

DOI:10.20035/j.issn.1671-2668.2024.01.006

基于有序 Logit 模型的高速公路追尾事故 持续时间分析*

陈喆¹, 巫立明¹, 杨恒旭², 温惠英², 赵胜²

(1.广东联合电子服务股份有限公司, 广东 广州 510075; 2.华南理工大学 土木与交通学院, 广东 广州 510640)

摘要: 追尾事故是高速公路常见事故类型之一。为探究高速公路追尾事故持续时间的特点及影响因素,根据持续时间离散性特征将追尾事故分为 3 个等级,分别采用有序 Logit 模型和多项 Logit 模型建立高速公路追尾事故持续时间分析模型,并通过广东省某高速公路近两年 925 起追尾事故数据对模型参数进行估计。结果表明,追尾事故占据车道数、受伤人数、事故发生时间、事故发生后 10 min 内路段平均速度及涉事车辆类型是影响追尾事故持续时间的主要因素,其中占据车道数及受伤人数越多、追尾事故发生于节假日和白天会增加追尾事故的持续时间,事故发生后 10 min 内路段平均速度每增加一个单位,一般追尾事故与重大追尾事故的概率分别降低 0.2%、0.1%;受伤人数每增加 1 人,一般追尾事故、严重追尾事故的概率分别提高 4.8%、1.3%;占据车道数每增加 1 条,一般追尾事故、严重追尾事故的概率分别提高 5.6%、1.5%;涉事车辆中含有大客车、货车及特种车会增加追尾事故的持续时间。

关键词: 交通安全;高速公路;追尾事故;事故持续时间;有序 Logit 模型;多项 Logit 模型;边际效应

中图分类号:U491.31

文献标志码:A

文章编号:1671-2668(2024)01-0035-05

高速公路由于车辆运行速度快、交通流量大,存在较大的事故隐患和风险。相关研究显示,高速公路的受伤率、死亡率及百公里事故率是普通公路的 3 倍以上^[1]。高速公路交通事故在造成人民群众生命财产损失的同时,还会导致高速公路交通拥堵甚至路网瘫痪^[2]。在高速公路交通事故中,常见的事故类型有追尾、翻车、撞击护栏、货物洒落、侧面碰撞等,追尾作为发生频率最高的高速公路事故类型之一,其处置时间过长,容易造成二次事故的发生,进而导致更大的生命财产损失与通行能力下降。

交通事故发生后,第一时间对其持续时间进行预测评估有利于后续相关救援疏散工作计划制定。近年来,众多学者开展了交通事故持续时间预测研究,形成了两类主要方法,分别为基于人工智能的方法和基于统计的方法。在人工智能学习模型方面,姬杨蓓蓓等采用贝叶斯方法建立决策树预测模型,对交通事故持续时间进行了预测^[3];何珂等基于山西省 2 115 起高速公路隧道交通事故数据,使用主成分分析和随机森林组合模型对隧道交通事故持续

时间进行了预测^[4];纪柯柯等结合多种文本分类算法,构建 V-Fisher 有序聚类模型,提升了对交通事故持续时间的预测效果^[5];Tang J. J.等基于华盛顿事件系统中 2 565 起交通事故数据,建立 K-Means 和 XGBoost 算法进行了事故持续时间预测^[6]。基于人工智能学习模型的事故持续时间预测具有较好的性能,但可解释性较差,且对样本量的要求较高,需要丰富的数据样本及标签。基于统计方法的交通事故持续时间预测模型主要分为概率分布模型、线性回归模型和基于风险的模型。Hou L.等建立非比例风险模型,基于西雅图 2 584 起高速公路交通事故样本,分析了事件、时间、地理、交通和运营对事件清除时间的时变影响,发现涉及重型卡车、夜间、周末等 18 个因素对清除时间有显著影响^[7];Ding C.等分别针对事故响应时间、清除时间建立二元概率模型和切换回归模型,分析了事故类型、时间、环境等因素的影响^[8];夏正丰应用多元回归方法建立了高速公路交通事故持续时间预测模型^[9];Zou Y. J.等利用分位数回归方法和 Copula 模型分析了美

* 基金项目:国家自然科学基金资助项目(52172345)

