山东科学
SHANDONG SCIENCE
第 39 卷 第 1 期 2026 年 2 月出版
Vol.39 No.1 Feb.2026
DOI:10.3976/j.issn.1002-4026.2025034
【交通运输】

考虑非均衡性的城市自行车事故骑行者伤害程度影响因素及异质性分析

王朝健[1,2],徐小金[1],冯斌[1],余松霖[1],张卫东[1]

(1.四川三河职业学院 工程技术学院,四川 泸州 646200;2.泸州市智能机电控制工程技术研究中心,四川 泸州 646200)

摘要:为探究城市自行车事故骑行者伤害程度的影响因素,同时降低数据异质性和非均衡性对因素量化的影响。基于 CRSS 数据库的 3 895 起自行车事故,提出了一种融合重采样、潜在类别分析(LCA)和贝叶斯网络(BN)的方法。首先,采用 LCA 将事故数据重新划分为若干组具有组内同质性和组间异质性的子事故群,减少数据异质性的影响;其次,采用随机过采样(ROS)、合成少数类过采样技术(SMOTE)和自适应合成过采样算法(ADASYN)对各事故群重采样,减少数据非均衡性的影响;最后,基于各类重采样后的事故群,分别搭配 2 种 BN 结构学习算法和 1 种参数学习算法,并依据 AUC 值评选每类事故群的最优 BN 模型,实现骑行者伤害程度影响因素的定量分析和异质性分析。研究结果表明:当整体事故数据被划分为 3 类同质子数据群时,LCA 模型的 Entropy 值较优,达 0.943。其中 C1 事故群、C2 事故群、C3 事故群和 OD 事故群分别被挖掘出 10、13、9 和 12 个影响骑行者伤害程度的关键因素;将 LCA 和重采样引入 BN,能显著提升 BN 模型的 G-mean 值、AUC 值和风险因素挖掘能力;时间段、骑行者性别、骑行者年龄和天气状况等因素在不同事故群中存在明显的异质性。

关键词:交通安全;自行车事故;伤害程度;潜在类别分析;Bayes 网络

中图分类号:U491.31　　**文献标志码**:A　　**文章编号**:1002-4026(2026)01-0088-12

开放科学(资源服务)标志码(OSID):

Analysis of factors influencing cyclist injury severity and heterogeneity analysis in urban bicycle accidents considering data imbalance

WANG Chaojian[1,2],XU Xiaojin[1],FENG Bin[1],YU Songlin[1],ZHANG Weidong[1]

(1. School of Engineering and Technology, Sichuan Sanhe College of Professionals, Luzhou 646200, China;
2. Luzhou City Research Center for Intelligent Electromechanical Control Engineering Technology, Luzhou 646200, China)

Abstract: To explore the factors influencing the injury severity of cyclists in urban bicycle accidents and mitigate the impact of data heterogeneity and imbalance on the quantification of these factors, this study proposes a method integrating resampling, latent class analysis (LCA), and Bayesian networks (BNs) based on 3 895 bicycle accidents from the CRSS database. First, LCA was used to reclassify accident data into several sub-accident clusters with intra-cluster homogeneity and inter-cluster heterogeneity to reduce the impact of data heterogeneity. Second, random over-sampling (ROS),

收稿日期:2025-04-03　**修回日期**:2025-04-30
基金项目:泸州市智能机电控制工程技术研究中心项目(ZNJDKT25-09);泸州市科技局项目(2024RCM238)
作者简介:王朝健(1997—),硕士生,助教,研究方向为交通安全。E-mail:549786670@qq.com,Tel:15008322602

synthetic minority oversampling technique, and adaptive synthetic sampling approach were used to resample each accident cluster to reduce the impact of data imbalance. Finally, based on various resampled accident clusters, two BN structure learning algorithms and one parameter learning algorithm were applied and the optimal BN model for each accident cluster was selected based on AUC values to enable quantitative and heterogeneity analyses of factors influencing the injury severity of cyclists. Results show that when the overall accident data were divided into three homogeneous sub-clusters, the LCA model achieved an increased entropy value of 0.943. For the C1, C2, C3, and OD accident clusters, 10, 13, 9, and 12 key factors influencing the injury severity of cyclists were identified, respectively. The introduction of LCA and resampling into the BN considerably improved the BN model's G-mean value, AUC value, and risk factor identification capabilities. Factors such as time period, cyclist's gender, cyclist's age, and weather conditions showed substantial heterogeneity across different accident clusters.

Key words : traffic safety; bicycle accidents; injury severity; latent class analysis; Bayesian networks

自行车出行低碳环保,具有良好的社会经济效益和个人健康效益。但骑行者作为弱势道路交通使用者易在碰撞中遭受致命伤害[1]。同时,随着自行车在城市道路的占有率和使用率的提升[2]、城市道路交通复杂性的加剧[3],导致自行车事故风险凸显。中国城市道路的里程仅占全国 7.5%,但城市交通事故数量却占交通事故总量的 45.8%[4]。因此,探究城市道路自行车事故骑行者伤害程度的影响因素,对减少城市道路自行车交通事故以及降低骑行者伤害程度具有重要意义。

现有研究通常采用具有内在假设性的 Logit、Probit 等统计学模型[5],难以有效处理非线性问题和充分挖掘风险因素[6]。因此,部分研究采用决策树[7]、随机森林[8]和贝叶斯网络(BN)[9]等机器学习模型。尽管机器学习模型的预先假设少、灵活性高且在处理非线性问题方面表现出色,但其固有的"黑箱"特性导致模型可解释性差。此外,事故数据中事故类别是非均衡的,这与机器学习模型隐含的类别均衡假设相悖。因此,基于此类非均衡数据构建的机器学习模型会导致模型的参数估计偏差、有效性降低,且难以揭示风险因素与伤害程度之间的真实关系。

为解决非均衡问题,目前主要有算法级和数据级两种方法。算法级方法通过定义"代价矩阵"为不同分类赋予错分类代价,但确定最佳错分类代价较为困难且依赖经验[10]。因此,多数研究人员会采用数据级方法,即数据重采样。近年来,国内外学者通过将多种重采样技术与 XGboost[11]、BN[12]等机器学习模型结合,探究了事故类别的风险因素。结果表明,重采样技术能显著提升模型的综合性能和风险因素的识别能力。

