

交通事件持续时间分布拟合及其加速消散模型

丛浩哲^{1,2}, 王俊骅¹, 方守恩¹, 童世鑫¹

(1. 同济大学 道路与交通工程教育部重点实验室, 上海 201804; 2. 公安部道路交通安全研究中心, 北京 100062)

摘要:对某高速公路交通事件管理系统中记录的 3 年交通事件信息进行了数据统计处理,对交通事件持续时间的随机分布进行了多种拟合分析和分布检验,构建了基于对数逻辑斯特分布的交通事件持续时间加速消散模型.该预测模型可以使用不同的拟合分布并接受缺失数据,样本检验能很好地预测交通事件持续时间及其累积结束概率.

关键词:高速公路; 交通事件; 持续时间; 分布拟合; 加速消散模型

中图分类号:U 491.31

文献标识码:A

Fitting Distribution of Freeway Incident Duration and Accelerated Failure Time Modeling

CONG Haozhe^{1,2}, WANG Junhua¹, FANG Shouen¹, TONG Shixin¹

(1. Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China; 2. Road Traffic Safety Research Centre of the Ministry of Public Security, Beijing 100062, China)

Abstract: Distribution model choice, parameters' estimation and measures of goodness of fit were carried out, and the best distribution fitting for incident duration data was Log-Logistic distribution. Accelerated failure time (AFT) modeling was employed based on survival analysis and Log-Logistic distribution. Then stepwise by Akaike information criterion (AIC) was used to find the best group of factors for the prediction of freeway incident duration. The test result indicates that AFT model can represent actual incident duration and accept different distribution and missing data.

Key words: freeway; incident; duration; fitting distribution; accelerated failure time

高速公路的拥堵与交通安全问题已经成为我国高速公路交通管理面临的严峻考验.高速公路突发的交通事件不仅会引发不同程度的交通拥挤,而且容易导致交通事故及二次事故的发生,是道路交通管理的主要影响因素.交通事件的快速处置对于交通事故的快速救援,交通拥堵的及时疏导,交通安全隐患的有效排除具有重要意义.

交通事件持续时间是交通事件管理系统中的一项重要决策依据,是实现先进的突发交通事件管理和驾驶员信息服务的基础.交通事件持续时间的预测能够帮助道路交通管理人员确定最佳的紧急救援方案、交通管控措施和安全隐患排除对策,从而有效地减少交通延误,减少交通运行风险,提高交通事件管理的水平.目前,道路交通事件持续时间预测方法主要包括:状态概率,回归分析^[1],时间序列,决策树^[2],非参数回归,模糊逻辑^[3],危险分析^[4]等方法.然而,国外相关研究缺乏具体分析过程.考虑了我国道路条件、车辆性能及管理体制等方面的差别,亟需对我国交通事件的持续时间特征进行深入分析,并构建符合国情的预测模型.

本文就某高速公路交通事件管理系统中记录的 3 年交通事件信息进行数据统计处理,对交通事件持续时间的随机分布进行了多种拟合分析和分布检验,然后构建了基于对数逻辑斯特分布的交通事件持续时间加速消散模型,最后对模型预测分析的效果进行了检验.

1 交通事件持续时间分布拟合

本文研究的高速公路交通事件属狭义的交通事件范畴,是指偶然发生的交通事件并影响交通流的正常运行,如车辆碰撞、刮擦、抛锚、炸胎、车辆着火

收稿日期:2010-11-01

基金项目:“十一五”国家科技支撑计划(2009BAG13A06);高等学校博士学科点专项科研基金(200802470028)

第一作者:丛浩哲(1982—),男,博士生,主要研究方向为道路交通安全以及道路规划与设计. E-mail:conghaozhe@126.com

通讯作者:王俊骅(1979—),男,讲师,工学博士,主要研究方向为道路交通安全以及道路规划与设计. E-mail:benwjh@163.com

及散落物等. 交通事件持续时间的定义通常是指交通事件从发生到消除事件地点中所有事件痕迹的时间. 美国《道路通行能力手册》将交通事件持续时间分为 4 个阶段: 事件检测时间, 事件响应时间, 事件清理时间和交通恢复时间.

