

高速公路桥梁路段交通安全风险评价及影响要素挖掘

赵晓华¹, 杨海益¹, 姚莹¹, 郭森¹, 亓航¹, 戴义博¹, 苏岳龙²

(1. 北京工业大学城市建设学部, 北京 100124; 2. 高德软件有限公司, 北京 102299)

摘要: 为了深入探究影响高速公路桥梁路段风险的显著因素, 论文应用交通秩序指数综合评估交通安全风险, 并以此作为因变量, 研究桥梁路段交通流、道路属性及外部环境等要素和交通安全风险间的关系。在构建安全风险识别模型的基础上, 通过个体条件期望图挖掘影响高速公路桥梁路段的安全风险关联因素。结果表明: 相较于梯度提升决策树模型, 随机森林模型对于交通安全风险识别的准确性更高; 此外, 在影响因素中, 拥堵是影响桥梁路段安全风险的重要交通流因素; 在能见度较低及不良天气情况下, 交通风险较高; 并且, 跨河桥上、下游过渡段是高速公路桥梁路段中交通安全风险最高的区域。研究结果为高速公路桥梁路段的交通安全风险识别及影响要素挖掘提供了新的思路和方法, 有助于交通管理部门精准实施靶向治理。

关键词: 交通安全; 高速公路桥梁路段; 风险识别; 安全风险关联要素挖掘; 个体条件期望图

中图分类号: U491

文献标志码: A

Traffic Safety Evaluation and Feature Analysis of Freeway Bridge Sections

ZHAO Xiaohua¹, YANG Haiyi¹, YAO Ying¹, GUO Miao¹, QI Hang¹, DAI Yibo¹, SU Yuelong²

(1. Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China; 2. AutoNavi Co., Ltd., Beijing 102299, China)

Abstract: In order to explore in depth the significant factors affecting the traffic risk of bridge sections of freeway, this paper taking traffic order index (TOI) as dependent variable, and identifies the relationship between TOI and traffic characteristics, road attributes, and external environment. Based on the safety evaluation model, the Plots of individual conditional expectation

(ICE) were applied to discover the contributing factors to traffic risk of bridge sections of freeway. The results indicate that compared with the gradient boosting decision tree model, the random forest (RF) models is more accurate for traffic safety risk evaluation. In addition, among the influencing factors, congestion is an important traffic flow factor affecting the safety risk of bridge sections. Under low visibility and adverse weather conditions, traffic risks are higher. The upstream and downstream of the river-crossing bridge are considered high-risk corridors in the bridge sections of the freeway. This paper can provide new ideas and methods for traffic safety risk evaluation and discovery of contributing factors to traffic risks of bridge sections, which is helpful for the traffic management department to accurately implement management.

Key words: traffic safety; freeway bridge sections; risk identification; discovery of related factors to traffic risks; plots of individual conditional expectation

相对高速公路的普通路段、隧道、服务区等, 桥梁路段因其所处位置特殊性、外部环境复杂性, 故其发生的交通事故严重性也最高^[1]。作为高速公路交通安全治理的重要路段, 有必要对其安全风险要素开展深入分析, 以识别高风险区域及其诱发因素, 为提升高速公路桥梁路段的安全水平提供指导。

当前, 面向桥梁路段交通安全风险相关研究, 国内外学者主要关注由桥梁结构特征所产生的风险, 例如陈丰等^[2]通过驾驶模拟实验研究桥上侧风对于行车稳定性的影响; 周晋冬等^[3]采用自然驾驶数据, 论证当桥头沉降差较大时, 桥头跳车会增加驾驶员

收稿日期: 2021-10-31

基金项目: 国家自然科学基金(52072012); 中国博士后科学基金资助项(2021M690272)

第一作者: 赵晓华(1971—), 女, 教授, 博士生导师, 工学博士, 主要研究方向为驾驶行为与交通安全。

Email: zhaoxiaohua@bjut.edu.cn

通信作者: 姚莹(1991—), 女, 讲师, 硕士生导师, 工学博士, 主要研究方向为交通大数据与交通安全。

Email: yaoying@bjut.edu.cn



论文
拓展
介绍

生理、心理上的负担;此外,也有学者围绕车-桥耦合振动、极端外力条件等对桥上行车安全性的影响作用开展了相关研究^[4]。而交通流(如流量、速度、拥堵指数)、道路属性(如跨河桥上、下游过渡区)及外部环境(如天气条件、风力等级等)要素及其耦合作用作为影响桥梁交通运行安全风险的关键,由于其位置的特殊性及数据采集的困难,相关研究受到制约。随着车载导航等智能移动终端的广泛使用,通过GPS(全球定位系统)卫星定位及陀螺仪等内置传感器能够精准采集驾驶人风险驾驶行为(如急加速、急减速、急转弯等)、流量、速度、拥堵指数等参数,一方面为从风险驾驶行为角度评价桥梁路段的安全性风险提供了可能,另一方面也为交通安全风险影响要素的深度解析提供了机遇。

针对交通安全风险评估问题,其解决方法主要以事故数据为因变量构建其与人、车、路、环境间的关联关系模型。例如You等^[5]基于支持向量机建立事故预测模型,分析速度标准差、流量等因素对事故发生频次的影响;马聪等^[6]通过建立基于负二项和非线性负二项回归模型的交通事故次数预测模型,指出道路湿滑、起雾是造成交通事故主要原因。事实上,受制于交通事故小概率、偶发性、滞后性的特点,近年来,也有研究开始关注驾驶行为、速度波动等变量与交通事故间的关联关系,并以风险驾驶行为和速度差等风险替代指标评估交通风险,挖掘其与影响要素的关系^[7],以期将交通安全风险的防控从事后分析转向事前预防,这些研究均取得一定的进展。由于基于事前替代指标的安全风险评价可实现全域的实时评估与诊断,突破事故数据的制约,以替代指标为核心的安全评价与关联要素挖掘研究已受到学者们的广泛关注^[8],其具有良好的借鉴价值。

