

# 基于风险分析的交通事件持续时间预测

康国祥, 方守恩

(同济大学 道路与交通工程教育部重点实验室 上海 200092)

**摘要:** 采用基于风险分析的参数估计方法, 利用嘉兴市辖区高速公路上 1 062 个交通事件持续时间数据, 建立交通事件持续时间的威布尔加速失效时间预测模型. 并用 265 个交通事件数据对模型的预测精度进行检验. 检验结果表明, 交通事故持续时间预测误差小于 30 min 的准确率为 78%, 而只考虑持续时间大于 10 min 的抛锚事件时, 其持续时间预测误差小于 10 min 的准确率为 70%. 基于风险分析的方法比仅基于贝叶斯或仅基于决策树算法的预测精度更高, 数据利用范围更大.

**关键词:** 高速公路; 交通事件; 事件持续时间; 持续时间预测; 基于风险的持续时间模型; 威布尔加速失效时间模型

**中图分类号:** U491

**文献标识码:** A

## A Hazard-based Analysis of Traffic Incident Duration Prediction

KANG Guoxiang, FANG Shouen

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China)

**Abstract:** This paper presents a model for incident duration prediction based on accurately recorded and large incident dataset from the Jiaxing city's freeway incident response center system. To develop the duration prediction model, this study utilizes the Weibull accelerated failure time (AFT) model and a 3-year incident duration dataset from 2006 to 2008. Specifically, the 1062 incident dataset is utilized to develop the prediction model and then, the 265 dataset are employed to test the accuracy of prediction results of the 1062 model. Although the duration prediction model has limitations such as large prediction error due to the individual differences of the incident treatment teams in terms of clearing similar incidents, the results show that the prediction accuracy of anchor duration whose duration are more than 10 min is 70% with an error of 10 min while that of accident duration is 78% with an error of 30 min. So the hazard-based method is more extensive for dataset utilization and accurate than the method

based on Bayesian method or decision tree respectively.

**Key words:** freeway; incident; incident duration; duration prediction; hazard-based duration models; accelerated failure time (AFT) model

交通事件(如交通事故、车辆抛锚、天气状况等)是造成高速公路交通拥堵的主要原因之一. 一些研究表明<sup>[1-2]</sup>, 高速公路交通事件引起的偶发性交通拥堵占总拥堵的 50%~75%. 实践证明, 交通事件造成的延误大小在很大程度上取决于事件处理过程, 事件发现时间、反应时间以及清除时间越短, 它对道路交通流的影响就越小. 然而, 对于交通管理者而言, 最困难的一步是要预估交通事件的持续时间, 预估准确, 才能有助于决定采取何种有效的管理措施和交通转移策略. 另外, 预估准确的交通事件持续时间信息的发布还可以使交通出行者改变行车路线或者重新规划他们的出行计划.

目前, 最常用的事件持续时间预测方法是基于风险分析的方法<sup>[3-5]</sup>、回归分析方法<sup>[6-8]</sup>和决策树方法<sup>[9-11]</sup>. 基于风险分析的交通事件持续时间预测是一种条件概率的方法, 相比于其他研究方法, 它可以更好地处理因事件结束时间删失导致的信息残缺数据, 扩大了可利用的交通事件数据的范围, 并且可以定量地分析观察指标的作用强度和方向. 本文首先介绍了交通事件持续时间的定义和构建基于风险分析的持续时间预测模型, 接着对所用数据进行了描述, 然后对建立的预测模型进行参数估计和预测精度检验, 最后进行了总结并提出了未来进一步研究的建议.

## 1 基于风险分析的交通事件持续时间预测模型

收稿日期: 2010-12-01

第一作者: 康国祥(1977—), 男, 博士生. 主要研究方向为道路交通安全, 道路规划与设计. E-mail: kangguoxiang@sina.com

通讯作者: 方守恩(1961—), 男, 教授, 博士生导师, 主要研究方向为道路交通安全, 道路规划与设计. E-mail: fangsek@tongji.edu.cn

### 1.1 交通事件持续时间

交通事件持续时间是指从交通事件发生到交通恢复到正常状态的总时间,可分为四个阶段,如图1所示.由于第四个阶段是不可控制的变量,受各种因素的影响而变得比较复杂,并且在实际较难获得关键信息,因此本文研究的交通事件持续时间仅包含前三个阶段,即从事件发生至现场清除完毕,救援车辆离开时所经历的时间.

