

融合时空信息的 BROC-LGBM 高速公路拥堵识别算法

钟欢¹, 邹复民¹, 廖律超¹, 罗永煜², 罗旭³, 马纪媛¹

(1. 福建理工大学 福建省汽车电子与电驱动技术重点实验室, 福建 福州 350118;

2. 福建省高速公路信息科技有限公司, 福建 福州 350001;

3. 电子科技大学 计算机科学与工程学院, 四川 成都 611731)

摘要: 为了解决现有研究在路网交通流刻画不足、数据处理成本高及模型可解释性差等问题, 提出了融合时空信息的 BROC-LGBM (borderline-ROC optimized light GBM) 高速公路拥堵识别算法。首先, 设计时间维度的区段分级延误比向量和空间维度的车型加权占有率向量, 解决了传统方法对路段时空特征刻画不足的问题, 实现对交通流特征的精准描述。然后, 引入 Borderline-SMOTE 生成边界合成样本, 增强模型对少数类(拥堵数据)的学习能力并优化类别分布。最后, 采用 ROC 曲线动态阈值优化策略, 解决传统分类模型在拥堵识别中因固定阈值导致的精度不足及误判率高问题。基于福建省高速公路 43.2 万余条 ETC 数据的实验结果表明, BROC-LGBM 在 G-Mean (95.2%)、召回率 (97.1%) 和误检率 (6.4%) 等指标上显著优于现有模型。

关键词: 高速公路拥堵检测; 时空特征; ETC 数据; Borderline-SMOTE; LightGBM

中图分类号: TP181; U49.14 **文献标志码:** A **文章编号:** 2097-3853(2026)03-0258-08

BROC-LGBM algorithm for highway congestion detection with spatiotemporal information fusion

ZHONG Huan¹, ZOU Fumin¹, LIAO Lüchao¹, LUO Yongyu², LUO Xu³, MA Jiyuan¹

(1. Fujian Provincial Key Laboratory of Automotive Electronics and Electric Drive Technology, Fujian University of Technology, Fuzhou 350118, China;

2. Fujian Expressway Information Technology Co., Ltd., Fuzhou 350001, China;

3. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: To address the limitations of existing research, such as insufficient characterization of traffic flow, high data processing costs, and poor model interpretability, a borderline-ROC optimized light GBM (BROC-LGBM) highway congestion detection algorithm is proposed, which integrates spatiotemporal information. First, a time-domain segmental delay distribution vector and a spatial-domain vehicle-type weighted occupancy rate vector are designed to overcome the shortcomings of traditional methods in characterizing spatiotemporal features of road segments, enabling precise traffic flow representation. Then, the Borderline-SMOTE technique is introduced to generate synthetic boundary samples, enhancing the model's ability to learn from minority classes (congestion data) and optimizing class distribution. Finally, a ROC curve-based dynamic threshold optimization strategy is employed to address the issues of insufficient accuracy and high false detection rates caused by fixed thresholds in traditional classification models. Experimental results based on over 432,000 ETC data records from highways in Fujian Province demonstrate that BROC-LGBM significantly outperforms existing models in key metrics, achieving a G-Mean of 95.2%, a recall rate of 97.1%, and a false detection rate.

Keywords: highway congestion detection; spatiotemporal features; ETC data; Borderline-SMOTE; LightGBM

收稿日期: 2025-02-18

基金项目: 福建省交通运输科技计划项目 (2025Y047)

第一作者简介: 钟欢 (1999—), 女, 四川广安人, 硕士研究生, 研究方向: 智能驾驶与交通安全。

通信作者: 邹复民 (1976—), 男, 湖南隆回人, 教授, 博士, 研究方向: 大数据技术、交通信息与控制工程。

近年来,我国机动车保有量和驾驶人数量持续快速增长,交通流量的急剧增加使得高速公路交通拥堵问题日益严峻^[1-2]。实现高速公路交通拥堵的精准检测是提升道路通行能力和保障通行安全性的关键问题。现有的检测方法主要依赖视频监控数据、结构化轨迹数据和仿真数据。其中,视频监控方法因其直观性广泛应用,但部署成本高、覆盖有限,且易受环境干扰^[3];结构化轨迹数据(如 GPS、ETC 交易数据)在反映交通流时空特性方面优于视频数据,但实时性欠佳,难以满足高速公路大范围、动态化的检测需求^[4];仿真数据作为补充,通过模拟交通流进行实时检测^[5],大大提升了拥堵检测方法的准确性。

