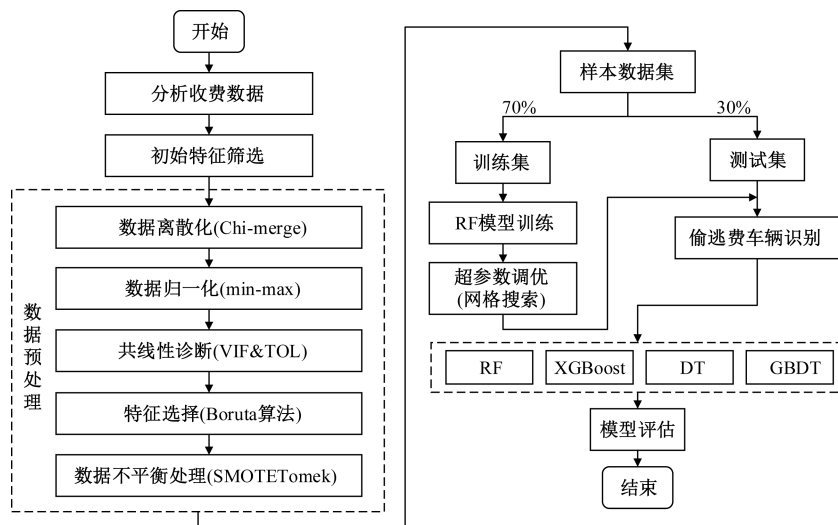


图 2 研究过程

Fig. 2 Research process

通过计算 ROC 曲线下的面积 (area under curve, AUC) 来确定最优离散区间。表 4 为对连续型特征“通行时间”进行离散化后的结果。

表 4 通行时间离散化结果

区间编号	区间段
1	[0, 0.23)
2	[0.23, 0.49)
3	[0.49, 0.99)
4	[0.99, 1.59)
5	[1.59, +∞)

2.2 数据归一化

由于高速公路收费数据中各特征的量纲差异较大,会影响模型的收敛和识别精度,例如特征“计费里程”,故对该类特征进行 min-max 归一化处理,将原始数据线性映射到 [0, 1] 区间,其计算公式为

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

式(3)中: x' 为归一化后特征的值; x 为归一化前特征的原始值; x_{\max} 为某特征中的最大值; x_{\min} 为某特征中的最小值。

2.3 共线性诊断

为简化模型计算的复杂度,采用方差膨胀因子 (variance inflation factor, VIF) 和容忍度 (Tolerance, TOL) 判断特征之间是否存在共线性,计算公式为

$$VIF = \frac{1}{1 - R^2} \quad (4)$$

式(4)中: R^2 为某个特征与其他全部特征的可决系数。

$$TOL = \frac{1}{VIF} \quad (5)$$

一般认为,当 TOL 小于 0.2 和 VIF 大于 10 时,自变量之间存在共线关系。通过计算各特征的 TOL 和 VIF,逐次剔除不满足条件的特征,根据计算结果,剔除掉“计费里程”,“入站日期”和“出站日期”3 个特征,剩余特征的 VIF 和 TOL 如表 5 所示,均满足条件,剩余特征之间不存在共线关系。

表 5 共线性诊断计算结果

特征名称	方差膨胀因子	容忍度
	VIF	TOL
最小费额里程	3.514	0.284
行驶方向是否一致	1.110	0.900
入站时段	4.273	0.234
出站时段	5.845	0.171
通行时间	1.354	0.738
是否超时行驶	1.197	0.835
入出站是否一致	1.040	0.961
是否为省内车牌	7.129	0.140

2.4 基于 Boruta 算法进行特征选择

Boruta 算法是一种基于随机森林的特征选择算法,其目标是筛选出所有与预测变量具有相关性的特征^[11]。与其他特征选择算法相比,该算法具有更强的鲁棒性,因此使用该算法来选择与高速公路变路径偷逃费行为相关的特征。

Boruta 算法的基本原理是通过计算原始特征和阴影 (重复) 特征的 Z-score 来确定特征是否重要^[12]。Z-score 的计算公式为

$$Z\text{-score} = \frac{M}{s} \quad (6)$$

式(6)中: M 为特征的平均精度损失 (mean decrease accuracy, MDA); s 为特征精度损失的标准差。

使用 Python 实现 Boruta 算法对有效特征的重要性分析,算法输出结果可分为 3 类:接受的特征 (Confirmed)、无法确定的特征 (Tentative) 和拒绝的特征 (Rejected)。经过多次迭代训练后,“行驶方向是否一致”“入出站是否一致”“通行时间”“最小费额里程”这 4 个特征被接受。表 6 为基于 Boruta 算法的特征选择结果,其中接受的特征排名为 1,暂定的特征排名为 2,拒绝的特征排名大于 2。