针对风险关联要素挖掘问题,相较传统统计模型、线性模型和因果推断模型,基于树模型的机器学习算法虽具有更强大的非线性分类和自学习能力,在识别多源要素耦合作用下的交通安全风险更具优势^[9],但可解释性的匮乏是阻碍其应用的重要因素。而部分依赖图(partial dependence plots, PDP)技术作为一种新的机器学习模型解析方法,可量化描述树模型的特征变量对交通安全风险的影响作用,为交通风险影响要素挖掘提供新的途径,也为交通管理部门及政策制定者开展精准的道路风险研判提供依据。

综上所述,本文针对高速公路桥梁路段的安全评价及关联因素挖掘问题开展研究,论文以风险驾

驶行为及交通运行状态为核心形成交通秩序指数以评价交通安全风险,突破交通事故数据不均衡的限制,及对于潜在的交通安全风险影响要素深度挖掘问题,分别建立基于机器学习模型(随机森林、梯度提升决策树)的高速公路桥梁路段风险识别模型,并通过个体条件期望图挖掘交通流特征、路段属性、外部环境条件、时间属性与高速公路桥梁路段风险等级间的量化关系,以期实现对多因素安全风险间耦合作用的深度挖掘,消除基于部分依赖图的影响要素解析中由数据异质性及非均匀效应的影响。研究结果为解决高速公路桥梁路段安全问题提供了新的视角,为主动安全防控和靶向治理方案提供理论支持。

1 数据来源与指标提取

本研究以鄂东长江大桥为研究对象,其是中国湖北省境内连接黄石市和黄冈市的重要过江通道。论文所选数据来源于高德导航软件的浮动车轨迹数据和当地交通管理部门的事故调查数据。高德导航软件数据集包含路段编号、交通流运行状态、外部环境及驾驶行为。事故调查数据集包含事故发生时间及所处的道路桩号。

1.1 路段单元划分

研究以同质性为划分原则,将鄂东长江大桥划分为6个区域,分别为高速公路路基直线段、高速公路路基曲线段、跨河桥上游过渡段、跨河桥下游过渡段、跨河桥段、跨线桥段。但由于跨河桥上、下游过渡区分别连接圆曲线和互通立交匝道,而部分学者已证明圆曲线和互通立交匝道对驾驶行为有较大负面影响^[10-11]。为防止这两个特殊区域对风险识别模型的影响,根据美国道路通行能力手册,入口匝道的影响缓冲区设置为450 m^[11]。不同路段的起终点桩号范围、路段长度详见表1。

1.2 关键参数提取

在原始数据集中,部分变量可直接用于风险影响因素解析,例如路段限速值、能见度等。其余变量需在原始数据的基础上提取获得新的特征变量,例如速度变异系数、延误系数等。具体定义如下。

1.2.1 激进驾驶行为频次

激进驾驶行为事件主要包含急加速、急减速、急左并道、急右并道、急左转和急右转。激进驾驶行为数据是由手机传感器采集,其可精准辨识用户在行车过程中实时发生的激进驾驶行为事件,并记录事件发生的

表 1 各路段单元的桩号范围

Tab. 1 Stake number of each unit partition

路段类型	上行起点桩号	上行终点桩号	下行起点桩号	下行终点桩号	路段长度/m
高速公路路基曲线段	K787+195	K788+994	K788+994	K787+195	1 799
跨河桥上游过渡段	K791+373	K791+670	K791+670	K791+373	297
跨河桥段	K791+670	K793+156	K793+156	K791+670	1 486
跨河桥下游过渡段	K793+156	K793+418	K793+418	K793+156	262
跨线桥段	K793+418	K796+315	K796+315	K793+418	2 897
高速公路路基直线段	K796+950	K798+400	K798+400	K796+950	1 450

时间和经纬度。而受用户隐私保护协议的限制,具体的判定算法无法给出。此外,因不同道路长度和某时段内导航软件用户数量的差异,原始数据无法在同一尺度下衡量路段风险程度。故需将原始激进驾驶行为发生的次数转化为集计数据后再开展分析,集计激进驾驶行为频次 FE_{ijk} 的计算方法如下:

$$FE_{ijk} = \frac{\text{sum}(\text{event}_{ijk})}{l_k \cdot UP_{ijk}} \quad (1)$$

式中: i 与 j 分别表示某天 and 某时 ($i=1, 2, 3, \dots, 62, j=0, 1, 2, \dots, 23$); k 表示路段编号; FE_{ijk} 为 i 日 j 时 k 路段发生的驾驶行为事件频次; $\text{sum}(\text{event}_{ijk})$ 为 i 日 j 时 k 路段发生的激进驾驶行为次数之和; l_k 表示路段长度; UP_{ijk} 为 i 日 j 时 k 路段的导航软件实际用户量。

1.2.2 速度变异指数

相关研究表明,高风险和较高的事故频次与某路段内交通流平均速度的离散程度和速度变异系数 (coefficient of speed variation, CSV) 有显著关联^[12]。速度变异系数是评价交通流速度离散程度的指标。速度变异系数越大,则表明交通流速度的分布越离散,交通运行状态越混乱,交通安全风险越大。速度变异系数 CSV_{ijk} 的计算公式如下:

$$CSV_{ijk} = \frac{\sigma_{v_{ijk}}}{\bar{v}_{ijk}} \quad (2)$$

式中: $\sigma_{v_{ijk}}$ 为 i 日 j 时 k 路段的交通流速度标准差; \bar{v}_{ijk} 为 i 日 j 时 k 路段的交通流速度平均值。

