

# 基于 CatBoost-SHAP 与贝叶斯网络的交通事故严重程度情景化分析

张豪森, 王景升

(中国人民公安大学交通管理学院, 北京 100038)

**摘要:** 为研究城市道路交通事故严重程度的关键致因及其联动效应, 基于深圳市 2016-2019 年 6 968 起一般程序处理的事故数据, 构建“CatBoost-SHAP 特征筛选、K2 结构学习、极大似然估计”的贝叶斯网络模型, 结合逆向推理与情景推演, 并以  $\Delta P_c$ 、 $R_c$  量化严重度迁移与风险放大。结果显示, 车辆类型、照明条件、机动车状态与车辆安全状况为关键致因, 据此构建的多个典型高风险情景表明, 多因素联合作用呈现出严重事故端概率差  $\Delta P_c > 0$  及放大指数  $R_c > 1$  的系统性上升; 模型经分层五折验证, 总体准确率为 80.37%, 三类事故类型的 AUC 为 0.839、0.815、0.774。该框架兼顾预测性能与因素间依赖关系的可解释性, 可揭示多因素联合作用的风险放大机制, 支撑情景化联合干预, 为城市交管采取管理措施提供量化参考依据。

**关键词:** 交通安全; 情景推演; 事故严重程度; CatBoost-SHAP; 贝叶斯网络; 多因素联合作用

中图分类号: U491.31

文献标志码: A

## Scenario-Based Analysis of Road-Traffic Accident Severity Using CatBoost-SHAP and Bayesian Networks

Haosen Zhang<sup>1</sup>, Jingsheng Wang<sup>1</sup>

(1. School of Traffic Management, People's Public Security University of China, Beijing 100038, China)

**Abstract:** To identify the key determinants of urban road-traffic accident severity and their interaction effects, we analyze 6,968 accidents handled under the general procedure in Shenzhen during 2016–2019 and develop a Bayesian-network framework that integrates CatBoost-SHAP feature screening, K2 structure learning, and maximum likelihood estimation (MLE). Coupling backward (inverse) inference with scenario analysis, we quantify severity migration and risk amplification using the probability difference ( $\Delta P_c$ ) and the amplification index ( $R_c$ ). The results indicate that vehicle type, road lighting, motor-vehicle compliance/inspection status (e.g., unresolved violation or expired inspection), and vehicle technical condition are principal contributors. Scenario analyses constructed from these determinants reveal systematic increases on the more-severe end ( $\Delta P_c > 0$  and  $R_c > 1$ ) driven by multi-factor coupling. Under stratified five-fold cross-validation, the model attains an overall accuracy of 80.37%, with AUCs of 0.839, 0.815, and 0.774 for the three severity classes (Damage, Injury, Fatal). The framework balances predictive performance with interpretable inter-factor dependencies, elucidates the risk-amplification mechanism of multi-factor interactions, and provides quantitative evidence to support scenario-based joint interventions and urban traffic-management measures.

**Key words:** traffic safety; scenario analysis; accident severity; CatBoost-SHAP; Bayesian Network; multi-factor synergistic effects

世界卫生组织发布的《全球道路安全状况报告》<sup>[1]</sup>显示,道路交通事故是全球健康问题和发展的主要原因之一,全球道路交通事故伤亡率日益上升。随着中国城市经济快速发展,车辆数量及交通量不断增加<sup>[2]</sup>,城市道路交通事故也随之增加。面临日益严峻的城市道路交通安全形势,充分利用事故数据,调查事故严重程度的影响因素,为交管部门采取针对性的预防策略,降低事故损失十分有必要。

国内对城市道路交通事故的研究可以追溯到2003年,陈宽民<sup>[3]</sup>等总结梳理了事故时空、人群分布特点,并针对特点进行了成因分析。近年来,学者对城市交通事故因素进行分析主要围绕着“人、车、路、环境”<sup>[4]</sup>四方面展开,汪文武等<sup>[5]</sup>从人的因素出发,发现城市道路交通事故中驾驶员操作状态具有较高依赖性,易受到其他因素的影响;严利鑫<sup>[6]</sup>等研究发现普通道路、车辆与非机动车事故、夏季对轻微事故的影响更为显著;Mehdi<sup>[7]</sup>等建立了各道路因素与交通事故严重程度之间影响关系模型;潘义勇<sup>[8]</sup>等从环境因素出发,探究不同光照条件下交通事故严重程度影响因素的异质性和可转移性。但是多数研究仍聚焦单因素,缺少系统性的依赖建模。在事故研究方法层面,研究者经常用的方法有关联规则分析法<sup>[9]</sup>、Logistic<sup>[10]</sup>、贝叶斯网络<sup>[11]</sup>、模糊Petri网<sup>[12]</sup>分析法等传统模型及其改进模型<sup>[13]</sup>,张洁<sup>[14]</sup>等基于二元Logistic模型对各类事故发生机理进行分析;张道文<sup>[15]</sup>等构建二元Lasso-Logistic回归分析模型,选取10个自变量进行了事故严重程度分析;赵丹<sup>[16]</sup>等建立双变量Probit模型,将各影响因素作为自变量,事故的严重程度和类型作为因变量,从而识别了影响因素并分析了它们之间的相互关系,但上述方法并不能清楚表示出各因素之间的依赖关系及依赖程度。由冰玉等<sup>[17]</sup>结合贝叶斯网络和故障树模型,探讨了山区高速公路交通事故发生的关键影响因素;李世杰<sup>[18]</sup>等建立了基于故障树的贝叶斯网络模型,并定性定量分析了道路交通事故影响因素,但是贝叶斯网络在交通安全应用中,结构学习常依赖专家经验或单纯数据驱动,缺少二者结合。

综上所述,事故严重程度研究已由单一因素走向多因素分析<sup>[19]</sup>,并逐渐考虑因素间的依赖结构。但是建模常偏于纯专家先验或纯数据驱动,二者融合并同时利用可解释先验的很少。为此,本文在人、车、路、环境四类因素上构建 CatBoost-SHAP 先验约束的贝叶斯网络,以 SHAP 重要度与交互关系生成可解释先验顺序与父节点候选集,并与领域专家知识及相关性约束共同收缩结构搜索空间、抑制伪相关,随后结合灵敏度分析与多因素情景推理,以严重度迁移与风险放大的定量指标揭示关键致因与联合作用机制,从而为城市交通安全管理提供可操作的理论依据与实证支撑。