关于交通事件持续时间分布的研究, Golob 等人研究了加州洛杉矶地区高速公路涉及卡车的事故数据, 发现其持续时间服从对数正态分布^[5]. 同时, Jones 等人分析了西雅图地区的事件数据, 发现其服从对数逻辑斯特分布^[6], 而 Nam 等人发现某些事件持续时间服从威布尔分布^[2]. 总之, 这些分布的共同特点是分布曲线都向左偏移, 持续时间较短的事件比例大, 符合实际情况.

1.1 事件持续时间数据总体描述

本文数据来源于浙江省某高速公路 2006—2008 年监控中心接处警系统记录的 1 198 条交通事件信息. 该高速公路为单向 4 车道, 全长 78 km, 设计车速 120 km · h⁻¹. 系统中记录的交通事件开始时间是接到报警的时间, 结束时间是事件清理结束时间, 没有记录交通事件实际发生时刻和事件清除完成到道路恢复正常交通状况的时间. 因此本文研究的交通事件持续时间实质上只包含了事件响应时间和事件清理时间 2 个主要阶段. 用于分布拟合的交通事件持续时间样本总体描述如表 1 所示.

表 1 交通事件持续时间样本总体描述

Tab.1 Sample summaries of incident duration

样本统计项	统计值	百分位/%	对应的持续时间值/min
样本数	1 198.00	0.1	2.00
范围	490.00	5.0	14.00
平均值	66.77	10.0	20.00
方差	3 752.40	25.0	33.00
标准差	61.26	50.0	49.00
变异系数	0.92	75.0	75.00
标准误	1.77	90.0	136.00
偏度	3.06	95.0	177.15
峰度	13.12	100.0	492.00

1.2 事件持续时间的分布拟合

事件持续时间属连续型随机变量, 为了进行分布特征分析, 需要进行分组后计算每组中事件出现的频数, 从而形成对应的频数分布直方图, 横轴表示数据分组, 纵轴表示频率. 事件持续时间的分布可以通过概率密度函数进行描述, 记作 $f(t)$, 如图 1 所示. 本文采用正态分布、对数正态分布、威布尔分布和对数逻辑斯特分布对事件持续时间进行分布

拟合.

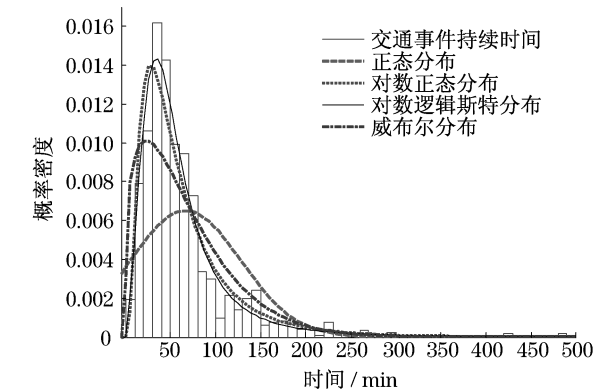


图 1 交通事件持续时间分布拟合
Fig.1 Fitting distribution of incident duration

1.3 分布拟合度检验

各种分布的参数以及 K-S 和 A-D 检验的统计值和 P 值见表 2, 只有对数逻辑斯特分布通过了 2 种检验. 本文事件持续时间服从对数逻辑斯特分布.

表 2 交通事件持续时间分布的估计参数

Tab.2 Estimated parameters of incident duration distribution

分布类型	位置	尺度	K-S 检验		A-D 检验	
			统计值	P 值	统计值	P 值
对数逻辑斯特	2.36	49.76	0.03	0.240	1.04	0.25
对数正态分布	0.77	3.91	0.05	0.005	3.54	0
正态分布	61.25	66.77	0.20	0	88.75	0
威布尔分布	1.63	70.81	0.09	0	24.55	0.01

2 事件持续时间加速消散模型

本文对交通事件持续时间的预测引用的是统计学中的生存分析理论^[7]. 生存分析最早应用于医学, 描述的是某个个体的生存时间, 即病人从某一观测时间点开始, 直到死亡的时间. 后应用于工程和社会经济等领域, 工程中称为可靠性理论, 社会学中称为延时延误模型. 对于交通事件持续时间而言, 不仅关心事件持续时间的分布及其回归模型, 而且关心事件已经持续了一段时间, 在后续一个较短的时间段内结束的概率是多少. 基于这种事件管理的需求, 本文构建了基于对数逻辑斯特分布的交通事件持续时间加速消散模型, 能够解决上述 2 种需求问题.

