

DOI: 10.20035/j.issn.1671-2668.2024.03.009

引用格式: 赵璟. 降雨条件下山区高速公路交通事故发生风险短时预测模型[J]. 公路与汽运, 2024, 40(3): 42-46.

Citation: ZHAO Jing. A short-term prediction model for the risk of traffic accidents on mountain highways with rainfall conditions[J]. Highways & Automotive Applications, 2024, 40(3): 42-46.

降雨条件下山区高速公路交通事故发生 风险短时预测模型^{*}

赵璟

(重庆交通大学 交通运输学院, 重庆 400074)

摘要: 为实现对降雨条件下山区高速公路交通事故发生风险的准确判别, 以云南省某典型山区高速公路 2019—2020 年交通事故、交通运行数据为基础, 解析降雨条件下山区高速公路交通事故风险影响因素的相互作用机理, 采用贝叶斯随机截距 Logistic 回归模型构建降雨条件下山区高速公路交通事故发生风险短时预测模型。结果表明, 降雨强度与交通流特征参数之间存在交互效应, 上游检测器平均交通量、上下游检测器平均运行速度差值和下游检测器运行速度标准差与事故发生风险显著相关; 引入降雨强度变量能提高 Logistic 回归模型在山区高速公路事故分析预测中的拟合优度; 基于贝叶斯随机截距 Logistic 回归模型构建的事故风险预测模型的预测准确率可达 84.3%。

关键词: 交通安全; 交通事故; 山区高速公路; 预测模型; 降雨条件; 交互效应

中图分类号: U491.31

文献标志码: A

文章编号: 1671-2668(2024)03-0042-05

云南是中国不良天气影响交通运营的气象灾害重点防控区域, 在复杂地形、不良天气条件下, 交通事故在连续弯坡、特长纵坡、桥隧密集区等高危路段表现出明显的聚集特征。据统计, 云南省约 35% 的交通事故发生在雨、雾、冰、雪等不良天气, 降雨天气尤为突出^[1]。雨水通过改变路面条件、行车环境、交通流参数等交通构成要素影响驾驶人对环境的感知, 从而影响驾驶人的驾驶行为。目前关于降雨天气对交通流特征参数和交通事故风险影响的研究集中于根据不同降雨强度下交通流特征参数构建事故风险模型, 以交通流特征参数变化作为媒介, 探究降雨强度对交通事故风险的影响。张存保等采用实测数据标定 Van Aerde 模型, 分析了降雨强度对高速公路交通流特性的影响^[2]。Eisenberg D. 采用负二项回归模型验证了降雨对交通风险的影响具有“滞后效应”^[3]。高银钧等研究山区高速公路交通事故时空分布特征, 发现季节变化带来的客观环境差异并不足以导致交通事故数明显的季度波动^[4]。杨文臣等综述与不利天气相关的交通安全及交通流问题, 指出不利天气会增加交通事故的发生次数, 不利

天气对交通事故伤害程度的影响则表现出不一致性^[5]。Malin F. 等研究发现尽管高速公路上总体风险较低, 但在恶劣天气和路况下高速公路上事故风险高于双车道和多车道道路^[6]。张诗等分析了山区高速公路交通事故时空分布、路段线形分布、事故类型分布、事故车型分布及事故严重程度分布特征, 从车辆、道路条件、运行环境和事故形态方面确定了山区高速公路交通事故严重程度影响因素^[7]。现有研究多将天气因素和交通流参数作为独立变量分析各因素变量与事故发生及严重程度之间的关系, 由于相对于其他变量天气因素对事故严重程度的影响并不明显, 现有研究忽视了不良天气事件和交通流、交通事故间的间接影响效应。贝叶斯网络 (Bayesian Network, BN) 可用于分析要素间关联性, 系统展示各因素对事件过程的影响, 被广泛应用于不确定性问题推理, 解释和预测事故发生的可能性^[8-9]。本文采集云南省某典型高速公路的交通流、交通事故和天气条件数据, 分析降雨天气特征与交通流动态特征之间的交互效应, 构建贝叶斯随机截距 Logistic 回归模型预测降雨条件下山区高速公路交通事故风

^{*} 基金项目: 云南省交通运输厅科技项目 (云交科教便 [2021] 90-2); 云南交投科技研发项目 (YCIC-YF-2023-03)

险,为山区高速公路事故风险防控提供理论支持。

1 数据处理

1.1 数据来源

统计 2019—2020 年云南省某典型山区高速公路交通事故、交通流和气象数据,筛选降雨条件下交通事故数据,剔除信息完整度低于 80% 的数据及由行人、非机动车等闯入高速公路引起的异常事故数据,共获得 639 起有效事故数据。按财产损失和人员伤亡情况将事故分为仅财产损失事故、受伤事故和死亡事故三类,各类事故的分布见图 1。