## 1 数据与方法

### 1.1 数据来源及变量处理

收集深圳市交管部门 2016-2019 年一般程序处理的道路交通事故数据,经过去重、空间与时间有效性校验及缺失阈值筛查后,得到 6 968 条事故样本数据,其中包括人、车、路、环境四大类下 12 个候选自变量,事故类型(Acc)为因变量。为提升可解释性并便于后续建模,对相关字段按业务标准与《道路交通事故统计规范》(GA/T497—2019)进行离散化与符号化编码,变量定义与取值见表 1,各取值分布见图 2,基于数据及研究目标构建的研究流程见图 1。

表 1 模型变量描述及离散化取值

Tab.1 Model Variable Description and Discretized Values

变量类别	变量名称	符号表示	变量 GeNIe 描述及离散化取值				均值	标准差
			1	2	3	4		
事故	事故类型 Acc	$Y$	财损事故 Damage	伤人事故 Injury	亡人事故 Fatal	—	1.84	0.754
人的因素	驾驶人性别 Gen	$X_1$	男 Male	女 Female	—	—	1.09	0.280
	驾驶人学历 Edu	$X_2$	初中 Junior	大学 College	高中 Senior	小学 Primary	1.83	0.993
车辆因素	车辆类型 Vt	$X_3$	摩托车 Motor	货车 Truck	客车 Bus	—	2.40	0.790
	机动车状态 St	$X_4$	事故未处理 Accident	违法未处理 Illegality	逾期未检验 Expired	正常 Normal	2.52	1.131
	车辆安全状况 Vs	$X_5$	正常 Normal	不良 Poor	—	—	1.20	0.403
道路因素	道路隔离措施 Iso	$X_6$	无隔离 None	中心隔离 Center	中心和机非 CNM	—	1.81	0.730
	道路线形 Agl	$X_7$	平直 Straight	非平直 Irregular	—	—	1.07	0.257
	路面材料 Mtl	$X_8$	沥青 Asphalt	水泥 Cement	—	—	1.39	0.487
环境因素	是否工作日 Wkd	$X_9$	否 no	是 yes	—	—	1.68	0.467
	是否平原地区 Pln	$X_{10}$	是 YES	否 NO	—	—	1.41	0.493
	照明条件 Lht	$X_{11}$	白天 Daytime	夜间无路灯 NWSL	夜间有路灯 WSLN	—	2.12	0.965
	天气状况 Wea	$X_{12}$	晴 Sunny	阴 Cloudy	雨 Rainy	—	1.27	0.615

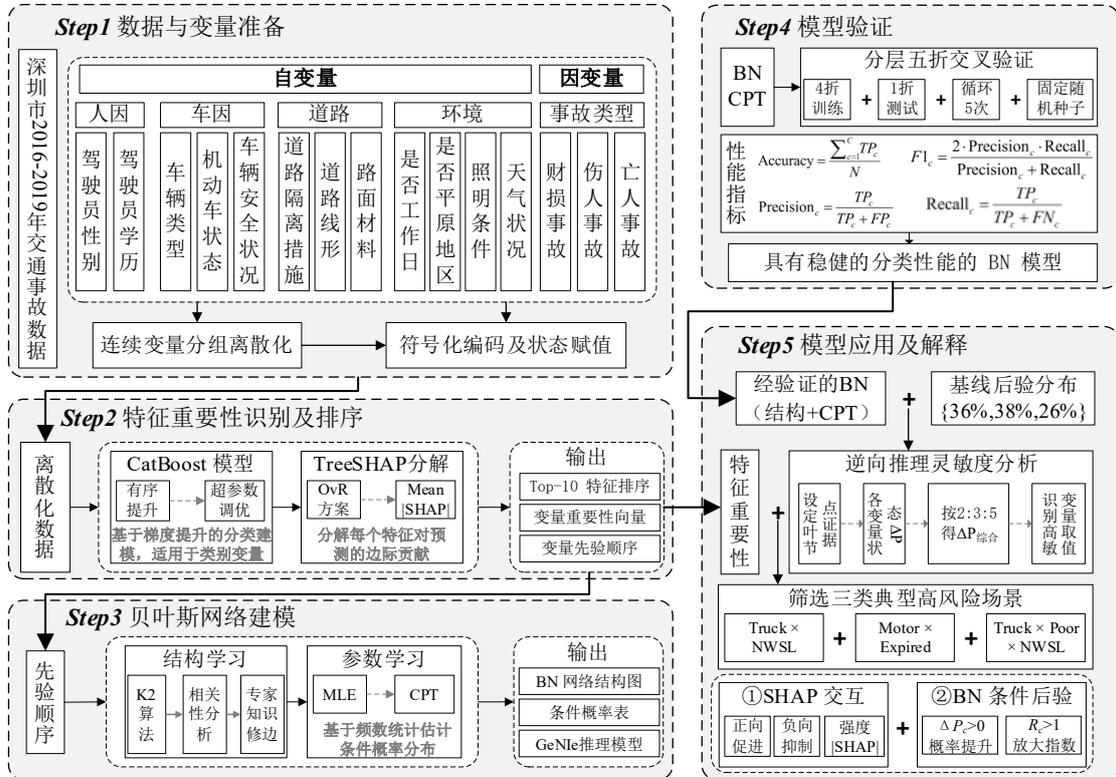


图 1 模型构建及应用结构层次图

Fig.1 Hierarchical Diagram of Model Construction and Application

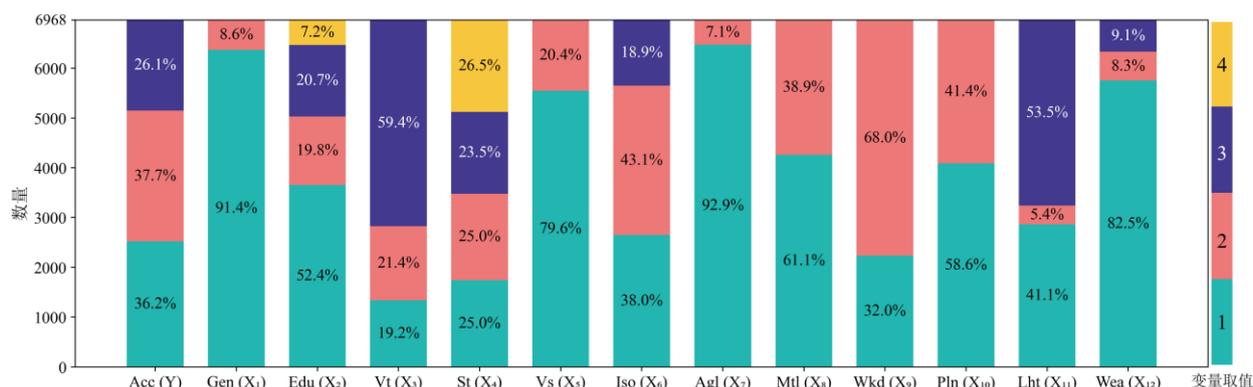


图 2 各变量的取值数量与比例堆叠图

Fig.2 Stacked Bar Chart of Variable Value Counts and Proportions

### 1.2 特征筛选与先验排序

为在保证可解释性的同时稳定筛选关键因子，本节采用CatBoost-SHAP完成特征筛选与先验排序，先将多类别任务拆解为一对其余（OvR）的三个二分类子模型，在训练折内拟合CatBoost并完成超参数设定；随后在每个子模型上以SHAP计算特征贡献，取各特征|SHAP|的样本均值作为该子模型的重要度，再跨子模型做宏平均以获得各特征一致性重要度，按由高到低排列得到先验排序。