2.1 理论基础

假设观测时间足够长, 使得每一起交通事件的持续时间 t_i 都能够被精确地观测到. 同时, 假设对

于第 i 起交通事件, 还知道与其持续事件 t_i 相关的解释性变量 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$. 需要量化生存时间 t_i 和解释性变量 x_i 之间的关系. 交通事件持续时间是个非负随机变量, 因而需要对这个因变量 t_i 作对数变换, 构造广义线性模型

$$\ln t_i = \mathbf{x}_i' \beta + w_i$$

式中: β 为回归参数; w_i 是服从某种分布的残差项, 亦可以理解为在没有协变量影响下的对数生存时间. 该模型可以变形为

$$\ln t_i = \mathbf{x}_i' \beta + \ln t_{i0} \Rightarrow t_i = t_{i0} \exp(\mathbf{x}_i' \beta)$$

这说明, 如果没有协变量的影响, 该个体的实际生存时间应该为 t_{i0} . 但是由于交通管控措施、救援清障措施等突发事件综合管理的介入, 致使交通事件的持续时间产生了一个加速消散的过程, 其生存时间被“加速”为 $t_{i0} \exp(\mathbf{x}_i' \beta)$. 因此, 本文称此模型为交通事件的快速消散模型.

2.2 交通事件解释性变量总体描述

首先对文字的记录信息进行数字化处理, 该记录信息中包含各类事件的以下条目: 周天、天气、报警时间、报警类型、到达现场时间、事件类型、占用车道数、涉及车辆数、大车数、受伤人数、死亡人数、最先到达现场的车辆、救援车辆数以及事件持续时间. 经数字化后的交通事件数据字段及变量说明见表 3, 其中类型 C 表示连续型变量, 共 7 项; D 表示离散型变量, 共 7 项. 然后将数字化后的 1 198 条事件信息组成的总体样本(完整数据样本), 随机抽取 900 条(约占总数 75%)事件信息组成建模样本组, 剩余的 298 条事件信息组成模型检验样本组. 同时, 事件记录信息中另有部分样本缺失了事件持续时间, 但交警到场时间、事发车辆靠边时间可以作为补充信息. 即使这种右截断(right censored)数据无法给出事件确切的结束时间, 但是能够说明事件至少存在了多长时间. 为了验证加速消散模型可以接受缺失数据, 将建模样本分成 2 组, 第 1 组是完整数据建模样本, 包含 900 条事件信息; 第 2 组是完整数据样本和缺失数据样本共同组成的建模样本, 包含 900 条完整数据和 336 条缺失数据, 共计 1 236 条事件信息. 在第 2 组建模样本数据中增加能够区分完整数据和右截断数据的事件状态字段, 1 表示结束, 即完整数据, 0 表示未结束, 即右截断的缺失数据. 完整数据建模样本用来构建加速消散模型和传统多元线性回归模型, 含缺失数据样本用来构建可接受缺失数据的加速消散模型.

