

doi: 10.3969/j.issn.1002-0268.2022.12.019

高速公路交通事故持续时间计算方法研究

吕路¹, 李杰², 郭忠印¹, 阎莹³, 高超⁴

(1. 同济大学 道路与交通工程教育部重点实验室, 上海 201804; 2. 山东高速信息集团有限公司, 山东 济南 250100;
3. 长安大学 运输工程学院, 陕西 西安 710064; 4. 陕西汇德通市政工程有限公司, 陕西 西安 710086)

摘要: 交通事故持续时间是衡量交通事故对交通影响程度的一项重要指标, 交通事故持续时间计算对制定事故下的交通分流管控方案和降低事故产生的不利影响具有重要作用。统计了贵州省高速公路不同交通事件频数及事件持续时间的分布。由高速公路交通事故数据集、交通流和气象观测资料、高速公路设计文件, 建立了交通事故持续时间初始变量集。通过极限梯度提升树进行了特征变量选择和事故持续时间计算模型训练, 并进行了特征变量敏感性分析。采用贵州省不同时期的交通事故数据进行了模型时间稳定性检验, 采用贵州省和山东省相同时期的交通事故数据进行了模型空间稳定性检验。对所建模型在实际交通环境中的应用方法进行了讨论。结果表明: 交通事故是最常发生的一类交通事件, 两车碰撞类事故较单方事故持续时间长 24%; 同回归模型相比, 优化后的集成学习算法可提高交通事故持续时间计算结果的准确率; 由于不同时期、不同地区交通事故持续时间影响因素可能存在差异, 模型时空稳定性分析结果表明, 所建模型具有时间稳定性, 但不具备空间稳定性; 事故持续时间的风险函数表明, 最常发生的一类事故持续时间为 29 min, 根据其特征变量取值, 结合变量敏感性分析结果, 交管部门可根据事故特征变量, 快速进行事故持续时间计算。

关键词: 交通安全; 交通事故; 事故持续时间; 计算方法; 极限梯度提升树

中图分类号: U491.4

文献标识码: A

文章编号: 1002-0268 (2022) 12-0155-08

Study on Calculation Method of Expressway Accident Duration

LÜ Lu¹, LI Jie², GUO Zhong-yin¹, YAN Ying³, GAO Chao⁴

(1. Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China;
2. Shandong Hi-speed Information Group Co., Ltd., Jinan Shandong 250100, China;
3. School of Transportation Engineering, Chang'an University, Xi'an Shaanxi 710064, China;
4. Shaanxi Huidetong Municipal Engineering Co., Ltd., Xi'an Shaanxi 710086, China)

Abstract: The duration of traffic accident is an important indicator to measure the impact of traffic accident on traffic. The calculation of the duration of traffic accidents plays an important role in formulating traffic diversion management and control scheme under accident and reducing the impact of accident. The distributions of frequency and duration of different traffic incidents on expressway in Guizhou Province are counted. According to the expressway accident dataset, traffic flow and meteorological observation data and the expressway design document, the initial variable set of traffic accident duration is established. The characteristic variables are selected and the accident duration calculation model training is carried out by using XGBoost, and the sensitivity analysis of the characteristic variable is carried out. The temporal stability of the model is tested by using the traffic accident data of Guizhou Province in different periods, and the spatial stability of the model is tested by using the traffic accident data of Guizhou Province and Shandong Province in the same period. The application method of the established model in the actual traffic

收稿日期: 2022-06-01

基金项目: 国家自然科学基金项目 (51978069); 山东省交通运输厅科技计划项目 (2019B32); 陕西省重点研发计划项目 (2021KWZ-09)

作者简介: 吕路 (1991-), 男, 陕西咸阳人, 博士研究生. (lvlu0426@tongji.edu.cn)

environment is discussed. The result shows that (1) Traffic accidents are the most common type of traffic incidents, and the duration of the accident between 2 drivers is 24% longer than the unilateral accident; (2) compared with the regression model, the optimized integrated learning algorithm can improve the accuracy of traffic accident duration calculation; (3) because the influencing factors of accident duration may differ in different periods and regions, the result of the spatial-temporal stability analysis shows that the model has temporal stability but does not have spatial stability; (4) the hazard function of accident duration shows that the duration of the most common accident type is 29 min, according to the value of its characteristic variables and the result of variable sensitivity analysis, the traffic management department can quickly calculate the accident duration based on the characteristic variables of the new traffic accident.