在方法论方面,机器学习和深度学习技术的发展为交通拥堵检测提供了强有力的工具。随机森林、LightGBM 等轻量级模型以高效性和特征表达能力被广泛应用^[6],而深度学习(如长短期记忆网络、Transformer 模型)通过强大的特征提取能力显著提升了检测精度^[7],但因复杂性和可解释性不足限制了其实际应用^[8]。为了弥合效率与精度的差距,轻量化模型和时空特征建模成为研究热点。

特征选择作为交通拥堵检测的关键环节,为模型性能优化提供了基础。速度特征直观反映交通动态变化^[9],时间特征(如延误率、行程时间可靠性)有效衡量交通流稳定性^[10],而密度特征则揭示了道路资源利用效率及拥堵扩散趋势^[11]。结合图卷积网络(GCN)和门控循环单元(GRU)的模型,不仅能够有效捕捉交通流的时空依赖性,还为交通预测提供了系统化解决方案^[12]。

基于上述背景,本研究针对现有高速公路拥堵检测交通流刻画不足、数据处理成本高及模型可解释性差等问题,提出一种基于电子不停车收费系统(ETC)数据融合时空信息的 BROC-LGBM 高速公路拥堵识别算法。

1 融合时空信息的 BROC-LGBM 高速公路拥堵识别算法

1.1 算法框架

所提 BROC-LGBM 高速公路拥堵识别算法框架如图 1 所示。算法基于 ETC 生成的仿真与路径预测数据,提取目标区段的时空关键信息,采用边界合成少数类过采样技术(borderline-synthetic

minority over-sampling technique, Borderline-SMOTE)对不平衡数据上采样,缓解类别分布不均问题。随后,利用 LightGBM 模型训练并识别拥堵区段,并基于 ROC 曲线(receiver operating characteristic curve)选择最佳阈值优化模型性能。

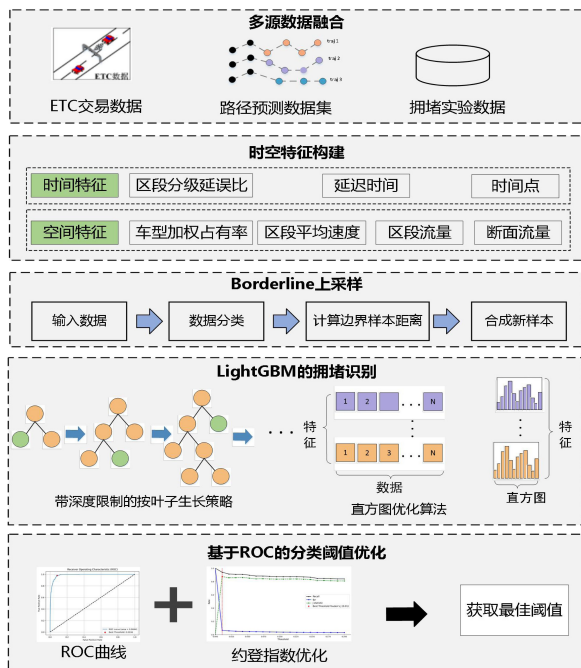


图 1 算法框架

Fig.1 Algorithm framework

1.2 融合时空信息的拥堵特征向量建模

为了更全面地描述高速公路区段的拥堵情况及动态变化,克服传统特征对交通流时空特性刻画不足的缺陷,本研究在传统区段流量、时间、速度特征向量的基础上,引入基于时间维度的区段分级延误比例向量(delayed distribution proportion vector, DDP)和基于空间维度的占有区段率特征向量(occupancy rate, OR)时空联合特征建模方法。