1.2.3 延误系数

拥堵会对交通安全产生负面影响^[12]。延误系数可有效衡量路段内的拥堵程度。延误系数 CI_{ijk} 的计算公式如下:

$$CI_{ijk} = \frac{v_k}{\bar{v}_{ijk}} \quad (3)$$

式中: v_k 表示 k 路段的自由流速度。

1.2.4 交通流量等级

为了表征交通流量在不同时段的差异性,使用 K 均值聚类算法,将导航软件的用户量离散化为三

类,分别为低峰期、平峰期、高峰期,并以聚类后各类的最大值、最小值设定区间划分阈值,具体定义如下所示。交通流量等级是一种状态参量,而非具体量值。

低峰期: $UP_{ijk} \in [0, 343)$

平峰期: $UP_{ijk} \in [343, 757)$

高峰期: $UP_{ijk} \in [757, +\infty)$

1.3 交通秩序指数

当前,交通安全分析主要以事故作为评价指标,但交通事故具有偶发性、局部性、滞后性等特质,无法覆盖全时空域的风险场景,易忽略事故发生时的潜在风险要素,实际应用过程中,难以开展全局动态交通安全评估与辨识。此外,由于事故数据数量和质量限制,传统统计分析及大数据机器学习算法均难以发挥真正效能,导致影响因素解析能力不足。借助导航软件,依托海量驾驶行为及交通流状态数据,为提出以风险替代指标为核心的交通安全风险评价方法提供了新的机遇。Yao 等^[13]提出了交通秩序指数 (traffic order index, TOI),并通过交通秩序指数评价城市道路的安全性和顺畅性。交通秩序指数是一种基于激进驾驶行为和速度变化特性的交通安全风险替代指标。Yao 等^[13]通过关联交通秩序指数与事故数间的关系发现,交通事故数多且事故持续时长较长的路段,交通秩序指数较低。总之,当路段内激进驾驶行为频发,且交通流紊乱时,其交通秩序指数较低,交通安全风险较高,秩序指数代表风险发生的概率,其计算方法详见文献^[13],简要计算流程如下所述。

交通秩序指数是由优劣解距离法 (technique for order preference by similarity to an ideal solution, TOPSIS) 及熵权法等算法将多类激进驾驶行为与速度变异系数加权而得。交通秩序指数的计算方法如式(4)所示。

$$TOI_{ijk} = \omega_{db} \cdot S_{db_{ijk}} + \omega_{csv} \cdot S_{csv_{ijk}} \quad (4)$$

式中: TOI_{ijk} 为 i 日 j 时 k 路段的交通秩序指数; $S_{db_{ijk}}$ 、 $S_{csv_{ijk}}$ 表示激进驾驶行为和车流速度波动的风险评

分; w_k 为激进驾驶行为和速度波动风险评分在交通秩序指数中所占的权重, $k=\{db, CSV\}$ 。

激进驾驶行为风险评分 $S_{db_{ijk}}$ 是由 TOPSIS 算法综合急加速、急减速、急左转等激进驾驶行为频次计算而得。TOPSIS 算法是一种多目标决策分析中常用的有效方法, 其已在交通安全分析领域获得广泛应用。而速度波动风险评分 $S_{CSV_{ijk}}$ 的计算方法如公式(5)所示:

$$S_{CSV_{ijk}} = \frac{y_{ijk} - y_{\min}}{y_{\max} - y_{\min}} \quad (5)$$

式中: $y_{ijk} = \frac{1}{CSV_{ijk}}$; y_{\max} 、 y_{\min} 分别表示 y_{ijk} 中的最大值和最小值。

激进驾驶行为和速度波动风险评分在交通秩序指数中所占的权重 w_k 由式(6)求得:

$$w_k = \frac{\sigma_k}{\sigma_{db} + \sigma_{CSV}} \quad (6)$$

式中: σ_k 表示激进驾驶行为或车流速度波动风险评分的样本标准差。

交通秩序指数是综合评判道路秩序程度的指标, 为了使其能研判路段内的风险水平, 故采用 K-均值聚类方法将 TOI_{ijk} 划为三个风险等级, 并以聚类后各类别的最大值、最小值设定区间划分阈值, 具体定义如下:

高风险道路: 交通秩序较差, 风险较高: $TOI_{ijk} \in [0, 0.3598)$

中风险道路: 交通秩序良好, 风险适中: $TOI_{ijk} \in [0.3598, 0.4107)$

低风险道路: 交通秩序较好, 风险较低: $TOI_{ijk} \in [0.4107, +\infty)$

1.4 数据库构建

以桥梁路段单元及时间戳为基础搭建时空网格, 并匹配所有类型数据, 构建桥梁路段安全分析数据库, 共计 17 856 条(6 个路段类型 \times 2 个方向 \times 62 d \times 24 h) 数据, 删除空缺及异常数据后, 共计 17 160 条数据。数据库中各变量的统计性描述如表 2 及表 3 所示。

表 2 分类变量的定义与统计

Tab. 2 Definition and statistics of categorical variables

风险类别	特征名称	类别说明	数量	占比/%	
因变量	桥梁路段风险水平	低风险道路	3 157	18.40	
		中风险道路	12 937	75.39	
		高风险道路	1 066	6.21	
交通流特征	交通流量等级	低峰期	8 775	51.14	
		平峰期	6 945	40.47	
		高峰期	1 440	8.39	
	路段类型	跨河桥	2 961	17.26	
		跨线桥	2 642	15.40	
		路基直线段	2 949	17.18	
道路属性	路段类型	路基曲线段	2 697	15.72	
		跨河桥上游过渡段	2 954	17.21	
	路段限速值	跨河桥下游过渡段	2 957	17.23	
		100 km·h ⁻¹	2 959	17.24	
		110 km·h ⁻¹	5 906	34.42	
自变量	天气状况	120 km·h ⁻¹	8 295	48.34	
		晴	8 578	49.99	
		多云	1 405	8.19	
	外部环境	天气状况	阴天	6 473	37.72
			雨	682	3.97
		风力等级	雾	22	0.13
			无风	3 855	22.46
			1级	7 169	41.78
	运行时段	工休日	2级	4 688	27.32
			3级	1 330	7.75
		照明情况	4级	118	0.69
			节假日	5 565	53.97
工作日			11 595	46.03	
	白天	9 261	32.43		
	黑天	7 899	67.57		

表 3 连续变量的定义与统计

Tab. 3 Definition and statistics of continuous variables

风险类别	特征名称	平均值	标准差	最大值	最小值
自变量	延误系数	1.054	0.284	0.908	22.602
	交通流特征 平均车流速度与限速值差/(km·h ⁻¹)	28.922	15.470	99.060	1.226
外部环境	降雨量/(mm·h ⁻¹)	0.887	4.626	51	0
	能见度/km	16.007	8.647	30	0.1