表 6 Boruta 算法特征选择结果

特征名称	特征选择结果	特征排名
最小费额里程	Confirmed	1
行驶方向是否一致	Confirmed	1
入站时段	Rejected	3
出站时段	Rejected	2
通行时间	Confirmed	1
是否超时行驶	Rejected	4
入出站是否一致	Confirmed	1
是否为省内车牌	Rejected	5

2.5 数据不平衡处理

SMOTETomek 综合采样是一种结合过采样方法 (SMOTE) 和欠采样方法 (Tomek Links) 的综合抽样技术,用于处理不平衡数据集的分类问题^[13]。它首先使用 SMOTE 过采样方法生成少数类样本,并在此基础上使用 Tomek Links 欠采样方法剔除类别边界上的噪声和重叠样本。这使得生成的合成样本能够更好地覆盖类别边界附近的空白区域,以在后期能够更好地训练模型,提高模型在不平衡数据集问题中的性能和精度。

在所使用的高速公路偷逃费数据中,U 型行驶、J 型行驶和正常行驶的车辆样本比例呈现极端不平衡,为典型的不平衡数据集,拟采用 SMOTETomek 综合采样方法进行处理,处理后得到 74 822 条正常

行驶车辆数据、74 843 条 U 型行驶车辆数据和 74 828 条 J 型行驶车辆数据, 各类别车辆数据数量得到均衡。

3 模型建立与评估

3.1 模型建立

3.1.1 随机森林原理

随机森林(random forest, RF)是一种基于决策树的集成学习算法^[14],它通过随机选择不同的样本子集来进行训练,以此来构建不同的决策树,最后综合多棵决策树的结果来获得更准确的高速公路变路径偷逃费行为识别结果。

随机森林具有很高的分类准确率和抗过拟合能力,能够有效地处理所研究偷逃费车辆数据集不平衡带来的问题,故使用随机森林建立高速公路变路径偷逃费行为识别模型。

3.1.2 超参数调优

机器学习模型中超参数的选择会导致不同的模型效果^[15]。为了提升模型的性能和泛化能力,采用网格搜索法(grid search)来确定模型的最优超参数组合。针对随机森林,选取对其效果影响最大的两个参数进行调优,包括基学习器个数($n_estimators$)和决策树最大深度(max_depth)。

采用均方误差(mean square error, MSE)作为指标来选取最优超参数,其结果如图 3 和图 4 所示。由图 3 和图 4 可知,最优的 $n_estimators$ 和 max_depth 超参数组合为 13 和 8。超参数调优细节及取值如表 7 所示,其余参数均选择默认值。