Key words: traffic safety; traffic accident; duration time; calculation method; XGBoost

0 引言

交通事故是指车辆在道路上因过错或者意外造成人身伤亡或者财产损失的事件^[1]。交通事故持续时间计算结果是划分事故等级,制定事故下的交通诱导、分流通控方案的重要依据。目前对交通事故持续时间的研究主要集中在事故影响因素分析^[2]、事故持续时间计算方法^[3]、降低交通事故损失^[4]等方面,然而现有研究主要针对城市道路交通事故,对高速公路交通事故持续时间计算方法的研究不足。

不同于城市道路,高速公路交通感知、管控设施较少,交通流密度低、流速大,事故形态相对单一,由此导致城市道路交通事故与高速公路交通事故持续时间的影响因素存在差异^[5],城市道路交通事故持续时间相关研究成果不适用于高速公路交通事故管理。除上述原因外,目前交通事故持续时间计算方法还存在计算复杂、模型稳定性差等问题^[6]。

根据交通事故发生和处理过程,通常将交通事故持续时间划分为事件检测、事件响应、事件清除3个阶段,其大小由事故类型和当地交管部门的应急服务水平决定^[7]。在建立交通事故持续时间计算模型时,将交通事故持续时间影响因素划分为交通事故特征、交通环境特征、交通状况特征以及道路特征^[8]4个部分,各个特征又包含不同变量。重要性较大的变量组合能够显著提高模型计算精度^[9]。随机森林和随机生存森林^[10]、贝叶斯网络^[11]、参数加速失效模型^[12]等方法均可用于特征变量选择。

交通事故持续时间计算方法的适用性决定了模型计算结果的准确性。按照统计方法,交通事故持续时间计算模型可分为概率模型和非概率模型。决策树^[13]、逻辑加速时间度量模型^[14]、离散选择模型^[15]是概率模型。线性回归模型^[16]、时间序列模型^[17]以及神经网络^[18]是非概率模型。模糊 Logistic

回归模型^[19]既可看作是概率模型,又可看作是非概率模型。受事故类型和交通环境差异影响,交通事故持续时间计算结果为一定误差范围内的估计值,对事故持续时间概率计算模型的应用研究越来越多。

本研究首先对贵州省高速公路交通事件持续时间分布进行了统计;其次,根据高速公路交通事故档案信息建立事故持续时间初始变量集,基于XGBoost方法对事故持续时间特征变量进行选择,建立了高速公路交通事故持续时间计算模型,并对各特征变量的敏感性进行了分析;最后,对模型性能和稳定性进行检验,建立了交通事故持续时间快速计算方法。

1 数据描述

统计贵州省路网中心2020年8月1日至2021年7月31日记录的6 582条高速公路交通事件信息。每条数据包包含事件初报和终报时间、事件地点、事件概况、事件类型、伤亡情况等要素。为了提高模型计算精度,对交通事件信息进行预处理。从描述性的事件概况中提取事件发生时间、事件发生位置以及车辆类型,消除事件发生和清除时间不明、事件类型等关键信息缺失的异常数据,删除发生在高速公路服务区、收费站站前广场等路外数据。为了避免异常值对计算结果产生影响,删除事件持续时间大于10 h的数据,最终得到6 332条有效数据。以50 min为间隔,绘制事件持续时间频数分布直方图,如图1所示。

图1显示,随着事件持续时间的增加,事件频数迅速下降。事件持续时间主要集中在100 min以内,其中频数最大值出现在50 min以内。按照事件类型和事件发生的主要原因,将交通事件划分为道路交通事故、地质灾害、气象灾害、交通设施安全事件。处理后的各事件信息均包括事件发生时间、

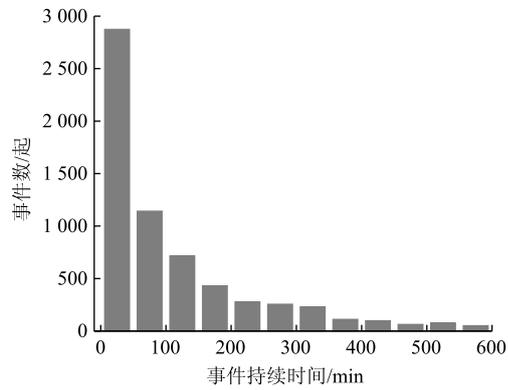


图 1 事件持续时间频数分布直方图

Fig. 1 Histogram of incident duration frequency distribution

事件清除时间、事件类别、事件类型、事件地点、伤亡情况、涉及车辆类型、事件发生位置、事件阻塞车道数等信息。

不同事件类别持续时间分布特征统计结果, 如表 1 所示。表 1 显示, 道路交通事故是平均持续时间最短的一类事件, 地质灾害事件和交通设施安全事件平均持续时间较长。除交通事故外, 其他事件占交通事件总数的 12.93%, 表明交通事故是最常发生的一类事件。以下主要对交通事故持续时间计算方法进行研究。

表 1 事件持续时间分布统计

Tab. 1 Statistics of incident duration distribution

事件类别	事件数量/起	事件持续时间均值/min	事件持续时间标准差/min	事件持续时间最小值/min	事件持续时间最大值/min
道路交通事故	5 513	37	12	17	236
地质灾害事件	75	354	91	114	583
气象灾害事件	713	95	63	45	174
交通设施安全事件	31	231	84	67	562