1.2.1 区段分级延误比向量

时间延误率(time delay rate, TDR)是衡量区段交通流稳定性的重要指标。本研究提出的预测交易时间延误指数如式(1)所示。

$$TDR = \frac{T_{e_i}}{\text{pre}(T_{e_i})} \quad (1)$$

式中, T_{e_i} 为边 e_i 的实际行程时间, s ; $\text{pre}(T_{e_i})$ 为预测车辆应到行程时间, s 。

基于此提出区段分级延误比向量 DDP,以严重延误车辆的比例为核心,通过分级分析揭示延

误分布特性,如式(2)所示。

$$DDP = \frac{\sum_i^n I(\text{TDR}_i > T_{\text{upper}})}{n}, T_{\text{upper}} \in \{1, 1.5, 2\} \quad (2)$$

式中, $I(x)$ 为指示函数, T_{upper} 为区段的延误上限, i 为车辆数, n 为该区段内的车辆总数,当 $\text{TDR}_i > T_{\text{upper}}$ 时,指示函数取值为1,否则为0。

1.2.2 车型加权占有率

为了更准确地反映车辆在道路空间实际占用情况,引入车型折算系数以精确表征车辆对道路空间的占用强度,提出车型加权的区段占有率特征向量OR,具体计算如式(3)所示。

$$\text{OR} = \frac{\sum_{k=1}^n (n_k \times w_k)}{L} \quad (3)$$

式中, L 为区段长度, n_k 为车型为 k 的车辆总数, w_k 为车型为 k 的车辆的折算系数。

本研究充分考虑不同车型对道路通行能力的影响,依据相关标准《公路工程技术标准》(JTG BO1-2014)及《收费公路车辆通行费车型分类》(JT/T 489-2019)车型划分标准,赋予各车型折算系数如表1所示,从而能更精确地表征车辆对道路空间的占用状态。

表1 交通车型折算系数表

Tab.1 Vehicle type conversion factors

客车		货车		专项作业车	
车型	折算系数	车型	折算系数	车型	折算系数
1	1.0	11	1.0	21	1.0
2	1.0	12	1.5	22	1.5
3	1.5	13	2.5	23	2.5
4	1.5	14	2.5	24	2.5
		15	4.0	25	4.0
		16	4.0	26	4.0

1.2.3 时空信息融合

为了联合表征交通流的时空动态特性,提出基于特征拼接的时空特征融合策略。

首先,对区段级特征归一化处理,对每个区段 i 的时间延误率(DDP)与车辆加权占有率(OR)进行分别进行标准化,如式(4)、式(5)所示。

$$D_i = \frac{\text{DDP}_i - \mu}{\delta} \quad (4)$$

$$\tilde{O}_i = \frac{\text{OR}_i - \mu}{\delta} \quad (5)$$

式中, DDP_i 为区段 i 的时间延误率, OR_i 为区段 i 的车辆加权占有率, μ 为特征均值, δ 为标准差, D_i 为标准化时间延误率, \tilde{O}_i 为标准化车辆加权占有率。

接着,进行特征拼接与矩阵构建,将每个区段的时空特征拼接为联合特征向量如式(6)所示,并构建全网特征矩阵如式(7)所示。

$$\mathbf{x}^{(i)} = [D_i, \tilde{O}_i] \quad (6)$$

$$\mathbf{X} = \begin{bmatrix} x^1 \\ x^2 \\ \dots \\ x^n \end{bmatrix} \quad (7)$$

式中, $\mathbf{x}^{(i)}$ 为第 i 个区段的归一化联合特征向量, n 为区段总数, $k+1$ 为每个区段特征维度, \mathbf{X} 为全网所有区段的联合特征矩阵。

1.3 自适应边界采样策略

高速公路交通数据集中,拥堵事件的发生是一个稀疏事件,并未在所有区段上普遍产生。为此,本研究选取了福建省某典型工作日的交通数据进行分析,揭示发生拥堵区段和未发生拥堵区段的样本分布情况。如图2所示,拥堵路段的比例极低,非拥堵路段占比极大,数据呈现出明显的类别不平衡现象。

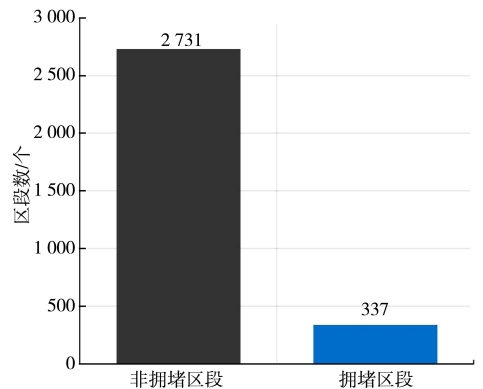


图2 高速公路交通数据类别分布

Fig.2 Distribution of highway traffic data categories

类别失衡会导致模型在训练过程中偏向多数类(非拥堵路段),从而削弱对少数类(拥堵路段)