在解决数据非均衡性问题的同时,数据异质性问题也同等重要。数据异质性是指:纳入模型的风险因素对伤害程度的影响呈现随机变化性。若在建模过程中忽略数据异质性,也会导致模型参数估计错误[13]。因此,能解释异质性的随机参数 Logit 被引入事故分析[14],但该方法会预先假设风险因素服从正态分布、对数正态分布等特定分布,忽略了风险因素遵循多态分布的情况。为克服这一缺点,基于聚类思想的潜在类别分析(latent class analysis,LCA)被引入事故分析。LIN 和 FAN[15]采用 LCA 将整体事故数据集划分为 7 份子数据集,随后采用偏比例优势模型分别对各数据集建模,发现子模型具有更高的拟合优度,且同一风险因素在不同组别中具有不同的影响效应。

综上所述,当前研究人员在探究骑行者伤害程度影响因素时,较少同时考虑数据异质性和非均衡性问题,易导致模型参数估计错误。此外,研究方法多采用 Logit 等统计学模型,其对自变量非共线性的要求,导致难以充分挖掘风险因素。因此,文章提出一种融合 LCA、重采样和 BN 的方法,旨在降低数据异质性和非均衡性对分析结果的影响,实现事故风险因素的量化分析。

1　数据来源及变量选择

1.1　数据来源

鉴于国内交通事故深度调查数据体系起步晚、数据共享机制不完善[16]。文章数据采用美国 2016—2020 年

的 CRSS 数据库[17]。虽国内外在交通环境和法规方面存在差异,可能导致研究结论并不完全适用于国内。但国内数据获取困难,且研究表明国内外在伤害程度的影响因素中有相似性[18],故采用国外数据仍合理。

研究对象为城市道路自行车事故,需进行样本筛选:①鉴于自行车单车事故样本少,且自行车-机动车事故(bicycle-motor vehicle crashes,BMV)是主要致死类型,为此仅筛选 BMV 事故。同时,为确保驾驶员与骑行者对应,降低其他潜在风险因素的干扰,仅筛选涉及一辆机动车(未计摩托车)和一辆脚踏自行车的碰撞事故。②剔除机动车自燃、驾驶员跳车、货物掉落等特殊事故。最终筛选出 3 895 起事故,其中骑行者未受伤 55 例、可能受伤 1 505 例、轻微受伤 1 865 例、严重受伤 428 例、死亡 42 例。由于骑行者伤害程度分类过多会降低模型的有效性并加剧复杂性,参照相关研究[19],将未受伤、可能受伤和轻微受伤合并为非严重伤害,其余事故类别合并为严重伤害。

1.2 变量选择

交通事故通常由人、车、路及环境等因素的配合失调引发[19]。人的方面包括年龄、性别、精神状态等心理和生理因素;车的方面包括车辆类型和车辆安全状况;道路的方面涉及路段类型、道路等级和路面状况等因素;环境通过影响人、车、路等方面而间接影响事故的发生,如照明条件、天气状况等。

结合以往研究[19-21]和 CRSS 数据库特征进行变量选取和离散。同时,为确保变量具有统计学意义,剔除缺失值比例超过 30%和单一取值事故占比超过 95%的变量[22],最终筛选出涵盖时间特征、驾驶员特征、骑行者特征、车辆特征、道路特征和环境特征等 6 方面的 24 个变量,具体变量名称及取值见表 1。

表 1 变量名称及其取值

Table 1 Names of variables and their values

特征分类	变量名称	变量取值				
		1	2	3	4	5
时间特征	季节	春季	夏季	秋季	冬季	
	事故日期	工作日	周末			
	时间段	早高峰	晚高峰	非高峰		
驾驶员特征	性别	男	女			
	年龄	18~25 岁	26~40 岁	41~65 岁	>65 岁	
	未按规定让行	是	否			
	视野被遮挡	是	否			
	分心驾驶	是	否			
骑行者特征	性别	男	女			
	年龄	<18 岁	18~25 岁	26~40 岁	41~65 岁	>65 岁
	初始行驶方向	顺向	逆向	横向		
	未按规定让行	是	否			
	违反交通控制装置	是	否			
	分心骑行	是	否			
车辆特征	车辆行驶状态	直行	左转	右转	起步/停车	其他
	车辆类型	乘用车	多用途车	卡车/公交车		
	碰撞位置	汽车正面	汽车左侧	汽车右侧		
	行驶年限	0~5 年	6~10 年	11~15 年	>15 年	

表1(续)

特征分类	变量名称	变量取值				
		1	2	3	4	5
道路特征	非道路/专用支路	干燥	潮湿			
	路段类型	交叉口	交叉口附近	非交叉口		
	道路横断面类型	车行道	自行车道/路肩	人行道/专用支路		
环境特征	天气状况	晴天	阴天	雨/雪等恶劣天气		
	照明条件	日间	夜间无照明	夜间有照明	黎明/黄昏	
	交通控制装置	无控制	交通信号灯	停车/让行标志		

2　方法

文章提出一种融合 LCA、重采样和 BN 的方法,主要有 4 个步骤。

步骤 1,数据聚类。采用 LCA 将事故数据重新划分为若干组具有组内同质性和组间异质性的子事故群,从而降低数据异质性对结果分析的影响。

步骤 2,数据重采样。采用 ROS(随机重采样)、SMOTE(合成少数重采样技术)和 ADASYN(自适应合成采样)方法对各类子事故群和原始事故群重采样,降低数据非均衡性的影响(每类事故群经重采样处理后,将分别获得 3 组均衡数据以及 1 组原始非均衡数据)。

步骤 3,构建 BN 模型。针对每一类事故群,基于 3 组采样后的均衡数据和 1 组未采样的原始非均衡数据,分别与 2 种 BN 结构学习算法和 1 种参数学习算法进行组合搭配,进而为每类事故群构建出 8 个 BN 模型。

