

DOI:10.19951/j.cnki.1672-9331.20230101002

文章编号:1672-9331(2023)02-0045-10

引用格式:林南亭,胡林,林森,等.基于时间序列聚类的交通事故黑点识别与分析[J].长沙理工大学学报(自然科学版),2023,20(2):45-54.

Citation: LIN Nanting, HU Lin, LIN Miao, et al. Black spot identification and analysis of traffic accidents based on time series clustering[J]. J Changsha Univ Sci Tech (Nat Sci), 2023, 20(2): 45-54.

基于时间序列聚类的交通事故黑点识别与分析

林南亭^{1,2}, 胡林^{1,2}, 林森³, 彭华⁴

(1.长沙理工大学 汽车与机械工程学院,湖南 长沙 410114;2.长沙理工大学 工程车辆安全性设计与可靠性技术
湖南省重点实验室,湖南 长沙 410114;3.中国汽车技术研究中心有限公司,天津 300300;
4.长沙市公安局交通警察支队,湖南 长沙 410006)

摘要:【目的】鉴别道路交通事故黑点以及探究事故高发区域的致因。【方法】结合地理信息软件与可解释性机器学习算法,提出一种考虑交通事故时空属性的事故黑点识别及致因分析的方法。【结果】事故高发区域主要聚集在大型商业圈、客运车站与工业区附近,其事故密度为事故低发区域的6.61倍。在事故高发区域,起决定性影响的因素是碰撞形态、天气、能见度、车道类型及两轮车类型。而在事故低发区域,碰撞形态、道路等级、能见度、路面材料以及两轮车类型为主要影响因素。此外,路口路段类型、车道类型、两轮车类型以及肇事逃逸等因素在不同事故区域的影响不同。【结论】交通事故在城市内存在事故黑点,且部分道路环境因素在不同事故区域的影响不一致。研究成果可为交管部门针对事故黑点区域制定防范措施提供指导。

关键词:交通安全;事故黑点识别;事故严重程度;两轮车事故;时间序列聚类;CatBoost

中图分类号:U491

文献标志码:A

0 引言

道路交通伤害已成为中国人除病因死亡外的主要死因,同时还是全球5至29岁人群的首要死因^[1]。其中,道路环境因素是影响道路交通伤害的重要因素之一。在不同的道路环境下,交通事故发生的频率、导致的伤害程度往往会表现出不一样的结果。道路环境在客观上因土地利用特征、城市发展程度和地理地形地貌的不同而产生差异。因而在城市内会存在一些事故密集发生的区域,也被称为事故黑点,这是治理道路交通事故的一个切入点。

目前,研究人员针对事故黑点识别开展了许多研究。陆化普等^[2]针对是否考虑路网密度两种情况,使用密度分析鉴别事故高发区域,发现郊区单位长度道路上的交通事故更为多发。HU等^[3]

使用核密度分析与多项式Logit方法对比了行人事故黑点与人口密度、路网密度、社会服务活动等的分布关系,发现行人事故在城市空间呈现出若干个集群分布,并与人口密度、路网密度以及社会服务活动存在显著相关性。龙科军等^[4]使用Categorical Boosting (CatBoost)算法并融合事故数据、异常驾驶行为数据和路段特征数据识别高速公路事故黑点。WAN等^[5]基于灰色Verhulst-经验贝叶斯,提出了一种快速识别和分类城市道路事故黑点的方法。戢晓峰等^[6]集成移动步长,并结合累计频率法、正态分布模型构建了山区摩托车事故黑点鉴别模型,发现曲线路段的风险高于直线路段的,而有坡路段的风险高于平直路段的。以上研究从不同视角使用不同方法在空间上找到了特定区域内的交通事故黑点,但忽略了道路交通事故的时间属性。从历史数据来看,道路交通事故存在一定的周期性规律,同时事故黑点在较

收稿日期:2023-01-01;修回日期:2023-02-27;接受日期:2023-03-03

基金项目:国家自然科学基金资助项目(52172399、52175088、52211530054);湖南省教育厅科研重点项目(21A0193);长沙市自然科学基金项目(KQ2208235);国家重点研发计划项目(2019YFE0108000);中汽中心指南类项目(21243421)

通信作者:林森(1986—)(ORCID:0009-0004-3904-2360),男,高级工程师,主要从事交通事故分析与交通安全评价的研究。E-mail:linmiao@catarc.ac.cn

长周期内可能会随城市的建设而发生变化。因此,在寻找城市事故黑点时考虑交通事故的时间属性十分重要。本研究为解决以上黑点识别方法存在的不足,提出了一种考虑交通事故时间属性与空间属性的黑点识别方法。该方法基于 ArcGIS Pro 时空分析工具箱中的时间序列聚类,可以在一段时期内的一定事故区域,根据周期内事故发生

数量识别事故黑点区域。

本文的研究路线如图 1 所示。首先提取事故时间特征与空间特征,再通过 ArcGIS Pro 地理信息工具生成包含若干个时间序列的时空立方体,之后使用时间序列聚类识别事故黑点区域,最后使用融合了 CatBoost 与 Shapley Additive exPlanations (SHAP)两种方法的模型识别风险因素。

