

基于车辆轨迹数据的交叉口危险驾驶行为预测

唐克双, 杨博文, 许 凯, 孙梓粟, 周 楠

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

摘要: 基于大量实证车辆轨迹数据建立了信号控制交叉口相位切换期间轨迹预测模型和驾驶员心理决策过程预测模型, 并在两个模型的基础上建立了驾驶员危险行为识别的规则, 三者共同构成了驾驶员危险行为预测方法。经检验, 预测方法的整体精度达到 88.89%。该预测方法将驾驶员临近交叉口时的隐藏心理决策过程作为一个重要变量运用到危险行为预测方法中, 并实现了对车辆个体动态化的轨迹预测。

关键词: 危险行为预测; 车辆轨迹预测; 隐藏心理预测; 危险行为识别

中图分类号: U121

文献标志码: A

Prediction of Driver's Risky Behaviors Based on Vehicle Trajectory Data at Signalized Intersections

TANG Keshuang, YANG Bowen, XU Kai, SUN Zili, ZHOU Nan

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

Abstract: A trajectory prediction model and a driver's hidden decision-making process prediction model were established based on the collection and analysis of vehicle trajectory data at intersection influenced areas. Moreover, based on these two models, the rules for recognizing driver's risky behaviors were built. The three together constitute the overall driver's risky behaviors prediction method at signalized intersections. It is found that the overall accuracy of this prediction method is 88.89%. This prediction method realizes vehicle's individual dynamic trajectory prediction, utilizing driver's hidden decision-making process as an important variable in prediction of driver's risky behaviors.

Key words: risky behaviors prediction; vehicle trajectory prediction; driver's hidden decision-making process prediction; risky behavior recognition

随着我国社会经济水平的迅速发展, 人均机动车保有量急速上升, 相应的道路交通事故量也随之增长。在这其中, 信控交叉口事故占据较大的比例。世界经济合作与发展组织(OECD)2014 年的统计数据显示^[1], 发生在交叉口及其周边的交通事故占整个道路交通事故的 30% 左右。交叉口的常见事故包括闯红灯、急停、追尾等。

当前对交叉口危险行为的研究主要包括以下三个方面:

一方面是基于两难区的驾驶员危险行为识别。Gazis 等^[2]于 1960 年首次提出关于车辆在黄灯时刻陷入两难区的定义, 建立了 Gazis - Herman - Maraduin 模型; 在之后的研究中, Moon 等^[3]提出了基于到达车辆变化的速度、加速度, 确定实时车辆两难区边界的方法; Zegeer 等^[4]在论文中研究了两难区规避系统实施前后的交叉口事故率的变化情况, 发现通过延长绿灯时间避免车辆在黄灯时刻陷入两难区的信号控制策略, 能够有效降低交叉口 54% 左右的交通事故, 成功验证了车辆在黄灯时刻陷入两难区是驾驶员的一种危险行为, 及其对交叉口交通事故的显著影响。但值得注意的是, 在这些基于信号控制策略的改善方案中, 对两难区的预测大多是通过大量数据获得恒定车辆参数经验值(包括车辆的速度、加速度等)的静态两难区预测方法, 这种计算方法很难真实地反映针对不同车辆个体的动态两难区边界变化情况, 在精度上存在较大误差。

第二方面是通过分析不同类型驾驶员对信号灯的认知和理解程度, 研究驾驶员在交叉口的危险行为。一部分学者探讨了绿闪信号对驾驶员可能做出危险决策的影响^[5-6], 而另外一部分学者则探讨了黄灯信号对驾驶员可能做出危险决策的影响^[7]。除了信号灯灯色的影响外, Shawarby 等^[8]通过研究发现, 车辆在黄灯时刻距离停车线的距离也会在很大

程度上影响驾驶员在交叉口做出的决策。

第三方面是研究驾驶员个人特征对驾驶行为的影响。Rakha 等^[9]研究发现驾驶员的年龄、性别会对驾驶员在黄灯时刻做出的决策产生显著影响。但当前对驾驶员个人特征进行的相关研究中,未给出如何将驾驶员的隐藏心理决策过程作为一个重要变量,运用在危险行为预测上的定量方法。

为了减少交叉口闯红灯、追尾等事故的发生频率,提高交叉口驾驶员的行驶安全性,本文从驾驶员的角度对可能发生的危险驾驶行为进行预测,开发了基于车辆轨迹的交叉口驾驶员危险行为预测方法,并基于预测信息做出提前预警,帮助驾驶员主动改变危险驾驶行为,提高驾驶安全性。

1 数据采集及处理

本课题数据采集自上海市郊区嘉定区曹安公路

的 3 个交叉口,以及市区四平路大连路交叉口。这 4 个交叉口作为上海市典型的公路交叉口和城市道路交叉口,能够较好地代表上海市两种不同类型交叉口的普遍状况。曹安公路的 3 个交叉口自西向东分别为:曹安公路-嘉松北路,曹安公路-曹丰路,曹安公路-翔江公路。4 个交叉口的基本情况如表 1 所示。交叉口数据采集时间为 2013 年 7 月至 2013 年 10 月间的 16 个工作日,天气情况均为良好。对每一个交叉口,分别由两台摄像机进行两组视频的拍摄:A 摄像机用于记录进口道上游区域的车辆,B 摄像机用于记录信号状态。