2 安全分析方法

2.1 高速公路桥梁路段安全风险识别模型

随机森林是一种以决策树为基础的算法,在训练模型的过程中引入随机属性选择,并采用引导聚集方法,最终结合多个弱分类器形成强分类器。由于随机森林模型结合了多个弱分类器的预测结果,故随机森林模型可以较好地控制过拟合问题。此外,随机森林模型相较于人工神经网络、支持向量机等,其在小样本集上拥有更加出色的分类识别能力^[14]。基于此,随机森林模型更适于数据量较小、数据可能存在异质性等情况。简易过程如下:

在已知训练集 $\{(x_1, y_1), \dots, (x_n, y_n)\}$, 模型将随机重复采样 B 次, 获得 $b = 1, \dots, B$ 的不同训练集 x' 用以训练弱分类器 f_b 。训练后的模型通过计算 B 个弱分类器的类别概率均值, 获得最终的分类结果, 如式(7)所示:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (7)$$

式中: \hat{f} 为结合 B 个弱分类器的预测值所形成的预测均值, 即为随机森林模型的最终预测结果。

本文选择梯度提升决策树模型为风险等级识别性能的对比模型。与随机森林模型不同在于, 随机森林是一种引导聚集算法, 该类算法会对样本重采样, 预测结果是各个分类器的平均值。而梯度提升决策树模型是一种提升算法, 基于上次迭代后预测器的分类结果更新样本权值和分类器权值, 因此随着模型迭代次数的增加, 模型的预测偏差会降低。这两种模型都可有效控制过拟合问题, 且在不同的应用场景和数据特质下有不同的性能表现。

此外, 在建立高速公路桥梁路段安全风险识别模型时, 均需要完成以下步骤:

(1) 基于随机分布原则按 7:3 的比例将数据集划分为训练集和测试集, 训练集用于训练识别模型, 测试集用于评估模型的性能;

(2) 训练模型时, 采用十折交叉验证方法将训练集的数据随机切分为 10 份, 每次训练会以 9 个子

本结合 1 个子样本验证的形式进行;

(3) 本文使用网格搜索方式进行参数调优, 实现指定参数值的穷举式搜索。

2.2 模型评价指标

为了评估安全风险识别模型的性能, 以机器学习模型中常用的分类评价指标准确率、召回率、精确率和 F_1 ·Score 作为模型评价指标。

对于多分类预测, 以各类别所占样本集的比例为权重, 计算各类别中每个评价指标的加权平均值。召回率 (Recall)、精确率 (Precision)、 F_1 ·Score 的相关定义如下:

$$\text{Recall} = \sum_{i=1}^M \omega_i \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (8)$$

$$\text{Precision} = \sum_{i=1}^M \omega_i \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \quad (9)$$

$$F_1 \cdot \text{Score} = \sum_{i=1}^M \omega_i \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (10)$$

式中: i 表示交通风险等级的类别编号 ($i = 1, 2, 3$); M 为交通风险等级的总类别数, 即 $M = 3$, ω_i 为第 i 类在全样本集所占的比例; TP_i 、 FN_i 、 FP_i 、 TN_i 的相关定义请见表 4。

表 4 i 类风险的混淆矩阵Tab. 4 Confusion matrix of i risk level

真实值	预测值	
	预测为 i 类风险	预测为非 i 类风险
真实为 i 类风险	TP_i	FN_i
真实为非 i 类风险	FP_i	TN_i

2.3 个体条件期望图

传统的部分依赖图可反映响应变量和一个或多个特征间依赖关系的平均值, 其可能掩盖由特征耦合作用产生的异质关系, 只有当特征间相关交互较弱时, 才可客观反映特征对自变量的影响关系。由于上述原因, Goldstein 等^[15]提出了一种全新的机器学习黑箱模型可视化工具——个体条件期望图 (plots of individual conditional expectation, ICE), 可用于监督学习算法的预测模型。ICE 通过计算多组测试集的依赖关系以分解部分依赖值, 有助于消除非均匀效应的影响, 并可通过绘制响应变量和个别

观测特征间的函数关系细化图形,便于深刻理解影响关系。

设 $S \in \{1, \dots, P\}$, 且 C 为 S 的补集。此时, S 测试样本集下, x_s 的依赖关系可以表达为

$$f_s = \mathbb{R}x_c[f(x_s, x_c)] = \int f(x_s, x_c) dP(x_c) \quad (11)$$

每个子集 S 都有其个体的依赖关系 f_s , 当特征 x_s 固定且 x_c 变化时, f_s 可给出此时 x_c 的边际贡献 $dP(x_c)$ 。由于无法直接求解 f 和 $dP(x_c)$ 的值, 可通过计算公式(12)推算公式(11)中的值。

$$f_s = \frac{1}{N} \sum_{i=1}^N f(x_s, x_{c_i}) \quad (12)$$

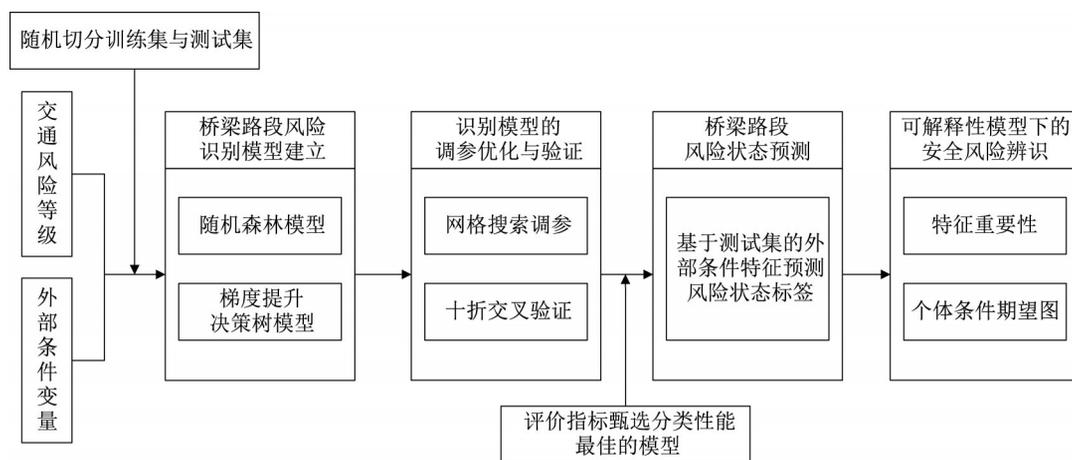


图1 模型构建及分析流程图

Fig. 1 Flowchart of modeling and feature analysis

3 模型构建与结果分析

3.1 模型性能对比

网格搜索优化调参及十折交叉验证后, 随机森林及梯度提升决策树模型的混淆矩阵如图2所示。经计算, 随机森林模型的准确率、召回率、精确率及 $F_1 \cdot \text{Score}$ 分别为 89.35%、89.35%、89.71%、89.50%, 均高于梯度提升决策树模型的 88.70%、88.70%、88.58%、88.60%。结果表明, 随机森林模型在风险等级识别方面优于梯度提升决策树模型, 其可更好地辨识高风险、中风险、低风险道路的特征。因此基于随机森林的安全风险识别模型将用于下一步分析。