表 3 交通事件数据字段及变量解释

Tab.3 Fields and variables explanation of incident data

字段名	类型	变量解释
周天	D	1—7 分别表示周一至周日
天气	D	1 晴天, 2 阴天, 3 雨天, 4 雪天, 5 雾天
报警时间	D	1 白天非高峰, 2 白天高峰, 3 夜间
报警类型	D	1 驾驶员, 2 交警部门, 3 高速公路公司, 4 交通部门
到场时间	D	1(0~15 min), 2(16~30 min), 3(31~60 min)
事件类型	D	1 追尾, 2 翻车, 3 撞中间护栏, 4 抛锚, 5 刮擦, 6 炸胎, 7 撞边护栏, 8 散落物, 9 着火
占用车道数	C	1 为 1 条车道, 5 为单向路段封闭(将硬路肩视为 1 个车道来处理)
涉及车辆数	C	单位: 辆
大车数	C	单位: 辆(包含大货车数和大客车数)
受伤人数	C	单位: 人
死亡人数	C	单位: 人
最先到场车辆	D	1 警车, 2 拖车, 3 修胎车, 4 医疗车, 5 路政, 6 管理处, 7 巡警
救援车辆数	C	单位: 辆
事件持续时间	C	单位: min

2.3 加速消散模型的构建

假定事件持续时间 T 的连续概率密度函数为 $f(t)$, 其分布函数 $F(t) = \int_0^t f(t)dt = P(T \leq t)$, 则生存函数为 $S(t) = 1 - F(t) = P(T \geq t)$. 那么, 当一起交通事件已经持续了时间 t , 在后续时间 Δt 结束的概率表示为条件概率 $P(t \leq T \leq t + \Delta t | T \geq t)$. 此条件概率可以刻画为风险率

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - f(t)}{\Delta t S(t)} = - \frac{f(t)}{S(t)}$$

令 $\sigma = 1/p$, $\delta_i = 1$ 代表事件状态为结束, $\delta_i = 0$ 代表事件状态为截断数据(缺失数据). 同时, 令

$$w_i = p \ln(\lambda_i t_i) = \frac{(\ln t_i - \mathbf{x}_i' \beta)}{\sigma}$$

记 w_i 的密度函数和生存函数为 $f(w_i)$ 和 $S(w_i)$, 那么 $\ln t_i$ 的密度函数和生存函数为

$$f(w_i) = \frac{e^{w_i}}{(1 + e^{w_i})^2}, S(w_i) = \frac{1}{1 + e^{w_i}}$$

基于样本(D 代表样本数据)的对数似然函数为

$$\ln L(\beta, \sigma | D) = \sum_{i=1}^n [\delta \ln f(\ln t_i | x_i, \beta, \sigma) + (1 - \delta_i) \ln S(\ln t_i | x_i, \beta, \sigma)]$$

通过极大化该似然函数估计, 并进行似然比检验.

在加速消散模型构建的过程中,解释性变量的选取也是一个关键问题.高速公路交通事件影响因素很多,如果在回归方程中忽略了对因变量显著影响的自变量,所建立的预测方程与实际就会有很大偏离,但变量选择过多,使用就不方便.因此,适当地选择变量来建立一个最优的回归方程十分重要,常用的方法有:一切子集回归法、前进法、后退法及逐步回归法等.本文采用逐步回归函数进行变量选择,并使得 AIC(akaike information criterion)值最小时

的变量组合为最佳.最优选择变量为:报警时间段,交警到达现场时间,事件类型,占用车道数,大车总数,涉及车辆总数,死亡人数及救援车辆数等 8 个变量.利用 2 组建模样本构建的 3 种模型参数见表 4.需要解释的是,模型截距项的含义是指报警时间为白天非高峰时段,到场时间在 15 min 以内,事件类型是追尾的交通事件持续时间预测的基准参照系数,其他类型的交通事件预测都是在此截距的基础上计算得到的.

表 4 交通事件持续时间加速消散模型参数表
Tab.4 Estimated results of AFT model for incident duration with or without missing data VS linear model