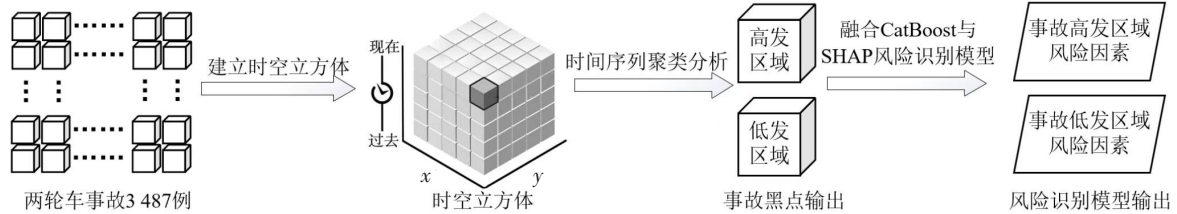


图1 研究路线流程图

Fig. 1 Research route flow chart

1 研究区域与数据源

本文的研究区域为长沙市,包含芙蓉区、天心区、岳麓区、开福区、雨花区和望城区,以及浏阳市、宁乡市和长沙县。研究所用的交通事故数据来源于长沙市公安局交通警察支队。该数据为2017年1月1日—2019年12月31日发生在研究区域的一般事故数据。本文基于以下规则对事故数据进行了筛选:①造成碰撞的车辆为两轮车,包括

摩托车、电动两轮车、电动自行车与自行车;②碰撞形式为行驶中的两轮车与汽车发生碰撞,静止状态下的碰撞被排除在外;③发生碰撞后至少一人受伤。

筛选后共获得了3 487个有效的两轮车碰撞事故数据。该数据包括事故描述、事故时间、道路设施、环境等60余字段。本文的研究重点为事故黑点区域道路环境因素对事故严重程度的影响,故主要保留了与事故道路环境相关的17个潜在风险因素,其字段描述见表1。

表1 长沙市2017—2019年两轮车事故数据统计与描述

Table 1 Statistical description of two-wheeler accident data in Changsha City from 2017 to 2019

影响因素及描述	高发区域占比/%	低发区域占比/%
受伤严重程度(因变量)(0非致命伤害/1致命伤害)	81.3/18.7	71.6/28.4
车道类型(1机动车道/2非机动车/3机非混合/4人行道/5其他)	79.0/4.1/13.3/1.6/2.0	70.3/6.2/19.9/1.1/2.5
中央隔离设施(1无/2绿化带/3混凝土/4护栏)	60.1/16.8/1.2/21.9	79.7/9.1/1.3/9.9
道路安全属性(1正常/2已治理但仍存在/3正在治理/4尚未治理)	89.1/8.0/2.3/0.6	90.9/6.0/2.5/0.6
交通信号方式(0无控制/1有控制)	22.6/77.4	28.9/71.1
道路物理隔离(1无隔离/2中心隔离/3机非隔离/4中心+机非隔离)	59.6/32.7/1.9/5.8	79.8/13.8/2.3/4.1
路面情况(1干燥/2潮湿/3其他)	79.4/19.0/1.6	82.7/16.2/1.1
路面材料(1沥青/2水泥/3其他)	87.5/12.0/0.5	76.4/20.6/3.0
路口路段类型(1三岔路口/2十字路口/3路段/4匝道或路段进出口)	6.8/25.1/66.0/2.1	9.0/11.8/76.7/2.5
道路线型(1平直/2弯道/3上下坡)	79.9/16.1/4.0	58.4/37.9/3.7
道路等级(1国道/2省道/3县乡道/4城市道路/5其他道路)	1.2/1.9/7.7/76.9/12.3	4.9/10.3/12.5/36.8/35.5
天气(1晴/2阴/3雨/4其他)	67.7/15.9/15.9/0.5	71.1/14.2/14.6/0.1
能见度(50 m以下/50~<100 m/100~<200 m/200 m及以上)	12.9/32.1/23.5/31.5	11.8/31.9/26.5/29.8

表1 (续)
Table 1 (Continued)

影响因素及描述	高发区域占比/%	低发区域占比/%
碰撞形态(1追尾/2正面碰撞/3同向刮撞/4对向刮撞/5侧面碰撞)	6.5/6.5/20.2/5.7/61.1	7.1/8.7/15.4/12.1/56.7
事故后是否逃逸(0否/1是)	97.3/2.7	97.5/2.5
照明条件(1白天/2夜间路灯开启/3夜间无路灯/4黄昏或黎明)	67.5/25.7/2.7/4.1	75.3/14.7/5.3/4.7
是否涉及卡车(0否/1是)	91.6/8.4	88.4/11.6
两轮车类型(1摩托车/2电动两轮车/3自行车)	42.9/50.4/6.7	56.4/38.6/5.0

2 时空特征分析

2.1 时间序列聚类原理

时间序列是将某种特定事件的某个统计指标在不同时间点上的数值,按时间的先后顺序排列的一组向量^[7]。本文基于事故的发生时间和事故地点的经纬度坐标,再借助 ArcGIS Pro 软件的时空模式挖掘工具,生成一个包含时间与空间特征的时空立方体,如图2所示。

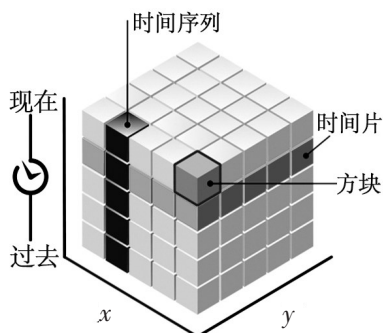


图2 时空立方体

Fig. 2 Space-time cube

每个方块表示在空间 (x, y) 和时间 (t) 上的特定位置,处于同一个方块的交通事故共用一个位置编号。在本研究中,方块的底面被设置为长度和宽度均为1.5 km的正方形区域,时间步长则为3个月。从时间维度来看,时空立方体由若干个时间序列(如图2中的黑色条柱所示)组成,时间序列之间的区别在于空间位置的不同。