的学习能力,影响拥堵检测的准确性。为缓解该问题,本研究采用 Borderline-SMOTE 方法通过 K 最近邻(K-nearest neighbors, KNN)分析将样本划分为安全样本、边界样本和噪声样本,选择少数类中的“边界样本”生成新的合成数据。与传统的 SMOTE 方法不同,Borderline-SMOTE 聚焦“边界样本”进行插值,生成更具代表性的新样本^[13],如图 3 所示,从而避免噪声点干扰,增强了少数类表征能力,提高了模型对拥堵特征的学习效果与鲁棒性。

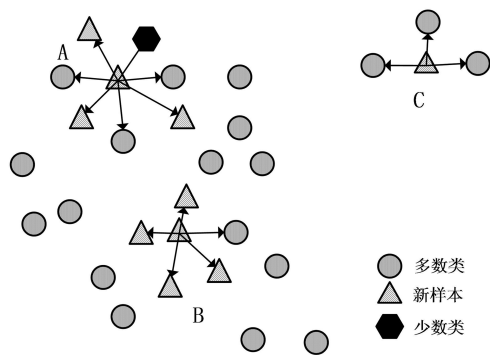


图3 基于 Borderline-SMOTE 的少数类样本合成
Fig.3 Synthesis of minority class samples based on Borderline-SMOTE

1.4 ROC 最佳阈值优化策略

在分类问题中,阈值是决定样本被分类为正类或负类的关键因素。通常,模型会采用默认的分类阈值,然而,在高速公路拥堵检测任务中,由于拥堵路段的比例相对较低,默认阈值往往导致模型对多数类(非拥堵路段)产生偏倚,从而削弱对少数类(拥堵路段)的检测能力^[14]。这种阈值偏置会显著影响模型的识别性能,尤其是在交通拥堵这一稀有事件的检测中,可能导致拥堵状态的漏检率较高,降低模型的实用性和可靠性。因此,合理调整分类阈值对于提升拥堵检测模型的性能至关重要。

ROC 曲线是评估二分类模型性能的重要工具,通过描绘假阳性率(false positive rate, FPR)与真正率(true positive rate, TPR)之间的关系,可以全面反映模型在不同阈值下的分类表现。本研究通过最大化约登指数(Youden's index)来优化模型阈值选择,在提升真正率的同时有效抑制假阳性率。约登指数 J 被定义为假阳性率与真正率之差,如式(8)。

$$J = \text{TPR} - \text{FPR} \quad (8)$$

1.5 LightGBM 分类算法

梯度提升决策树(gradient boosted decision trees, GBDT)是集成学习利用 Boosting 策略的具有代表性的框架,它通过组合弱学习器构建强分类器或强回归器常用的基学习器是分类与回归树(classification and regression tree, CART)。GBDT 是一种加法模型,它把所有 CART 树的预测值累加起来作为最终的预测值,通过沿着基学习器损失函数梯度下降的方向不断迭代训练模型。

假设训练集的输入样本特征向量为 \mathbf{x} , 输出向量为 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ 。GBDT 可以表示为决策树的加法模型,第 m 次迭代后模型的预测函数 $f_m(\mathbf{x})$ 如式(9)所示。

$$f_m(\mathbf{x}) = \sum_{m=1}^M T(\mathbf{x}, \theta_m) \quad (9)$$

式中, $T(\mathbf{x}, \theta_m)$ 表示第 m 棵决策树的预测函数, θ_m 为第 m 棵决策树的参数集合, M 为树的个数。即第 m 步迭代的模型表示如式(10)所示。

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + T(\mathbf{x}, \theta_m) \quad (10)$$