步骤 4,BN 模型评价与筛选。由于 AUC 值能反映 BN 模型的综合性能,即使数据处于非均衡状态也能做出合理评价,因此将各事故群中 AUC 值最高的 BN 模型视为最优 BN 模型,实现骑行者伤害程度影响因素的定量分析及异质性分析。

2.1　潜在类别分析 LCA

LCA 是一种基于概率模型的聚类分析方法,优点在于无需外显变量满足正交性和标准正态分布,且可根据拟合优度指标确定最佳聚类数[23]。LCA 有外显变量和潜在变量。假设事故数据集存在 C 个潜在类,γ_c 表示事故属于潜在类别 c($c=1,2,\cdots,C$)的概率,总和为 1。每起事故 i 包含 M 个属性,即 M 个外显变量,Z_{im} 表示第 i 起事故的第 m 个外显变量(分类变量)的水平数,$Z_{im}=1,2,\cdots,r_m$。条件概率 $\rho_{m,r_m|c}$ 表示在潜在类 c 中,第 m 个变量水平为 r_m 的概率,且每个外显变量各水平的条件概率总和为 1。则第 i 起事故在所有潜在类别集群下某个可能的概率为[24]:

$$P(Z=z)=\sum_{c=1}^{C}\gamma_c\prod_{m=1}^{M}\prod_{r_m=1}^{R_m}\rho_{m,r_m|c}^{I(Z_{im}=r_m)}, \tag{1}$$

其中,R_m 为第 m 个外显变量的总水平数;$I(Z_{im}=r_m)$ 为指示函数,当 $Z_{im}=r_m$ 时,$I=1$;否则 $I=0$。

2.2　重采样

重采样常用于处理数据非均衡问题,可分为欠采样和过采样。过采样通过对少数类样本进行复制或合成,使得少数类样本与多数类样本达到“均衡”状态 ,具有在不造成信息丢失的情况下提高模型性能的优点,是更常用的重采样方法[25],其常用代表方法有 ROS 采样、ADASYN 采样、SMOTE 采样[26]。DELEN D 等[27]指出处理数据非均衡问题没有最佳方法。为最大程度保证模型的有效性,文章选用以上 3 种简单且效果显著的过采样算法[25]。

ROS 采样通过随机抽取并复制少数类样本,解决了欠采样导致信息丢失的问题。ADASYN 采样根据少数类样本的学习难度为其赋予不同权重,具体原理详见 HE 等[28]的研究。SMOTE 采样结合最邻近算法和插

值法生成新的少数类样本,缓解了 ROS 采样带来的模型过拟合的问题,具体原理详见 CHAWLA 等[29]的研究。

2.3 贝叶斯网络 BN

BN 由代表随机变量的节点和描述随机变量依赖关系的有向边构成。BN 模型的构建包括结构学习和参数学习。结构学习从观测数据中学习最优网络拓扑,参数学习则量化变量间的依赖关系。文章为构建最优 BN 模型,结构学习采用运算高效且常用的爬山算法(hill climbing, HC)和三阶段依赖分析算法(three phase dependency analysis, TPDA),评分函数采用性能优异且无需遵循狄利克雷先验分布的 BDeu 评分[30]。

3 模型建立

3.1 数据聚类和重采样

为减少数据异质性的影响,采用 LCA 将整体数据集划分为组间异质性最大化的不同类群。为确定最佳聚类数,将聚类数从 1 类逐渐增加到 10 类,观察每类模型的 AIC、BIC、CAIC 和 Entropy 值,不同聚类数对应的各项指标见图 1。

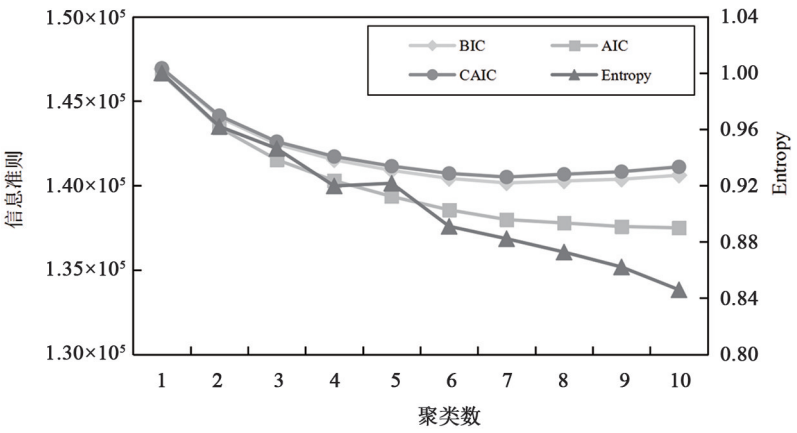

图 1 不同聚类数对应的信息准则和 Entropy 值

Fig.1 Information criterion and entropy values corresponding to different numbers of clusters

由图 1 可知,随着聚类数的增加,BIC、AIC、CAIC 值呈逐步下降趋势。若 IC 值下降率低于 1%,且 Entropy 值高于 0.9,则该聚类数被认为是最佳聚类数[24]。经计算,从聚类数为 4 开始,各 IC 值的下降率均低于 1%,且聚类数为 3 的 Entropy 值达 0.946 3,因此最佳聚类数确定为 3。原始事故群(OD 事故群)经划分后,C1 事故群占事故总量的 59.59%、C2 事故群占 29.46%、C3 事故群占 10.95%。同时,为减少数据非均衡的影响,采用 ROS、SMOTE 和 ADASYN 3 种方法分别对 OD 事故群、C1 事故群、C2 事故群和 C3 事故群进行重采样(采样结果见 OSID 开放科学数据与内容附件 1)。

3.2 BMV 模型的构建结果

依据第 2 节研究步骤,评选出 C1、C2、C3、OD 事故群的最优 BN 模型,分别命名为 BMV_C1 模型、BMV_C2 模型、BMV_C3 模型、BMV_OD 模型(BN 图见 OSID 开放科学数据与内容附件 2~5)。表 2 为各事故群最优 BN 模型与未经重采样 BN 模型的性能指标对比。

表 2 各事故群最优 BN 模型与未经重采样 BN 模型的性能指标对比

Table 2 Comparison of performance metrics of the optimal BN models for each accident cluster and the BN models built without resampling