3.2 特征重要性分析

特征重要性是一种衡量某特征对全局预测结果影响程度的指标, 特征重要性越高, 表明该特征对风险等级的识别结果影响作用越大。由图3所示, 特

式中: N 为 S 集中样本的个数; $\{x_{c_1}, \dots, x_{c_N}\}$ 为 S 集中 x_c 的不同取值。

2.4 模型框架及流程图

首先, 论文分别构建基于随机森林及梯度提升决策树的安全风险识别模型, 并通过网格搜索调参及交叉验证提升识别模型的准确性。基于样本权重的召回率、精确率等指标评估模型的识别性能, 并在后续分析中保留性能最佳的模型。其次, 基于基尼指数的特征重要性排序, 筛选对风险等级识别结果影响最大的7个影响要素。最后, 应用个体条件期望图解析这些变量与风险等级间的关联关系。模型构建及分析流程图如图1所示。

征重要性由大至小依次是延误系数、平均车流速度与限速值差、能见度、天气状况、交通流量等级、风力等级、路段类型。

3.3 个体条件期望图分析

个体条件期望值是综合衡量全样本集中, 仅当某单因素变化时导致分类结果变化程度的指标。若某因素水平下的个体条件依赖值均值 \bar{f}_s 越高, 则表明模型分类结果为高风险等级的概率越高。即在当前外部条件下, 交通秩序差, 安全风险高。

3.3.1 交通流特征相关变量

在交通流特征类变量中, 延误系数、平均车流速度与限速值差、交通流量等级会对风险等级的识别结果有较大影响。

延误系数方面, 由图4a可知, 随延误系数增长, \bar{f}_s 上升。即交通拥堵越严重, 安全风险越高, 具体而言, 延误系数每增加0.1, 高风险等级的发生概率将增加8%。这与大多数相关研究结论保持一致^[16],