假设 y_i 为第 i 个样本的真实值, $f_m(x_i)$ 为第 i 个样本的预测值,取损失函数为平方损失,表示为式(11)。

$$L(Y, f_m(x_i)) = \frac{1}{2} (y_i - f_m(x_i))^2 \quad (11)$$

式中, Y 为真实值, $L(Y, f_m(x_i))$ 为第 m 步模型的损失函数。通过最小化损失函数得到第 m 棵决策树的最优参数 $\hat{\theta}_m$ 如式(12)所示。

$$\hat{\theta}_m = \sum_{i=1}^n L(Y, f_m(x_i)) + T(\mathbf{x}, \theta_m) \quad (12)$$

经过多次迭代,更新回归树可以得到最终模型。

2 实验验证

2.1 实验数据集

实验数据来源于福建省电子与电驱动技术重点实验室与福建省高速公路信息科技有限公司合作开发的高速公路 ETC 门架数据关联分析系统,选用系统 2024 年 9 月 20 日生成的数据进行研究,共包含约 4 762 万条交易记录据,涉及约 147 万辆车。

2.2 模型参数优化

通过超参数优化框架 Optuna 优化 LightGBM 的核心参数,模型关键参数及其优化后的设置如表 2 所示。

表 2 LightGBM 模型参数优化结果

Tab.2 Optimization results of lightGBM model

parameters		
参数	参数名称	参数值
learning_rate	学习率	0.1
max_depth	最大深度	15
n_estimators	树的数量	300
num_leaves	决策树的叶子数	70

2.3 评价指标

传统分类问题的评估指标如精度、错误率不适用于极度不平衡数据集^[15],为更全面评估模型性能,本研究采用几何均值(geometric mean, G-Mean)、曲线下面积(area under curve, AUC)、召回率(Recall)、误检率(false detection rate, FPR)等评价指标。

G-Mean 能够衡量模型在不平衡数据集上的综合表现,反映正负类分类的平衡性,如式(13)所示。

$$G-Mean = \sqrt{\frac{TP}{(TP+FN)} \times \frac{TN}{(TN+FP)}} \quad (13)$$

式中,TP(true positive)为真阳性表示被正确分类的正类样本数,FN(false negative)表示被误判为负类的正类样本数,TN(true negative)为真阴性表示被正确分类的负类样本数,FP(false positive)为假阳性表示被误判为正类的负类样本数。

AUC 为 ROC 曲线下的面积,衡量模型区分正负类的能力,如式(14)所示。

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (14)$$

Recall 可以衡量模型对少数类(拥堵路段)的识别能力,如式(15)所示。

$$Recall = \frac{TP}{(TP+FN)} \quad (15)$$

FPR 表示负类样本被误分类为正类的比例,反映模型的误判情况,如式(16)所示。

$$FPR = \frac{FP}{FP+TN} \quad (16)$$

2.4 实验结果

2.4.1 ROC 曲线

受试者工作特征曲线(即 ROC 曲线)常用于衡量模型在各种阈值设置下的泛化性能,实验通过

计算模型在不同分类阈值下的真正率(TPR)和假正率(FPR),绘制了如图 4 所示的 ROC 曲线。

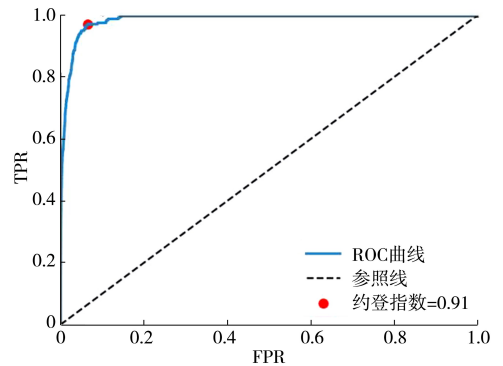


图 4 ROC 曲线

Fig.4 ROC curve

从图 4 的 ROC 曲线可见,模型表现出良好的分类性能。该曲线明显靠近图的左上角,且下方的面积即 AUC 值较大。曲线上的最优阈值对应点此时的约登指数(Youden's Index)为 0.91,该点位于曲线的左上角,表明在该最佳阈值下,模型能够在保持较高的真正率(TPR)的同时,维持较低的假正率(FPR),从而实现了最佳的分类效果。

2.4.2 不同模型的性能对比分析

为验证所提模型的有效性,将其与极限梯度提升(xtreme gradient boosting, XGBoost)、自然梯度提升(natural gradient boosting, NGBoost)等主流模型对比,实验结果如表 3 所示。