CatBoost是一种基于梯度提升决策树(GBDTs)并用于分类或回归任务的强大的机器学习算法，相较于传统GBDT，它在处理高维类别变量和非平衡样本方面表现出更强的泛化能力与鲁棒性，因此更适用于本文数据集。为获得最优预测与泛化表现，使用分组网格搜索与分层五折交叉验证对CatBoost关键超参数进行寻优，全部调参均在训练折内执行以防止信息泄露，图3为第一组调优过程，最终配置见表2，并据此用于后续分析。

表 2超参数分组与最优取值

Tab.2 Grouped Hyperparameters and Optimal Values

分组	参数	范围	最优值
第一组	iterations	100~400	200
	depth	3~6	3
	learning_rate	0.01~0.2	0.1
第二组	l2_leaf_reg	1~10	3
	od_wait	10~30	20

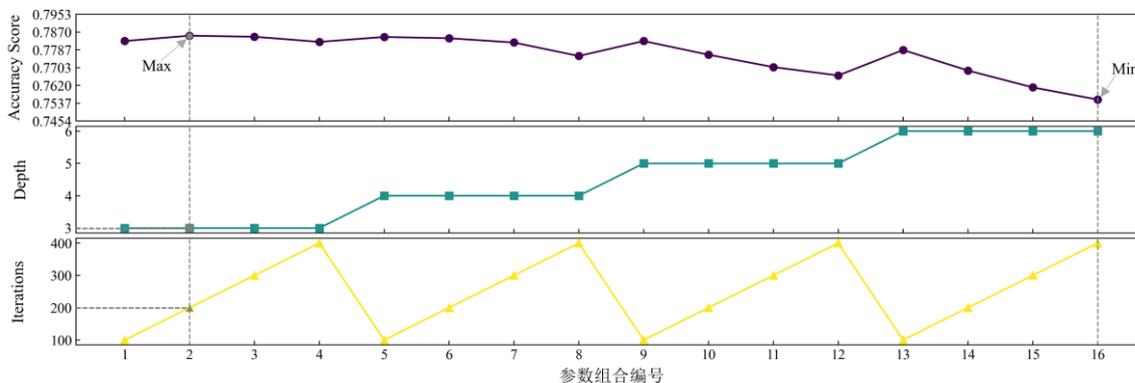


图 3 CatBoost 模型第一组参数寻优图

Fig.3 Parameter Optimization Diagram for the First Group of CatBoost Model

基于合作博弈论的 SHAP, 以 Shapley 值刻画各特征对单一样本预测  $f(x)$  的边际贡献, 满足局部准确性、一致性与加性分解等公理, 因而适合在多特征条件下进行稳定对比。其基本形式为:

$$\phi_i(x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \left[ f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right] \quad (1)$$

CatBoost 树模型通过 TreeSHAP 在多项式时间精确计算, 并对三组子模型的类别定向贡献取宏平均, 形成全局重要度:

$$G_i = \frac{1}{3} \sum_{c \in \{\text{Fatal, Injury, Damage}\}} E_x \left[ \left| \phi_i^{(c)}(x) \right| \right] \quad (2)$$

以此作为特征筛选和先验排序的依据, 用于后续结构学习的先验约束。同时, 结合 SHAP 汇总图、依赖图与交互图识别关键特征的单调性与强交互对, 从而在性能与可解释性之间取得平衡并为后续情景构建提供方向性证据。

### 1.3 贝叶斯网络模型学习

贝叶斯网络 (Bayesian Network, BN) 是一种概率图模型, 用于表示和推理随机事件之间的依赖关系。它通过有向无环图 (directed acyclic graph, DAG) 表示一组变量及其条件依赖关系。它通过条件概率表达变量之间的影响关系, 适用于解决多种不确定性问题, 是目前用于推理领域最有效的理论模型之一<sup>[20]</sup>。贝叶斯定理为模型提供了从已知信息中推断未知变量的数学基础, 在实际建模中, 贝叶斯网络通过对多个变量的联合分布进行分解, 具体形式如下:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}(X_i)) \quad (3)$$

其中,  $\text{Pa}(X_i)$  表示节点  $X_i$  的父节点集合。该公式说明, 只需定义每个变量在给定其父节点条件下的局部条件概率分布 (CPT), 即可构建整个系统的联合概率模型。

在结构学习阶段, 先依据 CatBoost-SHAP 的重要度确定节点先验顺序并将目标变量 Acc 置于末位, 在该顺序约束下采用 K2 贪婪加边搜索, 使用 BDeu 评分 (ESS=10, 综合本研究样本量  $N=6\ 968$  与变量离散度所设) 评估每次增边的边际增益, 同时将最大父节点数设为  $k=3$  以控制复杂度与抑制高阶伪相关, 从而在收缩搜索空间、降低误增边风险的前提下获得稳健且可解释的网络结构。由于数据完整、变量可观测, 采用极大似然估计法 (Maximum Likelihood Estimation, MLE) 进行参数拟合, 其形式如下:

$$\hat{\theta}_{\text{MLE}} = \operatorname{argmax}_{\theta} P(D | \theta) = \operatorname{argmax}_{\theta} \prod_{j=1}^N \prod_{i=1}^n P(x_i^{(j)} | \text{Pa}_i^{(j)}; \theta_i) \quad (4)$$

其中,  $D$  表示训练数据,  $\theta$  表示参数集合。该方法通过最大化观测数据在当前网络结构下的似然函数, 获得最优条件概率表 (CPT)。综上, 以 K2 与 MLE 的组合实现了结构、参数一体化学习, 能够对变量之间的依赖关系建模, 并进行推理与分析, 为交通事故严重程度建模提供坚实的理论基础。

