

城市快速路驶入匝道瓶颈车辆汇入行为

孙 剑, 蒋 舜, 欧阳吉祥

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

摘要: 基于上海延安高架两个驶入匝道(南线虹许路驶入匝道、虹井路驶入匝道)瓶颈的汇入行为视频, 对汇入车辆、当前和目标车道汇入交互车辆进行了轨迹和汇入行为参数提取, 共获得 416 个汇入行为样本; 应用分类回归树(CART)对 3 种汇入行为分别进行建模, 分析影响不同汇入行为的因素, 并用混淆矩阵对分类结果进行评价。结果表明, CART 能较好地预测 3 种不同汇入类型, 其分类准确率均达到了 75% 以上。CART 与经典离散选择模型和朴素贝叶斯分类结果对比表明, CART 的分类效果明显优于上述两类模型。

关键词: 城市快速路; 驶入匝道瓶颈; 汇入行为; 分类回归树; 交通流失效

中图分类号: U491

文献标志码: A

Modeling the Vehicle Merging Behaviors at Urban Expressway On-ramp Bottlenecks

SUN Jian, JIANG Shun, OUYANG Jixiang

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

Abstract: Based on the videos of traffic flow at two bottlenecks (Hongxu on-ramp and Hongjing on-ramp) on Yan'an Expressway in Shanghai, 416 empirical merging behavior samples were collected by extracting trajectories from merging vehicles, as well as each adjacent vehicles. The classification and regression tree (CART) was adopted for modeling three merging situations, the key parameters affecting different merging behaviors were analyzed and the confusion matrix was used to evaluate the result of the classification accuracy. The results show that CART performed well with these data. All the accuracies are over 75%. Moreover, a comparison among CART, classical discrete choice model and naive Bayes classifier was conducted, and the CART shows the best classification results.

Key words: urban expressway; on-ramp bottleneck; merging

behaviors; classification and regression tree; traffic breakdown

如果说交叉口是间断流交通设施的拥堵节点, 那么瓶颈点则是快速路等连续流交通设施的咽喉, 是导致快速路拥堵的直接诱因。快速路瓶颈包括常发性瓶颈和偶发性瓶颈, 而常发性瓶颈处的交通流失效是导致快速路拥堵的重要因素。现有研究表明^[1-3], 在驶入匝道这样的常发性瓶颈处, 驾驶人必须通过换道来进入主线, 而这种换道行为是触发瓶颈失效的因素之一。还有研究显示^[4], 主线车流与匝道车流会竞争下游汇入点的通行能力, 而这种相互作用会导致瓶颈的通行能力下降。上海的典型驶入匝道瓶颈研究表明, 快速路汇入行为特别是非自由汇入行为(压迫式汇入或协作式汇入)是导致驶入匝道瓶颈早发性失效的关键诱因^[5]。因此对个体车辆汇入行为的研究及建模能从微观层面上分析瓶颈失效的机理, 并为开发高精度微观仿真模型及设计实施相应的管理措施进而实现高效汇入提供理论和应用指导。

本文针对不同汇入行为, 分别用分类回归树(classification and regression tree, CART)进行了建模分析。此外, 为验证模型的精度, 本文还将 CART 分类精度与经典的离散选择模型及朴素贝叶斯分类精度进行了对比分析。

1 研究综述

驶入匝道瓶颈汇入行为是强制性换道行为, 而间隙接受模型最早被用来建立这类换道行为的模型^[6-9]。在这类模型中, 驾驶人汇入/换道与否取决于换道车辆与目标车道前车和后车之间的间隙, 如果

收稿日期: 2014-06-11

基金项目: 国家自然科学基金(51278362, 51422812)

第一作者: 孙 剑(1979—), 男, 工学博士, 教授, 博士生导师, 主要研究方向为交通仿真与实验, 交通流分析与优化。

E-mail: sunjian@tongji.edu.cn

间隙大于可接受的间隙,则进行汇入或换道,反之则继续在当前车道行驶.其中,可接受间隙为一些解释变量的函数.