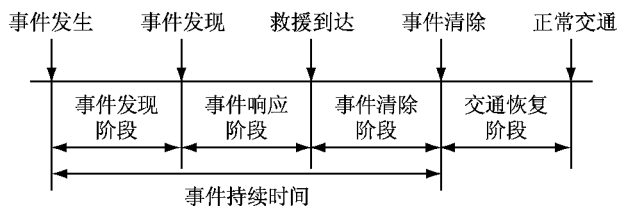


图1 交通事件持续时间的四个阶段

Fig.1 Components of accident duration

### 1.2 模型的建立

基于风险分析的持续时间是指从某个起始事件开始到某个终点事件发生所经历的时间,也称失效时间.某个具体交通事件的持续时间是一个随机变量,用  $T$  表示.基于风险的持续时间模型就是采用条件概率的思想,分析一个已经持续了  $t$  min 的事件在第  $t+\Delta t$  min 结束的概率,这个条件概率  $h(t)$  被称为风险函数(Hazard Function, HF).

基于风险的持续时间模型可以通过四个函数来描述:分布函数  $F(t)$ , 概率密度函数  $f(t)$ , 生存函数  $S(t)$ , 风险函数  $h(t)$ . 四个函数之间的关系如下:

一次交通事件的持续时间为  $T$ , 它小于一个定值  $t$  的概率为  $F(t)$ :

$$F(t) = P(T < t) = \int_0^t f(u) du \quad (1)$$

$F(t)$  是可微的,其概率密度函数为  $f(t)$ , 表示一次交通事件在  $t \sim t+\Delta t$  这一微时段内结束的概率,可表示为

$$f(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (2)$$

生存函数  $S(t)$  表示一次交通事件持续时间长于  $t$  的概率,可表示为

$$S(t) = P(T \geq t) = 1 - F(t) \quad (3)$$

风险函数  $h(t)$  指  $t$  时刻仍未结束的交通事件在  $t$  时刻后一瞬间结束的概率,为条件概率,即交通事件持续了  $t$  时刻的条件下在  $t \sim t+\Delta t$  这一微时段内结束的概率,其表达式为

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (4)$$

如果风险函数的导数  $dh(t)/dt$  大于 0, 则风险函数值随时间的增加而增加,表示若事件尚未结束,则预示着事件会很快被清除;如果风险函数值的导数  $dh(t)/dt$  小于 0, 则风险函数值随时间的增加而减小,预示事件可能仍将持续一段时间,不会立刻结束;如果风险函数值的导数  $dh(t)/dt$  等于 0, 则说明事件结束时间的概率与时间变量无关.

比例风险(proportional hazard, PH)模型和加速失效时间(accelerated failure time, AFT)模型是两种比较常用的基于风险的持续时间预测方法.两种方法都可以用来分析解释变量对持续时间概率的影响,前者不需要对理论分布做任何假设,是一种半参数的估计方法,后者是一种参数估计方法,通常假定持续时间服从特定的参数分布,根据分布的特点对影响持续时间的变量进行分析,常用的方法有指数分布法、Weibull(威布尔)分布法、Log-normal(对数正态)回归分析法和 Log-logistic(对数逻辑斯蒂)回归分析法等.为了进行持续时间预测,参数模型比半参数模型更为准确.如图2所示,将所用数据的概率分布图与特定分布的概率图进行比较发现,Weibull分布比较结果呈线性关系.因此,Weibull