数据处理过程如下:首先将视频数据导入视频数据处理软件 George 2.1;然后从 B 摄像机记录的视频中记录每个信号灯周期的绿闪时刻,再回到 A 摄像机记录的视频中所对应的时刻,跟踪记录绿闪前 2 s 至红灯亮起时的车辆轨迹,数据采集间隔为 0.12 s。

表 1 交叉口基本情况

Tab. 1 Basic situation at the intersection

| 交叉口名称 | 道路等级 | 相位 | 交叉口形状 | 限速/(km·h ⁻¹) | 黄灯时间/s | 全红时间/s | 绿闪倒计时 | 绿闪启亮时刻 |
|-----------|---------|-----|-------|--------------------------|--------|--------|-------|---------|
| 曹安公路-嘉松北路 | 主干路/主干路 | 四相位 | 十字 | 东西向 80 南北向 60 | 3 | 1 | 无 | 黄灯前 3 s |
| 曹安公路-曹丰路 | 主干路/支路 | 三相位 | 丁字 | 东西向 80 南北向 60 | 3 | 1 | 有 | 黄灯前 3 s |
| 曹安公路-翔江公路 | 主干路/次干路 | 四相位 | 十字 | 东西向 80 南北向 60 | 3 | 1 | 无 | 黄灯前 3 s |
| 四平路-大连路 | 主干路/主干路 | 四相位 | 十字 | 东西向 30 南北向 40 | 3 | 1 | 无 | 黄灯前 3 s |

剔除交叉口处的转弯、调头车辆的数据,将得到的交叉口处直行车辆数据导入 excel 中,提取的数据包括状态数据和轨迹数据两部分,共 1 025 条,其中跟车车辆数据 525 条,非跟车车辆数据 500 条。具体包括车辆编号、车辆类型、绿闪启亮时速度、绿闪启亮时至停车线距离、黄灯启亮时速度、黄灯启亮时至停车线距离、到达停车线时速度以及穿越停车线的时刻。

从 1 025 条车辆数据中分离出全部的 15 条闯红灯车辆和 3 条急停的车辆危险行为状态轨迹数据(共 18 条)用作后文中对整体预测方法精度验证的数据集 A,将剩余的 1 007 条数据作为后文中轨迹预测模型和驾驶员决策预测模型的数据集 B。

2 研究方法

危险行为预测方法的整体流程图如图 1 所示,

图中建立的轨迹预测模型、驾驶员隐藏心理决策过程模型和危险行为识别规则的具体方法和流程将在以下章节具体阐述。

2.1 轨迹预测

2.1.1 通过模型的建立

由于通过的车辆轨迹与停止的车辆轨迹存在明显差异,因此,首先需要预测车辆的“通过-停止”状况,得到用于被预测车辆后续算法的对应历史数据集。本文通过 SPSS 软件建立车辆通过与否的二元 Logit 模型^[10]。

本研究中,因所研究道路的车流特征有明显差异,所以需要两种通过模型,分别为跟驰状态下的通过模型和非跟驰状态下的通过模型。根据美国 1994 年版的《道路通行能力手册》规定^[11]:当车头时距 ≤ 5 s 时,车辆处于跟驰状态。而文献^[12]认为,跟驰行为发生在两车车头间距为 0~100 m 或 0~125 m 的范围内。结合实际视频观测,郊区曹安公路 3 个交叉口