时间序列聚类的目的在于将时空立方体的位置划分为多个组,其中组内的时间序列彼此之间的相似度较高,其相似性包括数值相似性与变化趋势相似性。本研究的主要目的是使用时间序列聚类找出整个周期内的事故黑点区域,即寻找数值聚集的区域,故采用数值相似性进行分析。在

本研究中,数值量即为一组时间序列下的事故发生数量。

对于给定的两个时间序列 $\Omega_m = (z_1^{(m)}, z_2^{(m)}, \dots, z_i^{(m)})$ 、 $\Omega_n = (z_1^{(n)}, z_2^{(n)}, \dots, z_i^{(n)})$,其数值相似性可被表示为:

$$\Delta = (z_1^{(n)} - z_1^{(m)})^2 + \dots + (z_i^{(n)} - z_i^{(m)})^2 \quad (1)$$

式中: $z_i^{(m)}$ 、 $z_i^{(n)}$ 分别表示第 m 、 n 个时间序列 Ω_m 、 Ω_n 在第 i 个时间步长内发生的两轮车事故数量。

2.2 时间序列聚类结果

为获得最佳的聚类簇数量,本文使用伪 F 统计量来衡量聚类的效果,如图3所示。伪 F 统计量可以直观地表示组内的相似度及组间的差异性。伪 F 统计量越大,则表明聚类效果越好,组间差异越大,组内相似度越高^[8]。

由图3可知,当聚类簇的数量为2时,伪 F 统计量最大,即该事故数据集最适合划分为两个区域。

时空立方体在长沙市范围内生成了897个有效区域,如图4所示。有效区域表示在过去的三年内该区域内至少发生一起两轮车交通事故。从地理分布上来看,两轮车事故沿路网自中心向外延伸,在市区的密度比在郊区内的低。时间序列聚类将897个有效区域划分为了两类,其中黑色区域共71块(占比7.92%),该区域内发生1 263起两轮车事故(占总事故量的36.22%),而条纹区域共826块(占比92.08%),该区域内发生2 224起两轮车事故(占总事故量的63.78%)。从事故总数上来看,黑色区域内的事故发生总量比条纹区域内的低,但其事故密度为条纹区域的6.61倍,故可以认为黑色区域为长沙市两轮车事故的高发区域,即交通事故黑点区域。

图5所示为不同区域内平均事故发生数量随

时间变化的曲线。由图5可知,事故发生数量随着时间的推移存在周期性变化,全年事故发生数量存在先增加后减少的趋势,事故的高发期为4月—6月。对比不同区域发现,事故高发区域在一个周期内事故发生数量基本大于1.00起,而在事故低发区域内一个周期的最高事故发生数量仅为0.31起。

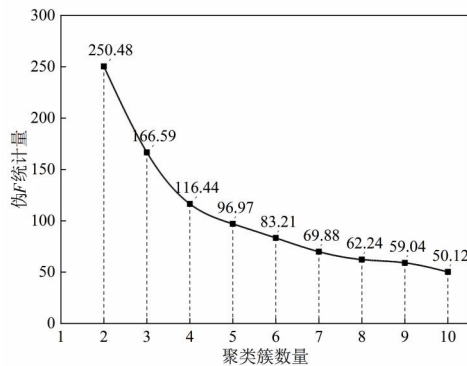


图3 伪F统计量与聚类簇数量关系图

Fig. 3 Relation between pseudo-F statistic and cluster number

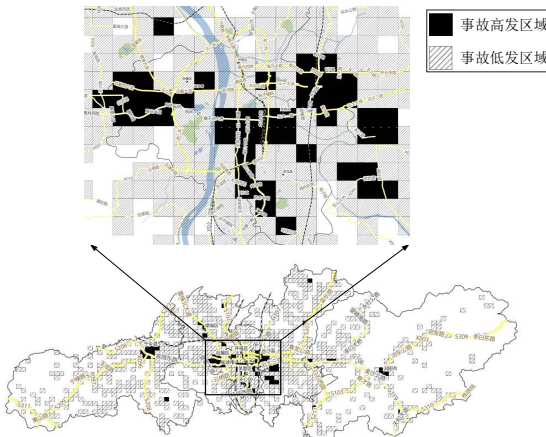


图4 时间序列聚类结果可视化

Fig. 4 Visualization of time series clustering results

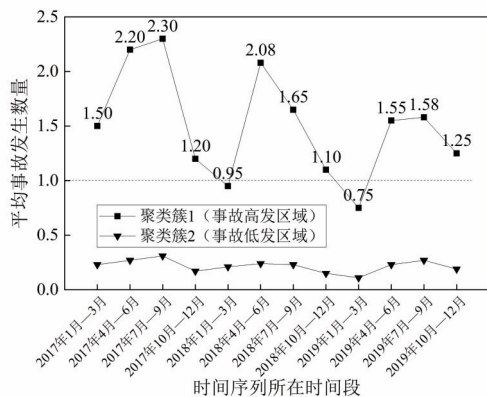


图5 不同聚类簇的平均事故发生数量

Fig. 5 Number of accidents in different clusters

3 不同事故区域致因分析

3.1 数据描述

为深入研究道路环境等因素对事故黑点区域及事故低发区域的影响,根据上节中的时间序列聚类结果,将3 487起两轮车事故划分为两类,其中事故黑点区域即事故高发区域发生1 263起,事故低发区域发生2 224起。选取的17个潜在道路环境风险因素的数据统计分析见表1。