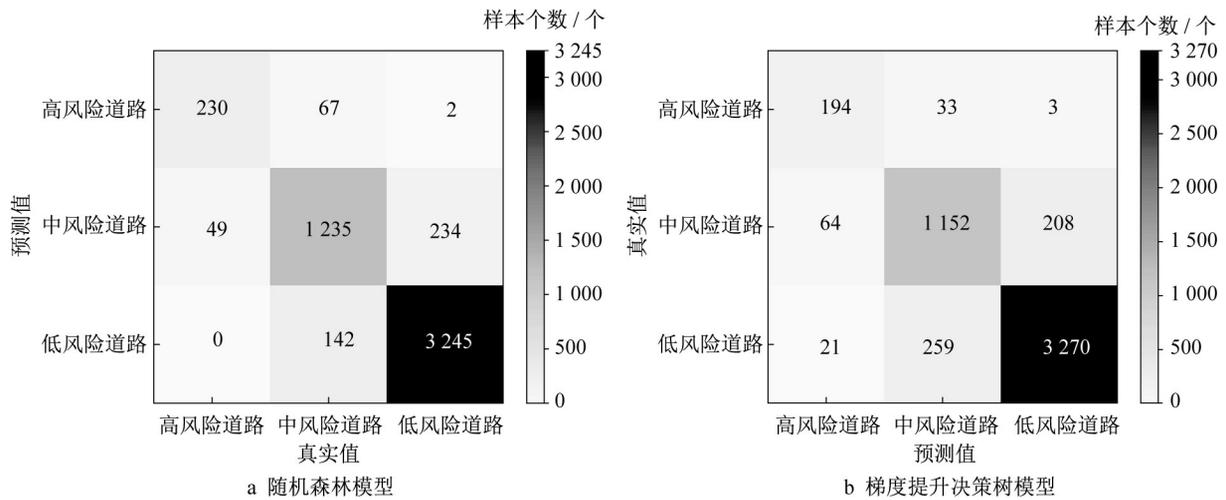


图 2 风险等级识别模型的混淆矩阵

Fig. 2 Confusion matrix of risk level identification models

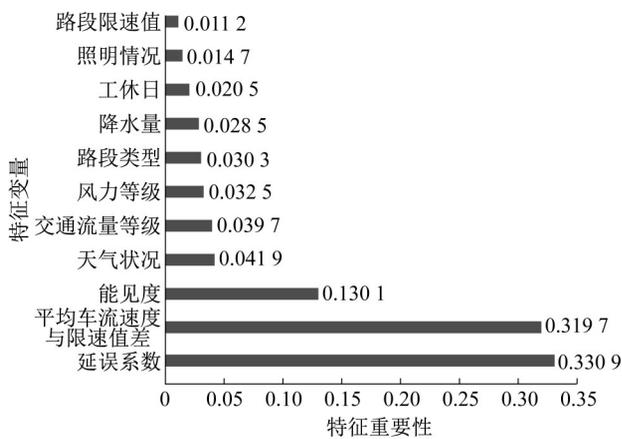


图 3 随机森林模型的特征重要性排序

Fig. 3 Feature importance score in RF model

交通拥堵严重程度越高,驾驶员的行为越倾向于激

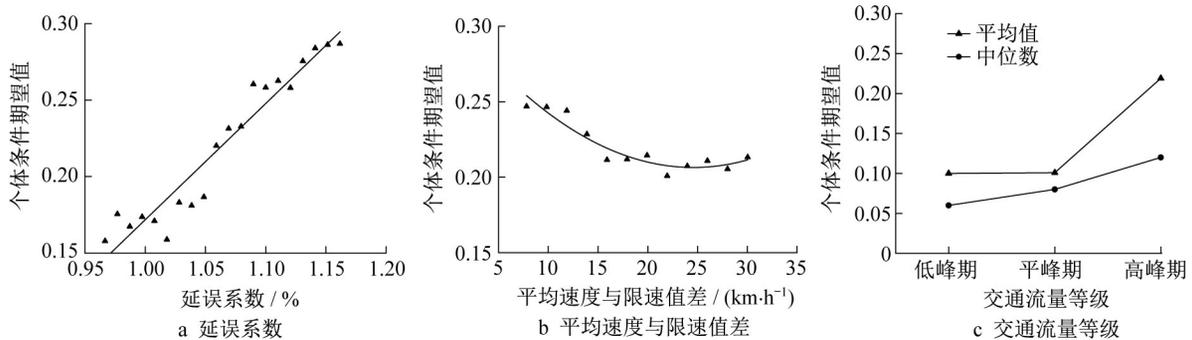


图 4 交通流特征的个体条件期望图

Fig. 4 Plots of individual conditional expectation of traffic flow factors

交通流量等级方面,图 4c 表明,高峰期间, \bar{f}_s 为 0.219, 高于低峰期 \bar{f}_s 的 0.100。文献[19]分析得出了类似的结果,交通流量的增长通常与事故发生频次正向相关。

进,对交通安全造成较大的负面影响。

平均车流速度与限速值差方面,由图 4b 可见,随平均车流速度与限速值差增大,该特征对道路安全的影响呈非线性变化趋势。在平均车流速度与限速值差增长时, \bar{f}_s 逐步降低,随其进一步增至 $24 \text{ km} \cdot \text{h}^{-1}$ 后, \bar{f}_s 反之上升,表明安全风险在增加。这两种情况可通过两种交通流状况解释。一方面,在平均车流速度与限速值差较小时,则说明此时平均车流速度较高, You 等^[17]研究得出在交通流较为顺畅时,驾驶员往往会更加激进,从而产生更高的追尾或侧面碰撞的风险。另一方面,随着二者速度差增加到一定限度,此时的情形与拥堵类似,过于密集的交通流会使驾驶员频繁切换车道,安全有序性较差^[18]。

3.3.2 外部环境相关变量

在外部环境类变量中,能见度、风力等级和天气状况会对风险等级的识别结果有较大影响。

天气情况方面,由图 5a 可见,不良天气如雾、雨

条件下, \bar{f}_s 较其他天气状况显著升高, 均值分别为 0.241、0.199, 表明在极端天气下, 安全秩序水平较低, 交通运行状态处于高风险的概率较大。相关研究发现了相同的结论。Yonas 等^[20]认为在雾天情况下, 驾驶员更容易出现风险驾驶行为, 且容易降低行车的稳定性。陈昭明等^[18]发现雨天时会增加事故中乘员受伤的概率。

风力等级方面, 由图 5b 可见, 在 4 级风的情况下, \bar{f}_s 值较高, 其表明安全有序性较差, 风险较高。与

其他文献研究结果类似, 陈丰^[2]的研究指出当风力较强时, 会影响驾驶员操纵车辆的能力, 降低行车的稳定性。但是 \bar{f}_s 增长趋势并不明显。相较于无风状态下, 最大风力等级的 \bar{f}_s 仅高 0.017。这可能与数据集中不同风力等级的数据比例有关, 3 级及以上的风力等级仅占数据集的 8.44%。结论表明: 该变量需要收集更大风力等级时的相关数据, 才可能得到更有意义的分析结果。

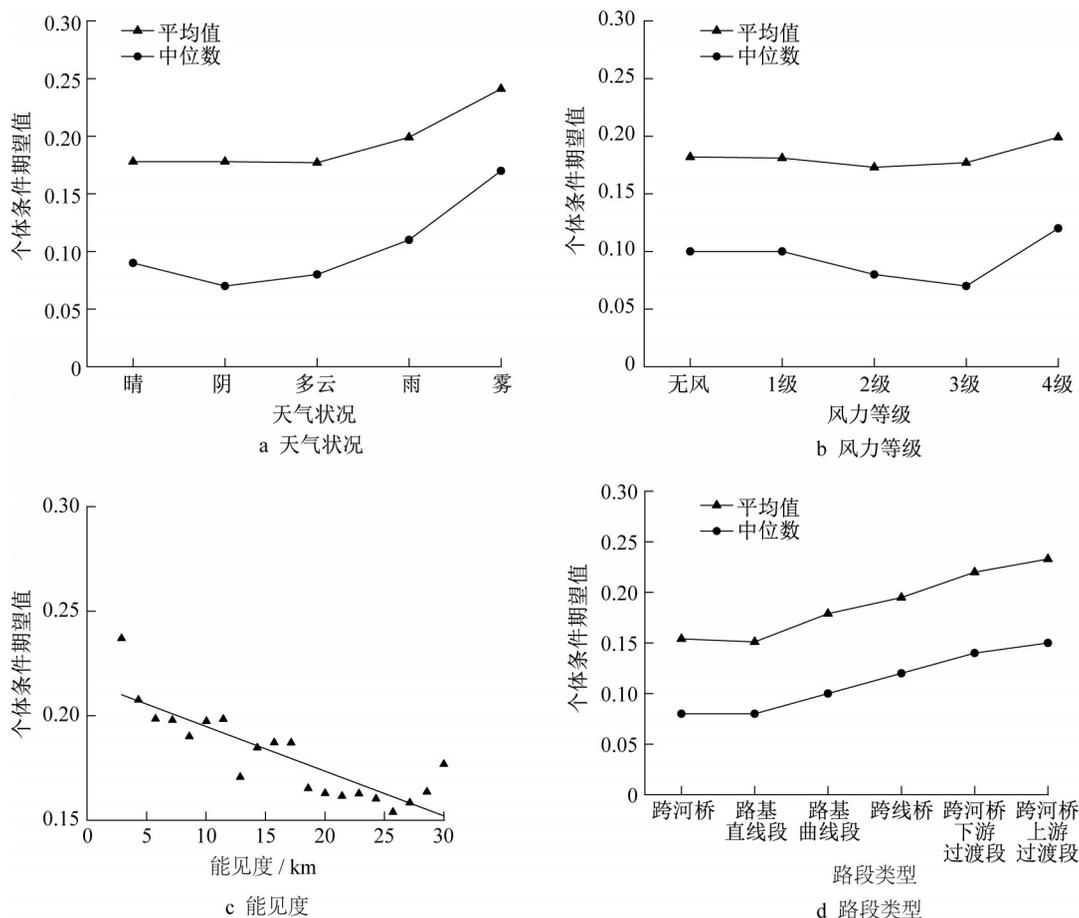


图5 外部环境及道路属性的个体条件期望图

Fig. 5 Plots of individual conditional expectation of external environment factors and road condition factors