随后,离散选择模型也被用于建立换道模型.其中最为常见的是二项 Logit 模型^[10-11],它是根据效用最大化理论所建立的,汇入效用作为解释变量的函数影响着驾驶人的换道行为.此外,McFadden 提出了多项 Logit 模型(multinomial logit model, MNL),使模型从二项选择变为了多项选择^[12].但 Logit 模型假设随机效用部分无法观测,并且服从 Gumbel 分布、各选择肢之间相互独立,这与实际情况有所出入.为了克服这一不足,Daly 和 Zachary 等提出了巢式 Logit(nested Logit, NL)模型^[12].Sun 等^[13]通过对汇入行为的分析,建立了 2 个 MNL 模型和 3 个 NL 模型,结果表明 2 层的 NL 模型(第 1 层为自由换道和非自由换道,第 2 层将非自由换道细分为协作和压迫换道)效果最好.

间隙接受模型和离散选择模型都是参数化的方法,所得出的结果均为解释变量的函数.近年来,机器学习被引入到换道行为的建模之中,如遗传模糊方法^[14]、贝叶斯分类^[15]等.这类方法是非参数化方法,且不需要假设分布形式.本文主要运用改进的决策树方法 CART 对汇入行为进行建模.考虑到城市

快速路驶入匝道非自由汇入行为对瓶颈失效的破坏性影响,与现有研究^[15]只将汇入行为分为汇入与不汇入两类不同,本文除了这一分类方法外,还分析了自由汇入与非自由汇入模型结果,以及自由汇入、非自由汇入与非成功汇入的多分类模型.此外,文本还将 CART 的结果分别与经典的离散选择模型及朴素贝叶斯分类结果进行了对比.

2 研究路段和数据

2.1 研究路段

本文所研究的驶入匝道瓶颈是由于驶入匝道汇入引起瓶颈点拥堵而下游保持畅通状态的独立瓶颈点.通过大规模现场踏勘,最终选择了上海市延安高架自西向东车流方向,虹井路以及其相邻的下游虹许路这 2 个独立瓶颈点作为分析对象.两处均为由驶入匝道汇入主线所构成的常发性瓶颈点,相距 1 912 m,其主线 3 车道,匝道 2 车道.图 1 给出了虹井路和虹许路驶入匝道瓶颈的道路几何线形.两个瓶颈点几何线形及尺寸基本一致,且处于同一快速路相邻路段,因此其汇入行为基本一致,样本数据可以合并处理.

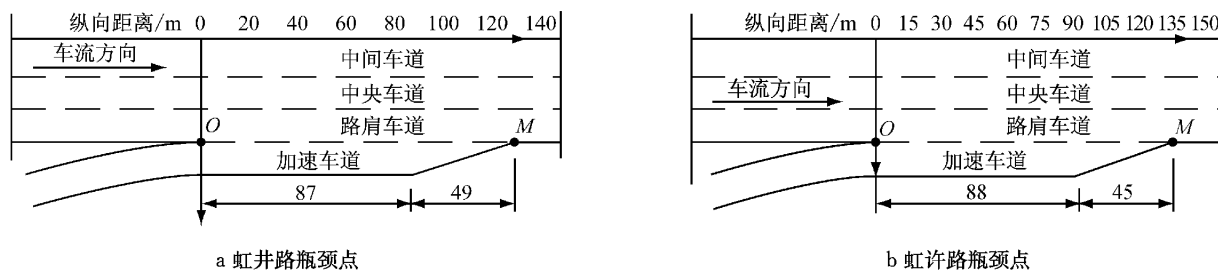


图 1 两个瓶颈点的道路线形和尺寸(单位:m)

Fig.1 The schematic diagrams of two on-ramp bottlenecks (unit:m)