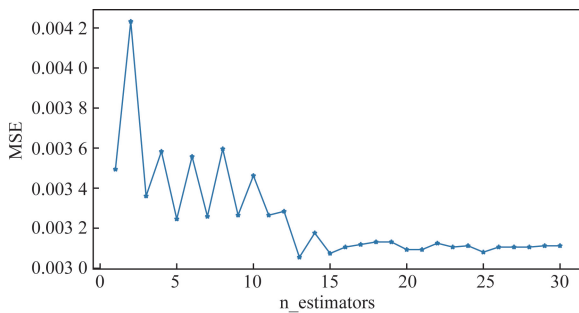


图 3 $n_estimators$ 对均方误差的影响
Fig. 3 Influence of $n_estimators$ on MSE

3.2 模型评价指标

为了验证不同模型的识别性能,选取基于混淆矩阵的精确率(P)、召回率(R)和 F1-score 作为评价指标。其计算公式为

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

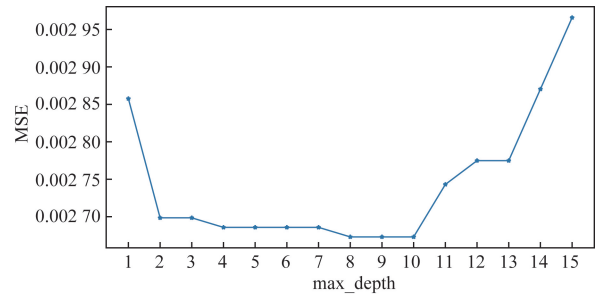


图 4 max_depth 对均方误差的影响
Fig. 4 Influence of max_depth on MSE

表 7 RF 模型的超参数调优细节及取值

Table 7 RF model hyperparameter tuning details and values

超参数	取值范围	搜索步长	取值
$n_estimators$	[1, 31)	1	13
max_depth	[1, 16)	1	8

$$F1\text{-score} = \frac{2PR}{P + R} \quad (9)$$

式中:TP 为被正确识别的偷逃费车辆数;FP 为被错误识别的偷逃费车辆数;FN 为被错误识别的正常车辆数。

考虑到所研究高速公路变路径偷逃费行为识别问题为多分类问题,且为数据集不平衡的多分类问题,因此引入 Macro-F1 分数对模型作进一步评价。Macro-F1 分数为各类别 F1-score 的平均值,其不受数据不平衡的影响,能够保证每个类别对模型最终评价的贡献相等。其计算公式为

$$Macro\text{-}F1 = \frac{1}{k} \sum_{i=1}^k (F1\text{-score})_i \quad (10)$$

式(10)中: i 为类别序号; k 为类别总数。

3.3 模型对比分析

对 RF、DT、XGBoost 和 GBDT 进行训练,并对同一测试集进行高速公路变路径偷逃费行为识别,得到不同分类模型识别结果的 P 、 R 、F1-score 和 Macro-F1,如表 8 所示。混淆矩阵能够快速分析各类别的误分类情况,绘制在测试集下不同分类模型识别结果对应的混淆矩阵,如图 5 所示。

分析表 8 可知,在正常行驶车辆的识别上,4 种模型的识别效果较为稳定,精确率和召回率均在 99% 以上,F1-score 保持在 0.9 以上。在 U 型行驶车辆的识别上,RF 模型识别的精确率最高,为 89.47%,召回率仅次于 DT 模型,但 F1-score 分数均高于其他模型。在 J 型行驶车辆的识别上,4 种模型的识别效果差异较小,精确率和召回率均在 95% 以上,F1-score 在 0.9 以上。同时分析不同模型的全局识别效果,RF、XGBoost、DT 和 GBDT 模型的 Macro-F1 分数分别为 0.966、0.943 1、0.956 3 和