表 3 不同模型对比实验

Tab.3 Comparative experiments of different models

模型	G-Mean	AUC	召回率/%	误检率/%
TableNet	0.761	0.785	59.3	2.3
XGBoost	0.853	0.862	73.9	1.4
SVM	0.877	0.878	83.4	7.7
CatBoost	0.871	0.876	78.4	3.2
NGboost	0.902	0.907	90.2	5.4
BROC-LGBM	0.952	0.953	97.1	6.4

如表 3 所示,BROC-LGBM 的 G-Mean(0.952)和 AUC(0.953)较最优对比模型 NGBoost 分别提升 5.0%和 4.6%;召回率达 97.1%,较 XGBoost 提升 23.2%;尽管误检率(6.4%)略高于 XGBoost(1.4%)、CatBoost(3.2%),但其在召回率上表现

更好,降低了交通拥堵漏检风险。实验表明,时空特征融合和动态阈值优化策略有效提升了模型对拥堵事件的敏感性和综合判别能力,验证了算法在高速公路场景下的优越性。

2.4.3 消融实验

为验证采样和 ROC 阈值选择组件的有效性,实验组合覆盖了两种核心因素的所有可能状态:是否应用数据平衡技术(有/无)以及是否采用 ROC 最佳阈值选择(有/无)。实验结果如表 4 所示。

表 4 消融实验模型性能结果

Tab.4 Performance results of ablation study models

模块		G-	AUC-	召回	误检
BL-SMOTE	ROC	Mean	ROC	率/%	率/%
		0.574	0.664	33.1	0.2
✓		0.621	0.692	38.6	0.1
	✓	0.842	0.844	78.4	9.5
✓	✓	0.952	0.953	97.1	6.4

从表 4 可见, Borderline-SMOTE 上采样技术显著改善了模型的对正样本的感知能力,召回率平均提升约 5%。ROC 最佳阈值选择对模型整体性能的优化最为明显,较仅使用采样模块模型 G-Mean 和召回率平均提升均超过 20%,且在综合使用时空矩阵、上采样和 ROC 阈值优化的情况下,模型性能达到最优状态,充分验证 Borderline-SMOTE 采样方法和阈值优化的有效性

2.4.4 融合时空信息有效性实验

为探索融合时空信息特征向量的有效性,设计了两组对比实验:第一组使用传统流量、时间、速度特征;第二组融合传统特征与时空特征,结果如表 5 所示。

表 5 融合时空特征消融实验模型性能对比

Tab.5 Performance comparison of spatiotemporal feature ablation experiments

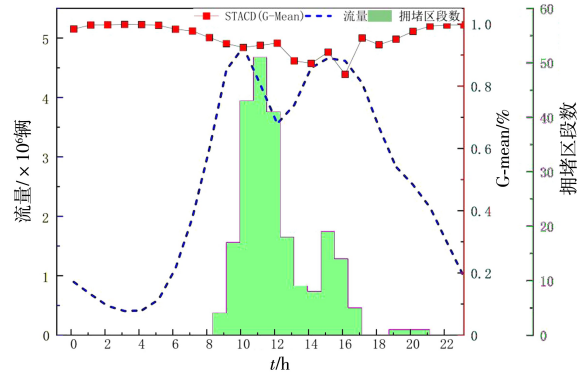
实验特征	G-Mean	AUC	召回率/%	误检率/%
传统特征向量	0.915	0.916	97.5	14.2
融合时空信息	0.952	0.953	97.1	6.4

实验结果表明,融合时空信息的特征向量显

著提升了模型性能,其 G-Mean 较使用传统向量模型提升了 3.7%,达到 95.2%,误检率从 14.2% 降至 6.4%,可有效减少对非拥堵路段的误判。

2.4.5 模型鲁棒性评估

为评估模型应对全天交通流量和拥堵状态变化的适应性与鲁棒性,实验分析了交通流量、拥堵区段数量及 BROC-LGBM 模型在 G-Mean 指标在 24 h 内的变化趋势,结果如图 5 所示。