2 分析方法

2.1 初始变量集构建

由处理后的贵州省高速公路交通事件数据筛选得到交通事故数据集, 按照交通事故特征、环境特征、交通状况特征以及道路特征构建事故持续时间的初始变量集。事故特征构成要素有事故类型、事故发生时段、受伤人数、死亡人数、涉及车辆类型、事故发生位置、剩余车道数; 环境特征要素有能见度、气温、雨量、路面湿滑程度; 交通状况特征要素有分车道流量、车型构成比例; 道路特征要素包括道路类型、道路几何特征。

根据事件信息记录表, 对各影响因素进行细分,

建立交通事故持续时间初始变量集, 其中环境特征和交通状况特征数据来源于高速公路沿线气象观测站和交通调查系统, 道路特征数据来源于公路工程设计文件, 结果见表 2。

表 2 交通事故持续时间初始变量集

Tab. 2 Initial variable set of traffic accident duration

影响因素分类	序号	事故影响因素	变量组成
交通事故特征	1	事故类型	1=单方事故, 2=两车事故, 3=多车事故
	2	事故发生时段	1=黎明 (3~7), 2=早晨 (7~11), 3=中午 (11~15), 4=下午 (15~19), 5=午夜 (19~23), 6=夜晚 (23~3) ^[20]
	3	受伤人数/人	1=0, 2=1, 3=2, 4=3+
	4	死亡人数/人	1=0, 2=1, 3=2, 4=3+
	5	事故发生位置	1=互通立交区, 2=其他
	6	涉及车辆类型	1=危险货物运输车, 2=营运客车, 3=货车, 4=小客车
	7	剩余车道数	1=0, 2=1, 3=2+
环境特征	8	能见度/km	事故发生前 10 min 的平均能见度
	9	气温/℃	事故发生前 1 h 内平均气温
	10	雨量/mm	事故发生前 1 h 内累积降雨量
	11	湿滑程度	事故发生前 10 min 道路平均湿滑程度
交通状况特征	12	服务水平	1=2 级及以上, 2=3 级, 3=4 级, 4=5 级及以下
	13	车型比例	事故发生前 10 min 断面交通流中的客车与货车比例
道路特征	14	道路类型	1=隧道, 2=桥梁, 3=路堤, 4=路堑
	15	道路几何特征	1=(10 直线, 11 平曲线), 2=(20 平坡, 21 上坡, 22 下坡)

2.2 基于 XGBoost 的交通事故持续时间计算方法

极限梯度提升树 (XGBoost) 是梯度提升决策树 (GBDT) 的一个变种, 属于顺序化的集成学习构建方法。同 GBDT 相比, 最大的区别是 XGBoost 通过对目标函数中的结构损失函数做二阶泰勒展开, 大大提高了模型的可扩展性、收敛速度和计算精度。

基于 XGBoost 的交通事故持续时间计算方法, 主要包括特征变量选择和模型构建与参数调优两个步骤。以处理后的初始交通事故数据构建样本库, 随机抽取 50% 的样本数据用于模型训练, 另一半数据用于模型测试。算法流程为:

(1) 对原始事故信息进行预处理, 构建初始变

量训练集；基于 XGBoost 进行特征变量重要程度排序及选择。

XGBoost 通过计算各特征变量给分裂节点带来信息增益的加权平均值，得到每个特征变量的重要性得分。特征变量重要性分数越高，其在模型构建中的价值越大。在特征变量按重要程度降序排列的基础上，依次计算由 1 个最重要特征变量至所有特征变量构建模型时的计算结果准确率，选择模型计算准确率最大时的特征变量作为变量筛选结果。

所得特征变量重要程度降序排列为：1, 7, 12, 3, 4, 6, 5, 2, 15, 13, 14, 8, 11, 10, 9, 表明事故特征对事故持续时间的影响最大。图 2 为模型计算准确率 R_k 随变量个数 N 的变化曲线，当变量个数由 15 减至 7 时，删除不重要变量能够减小冗余信息对算法带来的影响，模型计算准确率增加；当变量个数由 7 减至 1 时，随着重要特征变量被删除，模型计算准确率下降；当变量个数为 7 时，曲线有最大值。因此，事故类型、剩余车道数、服务水平、死亡人数、受伤人数、涉及车辆类型、事故发生位置即为筛选得到的交通事故持续时间特征变量。