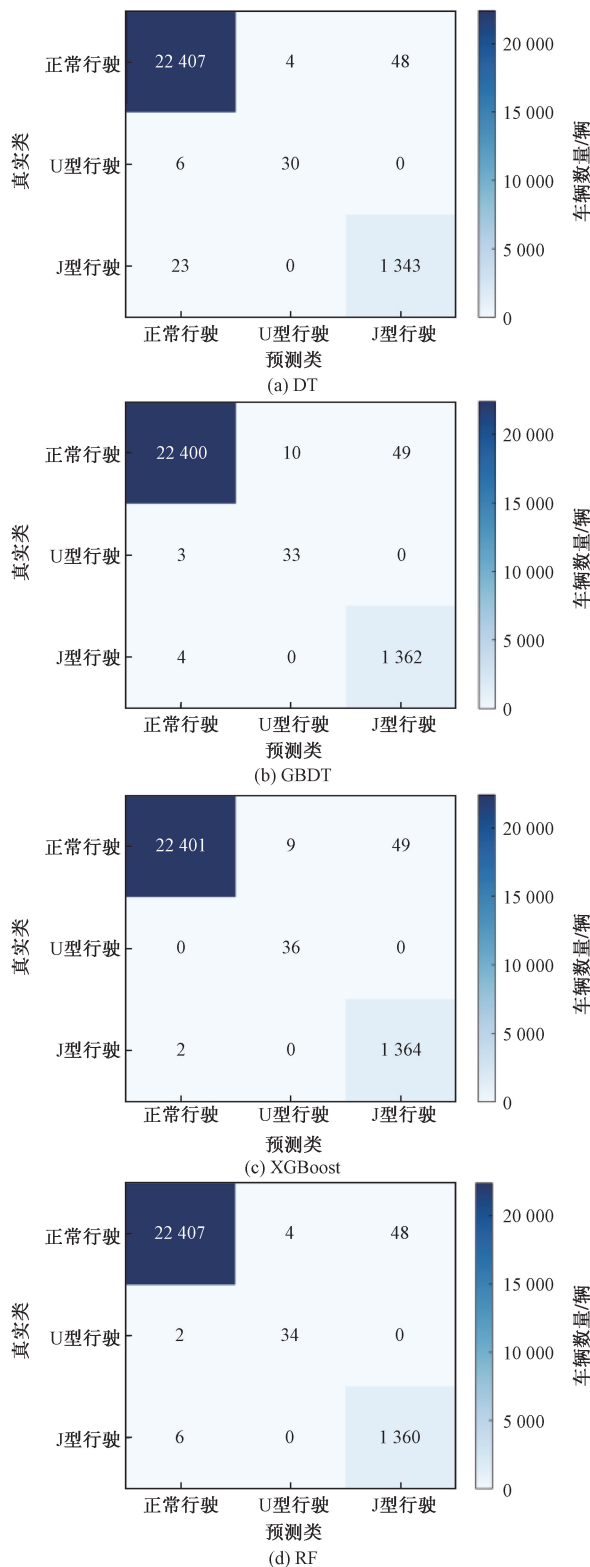


图5 不同模型的混淆矩阵

Fig. 5 The confusion matrices for different models

0.938 2, RF 模型的 Macro-F1 分数最高,表明该模型在改变路径偷逃费车辆数据集不平衡的情况下,展现出了最优的识别效果,识别结果较为全面和准确。

表8 不同模型评价指标分析

Table 8 Analysis of evaluation indexes of different models

模型	偷逃费类型	精确率/%	召回率/%	F1-score	Macro-F1
RF	正常行驶	99.96	99.76	0.998 6	0.966 0
	U型行驶	89.47	94.44	0.918 9	
	J型行驶	96.78	99.04	0.979 0	
XGBoost	正常行驶	99.87	99.76	0.998 2	0.943 1
	U型行驶	88.23	83.33	0.857 1	
	J型行驶	96.54	98.31	0.974 2	
DT	正常行驶	99.99	99.74	0.998 6	0.956 3
	U型行驶	80.00	100.00	0.888 8	
	J型行驶	96.53	99.85	0.981 6	
GBDT	正常行驶	99.96	99.73	0.998 5	0.938 2
	U型行驶	76.74	91.66	0.835 4	
	J型行驶	96.52	99.78	0.980 9	

4 结论

针对高速公路变路径偷逃费行为(U/J型行驶)进行研究,提出一种基于随机森林的高速公路变路径偷逃费行为识别模型,得出如下主要结论。

(1)为简化模型并提高模型识别准确度,对原始收费数据集进行离散化、归一化、共线性诊断和特征选取等处理。此外,利用 SMOTETomek 综合采样技术来解决改变路径偷逃费行为数据不平衡的问题。

(2)利用网格搜索法寻找影响随机森林识别效果的两个重要超参数(基学习器个数和决策树最大深度)的最优取值,提高模型的识别效果。

(3)基于预处理后的偷逃费数据集,分别训练 RF、XGBoost、DT 和 GBDT 模型来识别偷逃费行为,结果表明:RF 模型的 Macro-F1 分数为 0.966,优于其他模型,能够辅助高速公路管理部门稽查改变路径偷逃费行为车辆。

参 考 文 献

[1] 王莫凡,吴烈阳. 取消高速公路省界收费站后稽查方案探讨[J]. 中国交通信息化, 2020, 251(12): 37-38.
Wang Mofan, Wu Lieyang. Discussion on the plan to cancel the inspection after the expressway provincial boundary toll station[J]. China ITS Journal, 2020, 251(12): 37-38.

[2] 张凡,王宇轩,陈旻瑞,等. 基于 SMOTE-Logistic 的高速公路鲜活农产品绿色通道车辆稽查与分析[J]. 公路交通科技(应用技术版), 2019, 15(10): 303-307.
Zhang Fan, Wang Yuxuan, Chen Minrui, et al. Inspection and analysis of vehicles in green channel of fresh agricultural products on expressway based on SMOTE Logistic[J]. Road Traffic Technology (Applied Technology Edition), 2019, 15(10): 303-307.

[3] Zhao Y H, Shi H F, Ren X C, et al. A toll evasion recognition method based on Gaussian mixture clustering[J]. Communication in Statistics-Simulation and Computation, 2020, 2: 1-11.