BN 模型	Accuracy	Sensitivity	Specificity	AUC
C1_HC	0.907	1.000	0.000	0.575
C1_ROS_HC(C1 事故群最优)	0.892	0.841	0.943	0.960
C2_TPDA	0.801	0.999	0.000	0.645
C2_SMOTE_TPDA(C2 事故群最优)	0.731	0.707	0.756	0.811
C3_HC	0.937	1.000	0.000	0.425
C3_ROS_HC(C3 事故群最优)	0.963	0.925	1.000	0.995
OD_TPDA	0.879	1.000	0.000	0.617
OD_SMOTE_TPDA(OD 事故群最优)	0.743	0.730	0.755	0.817

注:模型名称格式为“A_B_C”,其中“A”代表模型所属事故群;“B”代表采用的重采样算法;“C”代表 BN 的结构学习算法。例如“C1_ROS_HC”表示出自 C1 事故群,其中 采样算法为 ROS,BN 的结构学习算法为 HC。

表 2 可知,基于均衡数据构建的 BN 模型相较于基于非均衡数据构建的 BN 模型,在 Accuracy 和 Sensitivity 上未显著提升,但 Specificity 和 AUC 值方面显著提高。这是因为机器学习模型通常假设类别均衡,基于非均衡数据构建的模型,其预测结果会偏向多数类,导致少数类精度欠佳,甚至模型无效。例如 C1_HC模型,Accuracy 和 Sensitivity 均高于 0.9,但 Specificity 为 0,即少数类精度为 0。这表明仅用 Accuracy 评估模型有效性是不够的,还需综合考虑模型的综合性能。

4　推理分析

4.1　BMV 模型的关键因素

为实现骑行者伤害程度风险因素的定量分析,应重点关注与骑行者伤害程度具有直接依赖关系的因素,并将其作为影响骑行者伤害程度的关键因素[12]。表 3 是根据文献[12]的方法识别出的影响骑行者伤害程度的因素。

表 3　每类事故群的关键因素

Table 3　Key factors for each accident cluster

因素		事故群			
		OD 事故群	C1 事故群	C2 事故群	C3 事故群
时间特征	季节				√
	事故日期		√		
	时间段		√		√
驾驶员特征	性别	√		√	√
	年龄	√		√	
	未按规定让行				√
	视野被遮挡	√		√	
骑行者特征	性别	√	√	√	
	年龄		√	√	√
	未按规定让行	√	√	√	
	违反交通控制装置		√		
	分心骑行				√
车辆特征	车辆行驶状态	√	√	√	
	车辆类型	√		√	√
	碰撞位置	√		√	
	行驶年限				√
道路特征	路面状况	√		√	
	路段类型	√			
	道路横断面类型			√	
环境特征	天气状况	√	√	√	√
	照明条件		√	√	
	交通控制装置	√	√		

注:“√”代表该因素是某事故群的关键因素。

表 3 可知,C1、C2、C3、OD 事故群中分别发现 10、13、9、12 个关键因素,且部分因素已被其他研究证实[5,19]。同时,经模型对比发现在同一 BN 算法下,基于均衡数据集构建的 BN 模型能发现更多影响伤害程度的关键因素(关键因素的对比结果见 OSID 开放科学数据与内容附件 7~10)。这说明非均衡数据会掩盖影响因素与伤害

程度的真实关系,导致误导性结论,因此事故分析建模时必须采用适当的重采样技术以减少此影响。

4.2 基于 BMV 模型的关键因素分析

鉴于严重伤害事故的误判代价远高于非严重伤害事故的误判代价,事故治理时需重点关注严重伤害事故。为探究关键因素的哪些取值会显著影响骑行者受严重伤害的概率(以下简称“严重伤害率”),基于不同事故群的最优 BN 模型,将每个关键因素的各取值分别作为“证据”,观察每个证据下伤害程度节点的严重伤害率。图 2 至图 5 分别展示了 BMV_C1、BMV_C2、BMV_C3、BMV_OD 模型中严重伤害率高于非严重伤害率的所有关键因素的取值。

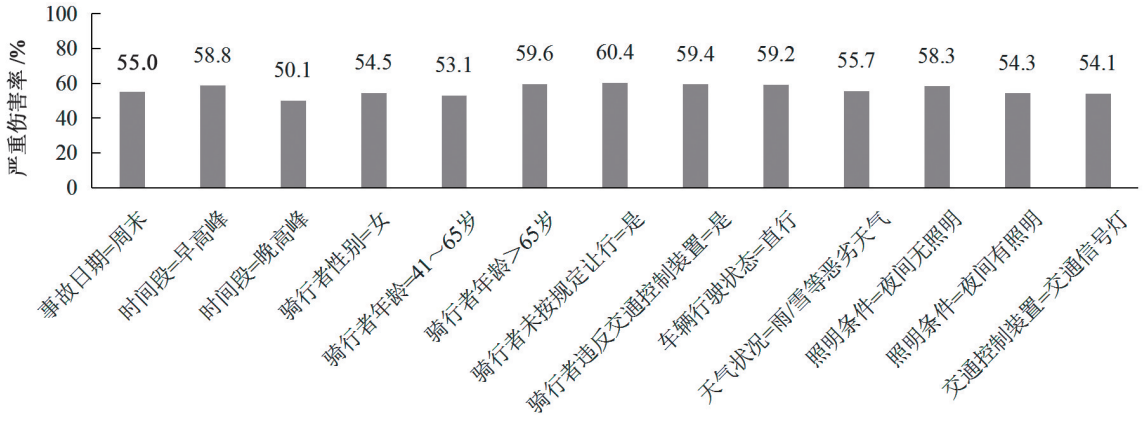

图 2 C1_ROS_HC 模型中严重伤害率高于非严重伤害率的关键因素及取值

Fig.2 Key factors and their values in the C1_ROS_HC model for which the severe injury rate exceeds the non-severe injury rate