图注:当数据集中仅包含单一类别样本时,由于无法计算模型在另一类别上的性能,G-mean 的值为该类别的准确率。

图 5 交通流量与准确率变化

Fig.5 Variation of traffic flow and accuracy

从图 5 可以看见,非高峰时段(0:00-7:00、19:00-23:59),交通流量较低,路网畅通,BROC-LGBM 算法表现优异,G-Mean 指标始终高于 0.95;在高峰时段(8:00-10:00、16:00-18:00),流量大幅增加,拥堵区段显著增多,但模型的 G-Mean 仅小幅下降,仍保持在 0.85 以上。

为进一步验证算法优异性,对 24 h 内使用不同模型对拥堵情况进行实验,结果如图 6 所示。

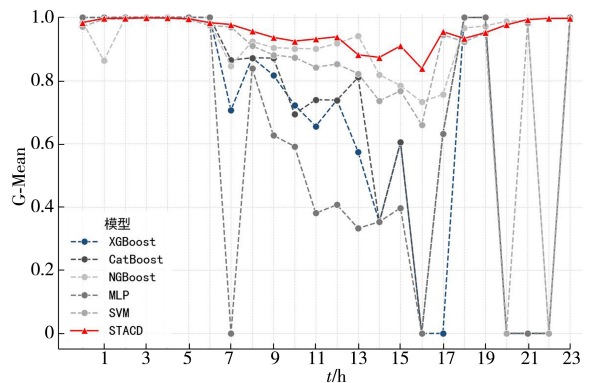


图 6 不同模型 G-mean 随时间变化情况

Fig.6 Variation of G-mean for different models over time

由图 6 可见,与其他模型相比,BROC-LGBM 的检测的准确率在全天内稳定性更强。无论高峰或非高峰期,其表现始终优于其他模型。

为验证模型在拥堵常发地区和其他区段的表现,筛选 50 个易拥堵区段与其他区段进行实验,结果如表 6 所示。

表 6 易拥堵区段和非拥堵区段模型表现情况

Tab.6 Model performance in congested and non-congested sections

区段类型	G-Mean	AUC	召回率/%	误检率/%
易拥堵区段	0.961	0.961	99.9	7.7
其他区段	0.942	0.943	97.5	8.9

从表 6 可见,BROC-LGBM 在易拥堵区段的 G-Mean 和 AUC 均达到 0.961,召回率为 99.9%,误检率仅为 7.7%;在其他区段,G-Mean 指数和 AUC 指数也达到 0.942 和 0.943,模型整体表现稳定。

综上,BROC-LGBM 在全天不同时间段及区域均展现出优异的检测性能,具有良好的鲁棒性、适应性及实际应用价值,为交通管理提供了可靠支持。

参考文献:

- [1] 中国政府网. 全国机动车保有量达 4.35 亿辆,驾驶人达 5.23 亿人[EB/OL]. [2024-01-11]. https://www.gov.cn/lianbo/bumen/202401/content_6925362.htm.
- [2] 交通运输部. 2023 年交通运输行业发展统计公报[EB/OL]. (2024-06-18)[2025-01-14]. https://xxgk.mot.gov.cn/2020/jigou/zhghs/202406/t20240614_4142419.html.
- [3] LI X Y,HAO T L,JIN X,et al. Fine traffic congestion detection with hierarchical description[J]. IEEE Transactions on Intelligent Transportation Systems,2022,23(12):24439-24453.
- [4] WANG W K,CHEN J C,WU Y J. The prediction of freeway traffic conditions for logistics systems[J]. IEEE Access,2019,7:138056-138061.
- [5] QU Z G,LIU X Z,ZHENG M. Temporal-spatial quantum graph convolutional neural network based on Schrödinger approach for traffic congestion prediction[J]. IEEE Transactions on Intelligent Transportation Systems,2023,24(8):8677-8686.
- [6] JASIM S, MOHAMMED M, ZAGHDEN N, et al. Improving detection and prediction of traffic congestion in VANETs: an examination of machine learning[J]. International Journal of Computing and Digital Systems 15.1(2024): 947-960.
- [7] MAKARA L A,MARIC P,PEKAR A. Public transport congestion detection using incremental learning[J]. Pervasive and Mobile Computing,2023,91:101769.
- [8] BAWANEH M,SIMON V. Novel traffic congestion detection algorithms for smart city applications[J]. Concurrency and Computation:Practice and Experience,2023,35(5):e7563.
- [9] 董家鹭,崔美琪,张慧杰,等. 基于图表示学习的交通拥堵传播模式挖掘与可视分析方法[J/OL]. 计算机辅助设计与图形学学报,2024:1-14. (2024-10-29). <https://link.cnki.net/urlid/11.2925,TP.20241029.1424.044>.
- [10] 邓红星,白伊杨. 考虑轨迹数据的公交运行状态识别[J]. 重庆理工大学学报(自然科学),2023,37(2):233-240.
- [11] OUMAIMA E J,JALEL B O,VÉRONIQUE V. A stochastic traffic model for congestion detection in multi-lane highways