### 1.4 逆向推理与情景推演

基于已建贝叶斯网络开展逆向推理, 即假设事故发生的类型已知 (将叶节点事故类型各类别概率依次设为 100%, 即  $\text{Acc}=c$ ), 逆向可推得其它节点的后验概率, 并通过观察其与各个先验概率的变化值来刻画变量取值对各事故类型的促进/抑制方向与影响强度, 从而形成单因素各节点状态的灵敏度:

$$\Delta P(X=x) = P(X=x | Acc=c) - P(X=x) \quad (5)$$

其中,  $(\Delta P)$ 正值表示促进作用, 负值表示抑制作用, 绝对值越大影响越强。为综合反映三类事故在社会成本差异下的影响强度, 对财产损失、伤人、死亡事故分别赋权2:3:5, 计算综合灵敏度:

$$\Delta P_{\text{综合}} = \frac{2|\Delta P_{\text{Damage}}| + 3|\Delta P_{\text{Injury}}| + 5|\Delta P_{\text{Fatal}}|}{2+3+5} \quad (6)$$

其次, 鉴于单因素难以完整揭示严重程度的复杂致因机制, 以逆向推理得到的单因素灵敏度为主线, 结合前述 CatBoost-SHAP 特征重要性分析, 从高敏取值中筛选并构建三类在现实交通中具有代表性的多因素高风险情景; 随后以 SHAP 交互效应提供方向性证据, 并以 BN 在组合证据下的条件后验变化进行量化验证, 构建“双证据链”交互分析框架, 用以刻画不同情景下事故类型分布的位移与风险放大机制。

评估以未施加证据的基线后验分布  $P_0(Acc)$  为参照, 记严重度类别  $C \in \{\text{Damage, Injury, Fatal}\}$  的基线概率为  $p_c^{\text{base}}$ , 情景下设定组合证据后的后验为  $p_c^{\text{scene}}$ , 定义概率差 (Probability Difference,  $\Delta P_c$ ) 为特定情景下某类事故严重度的后验概率与基线概率之差, 放大指数 (Amplification Index,  $R_c$ ) 为该后验概率与基线概率之比。如式 (7), 当  $\Delta P_c > 0$  且  $R_c > 1$  时, 判定该情景对严重度  $c$  具有系统性推升。

$$\Delta p_c = p_c^{\text{scene}} - p_c^{\text{base}}, R_c = \frac{p_c^{\text{scene}}}{p_c^{\text{base}}} \quad (7)$$

## 2 结果与讨论

### 2.1 特征筛选与排序结果

基于 CatBoost-SHAP 的综合重要度排序显示, 前十特征依次为: 车辆类型 (Vt)、照明条件 (Lht)、道路隔离措施 (Iso)、机动车状态 (St)、车辆安全状况 (Vs)、路面材料 (Mtl)、驾驶人学历 (Edu)、道路线形 (Agl)、是否工作日 (Wkd) 与天气状况 (Wea), 这些特征的累积 SHAP 重要性已超过总体的 95%, 表明它们捕捉了数据中绝大部分的预测信息。因此, 为在保证模型预测性能的同时, 构建一个不复杂、更具可解释性的贝叶斯网络结构, 本文最终确定这 10 个关键变量进行后续分析。

为深入理解这 10 个关键变量的作用机制, 图 4 的多维度 SHAP 分析结果表明, 各关键因素的影响强度与方向在不同事故严重等级下呈现出显著的差异。例如, 照明条件 (Lht) 是预测死亡事故的最关键特征, 但在伤人事故中其重要性却不如车辆类型 (Vt); 而且, 关键特征在不同取值下对事故严重程度的后验分布具有可区分的正/负向影响, 并呈现一定的非线性与交互趋势, 例如, “摩托车”的取值对推高伤人事故风险的贡献最为突出, 而“货车”的取值则在死亡事故中显示出更强的正向推动作用。这为后续构建高风险情景进行推演分析奠定基础。

### 2.2 结构学习与参数估计结果

本节以经 CatBoost-SHAP 排序的 10 个变量为特征变量与先验约束, 使用 MATLAB R2023a 软件中的 BNT 工具箱, 调用 K2 算法中的 learn\_struct\_K2.m 函数, 对样本集数据进行初步网络结构学习, 该结构经一个由三位交通安全学者与两位一线实战专家组成的专家组进行优化, 专家组以相关性热力图 (图 5) 为辅助, 诊断并一致认定初步结构中存在伪相关连接。例如, 从  $X_{12}$  (天气状况) 分别指向  $X_2$  (驾驶人学历) 与  $X_8$  (路面材料) 的边 (如图 6b 虚线所示) 被视为缺乏现实因果支撑的典型伪相关, 尽管数据中可

能存在微弱的统计关联。据此共识，在优化阶段手动移除了这些伪相关边，最终形成如图6c所示的优化网络。在此基础上，采用MLE方法学习各节点条件概率表。未置证据时，事故类型的基线后验为  $P_0(\text{Acc}) = \{\text{Damage } 36\%, \text{Injury } 38\%, \text{Fatal } 26\%\}$ ，作为情景推理的参照。为示例，仅列一组父节点取值的条件概率：当  $\text{Wea}=1$ 、 $\text{Lht}=1$ 、 $\text{Vt}=1$  时， $\text{Acc}$  为  $\{1,2,3\}$  的条件概率分别为  $\{0.00539, 0.70975, 0.28487\}$ 。