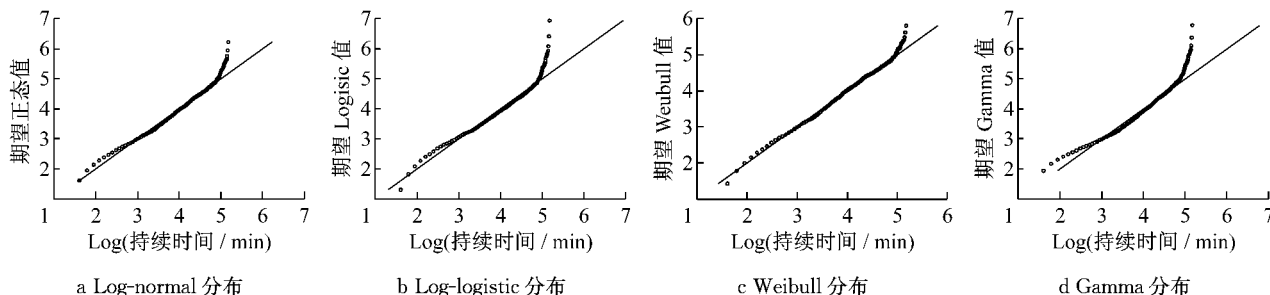


图2 交通事件持续时间概率分布图

Fig.2 Graphs on probability distribution of traffic incident duration

AFT(威布尔加速失效时间)模型更为适合本文数据进行交通事件持续时间的预测。

除已引入变量外,由于持续时间的影响因素较多,故引入异质性参数 $\gamma$ ,以验证是否仍存在未知因素可能会对风险函数形状及其协变量系数产生较大影响,一般假设 $\gamma$ 服从Gamma分布.考虑Gamma异质性的Weibull AFT模型的风险函数表达式为

$$h(t) = [1 - F(t)]^\theta (\lambda P)(\lambda t)^{P-1} \quad (5)$$

$$\lambda > 0, \gamma > 0, \theta > 0$$

式中: $P$ 为形状参数,决定分布函数的形态,当 $P=1$ 时,危险率是常数.当 $P>1$ 时,危险率随时间增大而增大, $P<1$ 时危险率随时间增大而减小; $\lambda$ 为尺度参数,决定分布函数的分散度,尺度参数 $\lambda = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m)$ , $X_i (i=1,2,\dots,m)$ 为危险因素, $\beta_0$ 为截距, $\beta_i$ 为危险因素的回归系数; $\theta$ 是异质性参数 $\gamma$ 的方差,当不存在异质性时,即不存在未知因素对模型产生较大影响, $\theta$ 值接近于0.

## 2 数据描述

所使用数据来源于浙江省嘉兴市高速公路应急救援指挥中心管理系统,该中心承担着辖区内高速公路上交通事件的报警、求助、施救、投诉、信息发布等工作,并详细记录高速公路养护、交通事件处置等相关信息.选择与目前处理事件的条件差异不大的数据来标定模型,以减少数据的离散性,避免了由于数据离散本身对模型精度的影响.所分析的交通事件数据是2006—2008年近3年双向8车道高速公路上发生的交通事件,见表1,选取了1 327组有效数据进行研究,将样本随机分为两部分,一部分用于风险模型的标定,标定数据为1 062个,占样本总数的80%,另一部分验证风险模型,共265个数据,占样本总数的20%.

表1 用于标定和验证的交通事件数据的描述性统计

Tab.1 Descriptive statistic of traffic incident data for validation and calibration

指标	标定数据	验证数据
事件数量	1 062.0	265.0
均值/min	50.3	57.7
中值/min	44.0	49.0
标准差/min	31.9	39.0
最小值/min	5.0	5.0
最大值/min	167.0	177.0
25%分位数	23.0	29.5
75%分位数	65.0	74.0

每组数据由事件的持续时间和可能影响事件持

续时间的因素构成,这些因素大致分为以下几类:事件时间信息,如发生时刻、日、月、删失状态;事件基本信息:事件类型、涉及车辆类型、严重程度(如涉及车辆数量、人员死亡和受伤人数)、发生位置;环境状况,如天气、道路表面状况、工作区、是否有车辆停靠在路肩;事件信息来源,如报警人类型;救援车辆信息,如是否需要救护车、牵引车、吊车等.由于影响因素较多,只有那些更直观、更易获得的显著信息才会被用到模型构建当中.取显著水平 $\alpha=0.05$ ,通过影响因素显著性分析最终获得除删失状态外的15个显著因素( $P<0.05$ ),而删失状态仅作为状态变量引入,不作为影响因素考虑,见表2,并将这15个因素作为初始变量进行模型的参数估计.

