

基于车道集计交通流数据的事故风险评估分析

杨奎, 余荣杰, 王雪松

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

摘要: 事故风险评估分析提取事故及对应非事故状况下的交通运行数据作为自变量, 以事故发生与否作为 0 或 1 因变量, 采用数理统计模型分析事故发生与交通流状态的关联性. 为探索上海市城市快速路的车道集计交通流数据用于事故风险评估分析的可行性, 基于条件 Logistic 回归, 使用该数据与事故数据构建事故风险评估模型, 并检验模型预测精度. 结果表明低运行速度、车道间流量的差异性会增加事故发生的概率; 上海市城市快速路的车道集计交通流数据可用于事故风险评估研究.

关键词: 城市快速路; 事故风险评估; 条件 Logistic 回归; 车道集计交通流; 可行性

中图分类号: U491

文献标志码: A

Application of Aggregated Lane Traffic Data from Dual-loop Detector to Crash Risk Evaluation

YANG Kui, YU Rongjie, WANG Xuesong

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

Abstract: In crash risk evaluation study, the relevance between crash occurrence and traffic conditions was analyzed through statistical model. In the model, traffic conditions prior to each crash and corresponding non-crashes were extracted as independent variables, while 0 or 1 was taken as dependent variable for crashes or non-crashes. For the purpose of searching the validity of aggregated lane traffic data from dual-loop detectors of Shanghai Urban Expressway System applied to crash risk evaluation, aggregated lane traffic data and crash data were used to develop crash risk evaluation model by condition logistic regression, and the predictive performance of the model was tested. The results show that lower speed and higher volume variation crossing lanes will increase the likelihood of crash occurrence; and the aggregated lane traffic data from dual-loop detectors of Shanghai Urban Expressway System can be applied to crash

risk evaluation study.

Key words: urban expressway; crash risk estimation; condition Logistic regression; aggregated lane traffic data; feasibility

快速路在城市道路系统中主要承担中长距离的快速出行需求, 具有大容量、高限速的特点, 是城市道路交通系统的主动脉. 城市快速路交通事故多发、对交通运行影响大; 2011 年上海市中心城快速路发生交通事故 19 164 起, 日均达 52.5 起; 由交通事故造成的偶发性拥堵占城市快速路拥堵的 50%~75%^[1]. 探索城市快速路事故的发生机理, 改善交通安全状况与交通运行效率已成为交通管理部门所面临的重大课题.

事故风险评估分析主要针对事故发生前的交通流紊乱状况, 采用数理统计模型对比分析事故、非事故状态下的交通流状况, 开展事故发生与交通流状态的关联性分析, 以发现影响事故风险的交通流因素, 并实现事故风险的预测^[2-4]. Oh 等^[2]通过整合先进的交通管理和信息系统(ATMIS), 利用基于非参数贝叶斯模型的统计分析, 用实时交通状况量化事故风险, 以此证明了判别导致事故的实时交通状况具有潜在能力.

建立事故风险评估模型的方法主要有病例对照 Logistic 回归^[3-4]、固定和随机参数的 Logistic 模型^[5]、贝叶斯随机常数的 Logistic 模型^[6]、二元概率模型(Binary Probability, BP)^[7]、概率神经网络(Probabilistic Neural Network, PNN)^[8]、基于多层感知器(MLP)和归一化径向基函数(NRBF)的混合神经网络模型^[9]、贝叶斯网络(Bayesian Belief Networks, BBN)^[10]、遗传规划模型(Genetic Programming, GP)^[11]、支持向量机(Support

收稿日期: 2015-09-15

基金项目: 国家自然科学基金(71401127, 51522810); 上海市科学技术委员会科研计划(15DZ1204800)

第一作者: 杨奎(1992—), 男, 博士生, 主要研究方向为交通安全、数据挖掘. E-mail: 1410725@tongji.edu.cn

Vector Machine, SVM)^[12]等,并可分为智能算法与传统算法两大类.神经网络模型、支持向量机等智能算法得到的模型精度相对较高,但犹如黑盒子,不能直接识别事故风险与交通流的关系^[11];而病例对照 Logistic 回归等传统模型能克服该缺陷,能更好地解释事故的发生机理.因此该研究使用病例对照 Logistic 回归作为研究方法.