3.2 研究方法

本文将CatBoost算法用于建立风险因素识别模型,SHAP方法用于对模型的输出进行解释。

3.2.1 Categorical Boosting (CatBoost)

CatBoost是由Yandex于2017年推出的一款开源、基于决策树的集成学习算法。针对传统目标统计方法中因目标泄露而产生的条件漂移问题,CatBoost提出了一种有序目标统计方法。该方法可以减小梯度偏差,克服预测偏移问题。此外,该算法直接处理类别特征,简化了数据处理的步骤^[9]。

假定有一组数据集 $D = \{(\mathbf{x}_k, y_k)\}_{k=1, \dots, n}$, 其中 $\mathbf{x}_k = (x_k^{(1)}, \dots, x_k^{(m)})$ 为一组包含 m 个特征变量的向量, y_k 为第 k 个样本的目标值。根据有序目标统计方法,类别特征 $x^{(i)}$ 的第 k 个训练样本的分类 $x_k^{(i)}$ 可以用等于某个目标统计 $\hat{x}_k^{(i)}$ 的数值型特征代替,其计算方法见式(2):

$$\hat{x}_k^{(i)} = \frac{\sum_{x_j \in D_i} \phi_{\{x_j^{(i)} = x_k^{(i)}\}} y_j + \alpha p}{\sum_{x_j \in D_i} \phi_{\{x_j^{(i)} = x_k^{(i)}\}} y_j + \alpha} \quad (2)$$

区别于贪婪目标统计,有序目标统计引入 σ 作为训练样本的随机序列。当计算训练集样本的目标统计时,令 $D_k = \{x_j: \sigma(j) < \sigma(k)\}$; 当计算测试集样本的目标统计时, $D_k = D$ 。对于 ϕ , 当 $x_j^{(i)} = x_k^{(i)}$ 时, $\phi = 1$; 当 $x_j^{(i)} \neq x_k^{(i)}$ 时, $\phi = 0$ 。此外, p 为增加的先验值,一般为数据集的平均目标值。 α 为权重系数,且为正数。利用先验值可以有效降低低频变量引入的噪声,避免过拟合现象。

3.2.2 Shapley Additive exPlanations (SHAP)

SHAP方法在2017年由Lundberg和Lee提出。它是一种事后归因解释方法,其核心在于计算特征的SHAP值,SHAP值反映了每个特征变量对预

测结果的贡献,这个贡献可能为正值,也可能为负值。其中正值代表该特征变量提升了预测值的结果,反之则降低了预测值的结果^[10]。

假定第*i*个样本为 x_i ,第*i*个样本的第*j*个特征变量为 x_{ij} ,模型训练后第*i*个样本的预测值为 y_i ,则单个样本中SHAP值的分布可以用式(3)表示:

$$y_i = y_{\text{base}} + f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ij}) \quad (3)$$

式中: y_{base} 为整个模型的基线,通常等于所有样本的预测均值; $f(x_{ij})$ 为第*i*个样本中第*j*个特征变量的SHAP值,其计算过程如式(4)所示:

$$f(x_{ij}) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left[f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S) \right] \quad (4)$$

式中: F 表示数据中包含的所有特征变量的集合; S 表示特征子集合; $S \subseteq F \setminus \{j\}$ 表示不包含特征*j*的所有子集; $f_{S \cup \{j\}}$ 表示通过特征子集*S*和特征*j*训练得到的模型输出;而 f_S 则表示使用特征子集*S*训练的模型输出; $x_{S \cup \{j\}}$ 表示特征子集*S*和特征*j*中的所有值; x_S 则表示特征子集*S*中的所有值。