的车辆可以认为是处于非跟驰(单车)状态,而市区四平路-大连路的交叉口必须认作是跟驰(跟车)状态。

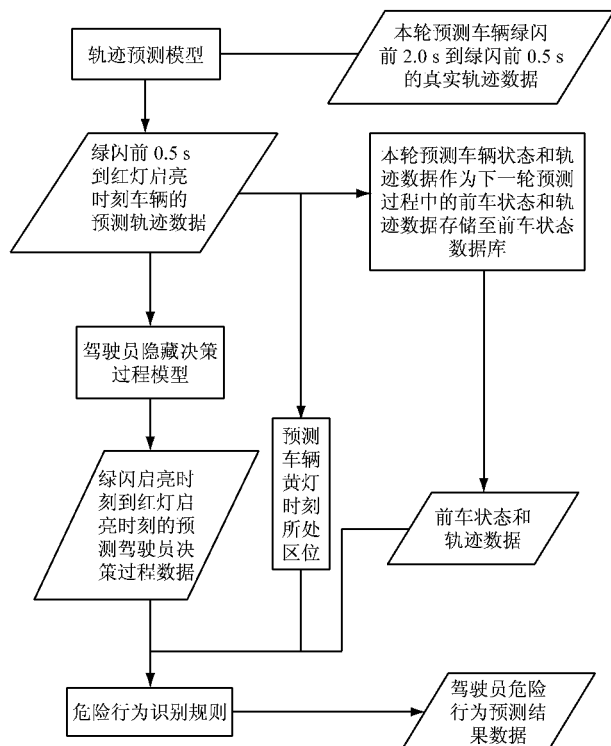


图 1 驾驶员危险行为预测方法整体流程图

Fig. 1 Flowchart of driver's risky behaviors prediction method

对单车状态下通过模型的建立,本文提取 498 辆位于曹安公路交叉口车辆在绿闪启亮时刻的速度、加速度、距离停车线的距离以及车型(大车为 1,小车为 0)这 4 个指标作为自变量集 X_i ,把车辆最终是否通过(通过为 0,不通过为 1)作为因变量 Y_i ,进行 Logit 模型的建立. 导入 SPSS 软件后,结果如表 2 中步骤 1 所示,车型的显著性水平(0.491)并不高,说明车型对车辆最终是否通过的影响是不显著的,因此可以考虑剔除车型这一变量,重新对模型进行标定,即进行步骤 2. 步骤 2 中 3 个自变量(速度、加速度、距离停车线的距离)都通过了显著性检验. 此时进一步对模型的自变量间的相关性检验,发现 3 个自变量间的相关性均为显著,因此考虑对自变量进行调整,将距离与速度合并为一个变量 T_s ,并剔除加速度变量,重新对模型进行标定,即进行步骤 3. 可以认为模型是合理的,同时,模型正确率高达 92.4%. 因此,最终的通过模型如下:

$$P = \frac{1}{1 + e^{-(3.481T_s - 5.618)}} \quad (1)$$

式中: P 为车辆不通过停车线的概率; T_s 为车辆在绿闪时刻距离停车线的距离(m)与车辆在绿闪时刻的速度($m \cdot s^{-1}$)的比值。

表 2 Logit 模型标定结果

Tab. 2 Logit model calibration results

| 步骤 | 变量 | 解释系数 | 标准误差 | 因变量显著性 | 自由度 | 自变量显著性 | 优势比 |
|------|-------|--------|-------|--------|-----|--------|--------|
| 步骤 1 | 加速度 | -1.460 | 0.466 | 9.824 | 1 | 0.002 | 0.232 |
| | 距离 | 0.135 | 0.028 | 23.962 | 1 | 0 | 1.144 |
| | 速度 | -0.237 | 0.054 | 19.456 | 1 | 0 | 0.789 |
| | 车型 | 0.582 | 0.845 | 0.474 | 1 | 0.491 | 1.790 |
| | 常量 | 2.742 | 1.768 | 2.407 | 1 | 0.121 | 15.524 |
| 步骤 2 | 加速度 | -1.356 | 0.429 | 9.997 | 1 | 0.002 | 0.258 |
| | 距离 | 0.130 | 0.026 | 25.522 | 1 | 0 | 1.139 |
| | 速度 | -0.231 | 0.052 | 20.160 | 1 | 0 | 0.793 |
| | 常量 | 2.993 | 1.728 | 2.999 | 1 | 0.083 | 19.953 |
| 步骤 3 | 距离/速度 | 3.481 | 0.622 | 31.312 | 1 | 0 | 32.491 |
| | 常量 | -5.618 | 0.989 | 32.237 | 1 | 0 | 0.004 |

对跟车状态下车辆通过模型的建立,也可以采取相似的方法. 提取剩余的 509 辆位于四平路-大连路交叉口车辆在绿闪启亮时刻的状态和轨迹数据,得到跟车状态下车辆的通过模型如下:

$$P = \frac{1}{1 + e^{-(2.734T_s - 4.415)}} \quad (2)$$

2.1.2 对应历史轨迹数据集的找寻过程

首先将用于轨迹预测模型和驾驶员决策预测模型训练和验证的数据集 B (共 1 007 条)一分为二:一半作为模型训练和矫正的历史数据子集 C (共 503

条数据),数据子集 C 根据车辆是否跟车以及车辆是否通过交叉口进一步分为 4 个历史数据组;另一半作为验证模型精度的历史数据子集 D (共 504 条数据)。

随后取被预测车辆在任意特定时刻的速度、加速度和距离停车线的距离作为通过模型的参数,建立模型预测车辆是否通过交叉口,结合该时刻该车与前车间的车头时距,判断其是否处于跟车状态,根据两者的结果找到该车辆在数据集 C 中对应的数据组。

2.1.3 轨迹预测模型的建立

本文首先考虑基于单纯 K-Nearest Neighbors (KNN) 算法^[13] 建立轨迹预测模型。

为被预测车辆找到对应的历史数据集分组后,根据被预测车辆预测的起始时间 t_0 、到停车线的距离 s (本文选取绿闪前 2 s 的车辆真实状态和轨迹数据作为预测起始点数据),在对应的历史数据组内找出同样处于时间 t 的样本点.若处于时间 t_0 的样本点数量大于 4,则选取距离 d 最小的 4 个样本点;若处于时间 t_0 的样本点的数量小于 4,则将时间范围扩大,即在 $[t_0 - 4l, t_0 + 4l]$ 的区间内 (l 代表步长,本项研究中每个步长取 0.12 s) 寻找样本点.根据寻找到的 4 个最接近的历史车辆状态和轨迹数据值,计算样本点之间的距离 d ,即 $d = \sqrt{s^2 + t_0^2}$,同时对距离进行离差标准化处理以忽略量纲对计算的影响,即 $d'_j = \frac{d_j - d_{\min}}{d_{\max} - d_{\min}}$ ($j=1,2,3,4$),然后对 4 个最接近的距离值的权重系数采用平方权函数,使最近的特征值获得最大权重,依次减小,即

$$w_i = \begin{cases} \frac{K^2 - (i-1)^2}{K(K+1)(4K-1)/6}, & i=1,2,\dots,K \\ 0, & i=K+1,K+2,\dots,N \end{cases} \quad (3)$$