国西雅图高速公路交通事故持续时间的影响因素^[10];Wali B.等对固定参数回归、随机参数回归与基于分位数回归的事故持续时间模型进行对比,结果显示随机参数模型的预测结果更准确^[11]。以往的研究多是基于历史事故数据对高速公路交通事故持续时间进行分析,较少考虑不同事故类型的事故持续时间的差异,且受限于数据采集难度,未充分考虑事故发生前后短时交通流参数变化对高速公路追尾事故持续时间的影响。基于高速公路追尾事故持续时间的有序性,本文将追尾事故持续时间分为3个等级,综合考虑追尾事故信息与事故前后短时交通流运行状况,运用有序 Logit 模型和边际效应分析高速公路追尾事故持续时间的影响因素,为高速公路交通管理部门制定追尾事故救援及事故后交通疏散方案提供依据。

1 数据描述

研究数据来自广东省高速公路主线 ETC(电子不停车收费系统)门架数据及交警记录的交通事故数据,基础数据结构完整、字段齐全,具有良好的精度。通过数据匹配与数据挖掘进一步提取交通流数据与事故信息,其中交通流数据主要字段包括高速公路 ETC 门架监测的交通流量、交通密度、区间速度等,交通事故数据主要字段包括事故类型、涉事车辆类型、涉事车辆数、事故持续时间、占据车道情况及事故发生地点车道数、上下游路段交通设施布局等。通过对上述交通流数据、事故数据及道路数据进行匹配,得到 925 起高速公路追尾事故数据及发生前后短时交通流参数,为高速公路追尾事故持续时间分析及预测提供数据基础。

如表 1 所示,所选数据集中高速公路追尾事故持续时间最大值为 721.00 min,最小值为 14.00 min,平均值为 88.25 min,均方根为 107.37 min。其中事故持续时间为 50~100 min 的追尾事故数量最多,约占全部追尾事故的 52%(见图 1)。

由于存在多个难以观测的变量影响事故持续时间,准确的追尾事故持续时间往往难以预测。高速

表 1 高速公路追尾事故持续时间统计

统计指标	事故持续时间/min	统计指标	事故持续时间/min
最大值	721.00	平均值	88.25
最小值	14.00	均方根	107.37

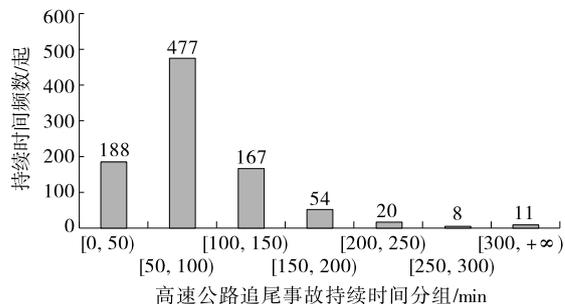


图 1 高速公路追尾事故持续时间频数分布

公路交通管理部门在进行事故救援及事故后交通疏散时并不需要获知事故准确的持续时间,而是需要根据预测的追尾事故持续时间长短将事故进行分级,以制定不同的事故处置方案。鉴于此,区别以往按照伤亡人数、财产损失对交通事故的分类,依据采集的 925 起追尾事故持续时间的分布及追尾事故对道路交通的影响,将追尾事故分为轻微追尾事故(事故持续时间小于 100 min)、一般追尾事故(事故持续时间为 100~200 min)和严重追尾事故(事故持续时间大于 200 min)。

高速公路追尾事故的持续时间与事故发生时间、涉事车辆类型、涉事车辆数、伤亡人数、占据车道情况及交通流状况等密切相关^[11]。因此,以追尾事故持续时间的 3 个等级为因变量,从事故发生时间、涉事车辆类型、伤亡情况、车道情况、事故发生前后交通流运行状况五方面选取 12 个自变量进行分析,变量定义见表 2。

表 2 变量编码及描述

变量名称	编码	描述性统计
追尾事故持续时间	0=持续时间<100 min	674
	1=持续时间为 [100,200] min	212
	2=持续时间>200 min	39
事故发生日期	0=工作日	546
	1=节假日	379
事故发生时间昼夜划分	0=夜晚	154
	1=白天	771
事故发生时段	0=非峰时段	734
	1=高峰时段	191
涉事车辆类型	0=仅小车	770
	1=仅货车	38
	2=货车与小车	9
	3=有客车	13
	4=有特种车	95

续表 2

变量名称	编码	描述性统计
占据路肩情况	0=不占据路肩	852
	1=占据路肩	73
涉事车辆数/辆	连续变量	均值 2.83
受伤人数/人	连续变量	均值 0.07
占据车道数/条	连续变量	均值 1.24
事故发生前 10 min 内路段平均速度/(km·h ⁻¹)	连续变量	均值 69.37
事故发生前 10 min 内到达上游门架车辆数/辆	连续变量	均值 761
事故发生后 10 min 内路段平均速度/(km·h ⁻¹)	连续变量	均值 54.82
事故发生后 10 min 内到达上游门架车辆数/辆	连续变量	均值 739