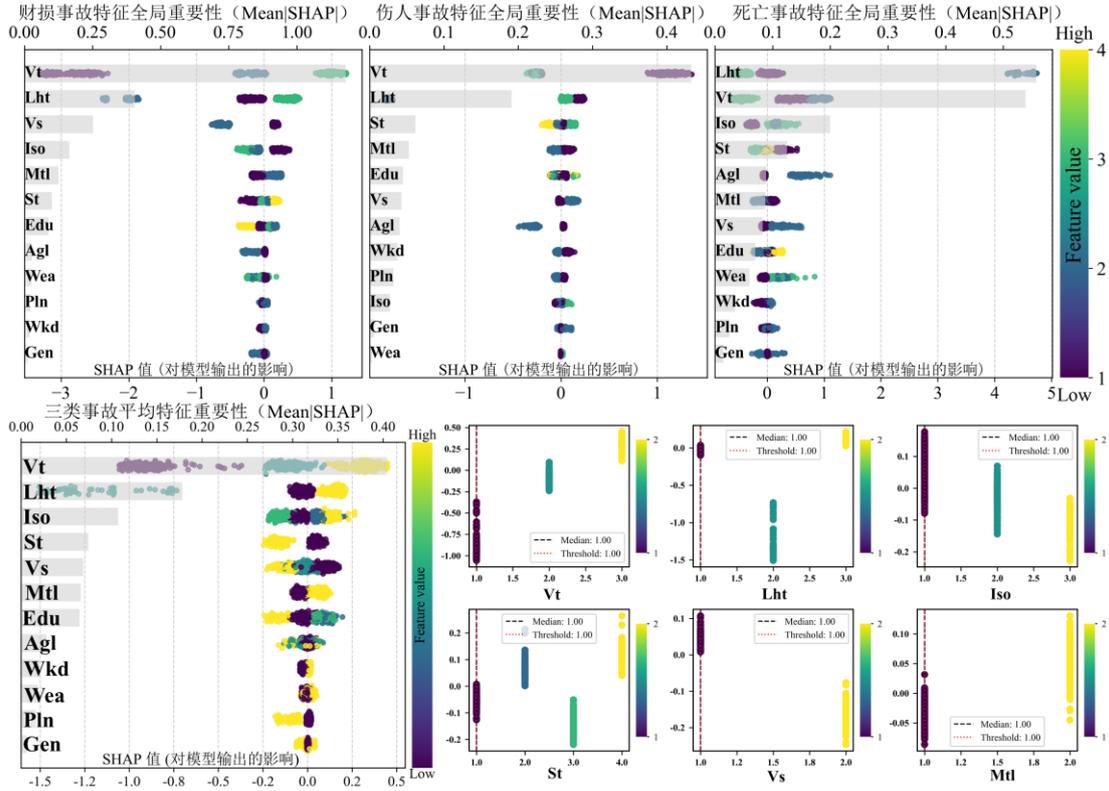


图 4 特征重要性排序及SHAP汇总图

Fig.4 Feature Importance Ranking and SHAP Summary Plot

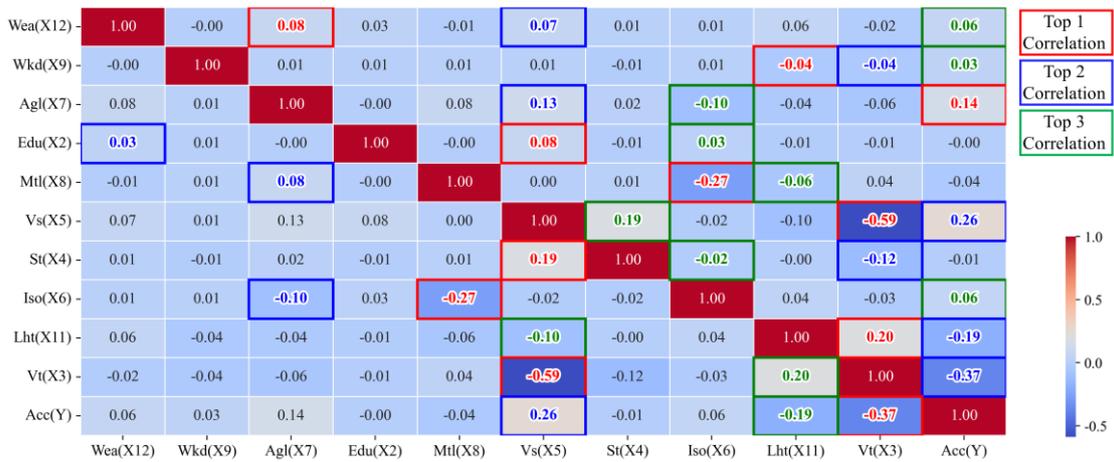


图 5 变量相关性热图与Top 3显著相关变量标识

Fig.5 Correlation Heatmap with Top 3 Significant Variable Identifications

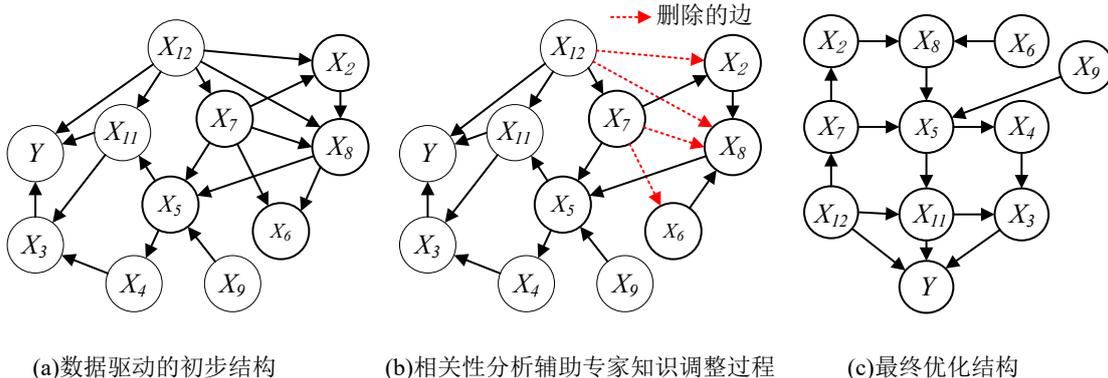


图 6 城市道路交通事故严重程度贝叶斯网络

Fig.6 Bayesian Network for Urban Road Traffic Accident Severity

### 2.3 模型性能评估

为了验证贝叶斯网络模型的有效性，采用分层五折交叉验证并结合OvR多类别方案，在保持类别比例一致的前提下，将数据划分为五折，循环以四折训练、一折测试。评估指标包括总体准确率、宏平均Precision/Recall/F1及各类别ROC-AUC，见图7(a)~(c)。模型总体准确率为80.37%，其中Damage、Injury、Fatal类别召回率分别为83.20%、78.50%、79.10%，各类别AUC分别为0.839、0.815、0.774，宏平均Precision、Recall与F1均约为80%。混淆主要集中在相邻严重程度之间，符合事故等级的连续性特征。综上表明，该性能表明模型具有良好的区分度与稳定性，能够支撑后续的逆向推理与情景推演。