式中: K 为样本点的个数,这里取 4; i 为样本点的排序,距离真实轨迹最近的历史轨迹样本点排序为 1,以此类推。

最后计算 4 个最相邻历史数据的加权平均数,得到下一时刻的预测车辆状态和轨迹值.之后将初始预测的下一时刻的车辆状态和轨迹值,作为下一时刻起始的车辆真实状态和轨迹值,用以预测再下一时刻的车辆状态和轨迹值,依次循环,直到黄灯时刻停止。

然而这种基于单纯 KNN 算法的轨迹预测方法,由于误差在每个步长的不断积累,导致后半段预测轨迹与真实轨迹间的差值较大,因此本文引入卡尔曼滤波算法对 KNN 算法存在的误差进行降噪矫正。

卡尔曼滤波算法^[14] 是用最小均方根误差作为最佳估计准则,利用前一时间步长内对变量的估计值,和当前时间步长内对变量的观测值综合得出当前时刻的最佳估计值.其核心公式如下:

$$\mathbf{X}_{(k|k-1)} = \mathbf{A}\mathbf{X}_{(k-1|k-1)} + \mathbf{B}\mathbf{U}_{(k)} \quad (4)$$

$$\mathbf{P}_{(k|k-1)} = \mathbf{A}\mathbf{P}_{(k-1|k-1)}\mathbf{A}^T + \mathbf{Q} \quad (5)$$

$$\mathbf{X}_{(k|k)} = \mathbf{X}_{(k|k-1)} + K_{g(k)}(\mathbf{Z}_{(k)} - \mathbf{H}\mathbf{X}_{(k|k-1)}) \quad (6)$$

$$K_{g(k)} = \mathbf{P}_{(k|k-1)}\mathbf{H}^T / (\mathbf{H}\mathbf{P}_{(k|k-1)}\mathbf{H}^T + \mathbf{R}) \quad (7)$$

$$\mathbf{P}_{(k|k)} = (\mathbf{I} - K_{g(k)}\mathbf{H})\mathbf{P}_{(k|k-1)} \quad (8)$$

式(4)~(8)中: $\mathbf{U}_{(k)}$ 为 k 时刻状态的控制量; $\mathbf{X}_{(k|k-1)}$ 为 $k-1$ 时刻对 k 时刻车辆状态和轨迹数据预测值; $\mathbf{X}_{(k-1|k-1)}$ 为 $k-1$ 时刻车辆状态和轨迹数据最优值; $\mathbf{X}_{(k|k)}$ 为 k 时刻状态的最优化估算值; $\mathbf{P}_{(k|k-1)}$ 为 $\mathbf{X}_{(k|k-1)}$ 对应的协方差矩阵; $\mathbf{P}_{(k-1|k-1)}$ 为 $\mathbf{X}_{(k-1|k-1)}$ 对应的协方差矩阵; $\mathbf{Z}_{(k)}$ 为 k 时刻车辆状态和轨迹数据的测量值; \mathbf{R} 、 \mathbf{Q} 为系统过程的协方差矩阵; K_g 为卡尔曼增益; \mathbf{I} 为单位矩阵; \mathbf{A} 、 \mathbf{B} 、 \mathbf{H} 为系统参数。

特别注意的是,在 KNN 结合卡尔曼滤波的算法中,转移系数 \mathbf{A} 不再是系统常数,而成为根据 KNN 预测轨迹位置变化的变量,即

$$A_{k-1} = \frac{s_k}{s_{k-1}}$$

式中: s_k 为 KNN 轨迹预测中车辆在 k 时刻的位置。

然而这种 KNN 结合卡尔曼滤波算法的缺点在于它需要实时的真实观测值的输入,这使得随着卡尔曼滤波对 KNN 算法预测结果矫正时间的延长,模型能够预测的时间长度减少.因此希望找到一个合适的临界点,作为 KNN 结合卡尔曼滤波算法和单纯 KNN 算法的交界,实现模型精度与预测时间长度间的平衡。

令绿闪前 2 s 到黄灯启亮时刻的时间段长度为 t ,依次取 0, 0.1 t , 0.2 t , ..., t 作为 KNN 结合卡尔曼滤波算法和单纯 KNN 算法的交界点,对交界点前的时间段采用 KNN 结合卡尔曼滤波算法进行轨迹预测,之后把交界点处最后得到的经过降噪矫正后的预测车辆状态和轨迹数据,作为后一阶段单纯 KNN 算法的起始真实车辆状态和轨迹数据,提高单纯 KNN 算法起点数据的精度。