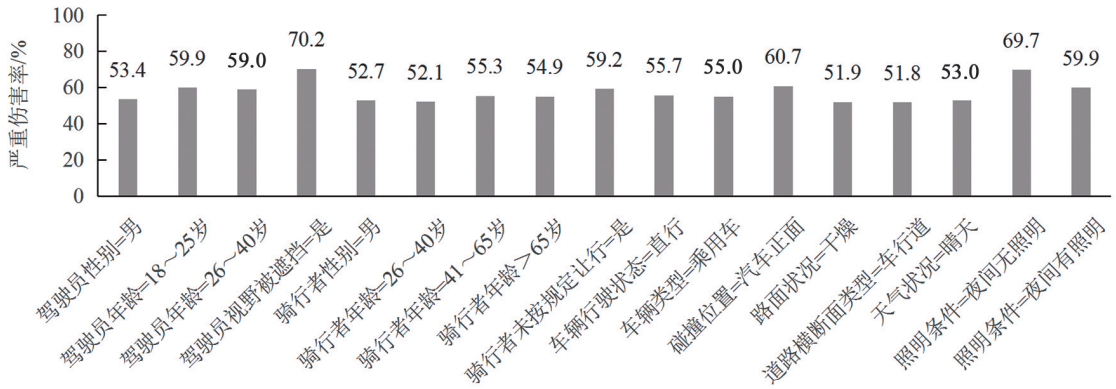

图 3 C2_SMOTE_TPDA 模型中严重伤害率高于非严重伤害率的关键因素及取值

Fig.3 Key factors and their values in the C2_SMOTE_TPDA model for which the severe injury rate exceeds the non-severe injury rate

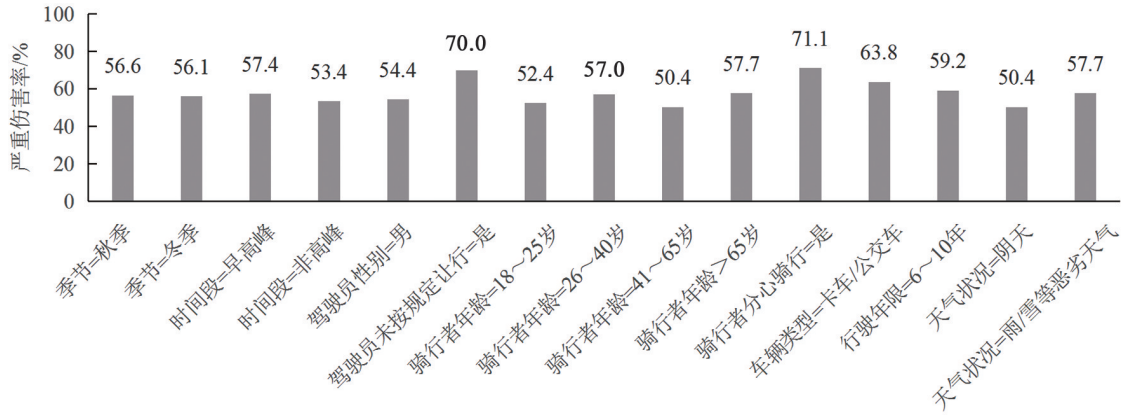

图 4 C3_ROS_HC 模型中严重伤害率高于非严重伤害率的关键因素及取值

Fig.4 Key factors and their values in the C3_ROS_HC model for which the severe injury rate exceeds the non-severe injury rate

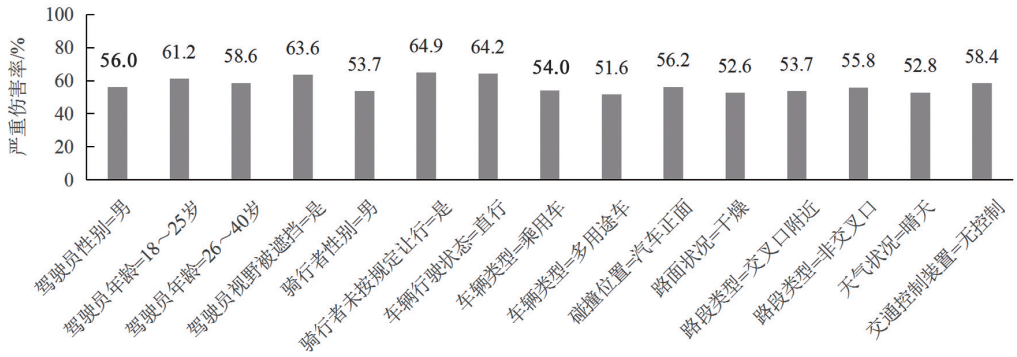

图 5　OD_ROS_TPDA 模型中严重伤害率高于非严重伤害率的关键因素及取值

Fig.5　Key factors and their values in the OD_ROS_TPDA model for which the severe injury rate exceeds the non-severe injury rate

基于图 2 至图 5,从“时间特征”、“驾驶员和骑行者特征”、“车辆特征”、“道路特征”和“环境特征”5 个角度对各类事故群的关键因素分析,可得到如下结论:

(1)时间特征

C1 事故群中发现:①周末事故会使严重伤害率增加 5%。这是因为周末人们的安全意识相对更低,因此骑行者更容易受到严重伤害[31]。②早高峰事故会使严重伤害率增加 8.8%。这是因为高峰时期的驾驶员和骑行者都倾向于更激进地驾驶和骑行。

C3 事故群中发现:①类似以往研究[32],秋季、冬季事故会使严重伤害率分别增加 6.6%和 6.1%。可能原因有:秋冬季节温度相对较低,导致车辆胎压较低、风险较高;秋冬季节骑行者常佩戴帽子,易被遮挡视线,从而增加受严重伤害的概率。②与 C1 事故群类似,早高峰事故会使严重伤害率增加 7.4%,不同的是,C3 事故群中还发现非高峰时段事故也会使严重伤害率增加 3.4%。结合 C3 事故群的特征看,大多数事故发生在专用支路,即商业中心、住宅、停车场等进入主干道的路段及路口。这类地点属于商业和休闲区域,在非高峰时段驾驶员和骑行者的活动也较为频繁,因此也容易引发严重伤害事故。