3.3 结果讨论与分析

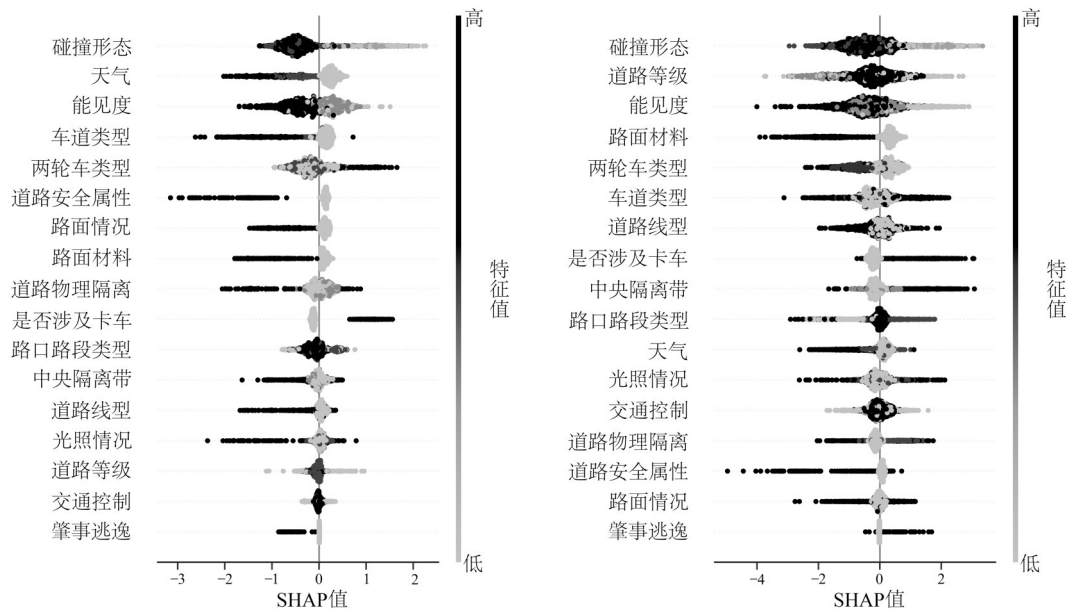
3.3.1 模型效果评估

本文使用准确率、精准率、召回率、F1分数和受试者工作特征(receiver operating characteristic,

ROC)曲线来评估模型的表现性能。对于ROC曲线,主要关注该曲线下的面积(area under the curve, AUC),即AUC的大小。AUC是一个聚合的度量,其值不受样本比例的影响,很适合用于评估二分类的效果。本研究在建模前将80.0%的数据用于训练,而剩余的20.0%用于测试。经测试集验证发现,由事故高发区域数据建立的模型1的准确率为91.8%,精准率为93.6%,召回率为94.9%,F1分数为0.942,AUC为0.959。由事故低发区域数据建立的模型2的准确率为93.9%,精准率为95.7%,召回率为95.1%,F1分数为0.954,AUC为0.976。综上可得,基于CatBoost建立的事故严重程度模型表现较好,可以接受。

3.3.2 风险因素重要性

SHAP方法被用于解释CatBoost模型的输出结果。为了更好地分析模型中风险因素对事故结果的具体影响,SHAP汇总图(图6)给出了根据风险因素归类的整体样本SHAP值分布。SHAP值汇总图很好地可视化了事故严重程度与潜在风险因素之间的总体关系。散点的颜色表示当前特征变量的取值大小,浅色代表低值,深色代表高值。纵坐标自上至下,依次表示风险因素对事故结果的贡献逐渐降低。横坐标为SHAP值,其大小代表特征对结果的影响程度,SHAP值为正代表该特征对致命事故的发生有正向作用,SHAP值为负代表该特征对致命事故的发生有负向作用。



(a) 事故高发区域SHAP汇总图

(b) 事故低发区域SHAP汇总图

图6 不同区域风险因素SHAP汇总图

Fig. 6 SHAP value summary plot in different areas

由图6可知,在不同事故区域,不同风险因素对事故的影响程度不同。在事故高发区域,起决定性影响因素的是碰撞形态、天气、能见度、车道类型和两轮车类型。而在事故低发区域,碰撞形态、道路等级、能见度、路面材料和两轮车类型为主要影响因素。虽然在不同事故区域内影响两轮车骑手受伤严重程度的主要影响因素存在一定差异,但部分因素对事故结果的影响趋势仍然一致。

对于碰撞形态,追尾和正面碰撞更容易使骑手受到致命伤害,这是由于正面碰撞与追尾碰撞的速度较快,碰撞动能大,从而对骑手的冲击过大。而与追尾碰撞和正面碰撞相比,侧面刮碰倾向于降低骑手受到致命伤害的可能^[11-13]。

在道路等级中,在高等级道路(国道、省道)发生的事故更容易造成骑手受到致命伤害,而城市道路与县乡道路倾向于降低骑手受到致命伤害的风险。这主要是因为高等级道路限速高,车辆的行驶速度较快,高速运动的车辆间碰撞后果更加严重^[14]。

与设置道路中央隔离带相比,无道路中央隔离带更容易造成致命事故,这主要是因为道路中央隔离带的设置限制了车辆超车与违规变道等行为,从而使事故率下降^[15]。

夜间无路灯与能见度低的道路环境会增加骑手受到致命伤害的可能。这是因为能见度越低或光照条件越差,人的视距范围越小,事故发生时骑手无法获得足够的时间和距离来避免事故,这与前人的研究一致^[16-18]。

涉及卡车的事例会显著增加骑手受到致命伤害的风险,这是因为卡车驾驶员的视野盲区大,在事故发生时无法及时采取措施,且卡车整备质量大、刹车延迟时间长,往往会造成严重的事故后果^[19]。

3.3.3 不同事故区域道路环境差异化影响

通过对比不同风险因素在两个模型中SHAP值的变化趋势,发现部分因素在两个模型中的影响不一致。为更好地对比其差异,下文使用SHAP的主效应图进行量化并分析其差异。

图7所示为路口路段类型影响因素下的SHAP主效应图。由图7可知,匝道或路段进出口的SHAP值在两个模型中均为负值,即降低了骑手受到致命伤害的可能。在事故高发区域,路口等因素对事故的影响无显著的趋向。而在事故低发