病例对照研究(Case-control Study)是经典的流行病学研究方法,常用于探索某疾病的危险因素^[13],以患有某疾病的病人作为病例,以不患该病但具有可比性的个体作为对照,测量其既往暴露于某个(或某些)危险因子的情况和程度,以判断危险因子与该病有无关联及关联程度大小.事故风险评估分析中常采用病例对照研究方法提取交通流数据,其中“病例”为单起交通事故,“对照”为没有发生交通事故时对应路段的“非事故”^[3-4].

感应线圈检测器是事故风险评估研究中交通流数据的主要来源,既有研究的线圈检测器数据记录时间间隔主要有 10 s^[2]和 30 s^[3-4,6,9,11],是高频率采集的微观交通流;而本文使用 5 min 车道集计流量、速度数据,是低频率集计的车道交通流.

已有研究^[3-4,6,9,11]发现流量、速度、流量标准差、速度标准差是影响交通事故风险的重要因素:流量是平均每车道的流量;速度、流量标准差、速度标准差分别是基于高频率采集的微观交通流,集计时间片段内车道间的平均速度、流量标准差、速度标准差(集计的目的在于消除由于交通流数据采集间隔较短而易出现的数据噪声^[3,9,14]).其平均速度变量反映的是集计时间片段内时间与空间上的整体平均水平;类似于速度标准差变量,其流量标准差变量是基于高频率采集的微观交通流,将流量在时间上的波动性与车道间的离散性汇集起来,整体反映道路断面上流量在时空上的波动程度.交通流数据的差异性会引起分析变量含义的不同,本文流量变量是断面总流量,更能反映具有不同车道数的事故地点附近的实际流量水平,和事故与非事故对应交通量间的差异,更能充分反映交通流量对事故风险的影响;平均速度变量是车道间平均速度,能单独反映空间上的平均水平;流量标准差变量反映车道间流量离散性,速度标准差变量反映车道间速度离散性,能单独反映空间上的波动水平.

为分析车道集计交通流数据开展事故风险评估分析的可行性,以上海市城市快速路为研究对象,基

于病例对照研究方法提取事故与非事故条件下的交通流数据,并将样本随机拆分为训练数据与测试数据.训练数据用于构建事故风险评估模型,测试数据则用于检验模型的预测精度.

1 数据准备

研究使用的数据包括上海市城市快速路道路线形数据、事故数据(2009 年 1 月和 3 月)和线圈检测交通数据(2009 年 1 月和 3 月的 5 min 集计的车道流量、平均车速).

1.1 事故数据与线圈数据的匹配

为分析交通流状态对事故风险的影响,建立事故风险评估模型,需获取事故发生前事故地点上、下游的交通流数据.由既有研究^[2-4,6,8-9,11]分析可知,与事故风险显著相关的交通流参数属于事故上、下游两个线圈检测器断面之内,因此本文选取事故地点上、下游各两个线圈检测器断面,通过匹配事故地点上、下游各两个线圈检测器断面的线圈编码提取交通流数据.

上海市城市快速路事故以路侧桩号定位(桩号是施工单位对设计基础桩进行统一的编号,号码不重复,分布于快速路两侧),基于线圈位置的桩号信息对每起事故地点上下游线圈检测器进行匹配.线圈断面从上游至下游依次命名为 U2, U1, D1, D2(见图 1).