不同状态的车辆在不同交界点条件下的轨迹预测均方根误差(RMSE)如表 3 所示.表 3 中 1)为采用全程单纯 KNN 算法;2)为采用全程 KNN 结合卡尔曼滤波算法。

其趋势变化图如图 2 所示。

由图 2 可以看出,当交界点位置位于 0.3 t 之后(包括 0.3 t),随着交界点位置的改变,模型精度变化很小,因此本文选取 0.3 t 作为 KNN 结合卡尔曼滤波算法与单纯 KNN 算法的交界位置.该点既最大程度满足了预测的时间长度,同时又实现了模型精度的优化。

2.1.4 对车辆未来黄灯时刻所处区位的判别

利用前一部分轨迹预测模型预测得到的车辆在黄灯时刻的速度 v 、加速度 a ,以及车辆自身加、减速

表3 不同交界点条件下轨迹预测的均方根误差(RMSE)

Tab. 3 Root mean square error (RMSE) of trajectory prediction under different conditions

| 车辆状态 | 均方根误差 | | | | | | | | | | |
|-----------|-------|-------|------|------|------|------|------|------|------|------|------|
| | 1) | 0.1t | 0.2t | 0.3t | 0.4t | 0.5t | 0.6t | 0.7t | 0.8t | 0.9t | 2) |
| 单车状态下停止车辆 | 26.91 | 12.97 | 6.17 | 5.28 | 4.89 | 4.82 | 4.80 | 4.77 | 4.76 | 4.74 | 4.74 |
| 单车状态下通过车辆 | 13.80 | 6.74 | 4.74 | 3.73 | 3.40 | 3.31 | 3.30 | 3.29 | 3.28 | 3.27 | 3.27 |
| 跟车状态下停止车辆 | 17.86 | 8.42 | 5.67 | 4.85 | 4.28 | 4.15 | 4.14 | 4.14 | 4.14 | 4.14 | 4.13 |
| 跟车状态下通过车辆 | 16.68 | 7.93 | 5.00 | 4.12 | 3.75 | 3.70 | 3.67 | 3.67 | 3.67 | 3.67 | 3.67 |

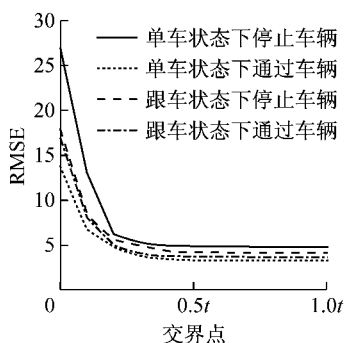


图2 不同条件下轨迹预测均方根误差

Fig. 2 Root mean square error chart for trajectory prediction under different conditions

度等车辆状态数据,计算该车辆的最小停车距离 X_c 和最大通过距离 X_0 。

根据之前对于车辆在信控交叉口两难区的研究^[2], X_c 和 X_0 的计算公式分别为

$$X_c = V_Y \delta + \frac{V_Y^2}{2d} \quad (9)$$

$$X_0 = V_Y \delta + V_Y(t - \delta) + \frac{1}{2}a(t - \delta)^2 - L_0 - L \quad (10)$$

式(9)、(10)中: V_Y 为黄灯启亮时车辆的速度, $m \cdot s^{-1}$; δ 为驾驶员反应时间,这里取 $1.5 s^{[15]}$; d 为车辆的减速度,这里取 $-0.35g^{[15]}$; t 为黄灯信号时长, s ; a 为车辆的加速度, $m \cdot s^{-2}$; L_0 为交叉口宽度, m ; L 为车辆长度, m 。

随后通过比较前一部分轨迹预测模型预测得到的车辆在黄灯时刻距离停车线的距离 X_v 与 X_c 和 X_0 之间的大小关系,得以判定车辆未来在黄灯时刻所处的区位,如图3所示。



图3 区位判定图

Fig. 3 Decision of location

(1) $X_c > X_0$

当 $X_v > X_c$, 车辆位于停止区。

当 $X_0 < X_v < X_c$, 车辆位于两难区。

当 $X_v < X_0$, 车辆位于通行区。

(2) $X_c < X_0$

当 $X_v > X_0$, 车辆位于停止区。

当 $X_c < X_v < X_0$, 车辆位于选择区。

2.2 驾驶员隐藏心理决策过程预测

本文尝试引入隐马尔科夫模型对驾驶员隐藏的心理决策过程进行预测。

建立的隐马尔科夫模型^[16]是一种特殊的马尔科夫过程,在这一模型中,假设不可观测状态(驾驶员隐藏心理状态)变化是服从马尔科夫过程的,而每一隐藏状态又对应于一个可观测状态(车辆的特定时刻的速度、加速度),而可观测状态可以通过采集的车辆状态和轨迹数据得到。利用隐马尔科夫模型的这一特性,可以对驾驶员潜在的危险认知进行预测。