2.2 数据

本文共处理了 9 h 的视频(包括交通流从畅通到拥挤的过程),采集了 416 条汇入行为数据.Hidas^[16]根据换道过程中前后车的相对位置和车间距,把换道行为分为自由换道、协作换道以及压迫换道三种类型.其中,自由换道是对目标车道车辆基本无影响的换道类型;协作换道是指车辆在汇入目标车道前,目标车道后车主动刹车礼让换道车进入的换道行为;压迫换道是换道车辆强行汇入目标车道而迫使目标车道后车被动减速让行的换道类型.基于协作换道与压迫换道都会对目标车道车辆产生影响,且在数据提取过程中对这两种换道行为的判

断存在一定主观性,故在本研究中将这两种换道行为进行了合并,统称为非自由汇入.因此,在 416 条汇入样本中,包括自由汇入(normal lane changing, NLC)数据 57 条、非自由汇入(non-normal lane changing, NNLC)数据 172 条、非成功汇入(unsuccessful lane changing, USLC)数据 187 条.其中,也将自由汇入与非自由汇入统称为成功汇入(successful lane changing, SLC),共计 229 条数据.而本文的非成功汇入样本是指:(1)有汇入动作却没有成功汇入的样本;(2)在汇入之前有一段压线骑行时间的样本.根据 Sun 等^[13]所建立的汇入行为离散选择模型研究结果,每个样本最多涉及 4 辆车

(汇入车辆、当前车道前车、目标车道前车和目标车道后车)以及 9 个因素,本文的数据采集要求与其相同,各关键参数变量说明如下: T_C 为快速路驶入匝道瓶颈点交通流状态(1-自由流状态,2-失效转变状态,3-拥堵状态), D 为汇入车辆距加速车道终点距离,m, V 为汇入车辆速度, $\text{km} \cdot \text{h}^{-1}$, T_{lead} 为汇入车辆与目标车道前车的车头时距,s, T_{lag} 为汇入车辆与目标车道后车的车头时距,s, T_{pre} 为汇入车辆与当前车道前车的车头时距,s, S_{lead} 为汇入车辆与目标车道前车的车头空距,m, S_{lag} 为汇入车辆与目标车道后车的车头空距,m, S_{pre} 为汇入车辆与当前车道前车的车头空距,m.

3 汇入行为建模

3.1 方法概述

CART 是由 Breiman 等在 1984 年提出的一种决策树分类方法^[17]. 其采用二分递归分割的技术,将当前的样本集划分成两个子样本集,也就是说,CART 所生成的决策树每个非叶节点都有两个分枝,是结构简单的二叉树,这样可以有效地处理缺失数据. CART 以 Gini 系数(以 G_{mi} 表示)的减少量为测度指标,选取使 Gini 系数减少量最大的属性为最佳特征变量,对训练样本数据集进行分组,构造二叉决策树. Gini 系数的数学定义为

$$G_{\text{mi}}(T) = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

式中: T 为未被分组前的训练样本数据集; k 为类别变量的个数; p_i 为类别 i 在 T 中出现的概率. G_{mi} 系数越小,样本的“纯净度”就越高,划分效果也越好. 若 T 被划分为 T_1, T_2 , 那么此时的 G_{mi} 系数为

$$G_{\text{mi}}(T) = \frac{S_1}{S} \cdot G_{\text{mi}}(T_1) + \frac{S_2}{S} \cdot G_{\text{mi}}(T_2) \quad (2)$$

式中: S 为 T 中的样本个数; S_1, S_2 分别为 T_1, T_2 中的样本个数. 由式(1)和(2)易得,将样本数据集分为两组后与原来的 Gini 系数相比,减少量为 $\Delta G_{\text{mi}}(T) = G_{\text{mi}}(T) \cdot$

$$\left[\frac{S_1}{S} \cdot G_{\text{mi}}(T_1) + \frac{S_2}{S} \cdot G_{\text{mi}}(T_2) \right] \quad (3)$$

Gini 系数的减少量说明将样本数据集分组后,两组子样本数据集类别变量取值异质性相加总和比未被分组前样本数据集类别变量取值异质性减少,即分组后样本类别变量尽量趋于相同类别值,其纯度增加. 通过反复计算保证决策树的每个节点均取 G_{mi} 系数减少量最大的特征变量,最终构建的二叉决

策树将是最理想的.