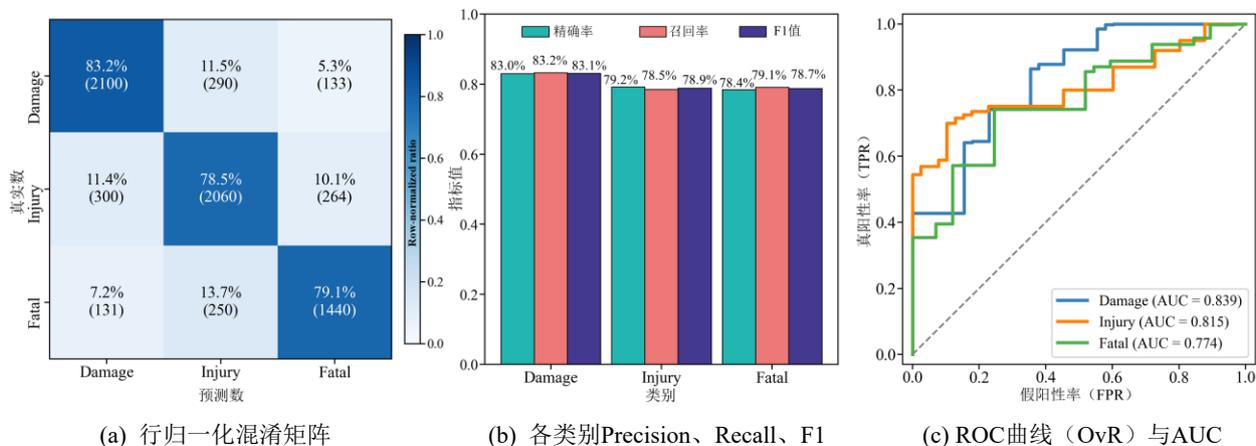


图 7 贝叶斯网络模型的5折交叉验证综合性能评估图

Fig.7 Comprehensive Performance of the Bayesian Network Model under 5-Fold Cross-Validation

### 2.4 灵敏度分析结果

基于经验证的贝叶斯网络，在Acc节点置证开展逆向推理与灵敏度分析，按式(5)~(6)计算 $\Delta P$ 及其2:3:5加权综合，结果见表3与图8。综合排名显示，车辆与道路环境相关取值贡献居前：Vt=Truck在Fatal端呈持续正向 ( $\Delta P > 0$ )，Vt=Motor在Injury端正向更为突出；Lht=NWSL (夜间无路灯) 对较高严重程度多为正向，而Lht=Day/Good多为负向；与技术及运行状态相关的不利取值Vs=Poor与St=Expired均位列综合前列，在Injury、Fatal端表现为显著的正向贡献。除此之外，Iso=None (隔离缺失) 在高严重程度端多为正向，隔离完备则为负向；Wea=Cloudy/Rainy (阴雨天) 在Injury端多为正向，晴天取值接近零或轻度负向；Agl、Mtl、Wkd、Edu的影响整体较弱、接近零，在表3中不再展示。高灵敏度因素在不同事故类型中作用显

著，应作为交通安全风险防控重点，也为后续多因素交互作用的情景构建提供了可靠依据。

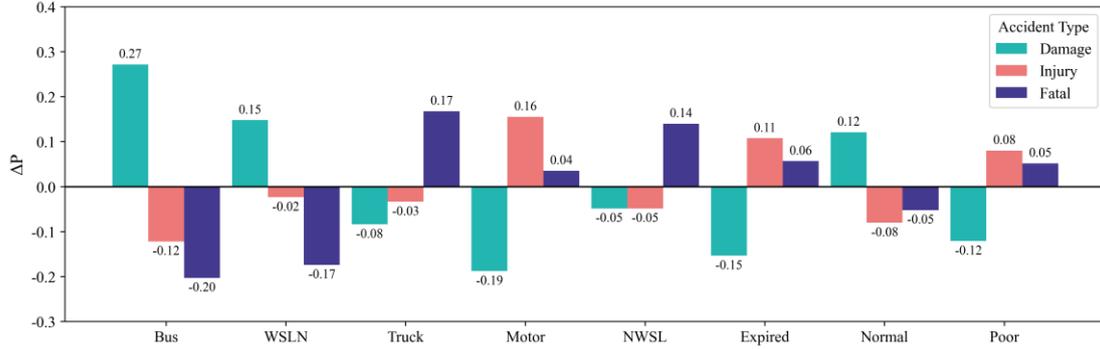


图 8 关键变量状态对事故类型节点的灵敏度影响

Fig.8 Sensitivity Effects of Key Variable States on Accident Type Nodes

表 3 高敏特征节点状态灵敏度排序

Tab.3 High-Sensitivity Ranking of Feature-Node States

变量名称	节点取值	先验概率	后验概率			灵敏度(ΔP)			灵敏度绝对值排序(影响力)			综合排序
			Damage	Injury	Fatal	Damage	Injury	Fatal	Damage	Injury	Fatal	
照明条件	Daytime	0.4126	0.3132	0.4845	0.4465	-0.0994	0.0719	0.0339	7	6	11	9
	NWSL	0.0497	0.0012	0.0011	0.1896	-0.0485	-0.0486	0.1399	11	7	4	5
	WSLN	0.5377	0.6856	0.5143	0.3638	0.1479	-0.0234	-0.1739	4	12	2	2
车辆类型	Motor	0.1913	0.0034	0.3468	0.2265	-0.1879	0.1555	0.0352	2	1	10	4
	Truck	0.2122	0.1284	0.1786	0.3799	-0.0838	-0.0336	0.1677	8	10	3	3
	Bus	0.5965	0.8683	0.4746	0.3936	0.2718	-0.1219	-0.2029	1	2	1	1
机动车状态	Accident	0.2507	0.2859	0.2206	0.2455	0.0352	-0.0301	-0.0052	12	11	16	14
	Illegality	0.2504	0.3159	0.2073	0.2220	0.0655	-0.0431	-0.0284	9	8	12	10
	Expired	0.2327	0.0792	0.3405	0.2897	-0.1535	0.1078	0.0570	3	3	5	6
车辆安全状况	Normal	0.2662	0.3190	0.2316	0.2428	0.0528	-0.0346	-0.0234	10	9	13	11
	Poor	0.7992	0.9201	0.7191	0.7472	0.1209	-0.0801	-0.0520	5	5	6	7
		0.2008	0.0799	0.2809	0.2528	-0.1209	0.0801	0.0520	6	4	7	8

### 2.5 情景推演结果

本节结合特征重要排序及灵敏度分析，选取由高敏状态构成的三类代表性场景： $S_1=Truck \times NWSL$ 、 $S_2=Motor \times Expired$ 、 $S_3=Truck \times Poor \times NWSL$ ，以给定的基线后验 $P_0(Acc)=\{36\%、38\%、26\%\}$ 参照，按照式(7)计算各严重度 $c$ 的 $\Delta P_c$ 及 $R_c$ ，结果显示， $S_1$ 在Fatal端出现放大， $\Delta P_c$ 为+72%， $R_c$ 为3.77； $S_2$ 以Injury端放大为主， $\Delta P_c$ 为+31%， $R_c$ 为1.82，同时Fatal端 $\Delta P_c$ 为+4%， $R_c$ 为1.15； $S_3$ 在Fatal端放大最强， $\Delta P_c$ 为+73%， $R_c$ 为3.81。总体比较，Fatal端放大强度 $S_3 > S_1 \gg S_2$ ，Injury端以 $S_2$ 为主，且均满足 $\Delta P_c > 0$ 且 $R_c > 1$ 。