3 结论

1)BROC-LGBM 引入基于时间维度的区段分级延误比向量和空间维度的车型加权占有率向量,克服了传统方法在时空特征刻画和交通流动态分析中的不足,增强了模型对交通流时空变化的精准捕捉能力,显著提升了模型的检测精度。

2)BROC-LGBM 通过有效处理不平衡数据问题和优化分类阈值,显著提升了模型的性能。针对高速公路拥堵检测中的数据不平衡问题,BROC-LGBM 采用了改进的 Borderline-SMOTE 方法,增强了模型对少数类(即拥堵路段)的学习能力。同时,结合 ROC 曲线的阈值优化策略,模型能够动态调整分类标准,从而有效避免了传统分类模型因固定阈值带来的精度不足和误判率较高的问题。在此优化下,模型在召回率达到 97.1%,且误检率仅为 6.4%,展现了卓越的精度和鲁棒性。

3)BROC-LGBM 在高峰时段和常发性拥堵路段条件下展现了出色的稳定性与适应性,保证了模型在复杂且动态变化的交通环境中的高效运行,具有较强的应用潜力,特别适用于高速公路大规模、实时交通管理系统。

[M]//Ad Hoc Networks. Cham:Springer International Publishing,2021:87-99.

- [12] ZHAO L, SONG Y J, ZHANG C, et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9):3848-3858.
- [13] HUSSEIN A S, LI T R, YOHANNESE C W, et al. A-SMOTE: a new preprocessing approach for highly imbalanced datasets by improving SMOTE[J]. International Journal of Computational Intelligence Systems, 2019, 12(2):1412-1422.
- [14] 杨绮梦, 张春平, 常耀文, 等. 基于 ROC 曲线的山区沟道提取最佳汇流量阈值分析[J]. 干旱区资源与环境, 2023, 37(10):93-101.
- [15] 宋玲玲, 王时绘, 杨超, 等. 改进的 XGBoost 在不平衡数据处理中的应用研究[J]. 计算机科学, 2020, 47(6):98-103.

(责任编辑:方素华)

(上接第 212 页)

- [8] LI Z J, WENG X L, CUI Y C. Study on the influence of pipe jacking construction on existing subway tunnels[J]. Environmental Earth Sciences, 2024, 83(23):651.
- [9] 张志伟, 李忠超, 梁荣柱, 等. 软土层矩形顶管掘进引起地表隆沉变形分析[J]. 岩土力学, 2022, 43(S1):419-430.
- [10] 焦义, 梁禹, 冯金勇, 等. 多因素影响下顶管施工引起土体变形计算研究[J]. 铁道科学与工程学报, 2021, 18(1):192-199.
- [11] 肖旦强, 张仕超, 胡智, 等. 大断面矩形顶管施工对近接斜交既有隧道影响研究[J]. 现代隧道技术, 2022, 59(S1):441-447.
- [12] 程伟峰, 江宏庆, 张敬磊, 等. 软土地区矩形顶管下穿高速公路施工全周期监测与分析[J]. 建筑结构, 2022, 52(S1):3131-3137.
- [13] 韩仲慧, 王梅. 浅埋矩形顶管施工对临近管线与地表的影响研究[J]. 都市轨道交通, 2024, 37(1):135-142.
- [14] 胡聪, 郝英奇. 双线平行顶管在不同间距下施工的模式试验与数值模拟分析[J]. 建筑结构, 2021, 51(S2):1854-1860.
- [15] 中华人民共和国住房和城乡建设部. 城市道路工程设计规范: CJJ37-2012[S]. 北京, 中国建筑工业出版社, 2016.

(责任编辑:马晓磊)