本课题中,隐藏状态是驾驶员从绿闪启亮时至黄灯结束的6 s期间在任意时刻决定通过/不通过的心理状态;可观测状态是驾驶员在这6 s期间的任意时刻的速度、加/减速度。

利用五元组对本课题中的隐马尔可夫模型进行描述,即

$$\lambda = (N, M, A, B, \pi) \quad (11)$$

五元组具体的矩阵形式如下所示:

$$N = \{q_1, q_2\} \quad (12)$$

$$M = \begin{Bmatrix} (v_1, a_1) & (v_1, a_2) & \cdots & (v_1, a_j) \\ (v_2, a_1) & (v_2, a_2) & \cdots & (v_2, a_j) \\ \vdots & \vdots & & \vdots \\ (v_i, a_1) & (v_i, a_2) & \cdots & (v_i, a_j) \end{Bmatrix} \quad (13)$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (14)$$

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ b_{21} & \cdots & b_{2m} \end{bmatrix} \quad (15)$$

$$\pi = \{\pi_1, \pi_2\} \quad (16)$$

式(11)~(16)中: N 为驾驶员隐藏心理状态(这里特指驾驶员心理通过与否的决策结果)随时间变化的有限集合; M 为车辆可观测的状态数据(包括车辆速度、加速度、到停车线距离等)的有限集合; A 为状态

转移概率矩阵, a_{ij} 是指驾驶员在 $t-1$ 时刻的心理状态(即通过/停止)转变为 t 时刻的心理状态(即停止/通过)的概率; B 为观察值概率分布矩阵, b_{jk} 是指每一个时间步 t , 当驾驶员的可观测状态为 (v_i, a_j) 时, 他的心理状态为通过/不通过的概率; π 为给定的初始观测状态下, 绿闪启亮时驾驶员心中决定通过/不通过的概率。

随后运用向前向后算法和维特比算法对隐马尔可夫模型进行求解。

运用向前向后算法训练得到准确的隐马尔可夫模型的步骤如下:

(1) 初始化. 根据文献[17]中给出的关于 π_i, a_{ij}, b_{jk} 的初值赋值方法(满足概率条件), 利用建立模型的 1 007 条车辆状态和轨迹数据, 得到初始 Hidden Markov Model(HMM) 参数 λ_0 , 设 $i=0$ 。

(2) 代入初始 HMM 参数 λ_0 和实际观测序列(这里选取之前建立的包含 503 条历史数据的预测数据集作为实际观测序列), 计算期望值 $\xi_i(i, j)$ 和 $\gamma_i(i)$ 。

(3) 用步骤(2)所得的期望值重新估计 π_i, a_{ij}, b_{jk} , 得到模型 λ_{i+1} 。

(4) 循环设计. 令 $i=i+1$; 重复步骤(2)和(3), 直至 π_i, a_{ij}, b_{jk} 值收敛。

在 MATLAB 中实现上述算法, 得到最终收敛的 A, B, π 分别为

$$A = \begin{bmatrix} 0.98 & 0.02 \\ 0 & 1.00 \end{bmatrix}$$

式中: 驾驶员隐藏心理状态由上一时刻通过转变为这一时刻通过的概率为 0.98; 驾驶员隐藏心理状态由上一时刻通过转变为这一时刻停止的概率为 0.02; 驾驶员隐藏心理状态由上一时刻停止转变为这一时刻通过的概率为 0; 驾驶员隐藏心理状态由上一时刻停止转变为这一时刻停止的概率为 1.00。

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.02 & 0.12 & 0 & 0.02 & 0.44 \\ 0.08 & 0.12 & 0 & 0.13 & 0.20 & 0.47 \\ 0 & 0 & 0.40 & 0 & 0 & 0 \end{bmatrix}$$

式中: 第 1 行表示每个时间步驾驶员心理状态为不通过的概率; 第 2 行表示预测时间段内每个时间步驾驶员心理状态为通过的概率。

$\pi = \{0.90 \quad 0.10\}$, 为驾驶员在绿闪启亮时刻心中决定通过和不通过的概率。

在得到训练后的隐马尔可夫模型后, 再次输入实际观测得到的车辆状态和轨迹数据 $M = \{o_1, \dots,$

$o_i\}$, 利用维特比算法求解驾驶员隐藏的心理决策状态, 具体的步骤如下:

(1) 初始化

$$\delta_1(i) = \pi_i b_i(o_1), \varphi_1(i) = 0, 1 \leq i \leq N \quad (17)$$

式中: $\delta_i(i)$ 为 Viterbi 变量, 即 t 时间步沿状态序列 q_1, \dots, q_t 且 $q_t = S_i$ 产生出 o_1, \dots, o_t 的最大概率; $\varphi_i(i)$ 为记忆变量, 记录概率最大路径上当前状态的前一个状态. 最后得到预测的驾驶员隐藏心理决策序列 $N = \{q_1, \dots, q_j\}$ 。

(2) 递归

$$\delta_t(j) = [\max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] b_j(o_t),$$

$$2 \leq t \leq T, 1 \leq j \leq N$$

$$\varphi_t(j) = [\operatorname{argmax}_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] b_j(o_t),$$

$$2 \leq t \leq T, 1 \leq j \leq N \quad (18)$$

(3) 终结

$$p^* = \max_{1 \leq i \leq N} [\delta_T(i)], q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)] \quad (19)$$