由表 2 可知:因变量是三分类的有序变量,且持续时间逐渐递增,而自变量中既有连续变量又有分类变量。其中二分类变量为事故发生日期、昼夜划分、时段划分及占据路肩情况,多分类变量是指多于 2 种选择的变量,将涉事车辆类型划分为仅小车、仅货车、货车与小车、有客车、有特种车五类。事故信息中,涉事车辆数、受伤人数及占据车道数为连续变量。考虑到事故发生前后交通流运行状况可能对事故清理造成影响,将事故发生前后 10 min 内路段平均速度及到达上游门架的车辆数也作为自变量,探究其对追尾事故持续时间的影响。交通流影响因素中,追尾事故发生前后 10 min 内路段平均速度及到达事故发生点上游门架的交通量均为连续变量。

2 模型构建

多项 Logit 模型可视为二元 Logit 模型的拓展,二者的差异在于,二元 Logit 模型的被解释变量只有 0 和 1 两个取值,而多项 Logit 模型涉及被解释变量有多个取值的情形。多项 Logit 模型基于效用理论拟合自变量与包含多个类别的分类变量间的内在关系。假定事故影响因素相互独立,且服从 Gumbel 分布(I 类广义极值分布),对于任一交通事故 i ,事故持续时间等级的效用函数为:

$$S_{i,k} = \beta_k \mathbf{X}_i + \varepsilon_{i,k} \quad (k=1,2,3) \quad (1)$$

式中: \mathbf{X}_i 为事故 i 对应的影响因素观测值组成的向量; β_k 为这些因素对应事故持续时间等级 k 的回归

系数组成的向量(包括常数项); $\varepsilon_{i,k}$ 为随机项。

有序 Logit 模型主要用于处理因变量为有序变量的多分类结果,属于多项 Logit 模型的拓展^[11],在多分类的交通事故等级预测中表现出较好的性能。传统有序 Logit 模型中,假定解释变量的系数是固定的,高速公路追尾事故持续时间第 j 个等级的有序 Logit 模型为:

$$P(Y_i > j) = \frac{\exp(\alpha_j + \mathbf{X}\boldsymbol{\beta})}{1 + \exp(\alpha_j + \mathbf{X}\boldsymbol{\beta})} \quad (j=1,2,\dots,J) \quad (2)$$

式中: Y_i 为第 i 起追尾事故的持续时间等级; \mathbf{X} 为自变量向量; $\boldsymbol{\beta}$ 为回归系数向量; α_j 为第 j 追尾事故等级的截距; J 为高速公路追尾事故持续时间划分的等级数量。

本文将因变量(追尾事故持续时间)分为 3 个等级,有序 Logit 模型为:

$$P(Y=1 | \mathbf{X}) = \frac{\exp(\alpha_1 + \sum_{k=1}^K \beta_k x_k)}{1 + \exp(\alpha_1 + \sum_{k=1}^K \beta_k x_k)} \quad (3)$$

$$P(Y=2 | \mathbf{X}) = P(Y \leq 2 | \mathbf{X}) - P(Y \leq 1 | \mathbf{X}) \quad (4)$$

$$P(Y=3 | \mathbf{X}) = 1 - P(Y \leq 2 | \mathbf{X}) \quad (5)$$

式中: x_k 为第 k 个自变量, $k=1,2,\dots,K$; β_k 为第 k 个自变量的回归系数。

边际效应可用于定量分析不同因素对因变量影响的强弱程度与变化趋势^[12],其表达式为:

$$E_{x_{ijk}}^{[P(Y_i > j)]} = \frac{[P(Y_i > j) | x_{ijk} = 1] - [P(Y_i > j) | x_{ijk} = 0]}{[P(Y_i > j) | x_{ijk} = 0]} \quad (6)$$

式中: $E_{x_{ijk}}^{[P(Y_i > j)]}$ 为第 i 起追尾事故的第 k 个变量在第 j 追尾事故等级的自变量 x_{ijk} 的边际效应值。

3 实例分析

3.1 参数结果分析

以广东省某高速公路 2020—2022 年 925 起追尾事故的持续时间为研究对象,考虑到事故持续时间划分等级具有有序性,通过回归分析建立追尾事故持续时间的有序 Logit 模型。采用统计分析软件 Stata16.0 进行求解,取显著性水平为 0.10,通过最大似然估计法估算得到有序 Logit 模型的参数结果(见表 3)。