能见度方面, 如图 5c 所示, 能见度与 \bar{f}_s 呈负相关关系, 即随能见度下降, 安全风险升高。这与 Yonas 等^[20]的研究相一致, 在低能见度下驾驶员识别接近和远离物体的能力有所下降, 无法精准辨识周围的道路及车辆状况, 容易错误判断当前的行车状况, 采取急刹车等行为, 车辆行驶的风险将大大增加。

3.3.3 道路属性相关变量

路段类型方面, 由图 5d 可见, \bar{f}_s 由大至小依次为: 跨河桥上游过渡段 (0.233) > 跨河桥下游过渡段 (0.220) > 跨线桥 (0.195) > 路基曲线段 (0.179) >

跨河桥 (0.154) > 路基直线段 (0.151)。

针对跨河桥上、下游过渡段, 可能造成其高风险的原因是: 首先, 在这两个路段内均存在桥梁伸缩缝。伸缩缝是一种桥梁特殊构造物, 会对行车造成负面影响, 其造成的桥头跳车可能导致驾驶员的心率失常、行车不舒适等^[21]。其次, 在跨河桥路段的限速为 $100 \text{ km} \cdot \text{h}^{-1}$, 而跨河桥上、下游过渡段的限速分别为 120 、 $110 \text{ km} \cdot \text{h}^{-1}$, 驾驶员看到限速标志后, 驾驶员会迫切改变自己的行车速度。此时, 驾驶员容易产生急加速或急减速等激进驾驶行为。而上下游的

速度差异性也是影响交通安全的重要因素之一,Cai等^[12]在研究中发现,当上下游的车流平均速度差较大时,发生交通事故的概率也同步增加。

根据 2016 年至 2019 年 4 月内不同路段类型的安全事故空间分布,如表 5 所示,可评估基于秩序安全分析与事故安全分析关联结果。单位公里事故数由大至小依次为:跨河桥上游过渡段(33.989) > 跨河桥下游过渡段(21.467) > 跨线桥(19.158) > 路基直线段(13.103) > 路基曲线段(13.063) > 跨河桥(12.876)。结果表明,以基于导航数据的交通秩序

指数作为风险替代指标,可在一定程度上反映道路安全风险程度。在宏观层面,与事故安全分析所得出的结论基本相同。但交通秩序指数与单位公里事故数间并非呈现线性关系,交通秩序指数更多地关注于激进驾驶行为和速度变化特性层面的运行风险,而非交通事故本身。未来研究中,将进一步寻找二者间深层次的关联关系。此外,也可拓展更多的研究对象,例如城市快速路、交叉口等,论证更多情景下交通秩序指数的有效性。

表 5 2016 年至 2019 年 4 月不同路段类型的事故空间分布

Tab. 5 Spatial distribution of accidents in different road sections from January 2016 to April 2019

	2016 年	2017 年	2018 年	2019 年 1 月—4 月	路段长度/km	单位公里事故数
跨河桥 上游过渡区	1	5	7	6	0.559	33.989
跨河桥 下游过渡区	2	4	3	3	0.559	21.467
跨线桥段	11	34	51	15	5.794	19.158
高速公路路基直线段	1	6	23	8	1.270	13.103
高速公路路基曲线段	3	17	15	12	3.598	13.063
跨河桥段	8	14	12	4	2.972	12.786

4 结语

论文主要依托导航软件所采集的高频、精准、海量的浮动车轨迹数据,采用基于激进驾驶行为和速度变化特性的交通秩序指数作为安全风险替代指标,主要贡献在于突破了传统事故分析中偶发性、局部性、滞后性等局限,实现全时空域下高速公路桥梁路段的交通安全风险辨识。此外,研究应用个体条件期望图描述及可视化黑箱模型的内部影响关系。相较于传统的部分依赖图,个体条件期望图可在一定程度上避免数据异质性及非均匀效应的影响。所获得的主要结论如下:

(1)道路属性方面,跨河桥上游及下游过渡区是高速公路桥梁路段交通秩序较差的区域,存在较大的交通安全风险;

(2)交通流方面,交通拥堵的严重程度与安全秩序水平呈负相关关系。延误系数每增加 0.1,高风险等级的发生概率将增加 8%。此外,平均车速与限速值差与安全风险间呈非线性关系,当二者差值约为 $24 \text{ km} \cdot \text{h}^{-1}$ 时,安全风险较低;

(3)外部环境方面,在能见度较低及不良天气情况下,交通风险较高;随桥梁侧风作用逐步增强,安全风险略微上升。

本文研究结果对于桥梁路段行车安全改善工程和风险防控措施具有参考意义,特别是在主动安全防控方案的设计具有理论支撑作用。在跨河桥上游过渡区及下游过渡区路段宜统一限速,并设置可变信息情报板、注意桥头跳车标志等。在桥梁路段可增设雾天行车诱导灯,一定程度上缓解由于能见度较低及不良天气条件导致的跟车距离难以判断、道路轮廓模糊等驾驶风险。

高速公路桥梁路段的交通安全风险还与桥梁的跨径、桥型、跨越对象、是否设置风屏障等结构特性有关,未来可采集不同类型的桥梁路段行车数据开展秩序分析,进一步深入探究桥梁结构因素对交通安全风险的影响。此外,本文仅简单分析事故与交通安全风险替代指标间的对应结果,未来可探究二者间的深层次关联关系。