(4) 路径回溯

$$q_t^* = \varphi_{t+1}(q_{t+1}^*), t = T-1, T-2, \dots, 1 \quad (20)$$

2.3 危险行为识别规则的建立

根据之前预测的轨迹, 可以判定车辆在黄灯时刻将处在通行区、停止区、选择区还是两难区. 依据前车是否通过交叉口(单车状态视为前车通过交叉口), 进一步把车辆的状态分为 8 个小类. 对这 8 个小类, 根据预测的车辆黄灯启亮时刻所处区位, 以及利用隐马尔可夫模型预测的驾驶员决策心理状态(通过/停止), 建立相应的危险驾驶行为预测识别规则. 具体的分类规则见表 4。

表 4 危险行为预测分类规则

Tab. 4 Classification rule of risky behavior prediction

| 预测所处区位 | 前车状态 | 隐马尔可夫模型预测心理状态 | 危险行为判别 | 现实中可能的危险行为 |
|--------|------|---------------|----------|----------------------|
| 通行区 | 停止 | 通过 停止 | 危险 危险 | 急停、与前车追尾 急停、与后车追尾 |
| | 通过 | 通过 停止 | \ 危险 | \ 急停 |
| 停止区 | 停止 | 通过 停止 | 危险 \ | 急停、与前车追尾 \ |
| | 通过 | 通过 停止 | 危险 \ | 闯红灯、超速 \ |
| 选择区 | 停止 | 通过 停止 | 危险 \ | 急停、与前车追尾 \ |
| | 通过 | 通过 停止 | \ \ | \ \ |
| 两难区 | 停止 | 通过 停止 | 危险 危险 | 急停、与前车追尾 急停、与前车追尾 |
| | 通过 | 通过 停止 | 危险 危险 | 闯红灯、超速 急停、与后车追尾 |

注: 表格中标“\”的表示不予考虑的情况

3 模型有效性验证

3.1 对轨迹预测模型的有效性验证

利用前文 2.1.2 节分离出来的预测历史数据集 $D(504)$ 对轨迹预测模型的精度进行验证. 在 504 条历史车辆数据中提取从绿闪前 2.0 s 到绿闪前 0.5 s 这一时间段内的车辆状态和轨迹数据, 代入模型中预测得到后续直到红灯结束时刻的车辆状态和轨迹数据, 并利用预测得到的车辆状态和轨迹数据对车辆在黄灯时刻所处的区位进行预测, 然后与真实情况进行对比, 得到轨迹预测部分模型的精度为 93.65%(472/504).

3.2 对隐藏心理决策过程预测模型的有效性验证

同样地, 利用之前特意分离出来的预测历史数

据集 $D(504)$ 条), 对驾驶员隐藏心理决策过程预测模型的精度进行验证. 提取数据集中 504 条历史车辆从绿闪前 2.0 s 到黄灯启亮这一时间段内的车辆状态和轨迹数据, 代入模型中预测得到相应时间段内的驾驶员隐藏的心理决策过程. 将这一决策过程和驾驶员实际的决策结果进行对比, 得到驾驶员决策过程预测部分模型的精度为 98.81%(498/504).

3.3 对整体驾驶员危险行为预测方法的有效性验证

为了对所建立模型的有效性进行验证, 在建立的数据库中提取存在危险行为但未参与模型训练的全部 18 辆车的轨迹数据, 提取每辆车从绿闪前 2.0 s 到绿闪前 0.5 s 这一时间段内的车辆轨迹数据, 利用建立的模型对 18 辆车可能的危险行为进行预测. 具体的验证结果见表 5.

表 5 模型验证结果

Tab. 5 Results of model validation

| 车辆编号 | 前车是否通过交叉口 | 预测黄灯时刻区位 | 实际黄灯时刻区位 | 预测驾驶员决策过程 | 驾驶员实际驾驶决策 | 危险行为预测结果 | 实际结果 |
|------|-----------|----------|----------|------------|-----------|----------|------|
| 1 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 2 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 3 | 是 | 停止区 | 停止区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 4 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 5 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 6 | 是 | 通过区 | 两难区 | 持续采取通过决策 | 通过 | 无危险行为发生 | 闯红灯 |
| 7 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 8 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 9 | 是 | 通过区 | 两难区 | 持续采取通过决策 | 通过 | 无危险行为发生 | 闯红灯 |
| 10 | 是 | 停止区 | 停止区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 11 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 12 | 是 | 两难区 | 两难区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 13 | 是 | 停止区 | 停止区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 14 | 是 | 停止区 | 停止区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 15 | 是 | 停止区 | 停止区 | 持续采取通过决策 | 通过 | 闯红灯 | 闯红灯 |
| 16 | 否 | 通过区 | 通过区 | 持续采取通过决策 | 急停(通过转停止) | 急停 | 急停 |
| 17 | 是 | 两难区 | 两难区 | 通过决策转向停止决策 | 急停(通过转停止) | 急停 | 急停 |
| 18 | 否 | 两难区 | 两难区 | 通过决策转向停止决策 | 急停(通过转停止) | 急停 | 急停 |

根据预测的结果, 对所建立的危险行为预测模型的误报率进行分析.