解释性变量	含缺失数据加速消散模型			完整数据加速消散模型			传统多元线性回归模型		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
截距	3.06	0.09	0	3.08	0.09	0	-2.10	7.96	0.79
报警时间段(白天高峰)	0.06	0.04	0.19	0.05	0.04	0.22	5.69	3.91	0.15
报警时间段(夜间)	0.25	0.05	0	0.25	0.05	0	18.11	4.38	0
到场时间段(16~30 min)	0.13	0.07	0.05	0.13	0.07	0.06	1.28	6.33	0.84
到场时间段(>30 min)	0.84	0.20	0	0.78	0.20	0	41.33	20.10	0.04
事件类型(翻车)	0.26	0.06	0	0.21	0.07	0	17.61	5.95	0
事件类型(撞中间护栏)	0.23	0.07	0	0.06	0.08	0.40	11.83	6.70	0.08
事件类型(抛锚)	-0.26	0.07	0	-0.39	0.08	0	-12.34	6.89	0.07
事件类型(刮擦)	0.17	0.11	0.14	0.16	0.12	0.16	15.46	9.65	0.11
事件类型(炸胎)	-0.01	0.12	0.56	-0.03	0.13	0.80	37.64	9.06	0
事件类型(撞边护栏)	-0.05	0.12	0.48	-0.06	0.12	0.60	4.74	11.05	0.67
事件类型(散落物)	-0.11	0.10	0.27	-0.17	0.11	0.13	5.41	8.93	0.54
事件类型(着火)	0.53	0.18	0	0.54	0.18	0.00	80.65	14.23	0
占用车道数	0.16	0.03	0	0.15	0.03	0	14.82	2.81	0
大车总数	0.25	0.03	0	0.22	0.03	0	18.77	2.91	0
涉及车辆总数	0.09	0.03	0	0.06	0.03	0.04	3.57	2.52	0.16
死亡人数	0.37	0.11	0	0.35	0.11	0	53.55	9.33	0
救援车辆数	0.16	0.02	0	0.18	0.02	0	11.36	1.56	0
尺度参数的对数值	-1.13	0.03	0	-1.14	0.03	0			

由表 4 可以看出,传统的多元线性模型截距项没有通过显著性检验,而加速消散模型的截距项则通过了显著相关的检验.另外,若以显著水平为 0.05 为界,加速消散模型的参数有 13 项通过了显著性检验,而线性模型只有 9 项通过了显著性检验.出于截距项是基准参照系数的考虑,加速消散模型总体上优于线性模型.加速消散模型可以针对不同的分布构建相应的极大似然函数,产生新的生存函数,来接受不同的拟合分布,同时加速消散模型当接受了包含缺失数据的部分信息后,模型参数的整体显著水平有了一定的提高,在模型的推广应用方面具有很好的可移植性和鲁棒性,这也是传统线性模型无法实现的.

根据构建的加速消散模型可以发现,交通事件持

续时间的主要影响因素分别是报警时间(夜间)、到场时间(>30 min)、占用车道数、大车总数、死亡人数及救援车辆数,这些因素的增多都会导致事件持续时间的延长.在事件类型中持续时间相对较长的是追尾(对应截距项的系数)、翻车、撞中间护栏、着火.相比之下,车辆抛锚、炸胎及散落物等单纯的清障事件要比交通事故的持续时间短(系数为负值).因此,从事件管理的角度分析,建议提高夜间事件的管理水平,加强夜间值班联动及其装备,重视早晚高峰时段交通事故的快速处置.争取在接到报警后 15 min 内到达现场,最迟不要超过 30 min.重点加强对车辆着火、翻车、追尾等交通事故的快速处置水平,完善应急处置预案.针对不同事件类型和占用车道情况,缩短事发车辆靠边时间,在减少影响车道数的同时,加快事件

的快速处置.加强涉及大货车和大客车的交通事件的处置,提高救援车辆(如拖车、驳车、吊车、救护车等)的资源优化分布及快速反应水平,实现高效的救援组织方式.根据事件的严重程度(尤其是重大伤亡事故),配置合理的处置装备,达到与救援和清障为目的的最优化匹配物资,在满足有效救援和清障的前提下,减少事件的持续时间.

3 模型效果检验

为了检验模型的效果,现在用 298 条事件信息组成模型检验样本组对 3 种模型的预测效果进行检验和对比分析.以事件类型为例,来说明模型预测效果.由图 2 可以看出,加速消散模型对追尾、翻车、撞中间护栏、炸胎、着火的事件持续时间预测效果最好,平均预测误差在 2~9 min,而车辆撞边护栏和炸胎的预测精度相对较低.当加速消散模型融入了缺失数据的部分信息后,该模型的预测精度也有了进一步的提高.3 种模型中,含缺失数据样本的加速消散模型总体预测精度最高.

