

# 快速路交通事件持续时间预测模型

杨超,汪超

(同济大学 道路与交通工程教育部重点实验室,上海 201804)

**摘要:** 针对城市快速路交通事件持续时间影响因素的复杂性和不确定性,结合贝叶斯网络和非参数回归方法,提出了一种新的快速路交通事件持续时间预测模型.采用上海市快速路监控中心数据,经过降噪处理,生成样本数据;在分析样本数据特征基础上,确定了贝叶斯网络的结构学习方法与参数学习方法;对贝叶斯网络模型的结果用非参数回归算法生成持续时间预测值.最后,对模型预测精度进行了验证,发现模型预测效果较好.

**关键词:** 快速路交通;事件持续时间;贝叶斯网络;非参数回归;预测模型

**中图分类号:** U491

**文献标志码:** A

## Traffic Incident Duration Forecast Model of Expressway

YANG Chao, WANG Chao

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract:** According to the complexity and uncertainty of impact fact of traffic event duration on urban expressway, a new forecasting model using Bayesian Network and non-parametric regression for traffic incident duration was proposed. A sample database, provided by Shanghai Expressway Monitoring Center, was generated by noise reduction. The algorithms of structure learning and parameter learning were determined based on data characteristics, and the forecast results with non-parametric regression were obtained. Finally, the forecasting model was tested with new data and the results verified the accuracy of the model.

**Key words:** expressway; traffic incident duration; Bayesian Network; non-parametric regression; forecast model

交通事件是指任何能够导致道路通行能力下降或交通需求不正常上升的各种事件.一般可分为可

预测性事件与突发性事件两类.美国道路通行能力手册<sup>[1]</sup>对交通事件的特征描述为:引起道路通行能力暂时下降与道路服务水平的降低.有关研究表明<sup>[2]</sup>,由于交通事件引起的突发性拥堵占城市快速路总拥堵的 50%~75%.拥堵意味着出行成本的增加以及路网效率的降低,交通管理者必须充分了解交通事件的属性,才能有效地管理交通事件.因此研究突发性交通事件持续时间对出行者和管理者具有现实意义.交通事件持续时间一般包括 4 个阶段<sup>[3]</sup>:事件发现、事件响应、事件清除和影响恢复.

本文采用的数据来源于上海市城市快速路监控中心所记录的事件.事件的起止时间分别是工作人员通过视频监控发现事件的时间和事件车辆撤离的时间.因此本文计算的交通事件持续时间包括事件响应与事件清除阶段的时间,不包括事件发现和影响恢复阶段.而事件发现阶段的时间与事件管理系统的设置有关,从上海市城市快速路监控中心对事件的检测方式来看,对某路段进行视频扫描的平均间隔为 2 min.因此,可以大致认为事件发现阶段的平均时间在 1 min 左右.影响恢复阶段的时间与事件的前 3 个阶段时间和事发时的交通状态等都有关系,可以通过相应的交通流模型计算得到.因此,本文将研究的重点放在事件响应和事件清除阶段时间的预测上.目前,预测交通事件持续时间的常用方法有:基于统计的预测方法、多元回归模型、时间序列模型、决策树模型、非参数回归模型、模糊逻辑模型、贝叶斯决策树模型以及基于风险分析的方法<sup>[4-6]</sup>.由于交通事件持续时间与影响因素之间存在着相关性、随机性和不确定性,而近年来贝叶斯网络已经成为不确定性领域的主要研究方法.

本文首先就上海市快速路交通事件记录进行统计分析,对分布进行拟合,通过专家知识与数据训练,完成贝叶斯网络构建与参数学习.在此基础上,

利用非参数回归模型,建立了快速路交通事件持续时间预测模型,并对模型精度进行了验证。

## 1 数据描述

交通事件数据来源于上海市城市快速路监控中心从2010年4月至2011年3月为期1年的32 470条数据。数据是经由监控中心利用视频人工观测的方式获取。监控中心的工作人员通过城市快速路布设的无盲点交通监控摄像装置和快速交通报警电话服务发现交通事件,并通知管理员,由其在系统中记录下事件的起止时间、事件类型、天气状况、影响车道数等信息。

为尽量提高预测精度,减少样本中人为因素,剔除重复录入记录,对数据进行噪声识别和降噪处理,消除异常数据,最终得到完整数据30 638组。使用2010年4月至2011年2月共11个月28 381组(占总样本的92.6%)数据用于标定持续时间预测模型。将2011年3月共2 257组(占总样本的7.4%)数据用于模型精度的检验。交通事件样本总体描述如表1所示。