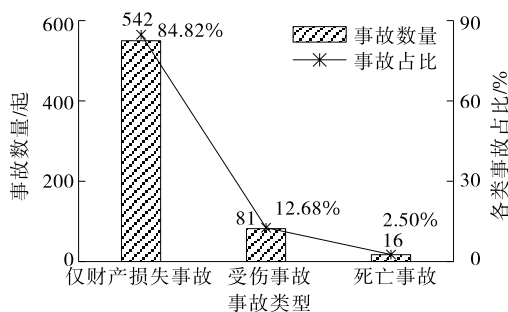


图 1 各类交通事故的分布

1.2 数据处理

由于交通流数据和气象数据的记录时间不一致,为深入研究降雨强度与交通流运行特征的交互效应,并减少数据噪声,将交通流和气象的原始数据汇聚为 15 min 间隔的数据,并根据记录时间将交通流与气象数据进行匹配关联形成新的分析样本集。交通流数据的清洗方式见表 1^[10]。

表 1 交通流数据的清洗方法

判别标准	处理方法	判别标准	处理方法
速度<0	删除	流量=0,速度>0	删除
占有率<0	删除	流量>0,速度=0	删除
占有率>1	删除	流量=0,占有率>0	删除
流量<0	删除		

采用配对式病例-对照法^[11]提取事故发生地点上下游检测器的交通流数据和气象数据,得到 639 条事故数据和 2 556 条非事故数据。匹配规则如下:1) 以有交通事故发生时交通流运行特征数据作为病例组,无事故发生时交通流运行特征数据作为对照组;2) 病例组和对照组的样本比例为 1:4^[12];3) 对每起事故,提取事故发生地点上下游检测器的交通流数据和气象数据,并按照病例-对照

法的原理从无事故日中随机抽取 4 个非事故案例作为对照组。

按降雨强度将天气条件划分为四级^[13],根据事故发生时间、地点匹配天气条件和交通流数据。不同降雨条件下交通事故分布见表 2。

表 2 不同降雨条件下交通事故分布

降雨强度	交通事故数据	
	频率/%	频次/次
无降水($I=0$)	79.8	518
小雨($0<I\leq 2.5$)	13.4	87
中雨($2.5<I\leq 8.0$)	3.2	21
大雨($I>8.0$)	3.6	13

注: I 为降雨量,单位为 mm/h。

2 模型构建

2.1 交互效应计算

交互效应是指两个相互独立的自变量对事故风险的影响不是独立的,其中一个自变量对事故风险的影响随着另一个自变量的变化而变化^[14]。为探究降雨强度和交通流变量的交互效应对事故风险的影响,避免特征变量之间的交互效应对事故风险预测结果产生影响,利用 Logistic 回归模型对降雨强度和交通流参数进行求积、求和形式的交互处理,对训练样本进行拟合。当考虑自变量 x_1 和 x_2 之间的乘积形式的交互效应时,二者乘积形式的交互效应可以通过在模型中增加 x_1 和 x_2 的乘积项来表示,此时 Logistic 模型可表示为:

$$\ln\left(\frac{P(x_1)}{1-P(x_1)}\right)=\ln(\text{odds})=w_0+w_1x_1+w_2x_2+w_3x_1x_2 \quad (1)$$

$$Mx_1=\frac{\text{odds}(x_1=1,x_2=0)}{\text{odds}(x_1=0,x_2=0)}=\frac{e^{w_0+w_1}}{e^{w_0}}=e^{w_1} \quad (2)$$

$$Mx_2=\frac{\text{odds}(x_1=0,x_2=1)}{\text{odds}(x_1=0,x_2=0)}=\frac{e^{w_0+w_2}}{e^{w_0}}=e^{w_2} \quad (3)$$

$$Mx_{1,x_2}=\frac{\text{odds}(x_1=1,x_2=1)}{\text{odds}(x_1=0,x_2=0)}=\frac{e^{w_0+w_1+w_2+w_3}}{e^{w_0}}=e^{w_1+w_2+w_3} \quad (4)$$

式中: P 为在自变量影响下各分类结果的条件概率; odds 表示发生的概率除以不发生的概率; w_0 、 w_1 、

w_2, w_3 为 Logistic 回归模型的参数; Mx_1, Mx_2 分别为特征变量 x_1, x_2 在事故风险预测中对事故风险的单独效应; $Mx_{1,2}$ 为特征变量 x_1 和 x_2 的联合效应。