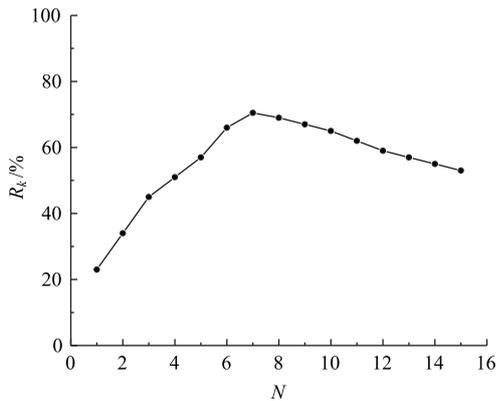


图 2 模型计算准确率与变量个数之间的关系

Fig. 2 Model calculated relationship between accuracy and number of variables

(2) 使用筛选得到的 7 个特征变量构建训练集来训练 XGBoost，通过粒子群算法 (PSO) 优化模型参数。

在给定的包含 n 个样本和 m 个特征变量的数据集 $D = \{(x_i, y_i) \mid |D| = n, x_i \in R^m, y_i \in R\}$ 中，集成树模型通过 K 步得到模型的计算值，如式 (1) 所示。

$$\hat{y}_i^{(t)} = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F. \quad (1)$$

式中， $\hat{y}_i^{(t)}$ 为第 i 个样本在第 t 次迭代时的计算结果； $\phi(x_i)$ 为第 i 个样本的最终计算结果。每次迭代通过 k 步完成，其中， $f_k(x_i)$ 为第 i 个样本在第 k 步迭代时的计算结果， $F = \{f(x) = w_{q(x)}\} (q: R^m \rightarrow T, w \in$

R^T) 是回归树空间， R 为实数集。即依次计算第 k 步的计算值与实际值的残差，通过残差的梯度下降最终得到真实值^[21]。XGBoost 模型的目标函数如式 (2) 所示。

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^p \Omega(f_k), \quad (2)$$

式中，右半部分的第 1 项为损失函数， $l(y_i, \hat{y}_i)$ 为第 i 个样本计算值与实际值的残差；第 2 项为抑制模型复杂度的正则项； f_k 为第 k 步迭代时的计算模型。

函数 $f(x + \Delta x)$ 在点 x 处的二阶泰勒展开式为 $f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)\Delta x^2$ ，通过类比可得到 XGBoost 的损失函数，如式 (3) 所示。

$$l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) = l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i), \quad (3)$$

式中， $f(x)$ 对应损失函数 $l(y_i, \hat{y}_i^{(t-1)})$ ； x 对应前 $t-1$ 棵树的计算值 $\hat{y}_i^{(t-1)}$ ； Δx 对应正在训练的第 t 棵树 $f_i(x_i)$ ； g_i 为损失函数的一阶导数； h_i 为损失函数的二阶导数，且均是对 $\hat{y}_i^{(t-1)}$ 求导。

通过枚举决策树每个叶节点中所有特征的可能分割方案，计算分割前后目标函数的收益，以收益最大的特征作为分割特征，用该特征的最佳分割点作为节点的分割位置，生成回归树模型。使用平均绝对百分误差 (MAPE) 评估模型计算准确率，MAPE 值越小，模型计算准确率越高。

采用 PSO 算法优化模型参数，训练得到的模型计算误差率随迭代次数的变化关系，如图 3 所示。当迭代次数为 56 时，模型计算误差率取得最小值 29.5%，此时模型参数达到最优。

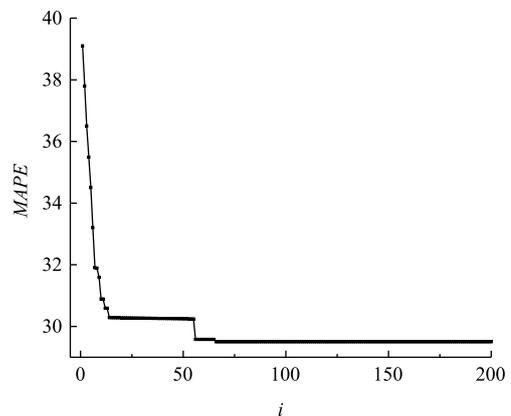


图 3 计算误差率曲线

Fig. 3 Calculation error rate curve

2.3 特征变量敏感性分析

通过 XGBoost 算法, 可得到事故持续时间特征变量的重要程度序列。对筛选得到的交通事故持续时间特征变量进行敏感性分析, 计算 PSO 算法优化后的各变量系数的指数值, 得到不同特征变量对事故持续时间的的影响程度, 即当特征变量的取值改变 1 个单元时, 事故持续时间变化的百分比, 并使用指数系数进行表示, 用于表征事故持续时间对特征变量的乘法效应估计。特征变量敏感性分析结果如表 3 所示。

表 3 特征变量敏感性分析

Tab. 3 Sensitivity analysis of characteristic variables

特征变量编号	特征变量*	系数	标准差	T 统计量	指标影响程度/%
a_1	事故类型: 1, 2, 3	0.215	0.052	3.05	24.0
a_2	剩余车道数: 1, 2, 3	-0.154	0.041	-3.12	-14.3
a_3	服务水平: 1, 2, 3, 4	0.132	0.039	2.75	14.1
a_4	死亡人数: 1, 2, 3, 4	0.097	0.072	4.13	10.2
a_5	受伤人数: 1, 2, 3, 4	0.056	0.075	7.24	5.8
a_6	涉及车型: 1, 2, 3, 4	-0.034	0.016	-6.38	-3.3
a_7	事发位置: 1, 2	0.024 4	0.047	1.75	2.5
	常数项	3.264	0.027		
	样本量		5 513		