表1 交通事件持续时间样本总体描述

Tab.1 Analysis of traffic incident duration samples						
样本个数/个	均值/s	标准差/s	偏度		峰度	
			偏度系数	标准误差	峰度系数	标准误差
30 638	446.891	383.687	1.174	0.014	0.656	0.028

交通事件持续时间属于连续型变量,将样本数据等距分组后,形成频率直方图,并对其分布函数进行拟合,其中 $R$ -方为0.9301,Kolmogorov-Smirnov检验中 $P$ 值为0.281,经检验样本数据服从威布尔分布,如图1所示。

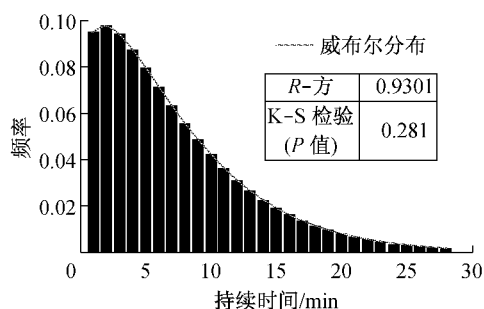


图1 交通事件持续时间频率直方图与拟合曲线

Fig.1 Frequency histogram and fitting curve of duration

## 2 贝叶斯网络和非参数回归结合的预测模型

贝叶斯网络(Bayesian networks, BN)又称信度网<sup>[7]</sup>,是一种对概率关系的有向图解描述。一方面用图论语意清晰的优点来揭示问题结构,另一方面用概率论的原则把复杂的联合概率分布分解成一系列简单模块,降低了知识获取难度和概率推理的复杂度。因此贝叶斯网广泛应用于复杂系统的不确定性推理。

### 2.1 贝叶斯网络节点确定及连续变量离散化

根据经验和数据情况,初步确定天气状况、发生日期、发生时刻、发生时交通流状态、事件类型、影响车道数等6个贝叶斯网络观测节点和持续时间1个预测节点。6个观测节点变量为离散变量,但预测节点(事件持续时间)为连续变量且不服从正态分布,如果贝叶斯网络模型直接用连续变量,计算复杂度高,除正态分布等极少数概率密度外的一般参数形式无法被采用,进而难以观察到连续变量间的关系<sup>[8]</sup>,故先对连续变量进行离散化处理。目前常用的离散方法有等距法、等频率法和层次法。本文采用等频率法将持续时间分隔成6个区间。节点描述与取值如表2所示。

表2 节点描述与值域

Tab.2 Nodes description and range	
节点描述	值域(代码)
事件发生时的天气	恶劣天气(1),晴(2)
事件发生所属的日期	节假日(1),工作日(2)
事件发生的时刻	平峰(1),高峰(2)
事件发生时交通流状态	拥挤(1),畅通(2),堵塞(3)
交通事件的类型	单车抛锚(1),两车碰撞(2),多车碰撞(3),其他(4)
事件影响的车道数	0(1),1(2),(2,6)(3)
事件持续时间/s	(3,103](1),(103,197](2), (197,332](3),(332,503](4), (503,826](5),(826,1 676](6)

### 2.2 贝叶斯网络结构学习及参数学习

贝叶斯网络模型分为两个部分,结构学习和参数学习。结构学习也叫网络构造,根据构造主体不同分为两种,一种是利用专家知识进行手工构造,另一种是利用计算机从数据集中学习。当数据量较大时,仅仅依赖于专家知识构造网络是费时费力甚至是不可能的,而完全利用计算机学习则效率不高<sup>[7]</sup>,本文采用以专家知识和机器学习相结合的方法构造贝叶

斯网络. 结构学习亦分两个部分: 一是模型选择, 即选取何种准则来评判不同模型结构的优劣, 常用的有 BIC(bayesian information criterion)评分准则和贝叶斯评分准则<sup>[7]</sup>; 二是模型优化, 即寻求最优的模型结构. 根据观察视角不同, 可将模型优化的方法分成两类, 基于约束满足的方法和基于评分搜索的方法. 在完备数据条件下, 基于评分搜索的方法常用的有 K2 算法、贪婪搜索算法和爬山算法<sup>[8]</sup>等. 参数学习是在完成或给定网络拓扑结构的情况下, 从观察的样本数据集学习概率参数, 利用先验知识, 确定贝叶斯网络模型各节点处的条件概率密度. 数据完备时学习参数的方法主要有两种: 最大似然估计(ML)和贝叶斯方法<sup>[9-10]</sup>.