3.2 生成树

CART 的生长过程是通过选择测试变量反复划分训练样本数据集的过程. 关键在于如何获取测试变量,即从给定的众多特征变量中选取最佳分组变量和从该分组变量的不同取值中选取最佳分割点.

在本研究中共涉及 9 个特征变量,暂将其设为 $X(1) \sim X(9)$. 首先,检查所有形如 $X(1) < C$ 的分化,这里 C 是一个常数,变化范围是 $X(1)$ 的最小值到最大值,所有 $X(1) < C$ 的被分化到左边,其余被分化到右边. 通过变化 C 值检查所有可能的分化,根据 Gini 系数选择其中最好的一个. 在找到 $X(1)$ 的最优化分化后,以同样的方式确定 $X(2) \sim X(9)$ 的最优化分化,这样就得到了 9 种不同的分化. 然后再用同样的评价准则,也就是 Gini 系数,从这些分化中挑选出个最好的分化. 接着再利用二叉树的结构,从根节点开始,在上述分组条件下对训练样本集再进行分组,在生成的每个左右子样本集上重复这个分组过程,直到无法继续这个过程时子样本集成为叶节点而终止,这时树生成完毕.

除了生成树外,还要为树上的节点标上类别. 因考虑到每个节点都可能成为叶节点,CART 算法对每个节点都分配类别,其方法可以是判断式(4):

$$\frac{C(j|i)p(i)N_i(t)}{C(i|j)p(j)N_j(t)} > \frac{N_i}{N_j} \quad (4)$$

式中: $p(i), p(j)$ 分别是类 i, j 的先验概率; N_i, N_j 分别是训练集中类 i, j 的样本个数; $N_i(t), N_j(t)$ 分别是节点 t 的数据集中类 i, j 的样本个数; $C(j|i)$ 是误分类成本,表示实际上属于 i 类却被错分成 j 类的概率, $C(i|j)$ 所表示的意思类似. 若对除类 i 外的所有类 j 都成立,则将 t 节点标记为类 i .

3.3 修剪枝

对所生成的树进行修剪可以控制树的复杂程度,避免其过于复杂和庞大,与此同时,还可以解决过度拟合的问题. 常用剪枝方法有预剪枝和后剪枝两种. 本研究是将两种方法相结合进行修剪.

预剪枝是在树生长完全前,按照某种标准停止树的生长,其优点是算法相对简单、效率高等. 本研究中对树的深度不进行限制. 结合相关文献^[18-19]及本研究中样本量的大小,规定对于自由/非自由分类来说,父节点中至少需包含 4 个样本、子节点中至少需包含 3 个样本,同时设定 3 个标准差为剪枝标准;对于自由/非自由/非成功分类来说,父节点中至少需包含 5 个样本、子节点中至少需包含 4 个样本,同时设定 2 个标准差为剪枝标准;对于成功/非成功分

类来说,父节点中至少需包含 6 个样本、子节点中至少需包含 5 个样本,同时设定 2 个标准差为剪枝标准。

后剪枝是对一棵生长完全的树,通过某种标准剪掉某些枝干。由于后剪枝法基于树的全局信息,因此通常优于预剪枝的方法。其方法有代价-复杂度剪枝(cost-complexity pruning, CCP)、降低错误剪枝(reduced error pruning, REP)、悲观错误剪枝(pessimistic error pruning, PEP)和最小错误剪枝(minimum error pruning, MEP)等,本文采用 CCP 方法^[20]。

4 模型结果

4.1 自由/非自由汇入分类

该分类方法共涉及自由汇入样本 57 条、非自由汇入样本 172 条,总计 229 条,按照训练:检验=1:1 的比例用 CART 算法进行分类,结果见表 1。

表 1 自由/非自由汇入分类结果

Tab.1 The classification results of NLC and NNLC

已观测	已预测			
	自由	非自由	正确百分比/%	
训练集	自由	26	9	74.3
	非自由	9	80	89.9
总计百分比/%				
	自由	16	6	72.7
测试集	非自由	9	74	89.2
	总计百分比/%			
	自由	23.8	76.2	85.7