注: * 表 3 中的特征变量与表 2 中的变量组成一致。

表 4 交通事故持续时间计算结果误差率

Tab. 4 Error rate of traffic accident duration calculation result

事故持续时间 T	0 < t ≤ 50 min					50 min < t ≤ 120 min				
	≤10	≤20	≤30	≤40	≤50	≤10	≤20	≤30	≤40	≤50
误差范围/min	≤10	≤20	≤30	≤40	≤50	≤10	≤20	≤30	≤40	≤50
MAPE	14	10	8	5	2	29	21	16	9	4
事故持续时间 T	120 min < t ≤ 300 min					300 min < t ≤ 600 min				
	≤10	≤20	≤30	≤40	≤50	≤10	≤20	≤30	≤40	≤50
误差范围/min	≤10	≤20	≤30	≤40	≤50	≤10	≤20	≤30	≤40	≤50
MAPE	92	79	66	57	49	94	85	79	71	67

依据。但当交通事故持续时间大于 120 min 时, 模型计算结果的 MAPE 大于 50%, 这主要是由于持续时间大于 120 min 的交通事故样本量小, 特征变量构成复杂等原因造成的。

为了验证 XGBoost 交通事故持续时间计算方法的性能, 采用邻近算法 (KNN) 和随机森林 (RF) 作为对比算法。不同模型在不同事故持续时间及误差范围内的计算准确率, 结果如图 4 所示。在不同事故持续时间和误差范围下, XGBoost 模型的计算准确率高于 KNN 和 RF。当交通事故持续时间大于 120 min 时, 模型计算准确率在各误差范围内均下降至较低值。通过扩大样本量, 对事故特征进行更加精细化的划分, 可提高此类事故持续时间的计算准确率。

表 3 显示, 剩余车道数的增加、因涉及车型变化可能导致事故产生的人员伤亡和财产损失的减少, 可降低交通事故持续时间, 其余变量的变化将增加事故持续时间。死亡人数增加 1 人, 引起的事故持续时间指数系数值为 $+e^{0.097}$, 即事故持续时间将增加 10.2%。同时, 剩余车道数的变化引起的事故持续时间指数系数值为 $e^{-0.154}$, 即 85.7%, 意味着剩余车道数每增加 1 条, 事故持续时间将减少 14.3%。

3 结果分析

3.1 模型性能比较

统计交通事故持续时间计算模型在一定误差范围内的平均绝对百分误差 (MAPE)。划分 10, 20, 30, 40, 50 min 共 5 个误差区间, 按照事故持续时间小于 50, 50~120, 120~300, 300~600 min 将交通事故划分为不同等级。基于 XGBoost 的交通事故持续时间计算结果误差率, 如表 4 所示。

当 MAPE 小于等于 20% 时, 可认为模型计算性能较好; 当 MAPE 位于 21%~50% 之间时, 可认为模型计算结果合理^[22]。由表 4 可知, 当交通事故持续时间小于 120 min 时, 此结果是可接受的, 计算结论可作为制订事故发生后交通分流与管控方案的决策

3.2 模型时空稳定性检验

受不同地区、不同时期交通运行环境差异影响, 交通事故持续时间计算模型的变量系数可能不同, 因此需要对模型时空稳定性进行检验。两个地区相同时期模型空间稳定性检验的对数似然比公式, 如式 (4) 所示。

$$LR_D = -2[LL(\beta_D) - LL(\beta_a) - LL(\beta_b)], \quad (4)$$

式中, $LL(\beta_a)$ 和 $LL(\beta_b)$ 分别为使用 a 地区和 b 地区的交通事故数据估计模型收敛时的对数似然; $LL(\beta_D)$ 为两个地区所有的事故数据估计模型收敛时的对数似然。 LR 近似服从 $\chi^2(k)$ 分布, 其自由度 k 与筛选得到的事故持续时间特征变量个数相同。当显著性水平 α 取 0.05 时, 查表可得模型区域稳定性对数似然比检验的临界值。