本文结合专家知识, 采用以 BIC 评分准则的 K2 算法进行结构学习, 在完成贝叶斯网结构学习之后, 采用最大似然估计进行参数学习. 利用 BNT 工具箱实现结构学习和参数学习. BNT 工具箱是 Murphy 基于 Matlab 语言开发的关于贝叶斯网软件包<sup>[11]</sup>, 分别得到了贝叶斯网结构(如图2所示), 并同时完成

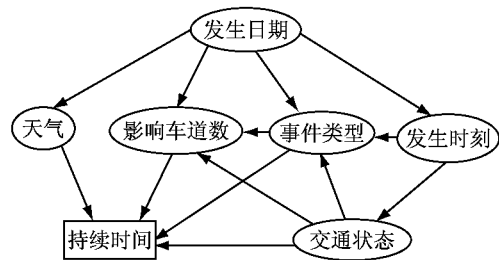


图2 贝叶斯网结构

Fig.2 Bayesian Network structure

了参数学习.

为验证贝叶斯网络模型参数学习的精度, 本文采取贝叶斯网络后验概率与实际计算结果进行对比分析, 表3列出了各个节点共同作用下事件持续时间概率分布与实际计算概率结果的误差. 通过对比分析发现, 由贝叶斯网络模型得到的后验概率的最大绝对误差为 10.32%, 因此贝叶斯网络模型具有较高的预测精度, 用贝叶斯网络进行数据处理与结果预测是可行的.

表3 部分事件持续时间的参数学习结果误差分析

Tab.3 Error analysis of Bayesian Network parameter learning(partial)

事件类别	事件属性代码						结果误差/%					
	天气	发生日期	发生时刻	交通状态	事件类型	影响车道	事件持续时间代码					
							1	2	3	4	5	6
1	1	1	1	1	1	1	3.48	-9.65	6.86	4.89	-2.57	-3.01
2	1	1	1	1	2	1	0.10	3.77	-3.82	5.89	1.40	-7.34
3	1	1	1	1	3	1	5.62	-7.45	-3.11	-0.01	-3.76	8.71
4	1	1	1	1	2	2	9.65	0.87	0.76	2.58	-5.48	-8.38
5	1	1	1	2	1	1	0.48	-2.19	0.21	8.78	0.05	-7.33
6	1	1	1	2	2	1	2.42	-2.09	-10.04	1.37	-1.98	10.32

注: 持续时间预测误差为实际概率与后验概率之差.

### 2.3 非参数回归模型

非参数回归<sup>[12]</sup>是一种适合不确定性、复杂性和非线性的动态系统的非参数建模方法, 它的一个优点是不需要对因变量的分布形式以及因变量与自变量之间的关系做出严格的假设, 只需在足够的历史数据中寻找与当前点相似“近邻”, 并用匹配的“近邻”预测当前点的值.

用非参数回归的方法对贝叶斯网络模型条件概率的结果进行改进, 即在事件持续时间 6 个区间内, 分别建立非参数回归模型, 将离散变量换算为连续变量, 预测模型流程示意图如图3所示.

非参数回归模型参数主要包含 4 部分内容, 即变量的定义与距离度量, 近邻值  $K$  的确定以及预测算法的构建. 为保持模型的一致性, 沿用贝叶斯网络模型中变量定义.

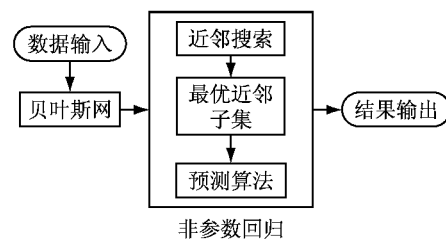


图3 预测模型流程示意图

Fig.3 Forecast model flow chart

距离度量是指当前事件与历史数据库中事件相匹配的程度. 事件所对应的变量是分类且离散的, 因此采用普通欧几里德距离来衡量匹配度是不合适的. 本文采用了基于两个事件各变量匹配的度量方法, 即当前事件所对应的每个变量与历史数据事件对应的变量相比较, 给定每个变量不同的权重系数, 它是该变量中两种状态所对应的交通事件持续时间

均值之差的绝对值. 以第6区间(即826~1 676 s区间内)为例,各变量权重系数如表4所示.