(2)驾驶员和骑行者特征

C1 事故群中发现:①女性骑行者比男性骑行者更容易受到严重伤害。这是因为 C1 事故群 99%都发生在交叉口及附近,而女性骑行者在路口更易遭遇危险冲突[33]。②骑行者年龄与严重伤害率正相关,其中 65 岁以上的骑行者因生理机能和风险感知能力退化,受严重伤害的概率会增加 9.6%。③骑行者未按规定让行和违反交通控制装置会使严重伤害率分别增加 10.4%和 9.4%。由此可见提高交叉口基础设施水平、加强老年人安全教育、鼓励 65 岁老人使用免费公交出行可能是降低此类事故风险的有效措施。

C2 事故群中发现:①18~25 岁驾驶员、26~40 岁驾驶员、男性驾驶员,会使严重伤害率分别增加 9.9%、9%和 3.4%。因为 C2 事故群多数发生在普通路段,而 18~40 岁驾驶员、男性驾驶员的驾驶行为更激进,且在普通路段的行驶速度更快[32]。②类似以往研究[28],驾驶员视野被遮挡会使严重伤害率增加 20.2%,原因是车辆在普通路段行驶速度较快,当驾驶员处于视野盲区时缺乏应急时间。③与 C1 事故群相反,C2 事故群发现男性骑行者发生严重伤害的概率会增加 2.7%。原因与之前类似,C2 事故群大多发生在普通路段,而男性骑行者在普通路段更易危险骑行[32]。由此可见有必要加强男性骑行者在普通路段的安全教育和车辆监控技术。这也体现了采用 LCA 探究 BMV 事故影响因素的优势。④骑行者未按规定让行会使严重伤害率增加 9.2%。

C3 事故群中发现:①男性驾驶员会使骑行者受严重伤害的概率增加 4.4%。②驾驶员未按规定让行、骑行者分心骑行会使严重伤害率分别增加 20%和 21.1%,它们是 C3 事故群发生严重伤害事故极为关键的两个因素。事实上 C3 事故群的大多事故都属于骑行者即将驶出专用支路,由于支路口道路比正常交叉口道路更狭窄,且 C3 事故群中 81.19%都没有交通控制装置,导致事故发生前行驶于主路车辆的驾驶员不容易发现骑行者。若此时骑行者分心骑行也未发现主干道的车辆或驾驶员未按规定让行,都极容易引发严重伤害事故。

(3)车辆特征

C1 事故群中发现:车辆直行事故会使严重伤害率增加 9.2%。这是因为相较于转向、起步/停车状态,直行车辆的行驶速度更快,碰撞产生的冲击能量更高。

C2 事故群中发现:①车辆直行事故会使严重伤害率增加 5.7%。②乘用车事故会使严重伤害率增加 5%,这是因为 C2 事故群多数发生在没有视野盲区的非交叉口,乘用车在这些路段行驶速度极快,导致双方难以避让。③碰撞位置位于汽车正面时会使严重伤害率增加 10.7%。因为正面碰撞更易撞毁自行车,尤其是车速较快的非交叉口路段[24]。针对这类事故,可以通过提高车辆的主动安全技术,例如 AEB 系统和引擎盖弹起技术,并强制要求骑行者佩戴头盔。

C3 事故群中发现:①车辆类型是卡车/公交车时,骑行者的严重伤害率会增加 13.8%,这是因为卡车/公交车的质量重、碰撞能量高。②行驶年限为 6~10 年的车辆会使严重伤率增加 9.2%,这可能是行驶年限在 6~10年的驾驶员经验丰富、风险阈值较高,容易过度自信而出现过激驾驶行为[34]。

(4)道路特征

C2 事故群中发现:①干燥路面事故会使严重伤害率增加 1.9%。这是因为干燥的非交叉口路段容易出现超速和冒险行为[4]。②车行道骑行事故会使严重伤害率增加 1.8%,而自行车道/路肩骑行事故会使严重伤害率下降。这与骑行者和汽车的交互有关,也可能与机非混行交通中骑行者的骑行速度相关[35]。因此,道路设计应尽量机非分离,并布置隔离带。

(5)环境特征

C1 事故群中发现:①雨/雪等恶劣天气事故会使严重伤害率增加 5.7%。这是因为恶劣天气的能见度低、制动和转向受限,导致碰撞更严重[36]。②夜间无照明和夜间有照明会分别使严重伤害率增加 8.3%和 4.3%。这是因为夜间光线不足,加上大多数自行车无灯光,容易被驾驶员忽视。因此建议骑行者要尽量避免夜间骑行或在夜骑时穿反光衣物,同时路政应定期检查路灯。③当交通控制装置是交通信号灯时,反而会使严重伤害率增加 4.1%。结合 C1 事故群的特征可知:C1 事故群中 91.13%的事故发生在交叉口,据王精滢[36]对机非交通事故的研究表明在路口路段存在交通信号装置时更容易造成严重事故,推断是因为交通流量大,导致骑行者的危险性增加。

C2 事故群中发现:①晴天会使严重伤害率增加 3%。因为 C2 事故群多发生在非交叉口,晴朗天气下驾驶员更容易出现超速行为。②与 C1 事故群类似,夜间无照明和夜间有照明均会增加严重伤害率。

C3 事故群中发现:雨/雪等恶劣天气会增加严重伤害率。综合三个事故群发现:在交叉口和专用支路路口,雨/雪等恶劣天气会增加严重伤害率,而在普通路段,晴天会增加严重伤害率。这解释了天气影响的异质性和以往研究争议。

(6)OD 事故群分析

OD 事故群与 3 个子事故群存在较多相似之处。与 3 个子事故群发现不同的是,路段类型仅在 OD 事故群被发现是关键因素,且结果显示事故发生在交叉口附近或非交叉口,会使严重伤害率分别增加 3.7%和 5.8%。这一结果印证了非交叉口危险性高于交叉口的研究结论[37],同时也揭示了路段类型的异质性效应。

4.3 关键因素异质性分析

为了观察不同事故群关键因素的异质性,对表 3 中 OD 事故群和 3 类子事故群关键因素进行对比分析,可得出以下 4 个结论:

(1)部分因素仅在特定子事故群中被识别为关键因素,但在 OD 事故群中未被识别为关键因素。例如“事故日期”和“骑行者违反交通控制装置”仅在 C1 事故群是关键因素。类似的“道路横断面类型”仅在 C2 事故群是关键因素;“季节”“驾驶员未按规定让行”“骑行者分心骑行”和“行驶年限”仅在 C3 事故群是关键因素。这些发现证实了使用 LCA 将事故数据划分为同质子数据的必要性和重要性。既可以挖掘隐藏在 OD

事故群中的关键因素以提供更全面的信息,还能探究不同事故群中关键因素的差异,从而揭示影响因素的异质性效应。

(2)部分因素在 OD 事故群以及特定子事故群中均被识别为关键因素。例如“驾驶员性别”和“车辆类型”在 OD、C2、C3 事故群中均为关键因素;“骑行者性别”“骑行者未按规定让行”和“车辆行驶状态”在 OD、C1、C2 事故群中均为关键因素。此外,部分因素仅在 OD 事故群和某一特定类子事故群中被识别为关键因素,例如“驾驶员年龄”“驾驶员视野被遮挡”“碰撞位置”等仅在 OD、C2 事故群中为关键因素。对于这些因素,借助 LCA 能剖析其对自行车事故的影响模式,这亦为子模型更全面地阐释骑行者伤害程度的影响因素提供了依据。

(3)尽管部分因素在 3 个子事故群中均被识别为关键因素,但它们对骑行者伤害程度的影响并非一致,甚至某些关键因素在不同子事故群中具有相反的效应。例如,在 C1 事故群中,“女性”骑行者更容易受到严重伤害,然而在 C2 事故群中相反;C1、C3 事故群均发现,在“雨/雪等恶劣天气”条件下,骑行者更容易受到严重伤害,但在 C2 事故群中却发现“晴天”时骑行者更容易受到严重伤害;C1 事故群发现“非高峰”时段骑行者不易受到严重伤害,而 C3 事故群发现“非高峰”时段骑行者更容易受严重伤害等。对于这些关键因素,在事故治理过程中若不加以区分,可能导致事故治理策略的失效或误导性决策。

(4)利用 LCA 将整体事故数据划分为同质子数据集揭示了一些新的关键因素,但 OD 事故群的关键因素也可能被忽略。例如,“路段类型”仅在 OD 事故群是关键因素。这表明分析事故数据时,需综合考虑整体数据与子数据集结果,以避免因数据划分而导致的关键因素遗漏。

以上这些差异主要归因于事故群之间具有的不同特征,因此在防治时,应结合不同事故群的特征实施针对性的措施。

5　结论

文章提出一种了融合重采样、LCA 和 BN 的方法,探究了数据异质性和非均衡性时城市自行车事故骑行者伤害程度的关键影响因素,并实现了因素量化分析及异质性分析,主要结论如下:

通过 LCA 将事故数据划分为 3 类同质子数据群,并结合重采样方法,显著提升了 BN 模型的综合性能和风险识别能力。在 C1、C2、C3 和 OD 事故群中分别挖掘出 10、13、9、12 个影响骑行者伤害程度的关键因素,其中时间段、骑行者性别等因素在不同事故群中存在异质性。

根据研究结果,建议提高交叉口基础设施水平、加强老年人安全教育、鼓励 65 岁老人使用免费公交出行。针对 C2 事故群,建议骑行者避免夜间骑行或在夜骑时穿反光衣物;同时,路政部门应定期检查非交叉口路段路灯并及时更换损坏路灯;此外,鼓励汽车搭载盲区监测系统和自动紧急制动系统。在商业中心、住宅、停车场等进入主干道的路段及路口,应加强交通执法并增加警示标语。

局限性分析:CRSS 数据库字段虽较全面,但缺少社会经济、人口特征和土地利用等宏观因素,导致因素选择受限。此外,国内外交通环境差异使得分析结果对国内的参考价值有限。未来国内交通事故数据共享机制完善后,可采用文章提出的方法进一步研究。

参考文献:

[1]王丙雨,邹俊,韩勇,等. 车辆和自行车碰撞事故中骑车人下肢损伤风险研究[J]. 振动与冲击, 2023, 42(11): 324-330. DOI: 10.13465/j.cnki.jvs.2023.11.038.

[2]魏晋,安实,张炎棠. 考虑建成环境交互影响的共享单车需求预测[J]. 科学技术与工程, 2023, 23(26): 11424-11430.

[3]赵琳娜,贾兴无,戴帅,等.中国城市道路交通安全特点解析[J]. 城市交通, 2018, 16(3): 9-14. DOI: 10.13813/j.cn11-5141/u.2018.0302.

[4]黎健侃,李泽炜,华汶雯,等. 城市道路交通事故统计分析[J]. 科技创新与应用, 2021, 11(21): 74-76. DOI: 10.19981/j.cn23-1581/g3.2021.21.023.

[5] ALNAWMASI N, MANNERING F. An analysis of day and night bicyclist injury severities in vehicle/bicycle crashes: A comparison of unconstrained and partially constrained temporal modeling approaches[J]. Analytic methods in accident research, 2023, 40: 100301. DOI: 10.1016/j.amar.2023.100301.

[6] WU J, RASOULI S, ZHAO J, et al. Large truck fatal crash severity segmentation and analysis incorporating all parties involved: A Bayesian network approach[J]. Travel Behaviour and Society, 2023, 30: 135-147. DOI: 10.1016/j.tbs.2022.09.003.

[7] LUAN S, LI M, LI X, et al. Effects of built environment on bicycle wrong Way riding behavior: A data-driven approach[J]. Accident Analysis & Prevention, 2020, 144: 105613. DOI: 10.1016/j.aap.2020.105613.

[8] DASH I, ABKOWITZ M, PHILIP C. Factors impacting bike crash severity in urban areas[J]. Journal of safety research, 2022, 83: 128-138. DOI: 10.1016/j.jsr.2022.08.010.

[9] YANG Z, YANG Z, SMITH J, et al. Risk analysis of bicycle accidents: A Bayesian approach[J]. Reliability Engineering & System Safety, 2021, 209: 107460. DOI: 10.1016/j.ress.2021.107460.