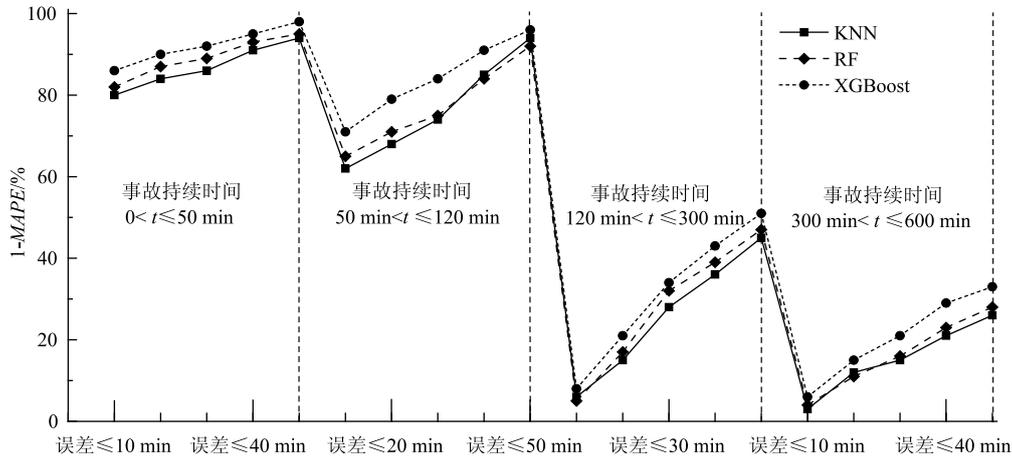


图 4 不同模型事故持续时间计算结果对比

Fig. 4 Comparison of accident durations calculated by different models

a 地区数据为上述研究使用的贵州省 2020 年 8 月 1 日—2021 年 7 月 31 日记录的 5 513 条高速公路交通事故信息, b 地区数据为山东省 2020 年 8 月 1 日—2021 年 7 月 31 日记录的 3 423 条高速公路交通事故信息。不同地区模型系数 $\chi^2(k)$ 的检验结果, 如表 5 所示。

表 5 算法空间稳定性检验

Tab. 5 Algorithm space stability test

检验事项	χ^2 分布	自由度 k	P 值
山东数据检验贵州模型系数	3.998	7	0.78
贵州数据检验山东模型系数	3.549	7	0.83
贵州 a 和山东 b 地区所有事故数据估计模型收敛时的对数似然	0.435	7	1.00

同一地区不同时期模型时间稳定性检验的对数似然比公式, 如式 (5) 所示。

$$LR_t = -2[LL(\beta_T) - LL(\beta_a) - LL(\beta_c)], \quad (5)$$

式中, $LL(\beta_a)$ 同式 (4), $LL(\beta_c)$ 为贵州省 2017 年 10 月 1 日—2018 年 9 月 30 日记录的 6 867 条高速公路事故数据估计模型收敛时的对数似然; $LL(\beta_T)$ 为两个时间段内所有事故数据估计模型收敛时的对数似然。不同时期模型系数 $\chi^2(k)$ 的检验结果, 如表 6 所示。

表 6 算法时间稳定性检验

Tab. 6 Algorithm time stability test

检验事项	χ^2 分布	自由度 k	P 值
贵州 a 时期数据检验 c 时期模型系数	1.564	7	0.98
贵州 c 时期数据检验 a 时期模型系数	1.997	7	0.96
贵州 a 和 c 时期的所有事故数据估计模型收敛时的对数似然	0.435	7	1.00

同一时期不同地区的模型系数随地区转移不具

有稳定性, 但同一地区不同时期的模型系数随时间转移具有稳定性。在模型空间稳定性检验时, 贵州交通事故数据检验山东模型系数的 P 值大于山东数据检验贵州模型系数的 P 值。这可能是因为贵州位于我国西南部, 其地形、气候、交通环境更加多样, 交通事故数据包含的信息更加全面, 事故持续时间计算模型的适应性更好。

3.3 交通事故持续时间计算方法

交通事故持续时间是指从交通事故发生至事故清除的时间, 这一概念符合生存分析对生存时间的定义^[23]。采用 log-logistic 函数对事故持续时间进行计算, 事故持续时间风险函数的 log-logistic 模型表达式, 如式 (6) 所示。

$$h(t) = \frac{\psi^{1/\gamma} t^{(1/\gamma-1)}}{\gamma[1 + (\psi t)^{1/\gamma}]}, \quad (6)$$

式中, $\psi = \exp(-\beta'X)$, X 为特征变量组成的列向量; β 为参数; γ 为示形参数, 当 $\gamma < 1$ 时, $h(t)$ 先增大后减小, 当 $\gamma \geq 1$ 时, $h(t)$ 单调递减。

单方事故、事故车辆停驶在硬路肩、路段服务水平为 1 级、死亡人数 0 人、受伤人数 0 人、涉及车型为小客车, 事发位置为基本路段的交通事故是最常见的一类事故, 可将此类事故看作基准事故。风险函数 $h(t)$ 是风险的一种度量, 在时刻 t_1 和时刻 t_2 之间, $h(t)$ 越大, 则事故清除 (失效) 发生的风险也越大。基准事故持续时间的失效风险如图 5 所示, 估算模型的拐点为 29 min。意味着当事故持续时间超过 29 min 时将被很快清除, 所以可认为基准事故的持续时间为 29 min。