作者贡献声明:

赵晓华:提出论点,文献综述,模型推导,初稿修改,校稿。

杨海益:方法论,模型编程实现及构建,初稿撰写。

姚莹:论文修改,数据分析,模型调参及优化。

郭淼:研究内容,论文写作与修改。

亓航:模型构建,文献总结。

戴义博:数据库构建,文献总结。

苏岳龙:需求调研,提供数据。

参考文献:

- [1] SUN Zongyuan, LIU Shuo, LI Dongxue, *et al.* Crash analysis of mountainous freeways with high bridge and tunnel ratios using road scenario-based discretization[J]. *Plos One*, 2020, 15(8): 1.
- [2] 陈丰, 彭浩荣, 马小翔, 等. 侧风作用下货车驾驶员反应行为模型[J]. 同济大学学报(自然科学版), 2020, 48(5): 702.
CHEN Feng, PENG Haorong, MA Xiaoxiang, *et al.* Model of driving behavior of truck driver under crosswind[J]. *Journal of Tongji University (Natural Science)*, 2020, 48(5): 702.
- [3] 周晋冬, 潘晓东, 马小翔. 桥头跳车对人体心率负荷强度影响因素的研究[J]. 公路工程, 2013, 38(6): 38.
ZHOU Jindong, PAN Xiaodong, MA Xiaoxiang. Research on the factors that affecting the human heart rate intensity of vehicle bumping at bridge-head [J]. *Highway Engineering*, 2013, 38(6): 38.
- [4] 《中国公路学报》编辑部. 中国桥梁工程学术研究综述·2021 [J]. 中国公路学报, 2021, 34(2): 1.
Editorial Department of China Journal of Highway and Transport. Review on China's bridge engineering research: 2021 [J]. *China Journal of Highway and Transport*, 2021, 34(2): 1.
- [5] YOU Jinming, WANG Junhua, GUO Jingqiu. Real-time crash prediction on freeways using data mining and emerging techniques [J]. *Journal of Modern Transportation*, 2017, 25(2): 116.
- [6] 马聪, 张生瑞, 马壮林, 等. 高速公路交通事故非线性负二项预测模型[J]. 中国公路学报, 2018, 31(11): 176.
MA Cong, ZHANG Shengrui, MA Zhuanglin, *et al.* Nonlinear negative binomial regression model of expressway traffic accident frequency prediction [J]. *China Journal of Highway and Transport*, 2018, 31(11): 176.
- [7] CHEN Shengdi, ZHANG Shiwen, XING Yingying, *et al.* The impact of truck proportion on traffic safety using surrogate safety measures in China [J]. *Journal of Advanced Transportation*, 2020, 2020(6): 1.
- [8] 陆建, 程泽阳. 道路交通安全网络风险辨识研究进展[J]. 东南大学学报(自然科学版), 2019, 49(2): 201.
LU Jian, Cheng Zeyang. Research and development of road traffic network security risk identification [J]. *Journal of Southeast University (Natural Science)*, 2019, 49(2): 201.
- [9] MAHMOUD Nada, ABDEL-ATY Mohamed, CAI Qing, *et al.* Vulnerable road users' crash hotspot identification on multi-lane arterial roads using estimated exposure and considering context classification [J]. *Accident Analysis & Prevention*, 2021, 159: 1.
- [10] 吴萌. 高速公路匝道口驾驶负荷特性研究[D]. 武汉: 武汉理工大学, 2013.
WU Meng. Study on driving workload characteristics of highway ramp area [D]. Wuhan: Wuhan University of Technology, 2013.
- [11] Transportation Research Board. Highway capacity manual [R]. Washington D C: National Research Council, 2000.
- [12] CAI Qing, ABDEL-ATY Mohamed, YUAN Jinghui, *et al.* Real-time crash prediction on expressways using deep generative models [J]. *Transportation Research Part C Emerging Technologies*, 2020, 117(1): 1.
- [13] YAO Ying, ZHAO Xiaohua, ZHANG Yunlong, *et al.* Development of urban road order index based on driving behavior and speed variation [J]. *Transportation Research Board*, 2019, 2673(7): 466.
- [14] MAFI S, ABDELRAZIG Y, DOCZY R. Analysis of gap acceptance behavior for unprotected right and left turning maneuvers at signalized intersections using data mining methods: A driving simulation approach [J]. *Transportation Research Board*, 2018, 2672(38): 160.
- [15] GOLDSTEIN A, KAPELNER A, BLEICH J, *et al.* Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation [J]. *Journal of Computational and Graphical Statistics*, 2015, 24(1): 44.
- [16] LI Guofa, LAI Weijian, SUI Xiaoxuan, *et al.* Influence of traffic congestion on driver behavior in post-congestion driving [J]. *Accident Analysis & Prevention*, 2020, 141: 1.
- [17] YOU Jinming, WANG Junhua, GUO Jingqiu. Real-time crash prediction on freeways using data mining and emerging techniques [J]. *Journal of Modern Transportation*, 2017, 25(2): 116.
- [18] 陈昭明, 徐文远, 曲悠扬, 等. 基于混合Logit模型的高速公路交通事故严重程度分析[J]. 交通信息与安全, 2019, 37(3): 42.
CHEN Zhaoming, XU Wenyuan, QU Youyang, *et al.* Severity of traffic crashes on freeways based on mixed logit model [J]. *Journal of Transport Information and Safety*, 2019, 37(3): 42.
- [19] CHEN Feng, CHEN Suren, MA Xiaoxiang. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data [J]. *Journal of Safety Research*, 2018, 65(6): 153.
- [20] YONAS A, LEE Z. Improving the ability of drivers to avoid collisions with snowplows in fog and snow [R]. Minnesota: Minnesota Department of Transportation, 2006.
- [21] 潘晓东, 杜志刚, 杨晓光. 桥头跳车对行车安全影响评价指标的研究[J]. 同济大学学报(自然科学版), 2006(5): 634.
PAN Xiaodong, DU Zhigang, YANG Xiaoguang. Evaluation indexes of the impact of vehicle bumping at bridge-head on driving safety [J]. *Journal of Tongji University (Natural Science)*, 2006(5): 634.