表2 交通事件持续时间的显著影响因素

Tab.2 Significant influence factors of traffic incident duration

变量名	变量赋值
删失状态	0=完整持续时间数据 1=删失持续时间数据
是否发生在夜晚	0=否 1=是
是否发生在高峰时段	0=否 1=是
报警人类型	0=驾驶员 1=交通管理者
事件类型	1=追尾 2=翻车 3=抛锚 4=碰撞固定物 5=着火
占用车道数	1,2,3,4
涉及车辆数	1,2,3,4,5+
受伤人数	0,1,2,3,4+
是否有人员死亡	0=无 1=有
是否有车辆翻车	0=无 1=有
是否需要救护车	0=不需要 1=需要
是否需要牵引车	0=不需要 1=需要
是否需要吊车	0=不需要 1=需要
是否需要驳载车	0=不需要 1=需要
是否需要消防车	0=不需要 1=需要
是否涉及货车	0=不涉及 1=涉及

## 3 模型的估计和检验

运用上述基于风险分析的方法,利用R软件进行Weibull AFT持续时间模型的参数估计,结果见表3.从表3中可以看出:

(1) 形状参数 $P>1 (P=1.8094)$ ,这表明危险率随时间增大而增大,见图3. $\theta$ 值接近于零,说明不存在未知因素对模型参数和风险函数产生较大影响.

(2) 只有截距、变量如夜晚、报警人类型、事件类型、占用车道数、涉及车辆数、人员死亡、救护车、牵引车、吊车、驳载车、涉及货车对持续时间影响有

非常显著性意义,其他变量均无显著性(显著性水平取  $\alpha=0.05$ )。

表 3 考虑 gamma 异质性的 Weibull AFT 持续时间模型估计结果

Tab.3 Estimated results of Weibull AFT model for incident duration with gamma heterogeneity

Variable	Estimated coefficient	Standard error	Pr > ChiSq
截距	3.080 6	0.191 4	<0.000 1
夜晚	0.214 6	0.035 8	<0.000 1
报警人类型	-0.086 2	0.039 7	0.029 9
追尾事件	-0.527 9	0.181 9	0.003 7
翻车事件	-0.484 0	0.184 5	0.008 7
抛锚事件	-0.800 0	0.181 8	<0.000 1
碰撞固定物事件	-0.080 3	0.184 4	0.039 1
着火事件(参照事件)	0	—	—
占用车道数	0.106 1	0.038 8	0.001 7
涉及车辆数	0.068 8	0.029 3	0.019 0
人员死亡	0.267 0	0.113 1	0.018 2
救护车	0.224 6	0.042 5	<0.000 1
牵引车	0.233 9	0.040 7	<0.000 1
吊车	0.332 1	0.057 0	<0.000 1
驳载车	0.426 3	0.066 7	<0.000 1
涉及货车	0.252 5	0.042 8	<0.000 1
P	1.809 4	0.040 3	—
$\theta$	0.131 0	—	—
似然函数对数值	-1 216.495 4	—	—
事件数量	106 2	—	—