事故持续时间的风险率随时间先增加后减小,

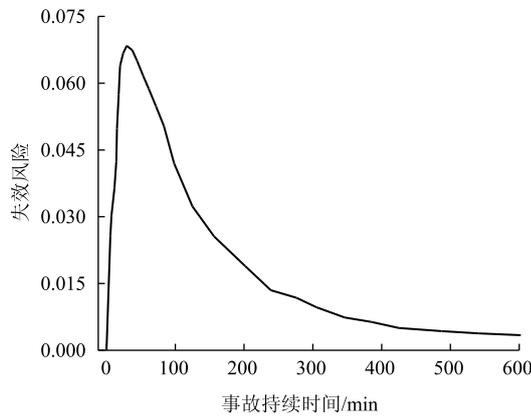


图 5 基准事故持续时间的失效风险

Fig. 5 Failure risk of base accident duration

表明形参数 γ 小于 1。根据交通事故持续时间特征变量敏感性分析结果, 类比基准事故特征变量以及基准事故持续时间, 建立交通事故持续时间计算模型, 如式 (7) 所示。

$$t = 29 \times (1 + 0.240(a_1 - 1) - 0.143(a_2 - 3) + 0.141(a_3 - 1) + 0.102(a_4 - 1) + 0.058(a_5 - 1) - 0.033(a_6 - 4) + 0.025(a_7 - 2)), \quad (7)$$

式中, $a_1 \sim a_7$ 的含义同表 3, 取值方法见表 2。当交通事故发生时, 高速公路管理人员可根据所发生的交通事故特征变量, 快速进行事故持续时间计算。例如, 事故类型为两车追尾、事发路段剩余 1 条通行车道、路段服务水平为 3 级、死亡人数 0 人、受伤人数 2 人、涉及车型为货车、事发位置为基本路段的事故持续时间计算方法, 如式 (8) 所示。

$$t = 29 \times (1 + 0.24 \times (2 - 1) - 0.143 \times (2 - 3) + 0.141 \times (2 - 1) + 0.102 \times (1 - 1) + 0.058 \times (3 - 1) - 0.033 \times (3 - 4) + 0.025 \times (2 - 2)) = 49 \text{ min}. \quad (8)$$

4 结论

本研究对高速公路交通事故持续时间计算方法进行了分析。利用贵州和山东省高速公路交通事件档案, 交通和气象观测站统计数据以及高速公路设计文件, 构建了包含 4 类 15 项事故持续时间影响因素的初始变量集。通过统计学习方法进行变量选择, 建立了包含事故类型、剩余车道数、服务水平、死亡人数、受伤人数、涉及车辆类型、事故发生位置等 7 个变量在内的极限梯度提升树交通事故持续时间计算模型, 并对模型性能进行了检验。得到的基本结论有:

(1) 将交通事件划分为道路交通事故、气象灾

害、地质灾害、公路设施安全事件, 交通事件发生频数和事件持续时间统计结果表明, 交通事故是平均持续时间最短、发生频次最高的一类事件。

(2) 特征变量筛选结果表明, 交通事故特征和交通状况特征对事故持续时间的影响程度高于道路特征和环境特征; 根据所发生的交通事故特征, 由变量敏感性分析结果、基准事故持续时间, 可快速进行事故持续时间计算。

(3) 由于不同省份不同时期交通事故持续时间影响因素存在差异, 所建交通事故持续时间计算模型具有时间稳定性, 但不具备空间稳定性。

参考文献:

References:

- [1] GB/T 39424.1—2020, 道路车辆 交通事故分析 第 1 部分: 术语 [S].
GB/T 39424.1—2020, Road Vehicles — Traffic Accident Analysis — Part 1: Vocabulary [S].
- [2] VORKO-JOVIC A, KERN J, BILOGLAV Z. Risk Factors in Urban Road Traffic Accidents [J]. *Journal of Safety Research*, 2006, 37 (1): 93–98.
- [3] GREIBE P. Accident Prediction Models for Urban Roads [J]. *Accident Analysis and Prevention*, 2003, 35 (2): 273–285.
- [4] NOLAND R B, OH L. The Effect of Infrastructure and Demographic Change on Traffic-related Fatalities and Crashes: A Case Study of Illinois County-level Data [J]. *Accident Analysis and Prevention*, 2004, 36 (4): 525–532.
- [5] 中国公路学报编辑部. 中国交通工程学术研究综述·2016 [J]. *中国公路学报*, 2016, 29 (6): 1–161.
Editorial Department of China Journal of Highway and Transport. Review on China's Traffic Engineering Research Progress: 2016 [J]. *China Journal of Highway and Transport*, 2016, 29 (6): 1–161.
- [6] 李维佳, 王建军, 白骅, 等. 路网环境下考虑大型车混入率的事故疏散诱导模型 [J]. *中国公路学报*, 2020, 33 (11): 275–284.
LI Wei-jia, WANG Jian-jun, BAI Hua, et al. An Accident Induction and Evacuation System Considering the Mixing Rate of the Large Vehicles Under the Freeway Network [J]. *China Journal of Highway and Transport*, 2020, 33 (11): 275–284.
- [7] 姬杨蓓蓓. 交通事件持续时间预测方法研究 [D]. 上海: 同济大学, 2008.
JIYANG Bei-bei. Research on Prediction Method of Traffic Incident Duration [D]. Shanghai: Tongji University, 2008.