表4 第6区间非参数回归变量权重系数表  
Tab.4 Weight coefficients of the 6th interval

变量	符号	权重系数 $W_x$	值域(代码)
天气	A	31.38	恶劣天气(1),晴(2)
发生日期	B	13.25	节假日(1),工作日(2)
发生时刻	C	2.95	平峰(1),高峰(2)
交通状态	D	$W_{D12}=18.61, W_{D13}=34.34, W_{D23}=15.73$	畅通(1),拥挤(2),堵塞(3)
事件类型	E	$W_{E12}=62.37, W_{E23}=49.85, W_{E13}=12.52, W_{E24}=61.44, W_{E14}=59.08, W_{E34}=71.59$	单车抛锚(1),两车相撞(2),多车相撞(3),其他(4)
影响车道数	F	$W_{F12}=27.56, W_{F13}=49.34, W_{F23}=21.78$	0(1),1(2),2-6(3)

在计算距离时,如果变量状态一致,则不产生距离,反之则在距离上增加相应权重. 定义距离公式如下:

$$d = W_A |A_1 - A_2| + W_B |B_1 - B_2| + W_C |C_1 - C_2| + W_D^* + W_E^* + W_F^* + m \quad (1)$$

式中: $d$ 为距离; $m$ 为给定的一个基准值,为常量,取 $m=1$ ;  $W_A, W_B, W_C$ 分别表示天气、事件发生日期、发生时刻3个变量的权重系数,  $W_D^*, W_E^*, W_F^*$ 分别表示交通状态、事件类型、影响车道数的权重系数集合.

近邻值  $K$  的确定:为选取最合适的近邻值  $K$ , 以平均绝对预测误差(MAPE)作为度量指标,计算公式如下:

$$M = \frac{\sum_{i=1}^K |P_i - T_i|}{K} \quad (2)$$

式中: $M$ 为平均绝对预测误差; $K$ 是近邻数; $P_i$ 是持续时间预测值; $T_i$ 是近邻子集  $i$  中的各交通事件的持续时间.

对第6区间内的样本进行试验发现最优近邻值为  $K_6=17$ (如图4所示),其原因是,当17超过后,会将相关性较小的交通事件样本持续事件划入待测的近邻中,从而影响预测模型的精度.

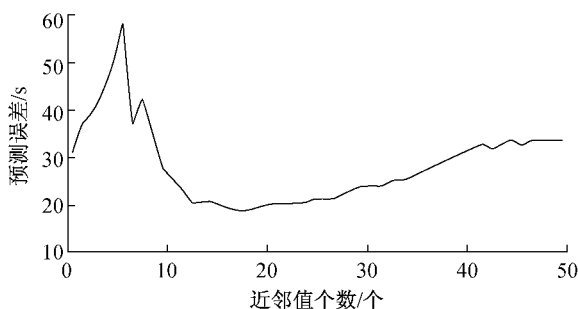


图4 不同近邻值  $K$  与预测误差关系曲线

Fig.4  $K$ -nearest neighbor and error curves

预测算法是从历史数据库选取的近邻子集和权重系数进行计算. 权重计算表达式以及预测持续时

间如下式:

$$W_i = \frac{1}{\sum_{i=1}^K \frac{1}{d_i}}, P_i = \sum_{i=1}^K W_i T_i \quad (3)$$

式中: $K$ 是近邻数; $W_i$ 是权重系数; $d_i$ 是匹配距离; $P_i$ 是持续时间预测值; $T_i$ 是近邻子集  $i$  中的各交通事件的持续时间.

式(3)含义为匹配距离越小的点,也就是越相似的点给予的权重越大.

## 2.4 算例分析

如果某个交通事件具有如下属性{1,1,1,1,2,2}(代码顺序与表2节点描述顺序一致),即该事件为在恶劣天气节假日平峰时段,在畅通状态发生两车相撞事件,影响1个车道,则该事件的贝叶斯网络后验概率分布为(0.113 0, 0.240 1, 0.403 0, 0.125 2, 0.070 6, 0.047 1). 再将该事件分别在对应的6个区间内进行匹配,结果为(44, 148, 223, 364, 542, 857),则预测持续时间为: $44 \times 0.113 0 + 148 \times 0.240 1 + 223 \times 0.403 0 + 364 \times 0.125 2 + 542 \times 0.070 6 + 857 \times 0.047 1 = 255$  s.

## 3 模型检验与分析

用2011年3月的2 257条交通事件记录来检验模型,预测模型结果如图5所示. 分别用绝对预测误差小于等于120, 180, 240, 300, 360, 420, 480, 500 s的事件个数占检验样本总数的百分比来表示.

模型预测结果表明随着预测误差范围的增大,预测精度有所上升. 当绝对预测误差 $\leq 120$  s时,预测精度为85.7%;当绝对预测误差 $\leq 300$  s时,预测精度为92.6%;当绝对预测误差 $\leq 500$  s时,预测精度为97.1%,模型精度基本符合实际预测要求. 同时发现在预测误差范围等距扩大时,精度的增量逐步缩小.