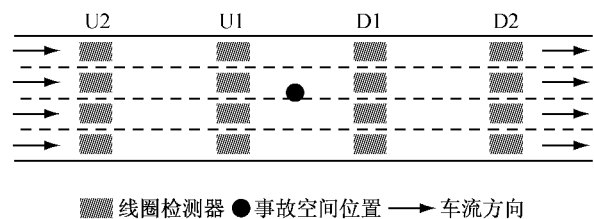


图 1 事故发生的空间位置

Fig.1 Location of crash

1.2 事故与非事故条件下线圈检测交通数据的提取

采用病例对照设计提取交通流数据,病例与对照的比例采用了较常用的 1 : 4 比例^[15],对照组数据的选取考虑了 5 个条件:①对照组日期与对应事故所在的日期不同;②与事故发生时间相同;③与事故发生地点相同;④对照组与事故发生对应一周中的同一天;⑤对照组对应事故时间前后各 1 h 内同地点无交通事故发生.

提取事故发生前 0~30 min 上下游最近两个线

圈断面的交通流数据;同时对应于每起事故,基于上述条件提取没有发生事故条件下的交通流数据.以 2009 年 3 月 28 日星期六 8:40 发生在内环高架内侧(桩号为 NN0312)的事例为例,提取 2009 年 3 月 28 日星期六 8:10~8:40 事故发生地点上、下游两个线圈断面的交通参数,作为一个病例的交通状况;提取 2009 年 3 月 21 日(星期六),2009 年 3 月 14 日(星期六),2009 年 3 月 7 日(星期六)(已经检查该地点该时间前后各 1 h 内无事故发生),8:10~8:40 事故发生地点上、下游两个线圈断面的交通参数,作为该病例的对照组交通状况.

将所提取的 30 min 交通流数据划分为 6 个 5 min 片段,依次命名为时间片段 1,2,3,4,5,6,时间片段 1 最靠近事故发生时间,依次越远.分别提取 6 个时间片段 4 个线圈断面的平均车速和总流量,并计算车道间速度和流量的标准差,共计 4 * 4 断面 * 6 个时间片段 = 96 个变量.变量的命名规则如图 2 所示,变量由 3 个英文字母和 2 个数字组成,其中首字母表示平均(A)、总和(T)或者标准差(S),第二个字母表示流量(V)或者平均速度(S).第三个字母与第一个数字组合表示断面编号(U2,U1,D1,D2).最后一个数字表示时间片段,值越小越靠近事故发生时间.例如 ASU12 表示上游最近断面 U1 在第 2 个时间片段(即事故发生前 5~10 min)的平均速度.为避免事故记录时间误差、获得采取措施的一定预警时间^[3-4,14],因此建模时不考虑时间片段 1 的变量.

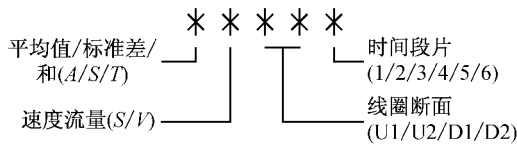


图 2 变量命名规则说明

Fig.2 Structure of variable naming

经匹配,分析样本中有 1 962 个病例对照组,其中包含 1 962 起事故(病例)和 5 925 起非事故(对照).以 7 : 3 的比例在事故与非事故层面(病例对照组层面)将最终样本随机拆分为训练数据与测试数据,前者用于事故风险评估建模,后者用于模型的预测精度检验.

$$P(y = 1 | x_{ij}) = \frac{1}{1 + \sum_{j=0}^m \text{EXP}[\beta_1(x_{ij1} - x_{i01}) + \dots + \beta_k(x_{ijk} - x_{i0k}) + \dots + \beta_n(x_{ijn} - x_{i0n})]} \quad (4)$$

以上是第 i 个配比组的条件概率,对于其他配

2 数学模型

2.1 病例对照 Logistic 回归

事故风险评估模型中,事故发生与否作为二分类因变量,比较常用的分析方法有病例对照 Logistic 回归^[3-4].设 y 为一个二分类因变量,事故发生时赋值 $y=1$,否则 $y=0$.用 p 表示事故发生的概率,则令

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \quad (1)$$

式中, $\text{logit} \in (-\infty, +\infty)$.