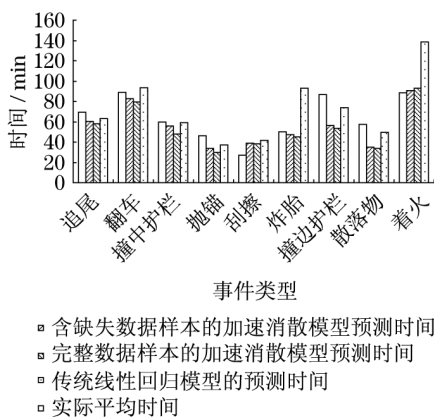


图 2 事件实际持续时间与预测持续时间的对比

Fig.2 Actual mean value VS predicted mean value of incident duration

另外,加速消散模型可以计算出持续时间的累积结束概率.例如已知事件持续时间为 t ,可以预测出其在下一个较短的时间间隔 Δt 后结束的概率是多少.根据事件的累积风险函数可以算出全部事件类型对应的累积风险概率,如图 3 所示.受篇幅所限,不再列表显示不同时间间隔的累积结束概率值.图中累积结束概率曲线上“+”符号代表的是缺失数据对累积结束概率的贡献点.

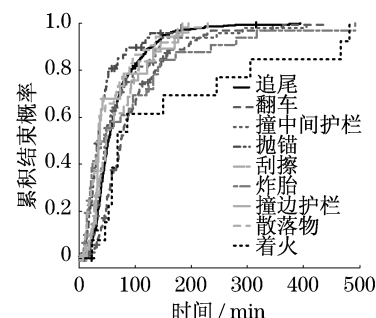


图 3 不同事件类型的持续时间累积结束概率

Fig.3 Accumulated ending probabilities of duration under different incident types

4 结论

本文就某高速公路 3 年的交通事件持续时间的随机分布进行了多种拟合分析和分布检验,发现事件持续时间服从对数逻辑斯特分布,此分布特征适用于我国东部平原区的高速公路交通事件持续时间的总体分布拟合.在此基础上,构建了基于对数逻辑斯特分布的交通事件持续时间加速消散模型,模型的检验效果表明,能够接受缺失数据的加速消散模型很好地预测了交通事件持续时间,可以满足事件管理对于事件持续时间预测的需要.今后,需要对模型在不同事件特征情况下的分布分别拟合,引入不同持续时间段记录的缺失数据进行验算,以便对模型的可移植性和鲁棒性作更好的证明和检验.

参考文献:

- [1] Garib A, Rawan A E, Al-Deek H. Estimating magnitude and duration of incident delays [J]. Journal of Transportation Engineering, 1997, 123(6): 459.
- [2] 姬杨蓓蓓. 交通事件持续时间预测方法研究[D]. 上海: 同济大学交通运输工程学院, 2008.
JIYANG Beibei. Research on prediction method of traffic incident duration [D]. Shanghai: Tongji University. College of Transportation Engineering, 2008.
- [3] CHOI Hoikyun. Predicting freeway traffic incident duration in an expert system context using fuzzy logic [D]. Los Angeles: University of Southern California. School of Urban Planning and Development, 1996.
- [4] Nam, Doohee, Fred Mannering. An exploratory hazard-based analysis of highway incident duration [J]. Transportation Research Part A, 2000, 34(2): 85.
- [5] Golob, Thomas F, Wilfred W Recker, et al. An analysis of the severity and incident duration of truck-involved freeway accidents [J]. Accident Analysis & Prevention, 1987, 19(4): 375.
- [6] Jones, Bryan, Lester Janssen, et al. Analysis of the frequency and duration of freeway accidents in Seattle [J]. Accident Analysis & Prevention, 1991, 23(4): 239.
- [7] Greene H. Econometric analysis [M]. 5th ed. Englewood Cliffs: Prentice Hall, 2002.