18 辆车中 16 辆车的预测结果为存在危险行为, 且根据危险行为规则判别的危险行为类别, 与实际发生的危险行为类别相吻合; 剩余 2 辆车的预测结果为不存在危险行为, 与实际驾驶员出现的危险驾驶行为(闯红灯)结果不符, 预测出现误差. 轨迹预测模型部分的准确率为 88.89%(16/18), 驾驶员决策过程预测准确率为 100.00%(18/18), 整体模型的精度为 88.89%(16/18).

从对模型误报率的结果分析来看, 可以认为利

用本文提出的危险行为预测模型, 对交叉口可能出现的驾驶员危险行为预测的准确性较高. 但是, 对模型预测失败的数据进行分析后, 发现造成模型失准的原因, 均为对车辆未来在黄灯时刻所处区位的预测错误. 这说明轨迹预测模型的算法精度仍需要进一步提高. 此外, 当前用于模型验证的事故数据量较小, 可能导致模型在有效性验证的精度上与真实值存在较大误差, 在未来的研究中还需要利用大量的交叉口事故车辆的数据, 对模型进行进一步的验证和改善.

4 结论

为了减少交叉口处的交通冲突,保障道路行车安全,基于车辆轨迹预测和危险驾驶行为识别开发出相应的交叉口危险驾驶行为预警方法.研究工作的主要成果如下:

(1)基于对车辆历史数据库的采集与分析,利用SPSS软件建立多元Logit通过模型,利用通过模型的数据筛选功能,再结合KNN和卡尔曼滤波的车辆轨迹预测算法,成功建立了车辆轨迹预测模型.

(2)引入两难区的概念,并结合运用两难区的区位概念及基于隐马尔科夫模型的预测方法,获得了驾驶员隐藏心理状态预测序列,建立了交叉口驾驶员决策过程预测模型.

(3)基于之前的车辆轨迹预测和驾驶员决策过程预测模型,确立了整体的交叉口驾驶员危险行为识别规则,进而建立了整体的危险行为识别预测模型.经过验证,整体模型的准确率为88.89%.

本项研究的创新之处在于:利用隐马尔科夫模型预测驾驶员隐藏心理状态,将其与车辆轨迹预测结合起来,用于识别交叉口驾驶员危险行为,大大提高了实际交通流条件下,对每个个体危险驾驶行为识别的准确率.

参考文献:

- [1] The Organization for Economic Cooperation and Development. Road injury accidents[DB/OL]. [2016-7-28]. <http://stats.oecd.org/>.
- [2] GAZIS D, HERMAN R, MARADUIN A. The problem of the amber signal light in traffic flow[J]. Operations Research, 1960, 8(1): 112.
- [3] MOON Y J, COLEMAN F. Dynamic dilemma zone based on driver behavior and car-following model at highway-rail intersections [J]. Transportation Research Part B: Methodological, 2003, 37(4): 323.
- [4] ZEGER C V, DEEN R C. Green-extension systems at high-speed intersection[J]. ITE Journal, 1978, 48(11): 19.
- [5] FACTOR R, PASHKER J N, MAHALE D. The flashing green light paradox[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2012, 15(3): 279.
- [6] GATES T J, NOYCE D A, LARACUENTE L, et al. Analysis of driver behavior in dilemma zones at signalized intersections [J]. Journal of the Transportation Research Board, 2007, 2030: 29.
- [7] SHAWARBY I, RAKHA H, INMAN V, et al. Effect of yellow-phase trigger on driver behavior at high-speed signalized intersections [C/CD] // IEEE Intelligent Transportation Systems Conference. [S.l.]: IEEE, 2006.
- [8] SHAWARBY I, AMER A, RAKHA H. Evaluation of driver stopping behavior on high speed signalized intersection approaches[J]. Journal of the Transportation Research Board, 2008, 2056: 60.
- [9] RAKHA H, AMER A, EL-SHAWARBY I. Modeling driver behavior within a signalized intersection approach decision-dilemma zone [J]. Journal of the Transportation Research Board, 2008, 2069: 16.
- [10] National Academies of Science. Highway capacity manual 2010 [R]. Washington D C: Transportation Research Board, 2010.
- [11] National Academies of Science. Highway capacity manual 1994 [R]. Washington D C: Transportation Research Board, 1994.
- [12] National Research Council. Traffic flow theory [R]. Washington D C: Transportation Research Board, 2006.
- [13] AGRETI A, KATERI M. Categorical data analysis[M]. New York: Wiley, 1990.
- [14] KALMAN R E. A new approach to linear filtering and prediction problems[J]. J Basic Eng Trans, ASME, 1960, 82(1): 35.
- [15] FAMBRO D B, FITZPATRICK K, KOPPA R J. Determination of stopping sight distances [R]. Washington D C: Transportation Research Board, 1997.
- [16] BAUM L E, PETRIE T. Statistical inference for probabilistic functions of finite state markov chains [J]. Annals of Mathematical Statistics, 1966, 37(6): 1554.
- [17] TANG K S, ZHU S F, XU Y Q, et al. Modeling drivers' dynamic decision-making behavior during the phase transition period: an analytical approach based on hidden markov model theory[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(1): 206.