设有 n 个自变量 $x_1, x_2, \dots, x_k, \dots, x_n$,用 $\text{logit}(p)$ 与自变量建立线性关系,经过简单变换即可得到 Logistic 模型,如.

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \dots + \beta_n x_n}} \quad (2)$$

式中, β_0 为常数项,表示自变量均为 0 时 $\text{logit}(p)$ 的值. β_k 为偏回归系数,表示当其他自变量不变时, x_k 每变动一个单位时 $\text{logit}(p)$ 的改变量.

病例对照研究包括配比和非配比两种方式,配比设计的 Logistic 回归通常也称为条件 Logistic 回归.相对地,非配比设计的 Logistic 回归称为非条件 Logistic 回归.使用病例对照设计的条件 Logistic 回归^[13],简称为病例对照 Logistic 回归.

设第 i 个配比组内包含 1 个事故和 m 个非事故,在此条件下第一个观察对象 $j=0$ 为事故,其余 ($j=1,2,3,\dots,m$) 为非事故的概率为

$$P(y = 1 | x_{i0}) \prod_{j=1}^m P(y = 0 | x_{ij})$$

因此任何一个观测对象 ($j=0,1,2,\dots,m$) 为事故,其余 ($j'=0,1,2,3,\dots,m, j' \neq j$) 为非事故的概率之和为

$$\sum_{j=1}^m P(y = 1 | x_{i0}) \prod_{\substack{j' \neq j \\ j'=0}}^m P(y = 0 | x_{ij'})$$

根据 Logistic 回归模型式(2),可构造出 $y=1$ (即事故)的条件概率为

$$P(y = 1 | x_{ij}) = \frac{\text{EXP}(\beta_1 x_{i01} + \dots + \beta_k x_{i0k} + \dots + \beta_n x_{i0n})}{\sum_{j=0}^m \text{EXP}(\beta_1 x_{ij1} + \dots + \beta_k x_{ijk} + \dots + \beta_n x_{ijn})} \quad (3)$$

式中, n 为变量个数,简化得

$$1 \quad (4)$$

以上也是第 i 个配比组的条件概率.根据概率的乘法原理,

多个独立事件同时发生的概率为各事件独立发生概率的乘积. 因此可以构造出所有匹配组第一个观测

$$L = \prod_{i=1}^k \frac{1}{1 + \sum_{j=0}^m \text{EXP}[\beta_1(x_{ij1} - x_{i01}) + \dots + \beta_k(x_{ijk} - x_{i0k}) + \dots + \beta_n(x_{ijn} - x_{i0n})]}$$

式中, $x_{ijk} - x_{i0k}$ 为事故与非事故在同一危险因素上的暴露水平之差.

对上述条件自然函数取自然对数后, 采用 Newton-Raphson 迭代法求得参数估计值及其标准误差. 本文采用 SAS9.3 中 PROC PHREG 模块^[13] 建立病例对照 Logistic 回归模型, 估计各危险因素对应的 β_k 值, 使用优势比进行评估预测.

2.2 模型预测精度检验

事故组的优势与非事故组的优势的比值称为优势比(又称比值比, OR), 用于说明某影响因素引起事故发生的危险度大小, 如

$$\text{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \quad (5)$$

将非事故状态下变量的平均值作为“正常交通流”, 将每起观测(事故或者非事故)相对于“正常交通流”的优势比^[3], 如式(6), 作为该观测预测发生事故的危险程度.

$$\text{OR}_{ij} = \text{EXP}[\beta_1(x_{ij1} - \bar{x}_{i01}) + \dots + \beta_k(x_{ijk} - \bar{x}_{i0k}) + \dots + \beta_n(x_{ijn} - \bar{x}_{i0n})] \quad (6)$$

式中, β_k 为变量 x_k 的偏回归系数, $(x_{ij1}, \dots, x_{ijk}, \dots, x_{ijn})$ 为第 i 个配比组中第 j 个观测的 n 个影响变量, $(\bar{x}_{i01}, \dots, \bar{x}_{i0k}, \dots, \bar{x}_{i0n})$ 为第 i 个配比组中非事故对照数据的 n 个影响变量的均值.