区域,三岔路口倾向于降低骑手受到致命伤害的风险,而十字路口会显著增加骑手受到致命伤害的风险。

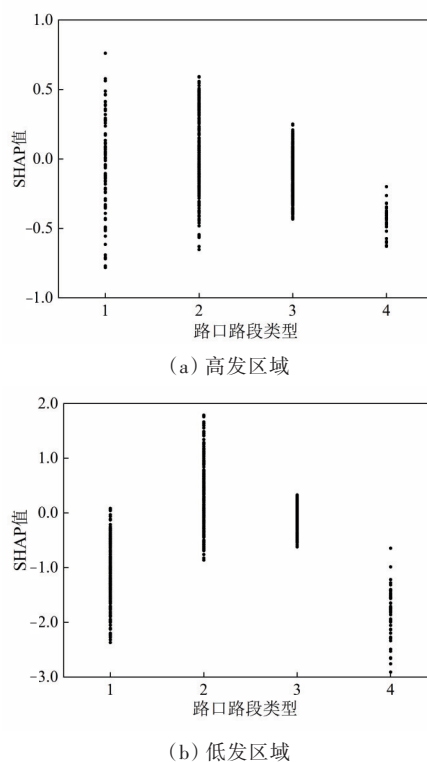


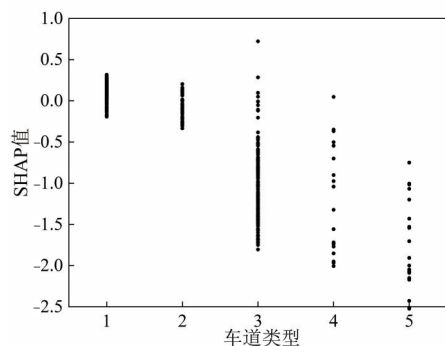
图7 路口路段类型SHAP主效应图

Fig. 7 SHAP main effect diagram of intersection type

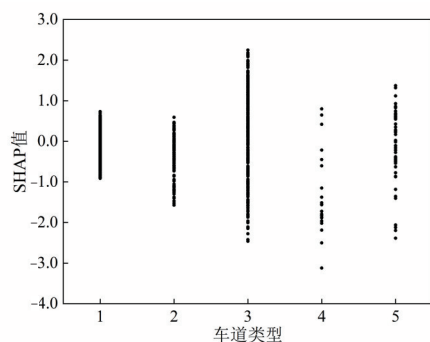
图8所示为车道类型影响因素下的SHAP主效应图。由图8可知,在事故高发区域,发生在机非混合车道的事故会降低骑手受到致命伤害的风险,但在事故低发区域,机非混合车道对事故结果无明显倾向影响。类似地,对于发生在机动车道的事故,在事故高发区域会增加骑手受到致命伤害的风险,而事故低发区域,该特征变量无明显倾向性影响。

图9所示为两轮车类型影响因素下的SHAP主效应图。由图9可知,两轮车类型在不同事故区域产生了相反的影响。对于事故高发区域,涉及摩托车与电动两轮车的事故会降低骑手受到致命伤害的风险,而在事故低发区域则相反。

图10所示为肇事逃逸影响因素下的SHAP主效应图。由图10可知,在以往的研究中发现肇事逃逸会加重事故严重程度,导致事故参与方遭受更严重的伤害^[20]。在事故低发区域的肇事逃逸对事故的影响与先前的研究结果一致。但在事故高发区域,肇事逃逸反而倾向于降低骑手受到致命伤害的风险。



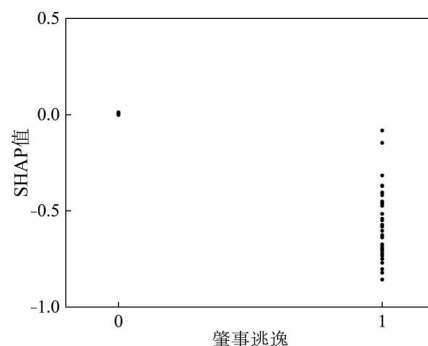
(a) 高发区域



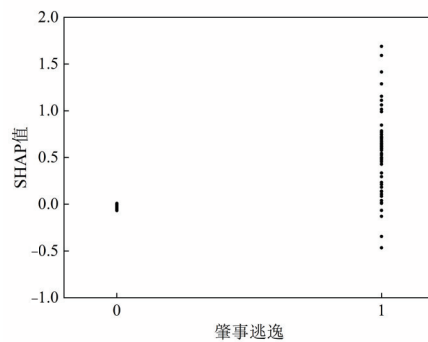
(b) 低发区域

图8 车道类型SHAP主效应图

Fig. 8 SHAP main effect diagram of lane type



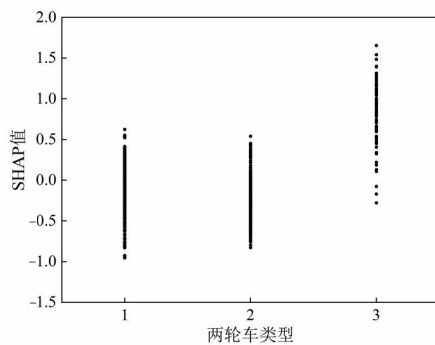
(a) 高发区域



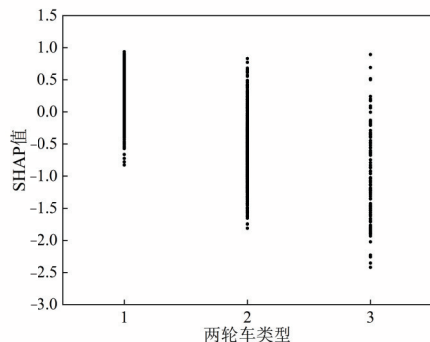
(b) 低发区域

图10 肇事逃逸SHAP主效应图

Fig. 10 SHAP main effect diagram of hit and run



(a) 高发区域



(b) 低发区域

图9 两轮车类型SHAP主效应图

Fig. 9 SHAP main effect diagram of two-wheeler type

4 结论

本文以长沙市2017年1月—2019年12月发生的3 487起两轮车事故为研究样本,以边长为1.5 km的正方形为单元区域,以3个月为时间步长,创建了基于时间序列聚类事故黑点识别模型。针对已识别到的事故黑点区域,使用CatBoost与SHAP方法研究了影响事故黑点区域与事故低发区域事故严重程度的环境道路因素。主要结论如下:

1) 基于时间序列聚类的长沙市两轮车事故黑点识别模型,得到了71个事故高发区域与826个事故低发区域。研究发现,事故高发区域的事故密度为事故低发区域的6.61倍。事故高发区域主要聚集在大型商业圈、客运车站与工业区附近,而事故低发区域多集中于远离市区的乡镇区域。

2) 在事故高发区域,起决定性影响因素的是碰撞形态、天气、能见度、车道类型以及两轮车类型。而在事故低发区域,碰撞形态、道路等级、能见度、路面材料和两轮车类型为主要影响因素。其中追尾、正面碰撞、高等级道路、低能见度、无道