由表 1 可以看出,该模型对于非自由汇入的预测精度较高,可达到近 90%,对于自由汇入的预测精度相对较低,但总体精度也能达到 85%以上。而且,检验数据的精度趋势与训练数据的精度趋势相同,且检验数据的总体精度要略高于训练数据,可见 CART 算法用于该分类情况较为合适。

4.2 自由/非自由/非成功汇入分类

该分类方法共涉及自由汇入样本 57 条、非自由汇入样本 172 条,非成功汇入样本 187 条,总计 416 条,把样本按照训练:检验=1:1 的比例用 CART 算法进行分类。其所生成的树及具体分类结果如图 2 及表 2 所示。图 2 中,Node0~Node12 均表示节点编号。

如图 2 所示,对于自由/非自由/非成功汇入行为的分类来说,所生成的决策树相对复杂。其所涉及的关键变量共有 5 个,分别为:距加速车道终点距离、汇入车辆与当前车道前车车头空距、汇入车辆与目标车道前车车头空距、汇入车辆与目标车道后车

车头时距和瓶颈点交通流状态。其中汇入车辆与当前车道前车的车头空距先后被判断了两次,与此同时,距加速车道终点距离这一变量是判断非成功汇入的首要条件。由此可以发现,是否成功汇入与当前车道的状况密切相关。

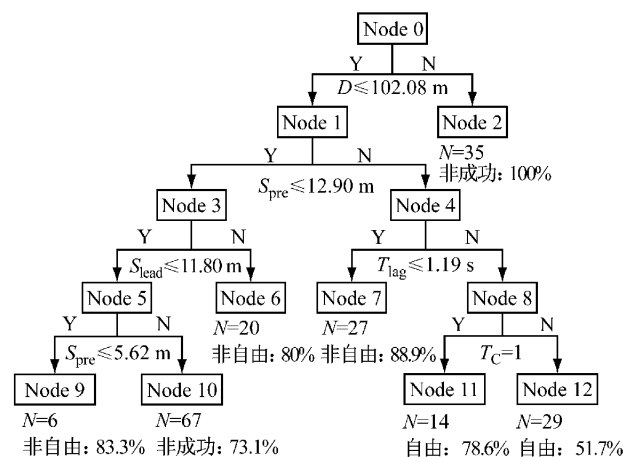


图 2 自由/非自由/非成功汇入的分类树

Fig.2 The classification tree of NLC, NNLC and USLC

表 2 自由/非自由/非成功汇入分类结果

Tab.2 The classification results of NLC, NNLC and USLC

已观测	已预测				
	自由	非自由	非成功	正确百分比/%	
训练集	自由	11	17	4	34.4
	非自由	3	60	14	77.9
	非成功	0	5	84	94.4
总计百分比/%					
	自由	7	14	4	28.0
测试集	非自由	3	72	20	75.8
	非成功	1	11	86	87.8
总计百分比/%					
	自由	5.0	44.5	50.5	75.7

由表 2 发现,该模型对于 3 种汇入情况的预测精度由高到底依次是非成功汇入、非自由汇入、自由汇入。对于精度最高的非成功汇入来说,其训练集和检验集的平均精度可达 90%以上,要比非自由汇入的平均精度高出约 15 个百分点,从中可以推测,非成功汇入的样本相对于其余两类样本来说更具可识别性。

4.3 成功/非成功汇入分类

该分类方法共涉及自由汇入样本 57 条、非自由汇入样本 172 条,非成功汇入样本 187 条,总计 416 条,按照训练:检验=1:1 的比例用 CART 算法进行分类。其所生成的树及具体分类结果如图 3 及表 3 所示。