由表 3 可知:1) 受伤人数、占据车道数、追尾事

表 3 基于有序 Logit 模型的参数估计

变量	优势比	参数估计	标准误差	p
涉事车辆数	0.969	-0.032	0.068	0.638
受伤人数	1.379	0.322	0.187	0.085*
占据车道数	1.455	0.375	0.185	0.042*
事故发生前 10 min 内 路段平均速度	1.005	0.005	0.005	0.294
事故发生前 10 min 内 到达上游门架车辆数	1.000	-0.001	0.001	0.329
事故发生后 10 min 内 路段平均速度	0.987	-0.013	0.005	0.010*
事故发生后 10 min 内 到达上游门架车辆数	1.000	0.000	0.001	0.994
节假日/工作日	1.321	0.278	0.155	0.072*
白天/夜晚	1.788	0.581	0.237	0.014*
高峰时段/非高峰时段	1.018	0.018	0.188	0.925
是否占据路肩	0.982	-0.018	0.334	0.956
涉事车辆类型	1.170	0.157	0.025	0.010*
截距 1	2.341	2.341	0.739	—
截距 2	4.545	4.545	0.756	—

注：* 表示在 90% 的置信水平上显著，无 * 表示该因素对因变量在统计学上无显著影响。

故发生时间为工作日还是节假日、追尾事故发生时间为夜晚还是白天、事故发生后 10 min 内路段平均速度及涉事车辆类型对高速公路追尾事故持续时间具有显著影响。占据车道数、受伤人数越多及追尾事故发生于节假日和白天都会导致追尾影响交通的持续时间更长。2) 占据车道数、受伤人数每增加一个单位，追尾事故持续时间等级上升的概率分别增加 38% 和 46%，验证了前人的研究成果^[13]。节假日发生的追尾事故持续时间等级更高的概率是工作日的 1.321 倍，白天发生的追尾事故持续时间等级更高的概率是夜晚的 1.788 倍。3) 相较于涉事车辆类型均为小车，涉事车辆中含有大客车、货车及特种车往往会造成更长事故持续时间，这点与高速公路交通事故规律相吻合，即有大型车辆参与的交通事故往往伴随着更大的事故影响范围与事故严重程度。而随着追尾事故发生后 10 min 内路段平均速度的增加，追尾事故持续时间等级更有可能下降，这可能是由于事故发生后上游驾驶员提高行车速度在短时间内驶离事故现场，以减少对现场交通的影

响，为事故现场清理及应急救援提供良好条件。

表 4 为 Pearson 统计量和 Deviance 统计量拟合优度检验结果。Pearson 和 Deviance 统计量的 p 值均大于 0.1，表明在显著性水平 $\alpha=0.10$ 的条件下模型的拟合效果较好。有序 Logit 模型平行性检验结果原假设是各回归方程互相平行，平行性检验 $p=0.271>0.10$ ，接受原假设，模型通过平行性检验，可进一步对高速公路追尾事故持续时间的影响因素进行分析。

表 4 有序 Logit 模型的拟合优度检验

检验方法	χ^2	自由度	p
Pearson	350.639	395	0.947
Deviance	330.164	395	0.992

分别计算有序多分类 Logit 模型和无序多分类 Logit 模型的赤池信息量 (Akaike Information Criterion, AIC) 和贝叶斯信息量 (Bayesian Information criterion, BIC)，结果见表 5。有序多分类 Logit 模型的 AIC 与 BIC 均低于无序多分类 Logit 模型，表明有序 Logit 模型的拟合优度更高。

表 5 模型对比

模型	对数似然值	赤池信息量	贝叶斯信息量
无序多分类 Logit 模型	-626.636	1 294.46	1 429.69
有序多分类 Logit 模型	-619.228	1 283.27	1 355.72

3.2 边际效应分析

结合边际效应对表 3 中显著的自变量进行参数解释，结果见表 6。

表 6 边际效应分析

变量	不同追尾事故类型的边际效应		
	轻微追尾事故	一般追尾事故	严重追尾事故
受伤人数	-0.061	0.048	0.013
占据车道数	-0.071	0.056	0.015
事故发生后 10 min 内路段平均速度	0.002	-0.002	-0.001
节假日/工作日	-0.052	0.041	0.011
白天/夜晚	-0.109	0.086	0.023
涉事车辆类型	-0.029	0.023	0.006