为进一步解释上述度量结果，见图9，各场景在在对应严重度维度上的SHAP交互证据与上述 $\Delta P_c$ 及 $R_c$ 的结论相对应： $S_1$ 中低照明条件下Truck的点云集中于SHAP>0的高值区，照明良好时趋近于零或转负； $S_2$ 中Motor且St=Expired的点云整体为正，St=Normal时接近零； $S_3$ 中Truck×NWSL×Vs=Poor形成最强正向簇，任一条件改善即明显回落。定量与可视化结果一致表明，车辆类型与照明条件、技术状态的联合作用可使严重度分布相对基线发生系统性迁移，在 $S_1$ 与 $S_3$ 下，由Damage、Injury向Fatal迁移，在 $S_2$ 下，主要由Damage向Injury迁移。

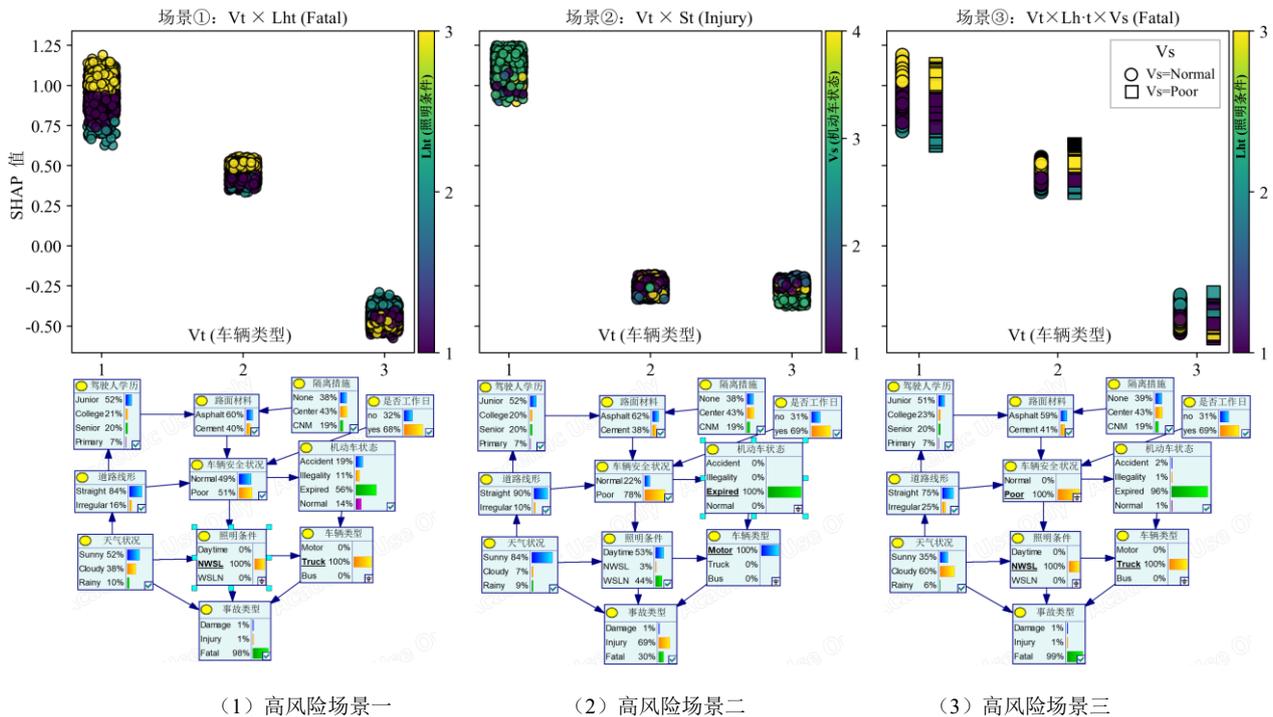


图 9 三类高风险场景的交互效应与事故严重程度推理

Fig.9 Interaction Effects and Accident Severity Inference under Three High-Risk Scenarios

### 3 结论

1) 围绕城市道路交通事故严重程度的预测与机理识别，构建“CatBoost-SHAP特征筛选—K2结构学习—MLE参数估计”的贝叶斯网络模型，该模型实现了从特征重要性识别、依赖关系建模到情境化推理的全流程分析。模型总体准确率达80.37%，三类事故的分类性能均衡且稳健，能够在多类别事故中识别不同严重程度的关键影响因素并支持多因素情景推演。

2) 贝叶斯网络模型应用结果表明，严重程度上行迁移主要由照明条件、车辆类型、车辆安全状况和机动车检验状态等因素耦合影响。灵敏度分析显示，夜间无路灯、货车、技术状况不良、逾期未检在死亡和重伤端贡献居前；摩托车对受伤端影响更为突出；隔离缺失与阴雨天气进一步加重严重程度。情景推演进一步验证：在货车行驶且处于夜间无路灯的情形下，死亡风险显著放大；摩托车逾期未检时，受伤风险明显上升；货车在技术状况不良且处于夜间无路灯的情形下，形成最强致死情境。围绕关键致因，城市交管部门可采取针对性措施，例如在管理上应优先完善夜间无路灯路段的照明与标志标线，对货车和摩托车实施区分时的差异化通行与路线引导，强化车辆技术状况检验与逾期治理，并在低能见度时段和隔离缺失路段加大巡查力度、严格限速与安全车距管理，使资源配置与高风险情境一一对应。

3) 城市道路交通事故由众多影响因素造成，而当前所用的交管业务数据不仅特征维度有限，且时空范围局限于单一城市。这导致了两个层面的限制：其一，模型更多揭示的是条件依赖而非严格的因果作用，可能引入混杂与偏差；其二，模型的泛化能力虽经内部交叉验证，但尚未在更广泛的跨时空外部数据（其他年份或城市）上得到检验。后续可融合多源异构数据进行因果发现与推理分析，进一步提升模型的决策支持能力与应用价值。