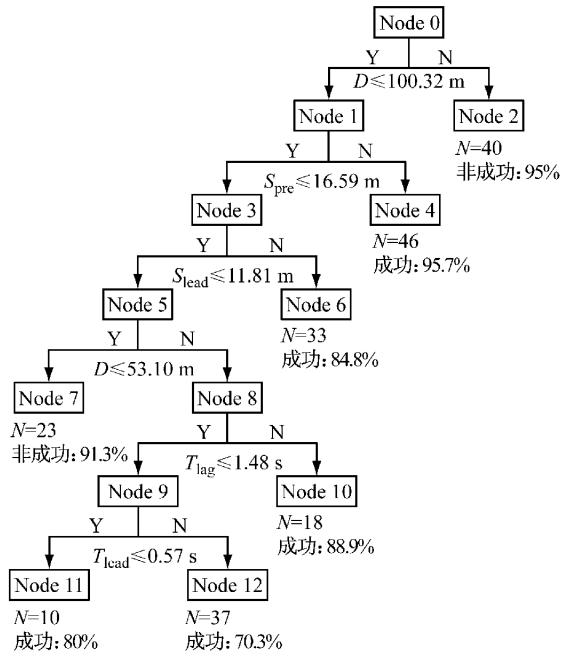


图 3 成功/非成功汇入的分类树

Fig. 3 The classification tree of SLC and USLC

表 3 成功/非成功汇入分类结果

Tab. 3 The classification results of SLC and USLC

	已观测	已预测		
		成功	非成功	正确百分比/%
训练集	成功	96	15	86.5
	非成功	11	85	88.5
	总计百分比/%	51.7	48.3	87.4
测试集	成功	99	19	83.9
测试集	非成功	8	83	91.2
	总计百分比/%	51.2	48.8	87.1

如图 3 所示,对于成功/非成功汇入行为的分类来说,所生成的决策树层数较前一种分类多,除最后一层外,每层都只判断出一种分类.其所涉及的关键变量也有 5 个,与第 4.2 节不同的是,其多了汇入车辆与目标车道前车车头时距,但少了瓶颈点交通流状态.由此可见,瓶颈点交通流状态是判断自由/非自由汇入行为的一个关键因素,对非成功换道来说则影响不大.

由表 3 的结果也可发现,在将自由汇入和非自由汇入合并后,其预测精度还是略低于非成功汇入.此外,该分类方法也是 3 种分类方法中精度最高的,但与此同时,其分类也是最粗糙的.

5 结果对比

结合表 1—3 的结果,表 4 给出了 CART 与朴素贝叶斯分类(naive bayes, NB)和离散选择模型(二项 logit, binary logit, BL 和 NL)在分类准确率方面

的比较结果.需要注意的是,在自由/非自由/非成功这一分类中,所用的为二层 NL 模型,第一层为成功与非成功汇入的判断,第二层将成功汇入细分为自由与非自由汇入.

表 4 不同算法的结果对比

Tab. 4 Comparison of several methods

	CART	NB	BL
自由/非自由			
训练集	85.5		83.50
测试集	85.7	72.21	85.96
自由/非自由/非成功			
训练集	78.3		62.98
测试集	75.7	68.38	
成功/非成功			
训练集	87.4		71.30
测试集	87.1	76.99	71.01

由表 4 可知,CART 表现出了较高的分类准确性,而朴素贝叶斯的准确性是最低的,因为本研究所使用的变量不满足 NB 一个属性值对给定类的影响独立于其他属性值的假设,故影响分类准确率.而对 CART 与离散选择模型这两种方法来说,在第一种分类中,两者无明显的差异,而在第二、第三种分类中,前者的准确性要高于后者 10%~20%,这说明,CART 能较好地处理分类数据.

6 结语

本文基于视频采集的匝道车辆汇入轨迹,通过决策树模型研究了快速路驶入匝道车辆的不同汇入行为以及影响因素,主要结论如下:

(1) 利用 CART 模型对快速路上匝道汇入行为建模的结果表明,3 种分类训练集与测试集的准确率均达到了 75%以上,其中自由/非自由和成功/非成功汇入的分类精度都达到了 85%以上,成功/非成功汇入的分类精度最高. CART 模型可以较准确地预测不同类型的汇入行为.

(2) 由 CART 模型所构建的判断条件可知,目标车道的状况是影响自由与非自由汇入的关键区分因素,而对非成功汇入的判断来说,当前车道的状况对其影响较大.