路中央隔离带、夜间无路灯、涉及卡车等因素在两个事故区域内均对致命事故起到积极作用,增加了两轮车骑手在事故发生时受到致命伤害的风险。而路口路段类型、车道类型、两轮车类型与肇事逃逸等因素在不同事故区域存在不同的影响。

以上结论表明,对不同的事故区域而言,道路环境因素对事故结果的影响不同,具体表现在起主要作用的因素不同和不同因素对事故的表现不同。这揭示了道路环境因素差异对交通事故的影响,即由于土地利用特征、区域定位的不同,道路环境建设存在一定的差异,这一差异反过来影响了交通事故的发生。虽然事故高发区域的事故密度大,但是其致死率显著低于事故低发区域的,这是因为事故高发区域多位于道路建设完善的城市区域。未来如何降低事故致死率,可以从道路建设方面入手,建设设施齐全、规范的道路将有助于降低事故发生率及事故严重程度。

本文的研究仍存在部分局限,即空间区域较大而时间跨度较小,在以后的研究中,将更加合理地划分事故区域的大小,并纳入更长周期、更广泛的事数据,建立更加精准的时空分析模型。

参考文献

- [1] World Health Organization. Global status report on road safety 2018: Summary [EB/OL]. (2018-12-24)[2023-01-01]. <https://www.who.int/publications/i/item/WHO-NMH-NVI-18.20>.
- [2] 陆化普, 罗圣西, 李瑞敏. 基于GIS分析的深圳市道路交通事故空间分布特征研究[J]. 中国公路学报, 2019, 32(8): 156-164. DOI: 10.19721/j.cnki.1001-7372.2019.08.014.
LU Huapu, LUO Shengxi, LI Ruimin. GIS-based spatial patterns analysis of urban road traffic crashes in Shenzhen[J]. China Journal of Highway and Transport, 2019, 32(8): 156-164. DOI: 10.19721/j.cnki.1001-7372.2019.08.014.
- [3] HU L, WU X H, HUANG J, et al. Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China [J]. Safety Science, 2020, 127: 104710. DOI: 10.1016/j.ssci.2020.104710.
- [4] 龙科军, 邹道兴, 何石坚. 基于多源数据融合的高速公路交通事故黑点判别[J]. 长沙理工大学学报(自然科学版), 2022, 19(1): 89-95. DOI: 10.19951/j.cnki.1672-9331.2022.01.011.
LONG Kejun, ZOU Daoxing, HE Shijian. Identification of black spots in highway traffic accidents based on multi-source data fusion [J]. Journal of Changsha University of Science & Technology (Natural Science), 2022, 19(1): 89-95. DOI: 10.19951/j.cnki.1672-9331.2022.01.011.
- [5] WAN Y, HE W Q, ZHOU J B. Urban road accident black spot identification and classification approach: a novel grey verhuls-empirical Bayesian combination method [J]. Sustainability, 2021, 13(20): 11198. DOI: 10.3390/su132011198.
- [6] 戢晓峰, 邓宇泽, 杨文臣, 等. 山区双车道公路摩托车事故黑点及成因集成分析[J]. 昆明理工大学学报(自然科学版), 2021, 46(2): 116-125. DOI: 10.16112/j.cnki.53-1223/n.2021.02.481.
JI Xiaofeng, DENG Yuze, YANG Wenchen, et al. Black spot and cause integrated analysis of motorcycle accidents on two-lane highways in mountainous areas [J]. Journal of Kunming University of Science and Technology (Natural Sciences), 2021, 46(2): 116-125. DOI: 10.16112/j.cnki.53-1223/n.2021.02.481.
- [7] 刘颖, 李惠迪, 谭博元. 基于多源数据深度融合的金融时间序列预测[J]. 统计与决策, 2022, 38(23): 52-56. DOI: 10.13546/j.cnki.tjyjc.2022.23.01.
LIU Ying, LI Huidi, TAN Boyuan. Financial time series prediction based on deep fusion of multi-source data [J]. Statistics & Decision, 2022, 38(23): 52-56. DOI: 10.13546/j.cnki.tjyjc.2022.23.01.
- [8] 刘文博, 梁盛楠, 余泉, 等. 基于伪F统计量的属性特征降维方法研究[J]. 东北师大学报(自然科学版), 2020, 52(1): 43-49. DOI: 10.16163/j.cnki.22-1123/n.2020.01.010.
LIU Wenbo, LIANG Shengnan, YU Quan, et al. Attribute feature dimensionality reduction method based on pseudo F statistic [J]. Journal of Northeast Normal University (Natural Science Edition), 2020, 52(1): 43-49. DOI: 10.16163/j.cnki.22-1123/n.2020.01.010.
- [9] CHEN S M, SHAO H P, JI X M. Insights into factors affecting traffic accident severity of novice and experienced drivers: a machine learning approach [J]. International Journal of Environmental Research and Public Health, 2021, 18(23): 12725. DOI: 10.3390/ijerph182312725.
- [10] PARSA A B, MOVAHEDI A, TAGHIPOUR H, et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis [J]. Accident Analysis & Prevention, 2020, 136: 105405. DOI: 10.1016/j.aap.2019.105405.
- [11] LELEDAKIS A, LINDMAN M, ÖSTH J, et al. A method for predicting crash configurations using counterfactual simulations and real-world data [J]. Accident Analysis & Prevention, 2021, 150: 105932.