选择合理阈值后, 优势比大于阈值时可认为是危险的, 即预测为事故, 优势比小于或等于阈值时则预测为非事故, 得到模型的事故与非事故预测精度.

3 结果与分析

后退法 (Backward Selection) 首先将全部自变量放入模型中, 回归计算变量的显著性 (P 值), 将最不显著的变量 (P 值最大) 剔除, 直至模型中变量的 P 值均不大于事先给定的显著水平临界值. 基于训练数据, 用后退法 (给定的显著水平临界值为 0.15) 筛选得到 19 个显著变量, 再结合变量间的相关性及显著性, 剔除了 16 个变量, 最终挑选出 3 个最显著变量, 使用病例对照 Logistic 回归建立事故风险评估模型, 显著变量统计描述如表 2 所示, ASU12 表示 U1 断面时间片段 2 的平均速度 ($\text{km} \cdot \text{h}^{-1}$), SVD12 表示 D1 断面时间片段 2 的流量标准差

对象为事故, 其余为非事故的概率, 即条件似然函数为

($1\text{PCU} \cdot (5 \text{ min})^{-1}$), TVD22 表示 D2 断面时间片段 2 的流量 ($1\text{PCU} \cdot (5 \text{ min})^{-1}$), 模型参数估计见表 3.

表 2 事故风险评估模型显著变量统计描述

Tab.2 Summary statistics for significant variables for crash risk evaluation model

变量	均值	标准差	最小值	最大值
ASU12	47.3	23.0	0	100.2
SVD12	18.6	16.6	0	157.3
TVD22	301.6	148.7	0	1 090.0

表 3 事故风险评估模型的参数估计

Tab.3 Parameters estimates for crash risk evaluation model

参数	参数估计	标准误差	卡方	$P_r >$ 卡方	危险比
ASU12	-0.034 63	0.002 31	224.331 4	<0.000 1	0.966
SVD12	0.008 57	0.003 56	5.801 9	0.016	1.009
TVD22	-0.001 5	0.000 562 1	7.138 4	0.007 5	0.998
模型参数 AIC	3 499.611				

最终模型有三个显著变量, ASU12 系数为负, 显示事故发生前 5~10 min 事故地点上游速度相对于非事故状况低; 运行速度越小, 事故发生风险越大, 这与已有研究^[3-4,11] 结果一致. SVD12 系数为正, 显示事故发生前 5~10 min 下游最近线圈断面的车道间流量的差异性相对于非事故状况大; 流量的差异性越大, 事故发生风险越大, 其每增加 1 pcu/(5 min)⁻¹, 则事故风险变为原来的 1.009 倍, 这与 Abdel-Aty 等^[4] 的研究结果不一致, 后者发现事故发生前下游道路断面流量在时空上的波动程度越小则事故风险越大. TVD22 的系数为负, 显示事故发生前 5~10 min 下游断面 D2 的流量相对非事故状态低; 该断面流量越低, 事故发生风险越大, 可能由于下游流量越低造成上下游密度差越大, 引起的交通流波动更大, 从而上游的事故风险增加, 这与 Abdel-Aty 等^[4] 的结果一致.

综合考虑模型结果: U1 断面附近可能处于拥挤状态, 低速 (ASU12 系数为负数), 而 D2 断面交通流量相对较低 (TVD22 系数为负数), 当车辆驶出拥挤交通流后, 容易出现车道间车速不一致现象, 引起车道间的流量不平衡性增加 (SVD12 系数为正), 出现紊乱交通流. 当车辆驶出拥挤交通流时, 易出现紊乱交通流, 车道间流量的不平衡增大了侧碰事故发生