(3) CART 与朴素贝叶斯分类、离散选择模型的结果对比表明,CART 模型在这三种方法中对汇入行为分类的准确度最高.

参考文献:

[1] Cassidy M J, Bertini R L. Some traffic features at freeway

- bottlenecks [J]. *Transportation Research Part B*, 1999, 33(1): 25.
- [2] Patire A, Cassidy M J. Lane changing patterns of bane and benefit: observations of an uphill expressway [J]. *Transportation Research Part B*, 2012, 15(4): 656.
- [3] Zheng Z, Ahn S, Chen D, *et al.* The effects of lane-changing on the immediate follower: anticipation, relaxation, and change in driver characteristics [J]. *Transportation Research Part C*, 2013, 26(1): 367.
- [4] Xuan Y, Coifman B. Identifying lane change maneuvers with probe vehicle data and an observed asymmetry in driver accommodation [J]. *Journal of Transportation Engineering*, 2012, 138(8): 1051.
- [5] Sun J, Zhao L, Zhang H M. Mechanism of early-onset breakdown at on-ramp bottlenecks on shanghai, China, expressways [J]. *Transportation Research Record*, 2014, 2421(1): 64.
- [6] Yang Q, Koutsopoulos H N. A microscopic traffic simulator for evaluation of dynamic traffic management systems [J]. *Transportation Research Part C*, 1996, 4(1): 113.
- [7] Ahmed K I. Modeling driver's acceleration and lane Changing behaviors [D]. Boston: Massachusetts Institute of Technology, 1999.
- [8] Hidas P. Modeling vehicle interactions in microscopic simulation of merging and weaving [J]. *Transportation Research Part C*, 2005, 13(1): 37.
- [9] Toledo T, Koutsopoulos H N, Ben-Akiva M. Integrated driving behavior modeling [J]. *Transportation Research Part C*, 2007, 15(2): 96.
- [10] Kita H. Effects of merging lane length on the merging behavior at expressway on-ramps [C] // *Proceedings of 12th International Symposium on the Theory of Traffic Flow and Transportation*. New York: Elsevier, 1993: 37-51.
- [11] Meng Q, Weng J. An improved cellular automata model for heterogeneous work zone traffic [J]. *Transportation research part C*, 2011, 19(6): 1263.
- [12] 陆化普. 交通规划理论与方法[M]. 北京: 人民交通出版社, 1998.
- LU Huapu. *Theory and method in transportation planning* [M]. Beijing: China Communications Press, 1998.
- [13] Sun J, Ouyang J, Yang J. Modeling and analysis of the merging behaviors at expressway on-ramp bottlenecks [J]. *Transportation Research Record*, 2014, 2421(1): 74.
- [14] Hou Y, Edara P, Sun C. A genetic fuzzy system for modeling mandatory lane changing [C] // *Proceedings of 15th International IEEE Conference on Intelligent Transportation Systems*. Anchorage: IEEE, 2012: 1044-1048.
- [15] Hou Y, Edara P, Sun C. Modeling mandatory lane Changing using Bayes classifier and decision trees [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 15(2): 647.
- [16] Hidas P. Modelling lane Changing and merging in microscopic traffic simulation [J]. *Transportation Research Part C*, 2002, 10(5): 351.
- [17] Breiman L, Friedman J, Stone C J, *et al.* Classification and regression trees [M]. New York: CRC Press, 1984.
- [18] Montella A, Aria M, D' Ambrosio A, *et al.* Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery [J]. *Accident Analysis & Prevention*, 2012, 49(1): 58.
- [19] Abellán J, López G, De OñA J. Analysis of traffic accident severity using decision rules via decision trees [J]. *Expert Systems with Applications*, 2013, 40(15): 6047.
- [20] 张宇. 决策树分类及剪枝算法研究[D]. 哈尔滨: 哈尔滨理工大学, 2009.
- ZHANG Yu. Study on decision tree classification and pruning [D]. Harbin: Harbin University of Science and Technology, 2009.