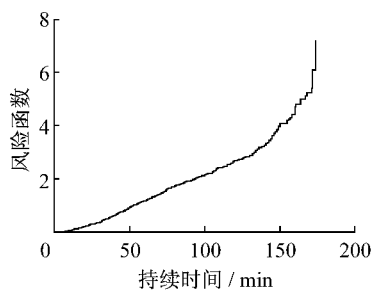


图 3 Weibull AFT 持续时间模型风险函数图

Fig.3 Hazard function plot of Weibull AFT model for incident duration

(3) 因事件类型是分类变量,它有 5 个水平,当拟合参数模型时,自动将最后一个水平(着火事件)作为参照分类对比,追尾事件、翻车事件、抛锚事件、碰撞固定物事件相比着火事件均具有显著性差异(显著性水平取  $\alpha=0.05$ ),并且这四种事件的持续时间都比着火事件短(其标志为回归系数为负值)。

(4) 变量报警人类型的回归系数为负值,表明该因素相对危险度在降低,即报警人为交通管理者(变量赋值为 1)时,相对于驾驶员报警(变量赋值为 0),减少了事件持续时间,这是因为交通管理者可直

接开展现场救援工作,略去了事件发现和响应阶段;其它变量的回归系数为正数,表明因素相对危险度增加,即该因素变量赋值为 1 时,事件严重程度或者施救难度增大,相比于变量赋值为 0,延长了事件持续时间。

为了验证模型的有效性,采用用于验证的 265 个交通事件数据进行检验,结果见图 4。从图 4 中可以看出:持续时间小于 10 min 的抛锚事预测精度较差,对抛锚事件整体预测精度有较大影响,因为持续时间小于 10 min 的抛锚事件一般是一些小的故障造成的,其持续时间受驾驶员维修能力等随机性因素影响较大;在只考虑持续时间大于 10 min 抛锚事件时,其持续时间预测误差小于 10 min 的准确率为 70%;而在交通事故中,持续时间小于 10 min 的事故比例较低,其预测误差对整体预测精度影响不大,可忽略不计,所以交通事故预测持续时间预测误差小于 30 min 的准确率为 78%。由于本文中交通事件发生在城市间高速公路,距离医疗急救点、医院、消防部门等较远,救援车辆施救距离较长,因此预测准确率相对于事件平均时间(交通事故平均持续时间为 56.9 min,抛锚事件平均持续时间为 31.1 min)来说是可以接受的。同时,这一预测精度比贝叶斯方法<sup>[12]</sup>和决策树方法<sup>[13]</sup>的预测精度有所提高。

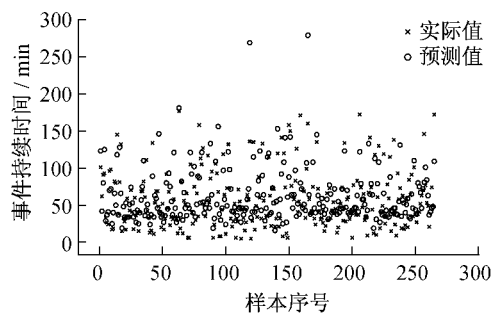


图 4 事件持续时间的实测值与预测值

Fig.4 Real value and prediction value of incident duration time

## 4 结语

对沪杭高速公路嘉兴段 1 327 个交通事件数据进行了分析,其中 1 062 个数据用于建立基于风险分析的 Weibull AFT 模型,265 个数据用于检验模型的预测精度。夜晚、报警人类型、事件类型、占用车道数、涉及车辆数、人员死亡、救护车、牵引车、吊车、驳载车、涉及货车等 11 个变量是影响交通时间持续时间的显著性因素。预测结果表明,持续时间小于 10 min 的抛锚事件的预测精度较差,其预测精度受驾

驶员维修能力等随机性因素影响较大;当只考虑持续时间大于10 min抛锚事件时,其持续时间预测误差小于10 min的准确率为70%;而交通事故持续时间预测误差小于30 min的准确率为78%。预测准确率相对于事件平均时间来说,是可以接受的。但由于事件处理人员的业务水平存在较大的差异,仍可能导致一些事件的预测值和实际值之间存在较大的误差。