似然比的计算公式如下:

$$\chi^2 = -2[\Pi(\beta_T) - \Pi(\beta_k)] \quad (5)$$

式中: $\Pi(\beta_T)$ 为不包含降雨强度变量的对数似然函数; $\Pi(\beta_k)$ 为包含降雨强度变量的对数似然函数。

2.2 贝叶斯随机截距 Logistic 回归模型构建

由于本文采集的事故样本量较少, 将传统 Logistic 回归模型扩展到贝叶斯模型的框架, 在模型中加入每个参数的先验分布, 并通过后验分布确定作为随机变量的参数, 以避免因训练数据数量有限而导致过拟合。考虑到数据集的异质性, 采用贝叶斯随机截距 Logistic 模型分析不同天气条件下实时交通事故风险。模型如下:

$$P(y_i = 1) = \frac{1}{1 + e^{-x_i'w}} \quad (i = 1, 2, \dots, n) \quad (6)$$

$$x_i'w = w_0 + w_1x_{1i} + \dots + w_kx_{ki} \quad (7)$$

式中: $P(y_i = 1)$ 表示在特定交通流状态和天气条件下发生碰撞的概率; $-x_i'w$ 为解释变量之间的多重线性组合, 即效用函数; w_k 为变量 k 的系数; x_{ki} 表示样本 i 的变量 k 值。

设定随机系数 u_0 呈正态分布, 均值为零, 方差为 σ^2 , $u_0 \sim N(0, \sigma^2)$ 。根据未知参数的非信息先验分布, 选择 $w_k \sim N(0, 10^6)$, $\sigma^2 \sim IG(0.001, 0.001)$ 表示对每个参数的先验分布, 并利用马尔科夫算法^[15]的模型后验概率之比比较模型 m_1 与模型 m_2 。

假设这两个模型的偏好相等, 则 m_1 模型的先验概率等于 m_2 模型。在这种情况下, 模型的后验概率比率可表示为:

$$P_{PO21} = \frac{f(m_2 | Y)}{f(m_1 | Y)} = \frac{f(Y | m_2) \prod m_2}{f(Y | m_1) \prod m_1} = \frac{f(Y | m_2)}{f(Y | m_1)} = B_{21} \quad (8)$$

$$f(Y | m) = \frac{1}{T} \sum_{i=1}^T [f(Y | w_m^{(t)}, m)]^{-1} \quad (9)$$

$$V_{\text{offset}} = -\ln\left(\frac{R_{\text{SRi}}}{R_{\text{PRi}}}\right) \quad (10)$$

$$L(w) = \prod_{i=1}^n \left[\left(\frac{1}{1 + e^{-x_i'w}} \right)^{y_i} \left(1 - \frac{1}{1 + e^{-x_i'w}} \right)^{1-y_i} \right] \quad (11)$$

式中: R_{PO21} 表示模型 m_2 与模型 m_1 的后验概率;

$f(m | Y)$ 表示模型 m 的后验概率; $f(Y | m)$ 表示模型 m 下数据的边际可能性; B_{21} 表示模型 m_2 与模型 m_1 的贝叶斯因子, 定义为 $f(Y | m_2)$ 与 $f(Y | m_1)$ 的比值; T 为迭代次数; V_{offset} 为截距的偏移值; R_{SRi} 表示样本中具有结果 i 的观测结果与其他观测结果的比值; R_{PRi} 表示总样本中具有结果 i 的观测结果与其他观测结果的比值; $L(w)$ 为 n 个观测数据集上 Logistic 回归模型的相应似然函数。

2.3 参数估计

采用逐步回归分析法从表 3 中选择模型解释变量^[16], 计算所选变量间的皮尔逊相关系数, 检测变量之间的相关性。如果两个变量之间存在较强的相关性, 对每个相关变量分别建立 Logistic 回归模型, 比较每个模型在收敛时的对数似然函数值。如果某个模型的对数似然函数值最小, 则保留该模型中含有的交通流变量, 去掉其他相关变量, 利用最终选择的交通流变量建立 Logistic 回归模型。

表 3 变量名称及定义

变量名称	变量定义
Qup, Qdw	上游、下游检测器平均交通量(辆/15 min)
Vup, Vdw	上游、下游检测器平均运行速度(km/h)
Oup, Odw	上游、下游检测器占有率(%)
STDQup, STDQdw	上游、下游检测器平均交通量标准差(辆/15 min)
STDVup, STDVdw	上游、下游检测器运行速度标准差(km/h)
STDOup, STDOdw	上游、下游检测器占有率标准差(%)
Qud	上、下游检测器平均交通量差值(辆/15 min)
Vud	上、下游检测器平均运行速度差值(km/h)
Oud	上、下游检测器占有率差值(%)
STDQud	上、下游检测器平均交通量标准差偏差(辆/15 min)
STDVud	上、下游检测器运行速度标准差偏差(km/h)
STDOud	上、下游检测器占有率标准差偏差(%)
Rain	降水强度