- DOI: 10.1016/j.aap.2020.105932.
- [12] HU L, LI H B, YI P, et al. Investigation on AEB key parameters for improving car to two-wheeler collision safety using in-depth traffic accident data [J]. IEEE Transactions on Vehicular Technology, 2023, 72(1): 113–124. DOI: 10.1109/TVT.2022.3199969.
- [13] 薛海涛, 李海波, 赵小羽, 等. 汽车-电动自行车与汽车-自行车碰撞中骑车人动力学响应对比研究[J]. 公路与汽运, 2021(6): 37–42. DOI: 10.3969/j.issn.1671-2668.2021.06.010.
- XUE Haitao, LI Haibo, ZHAO Xiaoyu, et al. Comparative study on dynamic response of cyclists in car-electric bicycle and car-bicycle collision [J]. Highways & Automotive Applications, 2021(6): 37–42. DOI: 10.3969/j.issn.1671-2668.2021.06.010.
- [14] AMARASINGHA N, DISSANAYAKE S. Gender differences of young drivers on injury severity outcome of highway crashes [J]. Journal of Safety Research, 2014, 49: 113–120. DOI: 10.1016/j.jsr.2014.03.004.
- [15] 金昊男, 李自豪, 赵亮. 基于GIS的章丘区公路交通事故风险诱因分析[J]. 山东交通学院学报, 2022, 30(3): 30–39. DOI: 10.3969/j.issn.1672-0032.2022.03.005.
- JIN Haonan, LI Zihao, ZHAO Liang. The risk inducement analysis of road traffic accidents in Zhangqiu District based on GIS [J]. Journal of Shandong Jiaotong University, 2022, 30(3): 30–39. DOI: 10.3969/j.issn.1672-0032.2022.03.005.
- [16] CASADO-SANZ N, GUIRAO B, LARA GALERA A, et al. Investigating the risk factors associated with the severity of the pedestrians injured on Spanish crosstown roads [J]. Sustainability, 2019, 11(19): 5194. DOI: 10.3390/su11195194.
- [17] RAHMAN M S, ABDEL-ATY M, HASAN S, et al. Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones [J]. Journal of Safety Research, 2019, 70: 275–288. DOI: 10.1016/j.jsr.2019.04.008.
- [18] HU L, WU X T, HU X T, et al. Injury severity analysis of electric bike crashes in Changsha, Hunan Province: taking different lighting conditions into consideration [J]. Transportation Safety and Environment, 2022, 4(3): tdac011. DOI: 10.1093/tse/tdac011.
- [19] 魏利华. 卡车右转弯-电动自行车碰撞事故再现及动力学响应研究[D]. 成都: 西南交通大学, 2020. DOI: 10.27414/d.cnki.gxnju.2020.000809.
- WEI Lihua. Study on accident reconstruction and dynamic response of the collision between right turn of truck and electric bicycle [D]. Chengdu: Southwest Jiaotong University, 2020. DOI: 10.27414/d.cnki.gxnju.2020.000809.
- [20] 柏伟, 范英飞. 逃逸事故伤害严重程度影响因素分析[J]. 综合运输, 2019, 41(5): 24–28.
- BAI Wei, FAN Yingfei. Analysis of influencing factors on injury severity in hit-and-run [J]. China Transportation Review, 2019, 41(5): 24–28.

Black spot identification and analysis of traffic accidents based on time series clustering

LIN Nanting^{1,2}, HU Lin^{1,2}, LIN Miao³, PENG Hua⁴

(1. School of Automotive and Mechanical Engineering, Changsha University of Science & Technology, Changsha 410114, China; 2. Hunan Province Key Laboratory of Safety Design and Reliability Technology for Engineering Vehicle, Changsha University of Science & Technology, Changsha 410114, China; 3. China Automotive Technology and Research Center Co., Ltd., Tianjin 300300, China; 4. Traffic Police Detachment of Changsha Public Security Bureau, Changsha 410006, China)

Abstract: [Purposes] Identify the black spots of road traffic accidents and explore the causes of the accidents in black spots area. [Methods] Combining geographic information software and interpretable machine learning algorithm, this paper proposes a method of black spot identification and analysis considering temporal and spatial attributes of traffic accidents. [Findings] The accident density in the area with high accident incidence was mainly concentrated in the vicinity of large commercial area, passenger station and industrial area. The accident density in the area with high accident incidence was 6.61 times that in the area with low accident incidence. In areas with high accident rates, collision patterns, weather, visibility, lane type and two-wheeler type play a decisive role. In the low accident area, collision form, road class, visibility, road material and two-

wheel vehicle type are the main influencing factors. In addition, factors such as intersection section type, lane type, two-wheeled vehicle type and hit-and-run have different impacts in different accident areas. **[Conclusions]** There are black spots in urban traffic accidents, and the influence of some road environmental factors is inconsistent in different accident areas. The results can provide guidance for the traffic management department to formulate preventive measures for the black spot area.

Key words: traffic safety; identification of accident black spot; accident severity; two-wheeler accident; time series clustering; CatBoost

Manuscript received: 2023-01-01; **revised:** 2023-02-27; **accepted:** 2023-03-03

Foundation item: Projects (52172399, 52175088, 52211530054) supported by the National Natural Science Foundation of China; Project (21A0193) supported by the Research Foundation of Education Bureau of Hunan Province; Project (KQ2208235) supported by Natural Science Foundation of Changsha; Project (2019YFE0108000) supported by the National Key Research and Development Program of China; Project (21243421) supported by Guide Project of China Automotive Technology and Research Center Co., Ltd.

Corresponding author: LIN Miao (1986—) (ORCID: 0009-0004-3904-2360), male, senior engineer, research interest: traffic accident analysis and comprehensive evaluation of traffic safety. E-mail: linmiao@catarc.ac.cn

(责任编辑:刘平;校对:石月珍;英文编辑:刘至真)