不同车道数的高速公路上发生交通事件时,由于车辆占用车道数量不同,导致高速公路路段剩余通行能力是不同的,这可能会影响模型的引入变量数量和预测精度的差异。本文所用的交通事件数据和建立的持续时间预测模型来自于双向8车道高速公路,由于浙江省嘉兴市辖区内的双向6车道高速公路上的交通事件数据由于数量较少,并未纳入到本文模型的构建和验证之中。虽然每条高速公路的环境条件以及信息来源不同,但是在方法论上对双向4车道、双向6车道高速公路上的事件持续时间预测研究仍具有普遍适用性。因此,双向4车道、双向6车道高速公路的风险分析模型将作为本研究进一步深化的方向,可与双向8车道风险分析模型进行比较研究。相比于其它的研究方法,基于风险分析的预测方法不仅可对具有完整持续时间的事件进行预测,还可处理因事件结束时间删失导致的持续时间残缺值数据,扩大了可研究的交通事件数据的范围,在定量分析观察指标的作用强度和方向的基础上,提高了预测的精度。同时,交通管理者只有在充分了解交通事件的性质、严重程度和道路状况,准确地预估交通事件持续时间的基础上,才能提出行之有效的处置方案,如人力、物力和技术资源的合理使用,以及交通控制策略,在最短的时间范围内完成事件管理的各项活动,减小事件的影响,使受到干扰的交通流恢复正常。

## 参考文献:

- [1] Giuliano Genevieve. Incident characteristics, frequency, and duration on a high volume urban freeway [J]. Transportation Research Board, National Research Council: Transportation Research Part A, 1989, 23(5): 387.
- [2] Stathopoulos A, Karlaftis M G. Modeling duration of urban traffic congestion [J]. Journal of Transportation Engineering, 2002, 128(6): 587.
- [3] Doohee Nam, Fred Mannering. An exploratory hazard-based analysis of highway incident duration [J]. Transp Res, 2000, 34A (2): 85.
- [4] Zografos K G, Konstantinos N A, George M V. A real-time decision support system for roadway network incident response logistics [J]. Transportation Research. Part C. 2002, 10 (1): 1.
- [5] CHUNG Younshik. Development of an accident duration prediction model on the Korean Freeway Systems [J]. Accident Analysis and Prevention. 2010, 42: 282.
- [6] Garib A, Radwan A E, Al-deek H. Estimating magnitude and duration of incident delays [J]. Journal of Transportation Engineering. 1997, 123(6): 459.
- [7] Edward C S. New model for predicting freeway incidents and incident delays [J]. Journal of Transportation Engineering. 1997, 123(4): 267.
- [8] 刘伟铭,管丽萍,尹湘源. 基于多元回归分析的事件持续时间预测[J]. 公路交通科技, 2005, 22(11): 126.  
Liu weiming, Guan liping, Yin xiangyuan. Prediction of Incident Duration Based on Multiple Regression Analysis [J]. Journal of Highway and Transportation Research and Development, 2005, 22(11): 126.
- [9] Ozbay K, Narayanan A. Jonnalagadda S. Wide-area incident management support system (WAIMSS) [C]//3rd Annual World Congress on ITS. Orlando: IT'S America. 1996: 14 - 18.
- [10] Ozbay K, Kachroo P. Incident management in intelligent transportation systems [M]. London: Artech House. 1999.
- [11] 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, 2008, 36(3): 319.  
JIYANG Beibei, ZHANG Xiaoning, SUN Lijun. Traffic incident duration prediction grounded on bayesian decision method-based tree algorithm. [J] Journal of Tongji University: Natural Science, 2008, 36(3): 319.
- [12] Boyles S, Fajardo S T, Waller. A naïve Bayesian classifier for incident duration prediction [C/OL]//[S. l.] Transportation Research Board Annual Meeting, 2007 [2006 - 7 - 31]. [http://www.uwyo.edu/sboyles/research/boyles\\_fajardo\\_waller\\_nbc.pdf](http://www.uwyo.edu/sboyles/research/boyles_fajardo_waller_nbc.pdf).
- [13] 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 [J]. Shanghai: College of Transportation Engineering of Tongji University, 2008.