利用 SPSS 统计分析软件实现 Logistic 模型, 并根据交互效应乘积形式的计算规则生成新的交互

变量,以数据中训练样本作为输入条件,构建事故风险预测模型。同时根据似然比的变化情况分析引入天气变量后模型拟合度的变化。

3 结果分析

3.1 交互效应分析

Logistic 模型的最终估计结果见表 4。去除相关性较大的变量后,模型中包含上游检测器占有率、下游检测器运行速度标准差、上下游检测器平均运行速度差值和降雨强度、降雨强度与上游检测器占有率的交互效应及降水强度与上下游检测器运行速度差值的交互相应。

表 4 Logistic 模型估计结果

变量	系数	标准误差	卡方值	显著性
截距	-1.019	0.320	22.273	0.000
Oup	1.106	0.184	17.485	0.001
STDVdw	-0.032	0.347	63.216	0.000
Vud	0.578	0.296	9.872	0.010
Rain×Oup	-0.125	0.036	18.943	0.001
Rain×Vud	1.478	0.945	37.877	0.000
Rain	1.175	0.022	22.157	0.001
拟合信息	$-2\Pi(\beta_T)=1\ 876.33;-2\Pi(\beta_k)=1\ 946.252$ $-2[\Pi(\beta_T)-\Pi(\beta_k)]=69.922$			

由表 4 可知:统计量为 69.922,大于 χ^2 临界值

63.21,似然比检验的显著性为 0.000,表明天气变量的加入显著提高了实时事故风险模型的拟合度。上下游检测器平均运行速度差值与事故风险存在正相关关系,下游检测器运行速度标准差与事故风险呈负相关关系。在天气条件参数中,降雨强度的系数为正数,表明随着降雨强度的增加,事故风险增大,发生交通事故的概率增大。

根据降雨强度与上游检测器占有率和上下游检测器平均运行速度差值的交互效应系数,上游检测器占有率下降会导致降雨强度对事故风险的影响程度上升,且随着降水强度等级的增加上下游运行平均速度差值增大,事故风险增大。分析其原因,占有率越小,车辆的运行速度越趋近于自由流速度,而在降雨天气条件下路面湿滑系数增大,能见度降低,存在更高的事故风险。由于上下游运行速度差值增大,为保证行车安全,驾驶员不断调整运行速度,随着降雨量的增大,路面附着系数降低,事故发生风险增大。

3.2 模型预测结果

降雨天气条件下贝叶斯随机截距 Logistic 回归模型的估计结果见表 5。由表 5 可知:在表 3 所示 19 个指标中,上下游检测器运行平均速度差值及上游检测器占用率与降雨天气下发生事故的可能性显著相关;降雨强度为正系数,表明雨天交通事故风险随降雨强度的增大而增大;上下游检测器平均运行速度差值为正系数,表明上下游速度差较大时,雨天发生交通事故的风险较高。

表 5 降雨天气下事故风险预测结果

变量	参数估计				优势比			
	均值	标准差	2.5%	97.5%	均值	标准差	2.5%	97.5%
截距	-3.019	0.420	-3.587	-2.765	—	—	—	—
截距分布标准偏差	0.398	0.257	0.256	0.688	—	—	—	—
Oup	0.270	0.062	0.259	0.296	1.326	0.068	1.195	1.487
Vud	0.078	0.046	0.033	0.102	1.058	0.088	1.032	1.098
Rain×Qup	-0.425	0.144	-0.877	-0.178	1.221	0.047	1.177	1.329
Rain×Vud	2.478	0.945	1.952	2.886	1.365	1.311	1.007	1.457
Rain	0.662	0.034	0.462	0.915	1.781	0.981	1.652	1.897
皮尔逊 χ^2 后验 P 值	0.437	0.635	0.000	1.000	—	—	—	—
边际似然对数	$\ln(f(Y m))=-889.63$							
截距调整量	-4.403							

利用 ROC(Receiver Operating Characteristic Curve)曲线对比 Logistic 模型和贝叶斯随机截距 Logistic 回归模型对降雨条件下交通事故风险的预

测性能,测试结果见图 2。由图 2 可知:贝叶斯随机截距 Logistic 回归模型的 ROC 曲线在普通 Logistic 回归模型 ROC 曲线的左上方,表明降雨条