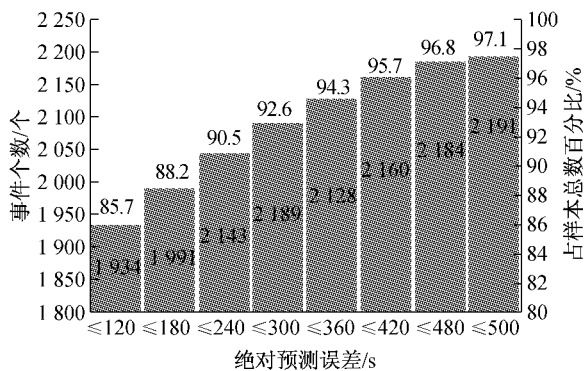


图5 预测模型检验结果

Fig.5 Test results of forecast model

模型发现事件类型、影响车道数、交通状态对事件持续时间影响较大。其中影响车道数越少,持续时间越短。对于交通状态而言,通畅状态的交通事件相对于拥挤与堵塞,事持续时间更短。因此管理部门应及时发布交通诱导信息,尽力减少影响车道数,以减少事件持续时间。

## 4 结语

本文就上海市快速路监控中心采集的数据进行分析,利用 K2 算法和专家知识结合的方法确定贝叶斯网络结构,最大似然估计进行参数学习,并在贝叶斯网络输出结果基础上,结合非参数回归的近邻搜索、最优邻域生成了交通事件预测模型,并对其进行了检验,结果基本符合实际预测要求。随着计算机技术的发展和各种算法的改进,交通事件持续时间预测模型将在信息发布、交通诱导等方面得到更广泛的应用。快速路交通事件持续时间预测模型有助于管理部门建立应急预案,协助事件管理者迅速做出决策,发布诱导信息以降低交通事件导致拥堵的恢复时间。由于持续时间影响因素的复杂性,今后需要继续探究更多影响因素,以便模型能够提高预测精度。

## 参考文献:

[1] Transportation Research Board. Highway capacity manual

- [M]. Washington D C: National Research Council, 2000.
- [2] Giuliano G. Incident characteristics, frequency and duration on a high volume urban freeway [J]. Transportation Research Part A, 1989, 23(5):387.
- [3] Nam D, Mannering F. An exploratory hazard-based analysis of highway incident duration [J]. Transportation Research Part A, 2000, 34(2): 85.
- [4] 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: College of Transportation Engineering of Tongji University, 2008.
- [5] 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.
- 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.
- [6] 康国祥, 方守恩. 基于风险分析的交通事件持续时间预测[J]. 同济大学学报: 自然科学版, 2012, 40(2):241.
- KANG Guoxiang, FANG Shouen. A hazard-based analysis of traffic incident duration prediction [J]. Journal of Tongji University: Natural Science, 2012, 40(2):241.
- [7] 张连文, 郭海鹏. 贝叶斯网引论[M]. 北京: 科学出版社, 2006.
- ZHANG Lianwen, GUO Haipeng. Introduction to Bayesian Networks [M]. Beijing: Science Press 2006.
- [8] 王飞, 刘大有, 薛万欣. 基于遗传算法的 Bayesian 网中连续变量离散化的研究 [J]. 计算机科学, 2002, 25(8):794.
- WANG Fei, LIU Dayou, XUE Wanxin. Discretizing continuous variables of Bayesian Networks based on genetic algorithms [J]. Chinese Journal of Computers, 2002, 25(8):794.
- [9] 宫秀军. 贝叶斯学习理论及其应用研究[D]. 北京: 中国科学院研究生院, 2002.
- GONG Xiujun. Research on Bayesian learning theory and its application [D]. Beijing: Institute of Computing Technology of Chinese Academy of Sciences. 2002.
- [10] 贾海洋. 贝叶斯学习若干问题研究[D]. 长春: 吉林大学交通学院, 2008.
- JIA Haiyang. Research on some problems in learning Bayesian Network [D]. Changchun: College of Transportation Engineering of Jilin University. 2008.
- [11] Murphy P K. The Bayes net toolbox for Matlab [J]. Computing Science and Statistics, 2001, 33: 1024.
- [12] 沃塞曼. 现代非参数统计[M]. 北京: 科学出版社, 2008.
- Wasserman. All of nonparametric statistics [M]. Beijing: Science Press, 2008.