由表 6 可知:1) 在 90%置信水平上,受伤人数每增加 1 人,追尾事故为一般事故、严重事故的概率分别增加 4.8%、1.3%,而轻微追尾事故的概率降低 6.1%;占据车道数每增加 1 条,一般追尾事故、严重追尾事故的概率分别增加 5.6%、1.5%,轻微事故概率降低 7.1%;事故发生后 10 min 内路段平均速度每增加一个单位,一般追尾事故与重大追尾事故的概率分别降低 0.2%、0.1%。2) 相较于工作日和夜晚,节假日与白天容易发生持续时间更长、严重程度更大的追尾事故。节假日期间一般追尾事故、严重追尾事故的概率分别增加 4.1%、1.1%,白天时段一般追尾事故、严重追尾事故的概率分别增加 8.6%、2.3%。这是由于节假日和白天高速公路交通量更大,给事故清理与应急救援带来更大干扰,造成更长的事故持续时间。3) 在涉事车辆类型方面,涉事车辆中含有大客车、货车及特种车会增加追尾事故的持续时间。

4 结论

(1) 以广东省某高速公路 2020—2022 年 925 起追尾事故为研究对象,从事故发生时间、涉事车辆类型、伤亡情况、车道情况、事故发生前后交通流运行状况 5 个方面选取 12 个自变量,以事故持续时间为特征将追尾事故划分为 3 个等级,并将其作为因变量,分别构建无序多分类 Logit 模型和有序多分类 Logit 模型。结果表明,相较于无序多分类 Logit 模型,有序多分类 Logit 模型对高速公路追尾事故持续时间的拟合效果更优。

(2) 高速公路追尾事故占据车道数、受伤人数、事故发生时间、事故发生后 10 min 内路段平均速度及涉事车辆类型对事故持续时间具有显著影响。

(3) 占据车道数和受伤人数越多、追尾事故发生于节假日和白天及涉事车辆中含有大客车、货车及特种车辆会增加追尾事故的持续时间。

(4) 追尾事故持续时间随着事故发生后 10 min 内路段平均速度的增大而减少,事故发生后 10 min 内路段平均速度每增加一个单位,一般追尾事故与重大追尾事故的概率分别降低 0.2%、0.1%。

参考文献:

[1] 贺玉龙,陈兵硕,高鹏旺,等.基于未确知测度理论的高速公路交通安全评价[J].重庆交通大学学报(自然科学版),2021,40(2):8-12.

- [2] 刘鑫,肖乐,金思雨.基于 NAIS 的高速公路严重交通事故特征及致因分析[J].公路与汽运,2022(5):32-37.
- [3] 姬杨蓓蓓,张小宁,孙立军.交通事件持续时间预测方法综述[J].公路工程,2008,33(3):72-79+141.
- [4] 何珂,杨顺新,郜勇刚.基于 PCA-RF 组合模型的隧道交通事故持续时间预测[J].交通信息与安全,2019,37(5):26-32.
- [5] 纪柯柯,陈坚,肖思瑶,等.文本数据驱动下的高速公路事故持续时间预测模型[J].交通信息与安全,2020,38(6):9-16.
- [6] TANG J J,ZHENG L L,HAN C Y, et al. Traffic incident clearance time prediction and influencing factor analysis using extreme gradient boosting model [J]. Journal of Advanced Transportation,2020,2020:1-12.
- [7] HOU L, LAO Y T, WANG Y H, et al. Time-varying effects of influential factors on incident clearance time using a non-proportional hazard-based model [J]. Transportation Research Part A: Policy and Practice, 2014,63:12-24.
- [8] DING C, MA X L, WANG Y H, et al. Exploring the influential factors in incident clearance time: disentangling causation from self-selection bias [J]. Accident Analysis & Prevention, 2015,85:58-65.
- [9] 夏正丰.高速公路交通事故持续时间概率预测模型[J].公路与汽运,2016(3):52-55.
- [10] ZOU Y J, YE X, HENRICKSON K, et al. Jointly analyzing freeway traffic incident clearance and response time using a copula-based approach [J]. Transportation Research Part C: Emerging Technologies, 2018, 86:171-182.
- [11] WALI B, KHATTAK A J, LIU J. Heterogeneity assessment in incident duration modelling: implications for development of practical strategies for small & large scale incidents [J]. Journal of Intelligent Transportation Systems, 2022,26(5):586-601.
- [12] 秦雅琴,谢碧珊,杨文臣,等.山区双车道公路摩托车事故严重度分析模型比较[J].安全与环境学报,2021,21(4):1397-1404.
- [13] 马景峰,任刚,李豪杰,等.电动自行车与机动车事故严重性影响因素分析[J].交通运输系统工程与信息,2022,22(2):337-348.
- [14] ISLAM N, ADANU E K, HAINEN A M, et al. A comparative analysis of freeway crash incident clearance time using random parameter and latent class hazard-based duration model [J]. Accident Analysis & Prevention, 2021,160:106303.