件下贝叶斯随机截距 Logistic 回归模型的预测精度高于普通 Logistic 回归模型。

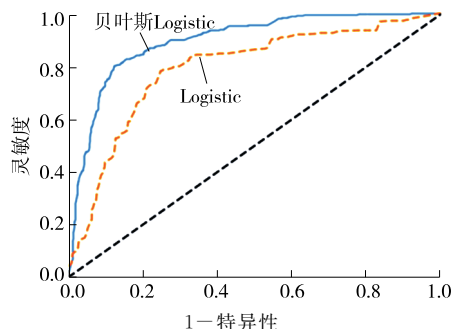


图 2 降雨天气下交通事故风险预测模型的 ROC 曲线

利用 ROC 曲线下面积估计模型的预测性能,在虚报率(假阳率)为 20% 的条件下,Logistic 模型的预测准确率为 72.3%,贝叶斯截距 Logistic 模型的预测准确率为 84.3%。

4 结论

(1) 降雨强度与交通流特征参数之间存在交互效应,其中上游检测器平均交通量和上下游检测器平均运行速度差值与事故风险存在正相关关系,下游检测器运行速度标准差与事故发生风险呈负相关关系。

(2) 利用 Logistic 回归模型对降雨强度和交通流参数进行求积和求和形式的交互处理,根据 Logistic 回归模型的预测结果,引入降雨强度变量可提高 Logistic 回归模型在山区高速公路事故分析预测中的拟合优度。

(3) 综合考虑降雨强度与交通流的交互效应及对事故风险预测的可靠性,基于贝叶斯随机截距 Logistic 回归模型构建事故风险预测模型,与传统 Logistic 回归模型相比,其拟合优度更佳,预测准确率可达 84.3%。

(4) 由于天气条件与交通安全存在明显的地域异质性,本文研究成果仅适用于云南省山区高速公路交通流和降雨强度对事故风险及事故严重程度的影响分析,模型的时空移植性有待进一步研究。

参考文献:

- [1] 李长城.不良天气下的高速公路交通流特性及引导控制研究[D].北京:北京工业大学,2015.
- [2] 张存保,万平,梅朝辉,等.雨天环境下高速公路交通流

特性及模型研究[J].武汉理工大学学报,2013,35(3):63-67.

- [3] EISENBERG D.The mixed effects of precipitation on traffic crashes[J].Accident Analysis and Prevention,2004,36(4):637-647.
- [4] 高银钧,刘德敬,熊昌安,等.山区高速公路交通事故特征分析与动态风险评估[J].公路与汽运,2023(6):44-47.
- [5] 杨文臣,李立,胡澄宇,等.不利天气对公路交通安全及交通流的影响研究综述[J].武汉理工大学学报(交通科学与工程版),2019,43(5):843-849+854.
- [6] MALIN F,NORROS I,INNAMAA S.Accident risk of road and weather conditions on different road types[J].Accident Analysis and Prevention,2019,122:181-188.
- [7] 张诗,彭利飞,黄杰,等.艰险山区高速公路交通事故特征及严重程度影响因素分析[J].公路与汽运,2024,40(1):40-48+54.
- [8] 孟庆森,韩皓,李易.基于贝叶斯优化随机森林的高速公路二次事故预测研究[J].中国安全生产科学技术,2023,19(7):205-210.
- [9] 刘志强,王玲,张爱红,等.基于贝叶斯模型的雾霾天高速公路交通事故发生机理研究[J].重庆理工大学学报(自然科学),2018,32(1):43-49.
- [10] HUANG Z.Extensions to the k-Means algorithm for clustering large data sets with mixed numeric and categorical values[J].Data Mining and Knowledge Discovery,1998,2(3):283-304.
- [11] 王殿海.交通流理论[M].北京:人民交通出版社,2002.
- [12] 杨洋.考虑区域类型差异的高速公路事故风险识别与交通安全评价研究[D].北京:北京交通大学,2020.
- [13] 江苏省气象科学研究所(南京交通气象研究所).高速公路交通气象条件等级:QX/T 111—2010[S].北京:气象出版社,2010.
- [14] XU C C,WANG W,LIU P.Identifying crash-prone traffic conditions under different weather on free-ways[J].Journal of Safety Research,2013,46:135-144.
- [15] 韦学武,朱海峰,刘彦斌,等.基于时序马尔科夫模型的电子警察采集数据异常识别[J].公路与汽运,2020(2):42-46+126.
- [16] 李甜,李瑞玲,张萌萌,等.基于网格模型与 K-Means 算法的交通状态演变特征[J].山东交通学院学报,2021,29(1):15-20.

收稿日期:2024-04-16