的风险,拥挤时反复的启动与急停则增加追尾碰撞事故的风险。

为检验所构建的事故风险评估模型的预测精度,使用测试数据计算优势比,进行检验。模型正确预测事故与非事故的累计频率随优势比的变化图如图 3 所示(图中只表示优势比不大于 5 的部分)。正确预测事故的累计频率随着优势比的增大而降低,而正确预测非事故的累计频率随优势比的增大而增大。选取 1 作为阈值,举例说明模型的预测精度,优势比大于 1 的观测可认为是危险的,即预测为事故,优势比小于或等于 1 即预测为非事故^[3-4]。结合每起观测的实际与预测,得到模型的预测精度如表 4 所示。

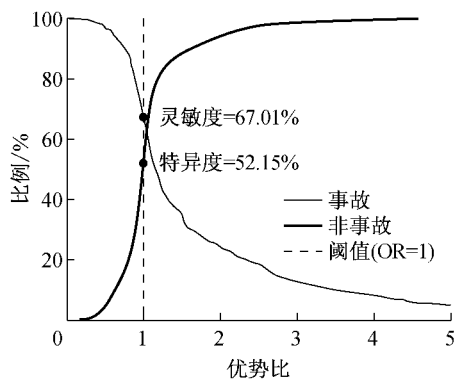


图 3 模型的预测性能

Fig. 3 Predictive performance of model

表 4 模型的预测精度

Tab. 4 Predictive accuracy for model

实际分类	预测分类		合计
	非事故	事故	
非事故	935(特异度:52.15)	858(特异度:47.85%)	1 793(100%)
事故	194(32.99%)	394(灵敏度:67.01%)	588(100%)
合计	1 129(47.42%)	1 252(52.58%)	2 381(100%)

注:括号中为行百分比

当优势比等于 1 作为阈值时,灵敏度为 67.01%,即事故样本中实际事故被预测为事故的比例,表明该模型可正确识别 67.01%的事故;特异度为 52.15%,即非事故样本中实际非事故被预测为非事故的比例,表明该模型可正确识别 52.15%的非事故;误判率为 47.85%,即非事故样本中实际非事故被判断为事故的比例,表明该模型可错误将 47.85%的非事故预测成了事故;准确率为 39.27% + 16.55% = 55.82%,即该模型可正确预测总样本中 55.82%的观测。

早期事故风险评估模型事故预测准确率在 50%~60%之间^[2,4,9];研究人员通过统计分析模型

的改进将事故预测准确率提高至 60%~70%^[3,11](见表 5)。从事务预测精度来看,该研究为 67.01%,与现有研究相近;该研究采用了常用的数据处理和统计分析方法,也可通过改进算法(如考虑事故类型的差异性、空间相关性等)、建模技术(如支持向量机等智能算法)、数据处理方法等手段获得更高精度的模型。因此,基于 5 min 车道集计交通流数据可用于构建事故风险评估模型。通过阈值的变化(如取不同的比值比),获得不同阈值下的灵敏度和误判率,以灵敏度为纵轴,以误判率为横轴,连接各点即为受试者工作特征曲线(简称 ROC 曲线),AUC 值是 ROC 曲线与横轴间的面积(介于 0.5~1 之间)。AUC 值越大,判断价值越高,模型的预测性能越好^[6,11-12]。

表 5 对比部分研究的事故风险评估模型精度

Tab. 5 Comparison of predictive performance of crash risk evaluation models to previous studies

作者	事故预测精度/%	非事故预测精度/%	误判率/%	事故样本	非事故样本
Oh 等 ^[2]	55.80	72.10	27.90	52	4 787
Abdel-Aty 等 ^[3]	69.40	52.80	47.20	375	1 875
Abdel-Aty 等 ^[4]	56.00	80.00	20.00	1 528	6 112
Abdel-Aty 等 ^[8]	73.90	71.30	28.70	377	2 857
Pande 等 ^[9]	57.00	70.00	30.00	162	3 650
Xu 等 ^[11]	61.00	80.00	20.00	807	8 070
Yu 等 ^[12]	AUC=0.75			265	1 017
本文	67.01	52.15	47.85	1 962	5 925