## 参考文献:

- [1] World Health Organization. World report on road traffic injury prevention[R].2023.
- [2] 中华人民共和国 2024 年国民经济和社会发展统计公报[J].中国统计,2025,(03):4-21.  
Statistical Bulletin of the People's Republic of China on National Economic and Social Development 2024[J]. China Statistics, 2025(03): 4-21. (in Chinese)
- [3] 陈宽民,王玉萍.城市道路交通事故分布特点及预防对策[J].交通运输工程学报,2003,(01):84-87.  
CHEN Kuanmin, WANG Yuping. Distribution characteristics and countermeasures of urban traffic accidents[J]. Journal of Traffic and Transportation Engineering, 2003(01): 84-87. (in Chinese)
- [4] 韦漠.基于 CiteSpace 的我国道路交通事故影响因素文献计量与可视化知识图谱分析[J].安全与环境工程,2024,31(02):17-33.  
WEI Mo. Bibliometric and visual knowledge graph analysis of road traffic accident influencing factors in China based on CiteSpace[J]. Safety and Environmental Engineering, 2024, 31(02): 17-33. (in Chinese)
- [5] 汪文武.城市道路交通事故影响因素分析及交通安全评价[D].长安大学,2023.  
WANG Wenwu. Analysis of influencing factors of urban road traffic accidents and traffic safety evaluation[D]. Xi'an: Chang'an University, 2023. (in Chinese)
- [6] 严利鑫,胡鑫辉,刘清梅,等.道路交通事故严重程度预测及致因分析[J].华东交通大学学报,2024,41(05):65-73.DOI:10.16749/j.cnki.jecjtu.20240408.002.  
YAN Lixin; HU Xinhui; LIU Qingmei; *et al.* Road traffic accident severity prediction and cause analysis [J]. *Journal of East China Jiaotong University*, 2024, 41(5): 65-73. doi:10.16749/j.cnki.jecjtu.20240408.002. (in Chinese)
- [7] Hosseinpour M, Yahaya A S, Sadullah A F. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads[J]. Accident Analysis & Prevention, 2014, 62:209-222.
- [8] 潘义勇,朱梦.不同光照条件下交通事故严重程度影响因素的可转移性分析[J].交通运输系统工程与信息,2025,25(03):346-357.  
PAN Yiyong, ZHU Meng. Transferability Analysis of Factors Influencing Severity of Traffic Accidents Under Different Lighting Conditions[J]. Journal of Transportation Systems Engineering and Information, 2025, 25(03): 346-357. (in Chinese)
- [9] LEE H, KANG M, HWANG K, *et al.* The typical AV accident scenarios in the urban area obtained by clustering and association rule mining of real-world accident reports[J/OL]. Heliyon, 2024, 10(3): e25000.
- [10] MADUSHANI J P S S, SANDAMAL R M K, MEDDAGE D P P, 等. Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers [J/OL]. Transportation Engineering, 2023, 13:100190.
- [11] KAPTAN M, BAYAZIT O. Data-driven Bayesian risk assessment of factors influencing the severity of marine accidents in port areas [J/OL]. Process Safety and Environmental Protection, 2024, 192: 1094-1109.
- [12] TANG J, PIERA M A, GUASCH T. Coloured Petri net-based traffic collision avoidance system encounter model for the analysis of potential induced collisions[J/OL]. Transportation Research Part C: Emerging Technologies, 2016, 67: 357-377.
- [13] AGYEMANG E F, MENSAH J A, OCRAN E. Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential change points[J/OL]. Heliyon, 2023, 9(12): e22544.
- [14] 张洁,张萌萌,李虹燕.基于二元 Logistic 模型的城市道路交通事故严重程度分析[J].交通信息与安全,2022,40(05):70-79.  
ZHANG Jie, ZHANG Mengmeng, LI Hongyan. An Analysis of Severity of Traffic Accidents on Urban Roadways Based on Binary Logistic Models [J]. Journal of Transport Information and Safety, 2022, 40(05): 70-79. (in Chinese)
- [15] 张道文,母尧尧,王朝健,等.城市道路交通事故特性及严重程度研究[J].安全与环境学报,2022,22(02):599-605.  
ZHANG Daowen, MU Yaoyao, WANG Zhaojian, *et al.* Research on characteristics and severity of urban road traffic accidents[J]. Journal of Safety and Environment, 2022, 22(02): 599-605. (in Chinese)

- [16] 赵丹,马社强,张雨萌,韩凤春.农村公路交叉口交通事故特征关联性与风险因素分析[J].中国安全科学学报,2020,30(07):146-151.  
ZHAO Dan, MA Sheqiang, ZHANG Yumeng, HAN Fengchun. Correlation and risk factors analysis of traffic crash at intersections on rural highways [J]. China Safety Science Journal, 2020, 30(07): 146-151. (in Chinese)
- [17] 由冰玉,廉福绵,孟祥海.基于故障树贝叶斯网的山区高速公路事故成因分析[J].交通信息与安全,2019,37(04):44-51.  
YOU Bingyu, LIAN Fumian, MENG Xianghai. An Analysis of Crash Factors for Freeways in Mountain Areas Based on Fault Tree and Bayesian Network[J]. Journal of Transport Information and Safety, 2019, 37(04): 44-51. (in Chinese)
- [18] 李世杰,王景升.基于 FTA- BN 模型的机动车与行人交通事故成因分析[J].交通工程,2022,22(06):47-54.  
LI Shijie, WANG Jingsheng. Cause analysis of vehicle-pedestrian traffic accidents based on FTA-BN model[J]. Journal of Transportation Engineering, 2022, 22(06): 47-54. (in Chinese)
- [19] 陈英,朱加勋,田毅,等.基于 K-Means-XGBoost-SHAP 集成化框架的交通事故严重程度致因分析[J].中国公路学报,2025,38(05):209-223.  
CHEN Ying, ZHU Jiayun, TIAN Yi, et al. Analysis of Factors Contributing to Traffic Accident Severity Based on an Integrated Framework of K-Means-XGBoost- SHAP[J]. China Journal of Highway and Transport, 2025, 38(05): 209-223. (in Chinese)
- [20] 成卫,马铭伟,张小龙.基于贝叶斯网络的高速公路交通事故严重程度预测及致因分析[J].重庆交通大学学报(自然科学版),2023,42(07):87-95.  
CHENG Wei, MA Mingwei, ZHANG Xiaolong. Prediction and Cause Analysis of Freeway Traffic Accident Severity Based on Bayesian Network[J]. Journal of Chongqing Jiaotong University (Natural Science Edition), 2023, 42(07): 87-95. (in Chinese)



第一作者:张豪森(2002—),男,硕士研究生,研究方向为交通安全。E-mail:zhs13172686187@foxmail.com。



通信作者:王景升(1970—),男,副教授,硕士,硕士生导师,研究方向为智能交通及交通数据分析等。E-mail:20051092@ppsuc.edu.cn。