[10]杨园园,鲁统宇,崔俊,等. 考虑错分代价的 ADASVM-CSLINEX 模型及应用[J]. 计算机工程与应用, 2024, 60(3): 348-356. DOI: 10.3778/j.issn.1002-8331.2210-0379.

[11]谭鼎. 基于集成学习的慢行交通事故严重程度预测及致因分析[D]. 兰州:兰州交通大学, 2024. DOI: 10.27205/d.cnki.gltec.2024.001534.

[12] YAHAYA M, GUO R, FAN W, et al. Bayesian networks for imbalance data to investigate the contributing factors to fatal injury crashes on the Ghanaian highways[J]. Accident Analysis & Prevention, 2021, 150: 105936. DOI: 10.1016/j.aap.2020.105936.

[13]潘义勇,李烁.建成环境对交叉口行人事故严重程度异质性影响[J].重庆交通大学学报(自然科学版),2024,43(6):87-93. DOI: 10.3969/j.issn.1674-0696.2024.06.12.

[14] HOSSEINI S H, DAVOODI S R, BEHNOOD A. Bicyclists injury severities: An empirical assessment of temporal stability[J]. Accident Analysis & Prevention, 2022, 168: 106616. DOI:10.1016/j.aap.2022.106616.

[15] LIN Z, FAN W D. Exploring bicyclist injury severity in bicycle-vehicle crashes using latent class clustering analysis and partial proportional odds models[J]. Journal of safety research, 2021, 76: 101-117. DOI: 10.1016/j.jsr.2020.11.012.

[16]马硕.考虑时空效应的自行车事故特征及影响因素研究[D].北京:北京建筑大学, 2024. DOI: 10.26943/d.cnki.gbjzc.2024.000816.

[17] SONG Y, CHITTURI M V, NOYCE D A. Intersection two-vehicle crash scenario specification for automated vehicle safety evaluation using sequence analysis and Bayesian networks[J]. Accident Analysis & Prevention, 2022, 176: 106814. DOI: 10.1016/j.aap.2022.106814.

[18]邓佳.人—车交通事故严重程度 Probit 预测模型构建及实证研究[D].西安:长安大学,2019.

[19]晏钰棚. 自行车事故伤害严重程度影响因素相对重要性和时空异质性研究[D]. 成都:西南交通大学, 2021. DOI: 10.27414/d.cnki.gxnju.2021.001225.

[20]布和.道路交通事故的成因分析及预防研究[J].武汉公安干部学院学报, 2019, 33(2): 16-20.

[21]申昕, 沈金星, 郑长江, 等. 基于 Multinomial Logit 模型的美国北卡罗莱纳州慢行交通事故严重程度分析[J]. 交通与运输, 2021, 37(5): 24-28.

[22]孙少峰. 基于集成学习的新能源车辆与内燃机车辆交通事故严重程度影响因素对比研究[D]. 西安:长安大学, 2021. DOI: 10.26976/d.cnki.gchau.2021.001818.

[23]翟洪军,陈启光,申春悌,等. 基于潜在类别分析对不同年龄组患者新冠肺炎病因病机证候研究[J].世界科学技术-中医药现代化, 2021, 23(3): 866-873. DOI: 10.11842/wst.20200622001.

[24]焦朋朋,李汝鉴,王健宇,等. 考虑潜在类别的老年行人交通事故严重程度致因分析[J].交通运输系统工程与信息, 2022, 22(5): 328-336. DOI: 10.16097/j.cnki.1009-6744.2022.05.034.

[25]李昂, 韩萌, 穆栋梁, 等. 多类不平衡数据分类方法综述[J]. 计算机应用研究, 2022, 39(12): 3534-3545. DOI: 10.19734/j.issn.1001-3695.2022.03.0198.

[26]ABD ELRAHMAN S M, ABRAHAM A. A review of class imbalance problem[J]. Journal of Network and Innovative Computing, 2013, 1: 9-9.

[27]DELEN D, TOMAK L, TOPUZ K, et al. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods[J]. Journal of Transport & Health, 2017, 4: 118-131.

[28]HE H, BAI Y, GARCIA E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]//2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). Hong Kong: IEEE, 2008: 1322-1328.

[29]CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357. DOI: 10.1613/jair.953.

[30]HECKERMAN D, GEIGER D, CHICKERING D M. Learning Bayesian networks: The combination of knowledge and statistical data[J]. Machine learning, 1995, 20: 197-243. DOI: 10.1007/BF00994016.

[31]SIVASANKARAN S K, BALASUBRAMANIAN V. Exploring the severity of bicycle-vehicle crashes using latent class clustering approach in India[J]. Journal of safety research, 2020, 72: 127-138. DOI: 10.1016/j.jsr.2019.12.012.

[32]SUN Z, XING Y, WANG J, et al. Exploring injury severity of bicycle-motor vehicle crashes: A two-stage approach integrating latent class analysis and random parameter logit model[J]. Journal of Transportation Safety & Security, 2022, 14(11): 1838-1864. DOI: 10.1080/19439962.2021.1971814.

[33]SAMEREI S A, AGHABAYK K, SHIWAKOTI N, et al. Using latent class clustering and binary logistic regression to model Australian cyclist injury severity in motor vehicle-bicycle crashes[J]. Journal of Safety Research, 2021, 79: 246-256. DOI: 10.1016/j.jsr.2021.09.005.

[34]孙晴.考虑时间不稳定性的货车-小汽车事故严重程度影响因素分析[D].西安:长安大学, 2021. DOI: 10.26976/d.cnki.gchau.2021.000866.

[35]MYHRMANN M S, JANSTRUP K H, MØLLER M, et al. Factors influencing the injury severity of single-bicycle crashes[J]. Accident Analysis & Prevention, 2021, 149: 105875. DOI: 10.1016/j.aap.2020.105875.

[36]王精滢.考虑空间异质性的机非交通事故严重程度分析[D]. 成都:西南交通大学, 2020. DOI: 10.27414/d.cnki.gxnju.2020.000850.

[37]丁晶玉.考虑关联因素异质性及非线性的骑行者碰撞事故受伤严重程度研究[D]. 大连:大连交通大学,2024. DOI: 10.26990/d.cnki.gsltc.2024.000296.