4 结论

(1)使用上海市城市快速路事故数据和线圈检测交通流数据,构建了事故风险评估模型,探索了交通运行状况与事故风险的关系,量化了交通变量对事故风险的影响。研究发现上游平均速度、下游流量及其标准差对事故发生有显著影响;低运行速度、车道间流量的差异性会增加事故发生的概率。

(2)与既有研究的事故风险评估模型预测精度进行对比,验证了基于 5 min 车道集计交通流数据具有构建事故风险评估模型的可行性,可进一步改进算法(如考虑事故类型的差异性、空间相关性等)、建模技术(如支持向量机等智能算法)、数据处理方法等获得更高精度的模型。

本文结论可为交通管理部门针对性交通安全管理提供依据,也可为基于上海市城市快速路事故数据和线圈检测交通流数据研究主动交通管理策略(如可变限速)应用提供基础。该研究仅针对总体事故构建了事故风险评估模型,后续可进一步考虑快

速路在不同时段交通运行的差异性.

参考文献:

- [1] Giuliano G. Incident characteristics, frequency, and duration on a high volume urban freeway[J]. *Transportation Research Part A: General*, 1989, 23(5): 387.
- [2] Oh C, Oh J, Ritchie S, *et al.* Real-time estimation of freeway accident likelihood[C]//TRB 80th Annual Meeting Compendium of Papers DVD. Washington D C: Transportation Research Board of the National Academies, 2001:1-3445.
- [3] Abdel-Aty M, Uddin N, Pande A, *et al.* Predicting freeway crashes from loop detector data by matched case-control logistic regression[J]. *Transportation Research Record*, 2004, 1897(1): 88.
- [4] Abdel-Aty M, Uddin N, Pande A. Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways[J]. *Transportation Research Record*, 2005, 1908(1):51.
- [5] Anastasopoulos P, Mannering F. An empirical assessment of fixed and random parameter logit models using crash-and non-crash-specific injury data[J]. *Accident Analysis & Prevention*, 2011, 43(3): 1140.
- [6] Xu C, Wang W, Liu P. Identifying crash-prone traffic conditions under different weather on freeways[J]. *Journal of Safety Research*, 2013, 46(1): 135.
- [7] Christoforou Z, Cohen S, Karlaftis M. Identifying crash type propensity using real-time traffic data on freeways[J]. *Journal of Safety Research*, 2011, 42(1): 43.
- [8] Abdel-Aty M, Pande A. Identifying crash propensity using specific traffic speed conditions [J]. *Journal of Safety Research*, 2005, 36(1): 97.
- [9] Pande A, Abdel-Aty M. Assessment of freeway traffic parameters leading to lane-change related collisions [J]. *Accident Analysis and Prevention*, 2006, 38(5): 936.
- [10] Hossain M, Muromachi Y. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways [J]. *Accident Analysis and Prevention*, 2012, 45(1): 373.
- [11] Xu C, Wang W, Liu P. A genetic programming model for real-time crash prediction on freeways [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 14(2): 574.
- [12] Yu R, Abdel-Aty M. Utilizing support vector machine in real-time crash risk evaluation [J]. *Accident Analysis and Prevention*, 2013, 51(1):252.
- [13] 冯国双, 刘德平. 医学研究中的 logistic 回归分析及 SAS 实现 [M]. 北京: 北京大学医学出版社, 2012.
FENG Guoshuang, LIU Deping. Logistic regression analysis and SAS realization in medical research [M]. Beijing, Peking University Medical Press, 2012.
- [14] Yu R, Wang X, Yang K, *et al.* Real-time crash risk analysis for Shanghai urban expressways: a Bayesian semi-parametric approach [C] // TRB 94th Annual Meeting Compendium of Papers DVD. Washington D C: Transportation Research Board of the National Academies, 2015:2015-2097.
- [15] Zheng Z, Ahn S, Monsere C. Impact of traffic oscillations on freeway crash occurrences [J]. *Accident Analysis and Prevention*, 2010, 42(2):626.