- [8] CHUNG Y, YOON B. Analytical Method to Estimate Accident Duration Using Archived Speed Profile and Its Statistical Analysis [J]. *KSCE Journal of Civil Engineering*, 2012, 16 (6): 1064-1070.
- [9] 商强, 林赐云, 杨兆升, 等. 基于变量选择和核极限学习机的交通事件检测 [J]. *浙江大学学报 (工学版)*, 2017, 51 (7): 1339-1346.
SHANG Qiang, LIN Ci-yun, YANG Zhao-sheng, et al. Traffic Incident Detection Based on Variable Selection and Kernel Extreme Learning Machine [J]. *Journal of Zhejiang University (Engineering Science Edition)*, 2017, 51 (7): 1339-1346.
- [10] 高珍, 柯阿香, 余荣杰, 等. 基于随机生存森林的交通事件持续时间预测 [J]. *同济大学学报 (自然科学版)*, 2017, 45 (9): 1304-1310.
GAO Zhen, KE A-xiang, YU Rong-jie, et al. Urban Expressway Traffic Incident Duration Prediction Based on Random Survival Forests [J]. *Journal of Tongji University (Natural Science Edition)*, 2017, 45 (9): 1304-1310.
- [11] 杨超, 汪超. 快速路交通事件持续时间预测模型 [J]. *同济大学学报 (自然科学版)*, 2013, 41 (7): 1015-1019.
YANG Chao, WANG Chao. Traffic Incident Duration Forecast Model of Expressway [J]. *Journal of Tongji University (Natural Science Edition)*, 2013, 41 (7): 1015-1019.
- [12] HOJATI A T, FERREIRA L, WASHINGTON S, et al. Hazard Based Models for Freeway Traffic Incident Duration [J]. *Accident Analysis and Prevention*, 2013, 52: 171-181.
- [13] 姬杨蓓蓓, 张小宁, 孙立军. 基于贝叶斯决策树的交通事件持续时间预测 [J]. *同济大学学报 (自然科学版)*, 2008, 36 (3): 319-324.
JIYANG Bei-bei, ZHANG Xiao-ning, SUN Li-jun. Traffic Incident Duration Prediction Grounded on Bayesian Decision Method-based Tree Algorithm [J]. *Journal of Tongji University (Natural Science Edition)*, 2008, 36 (3): 319-324.
- [14] CHUNG Y. Development of an Accident Duration Prediction Model on the Korean Freeway Systems [J]. *Accident Analysis and Prevention*, 2010, 42 (1): 282-289.
- [15] LIN P W, ZOU N, CHANG G L. Integration of a Discrete Choice Model and a Rule-based System for Estimation of Incident Duration: a Case Study in Maryland [C]//CD-ROM of Proceedings of the 83rd TRB Annual Meeting. Washington, D. C. : TRB, 2004.
- [16] VALENTI G, LELLI M, CUCINA D. A Comparative Study of Models for the Incident Duration Prediction [J]. *European Transport Research Review*, 2010, 2 (2): 103-111.
- [17] KHATTAK A J, SCHOFFER J L, WANG M H. A Simple Time-Sequential Procedure for Predicting Freeway Incident Duration [J]. *IVHS Journal*, 1995, 2 (2): 113-138.
- [18] WANG W, CHEN H, BELL M C. Vehicle Breakdown Duration Modelling [J]. *Journal of Transportation and Statistics*, 2005, 8 (1): 75-84.
- [19] KIM H J, CHOI H K. A Comparative Analysis of Incident Service Time on Urban Freeways [J]. *IATSS research*, 2001, 25 (1): 62-72.
- [20] MENG F, WONG W, WONG S C, et al. Gas Dynamic Analogous Exposure Approach to Interaction Intensity in Multiple-vehicle Crash Analysis: Case Study of Crashes Involving Taxis [J]. *Analytic Methods in Accident Research*, 2017, 16: 90-103.
- [21] CHEN T, GUESTRIN C, ASSOC COMP M. XGBoost: A Scalable Tree Boosting System [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco: [s. n.], 2016.
- [22] JERALD F. Lawless. *Statistical Models and Methods for Lifetime Data* [M]. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2003.
- [23] CHUNG Y. Development of an Accident Duration Prediction Model on the Korean Freeway Systems [J]. *Accident Analysis and Prevention*, 2010, 42 (1): 282-289.