[4] 李松江,周舟,李岩芳,等. 基于 IGA-IBP 算法的高速公路逃

- 费预测[J]. 计算机工程与设计, 2018, 39(12): 3840-3845.
- Li Songjiang, Zhou Zhou, Li Yanfang, et al. Prediction of highway escape cost based on IGA-IBP algorithm[J]. Computer Engineering and Design, 2018, 39(12): 3840-3845.
- [5] 刘昱岗, 郑帅, 徐旭东, 等. 基于历史通行数据的假冒绿通车逃费行为预测[J]. 公路交通科技, 2021, 38(4): 92-102, 141.
- Liu Yugang, Zheng Shuai, Xu Xudong, et al. Prediction of fake toll-free vehicles based on historical traffic data[J]. Journal of Highway and Transportation Research and Development, 2021, 38(4): 92-102, 141.
- [6] 向红艳, 杨朋涛, 伊佳佳. 基于 RF-LR 的高速公路逃费车辆状态预测模型[J]. 重庆师范大学学报(自然科学版), 2020, 37(1): 75-80.
- Xiang Hongyan, Yang Pengtao, Yi Jiajia. State prediction model of expresway escaping vehicles based on RF-LR [J]. Journal of Chongqing Normal University (Natural Science), 2020, 37(1): 75-80.
- [7] 郑覃, 潘军, 蒋立军, 等. 基于决策树判别的高温目标遥感识别方法[J]. 科学技术与工程, 2019, 19(11): 236-241.
- Zheng Qin, Pan Jun, Jiang Lijun, et al. High temperature targets remote sensing recognition based on decision tree discrimination[J]. Science Technology and Engineering, 2019, 19(11): 236-241.
- [8] 周建亮, 胡飞翔, 邢艳冬, 等. 基于人格特质和机器学习分类算法的建筑工人不安全行为识别[J]. 科学技术与工程, 2022, 22(29): 13013-13020.
- Zhou Jianliang, Hu Feixiang, Xing Yandong, et al. Identification of unsafe behaviors of construction workers based on personality traits and machine learning classification algorithms[J]. Science Technology and Engineering, 2022, 22(29): 13013-13020.
- [9] 于翔海, 白佃国, 于光, 等. 基于 XGBoost 模型的城市道路实时交通事故风险预测研究[J]. 公路交通科技, 2023, 40(4): 237-247.
- Yu Xianghai, Bai Dianguo, Yu Guang, et al. Study on predicting real-time traffic accident risk of urban road based on XGBoost model [J]. Journal of Highway and Transportation Research and Development, 2023, 40(4): 237-247.
- [10] Ropero R F, Renooij S, Van G L C. Discretizing environmental data for learning Bayesian-network classifiers[J]. Ecological Modelling, 2018, 368: 391-403.
- [11] Cao Y, Ashuri B, Baek M. Prediction of unit price bids of resurfacing highway projects through ensemble machine learning[J]. Journal of Computing in Civil Engineering, 2018, 32(5): 04018043.
- [12] Jamei M, Maroufpoor S, Aminpour Y, et al. Developing hybrid data-intelligent method using Boruta-random forest optimizer for simulation of nitrate distribution pattern[J]. Agricultural Water Management, 2022, 270: 1-18.
- [13] Wang Z, Wu C, Zheng K, et al. SMOTETomek-based resampling for personality recognition[J]. IEEE Access, 2019, 7: 129678-129689.
- [14] 张志刚, 徐莹, 张锦秋, 等. 基于随机森林的公路隧道 CO 气体浓度预测模型[J]. 科学技术与工程, 2022, 22(26): 11729-11735.
- Zhang Zhigang, Xu Ying, Zhang Jinqiu, et al. Prediction model of CO concentration in highway tunnel based on random forest [J]. Science Technology and Engineering, 2022, 22(26): 11729-11735.
- [15] 杨艳艳, 李雷孝, 林浩, 等. 参数并行: 一种基于群启发式算法的机器学习参数寻优方法[J]. 科学技术与工程, 2022, 22(5): 1972-1980.
- Yang Yanyan, Li Leixiao, Lin Hao, et al. Parallel parameters: an optimization method of machine learning parameters based on swarm heuristic algorithm[J]. Science Technology and Engineering, 2022, 22(5): 1972-1980.