

数据驱动的信号交叉口排队 尾车驶离状态预测

卢凯明^{1,2}, 陈艳艳^{1,2}, 仝瑶^{1,2}, 张健^{1,2}, 李永行^{1,2}, 罗莹^{1,2}

(1. 北京工业大学 城市建设学部, 北京 100124; 2. 北京工业大学 交通工程北京市重点实验室, 北京 100124)

摘要: 针对传统排队尾车驶离状态预测模型难以适应排队消散的不确定性问题, 提出了一种轨迹数据驱动的排队尾车驶离状态预测模型。首先, 分析排队消散轨迹形态及潜在影响因素, 以揭示排队尾车驶离状态的不确定性。然后, 从排队等待和车辆启动两个阶段入手, 提出排队尾车驶离状态影响特征集。最后, 基于极端梯度提升(XGBoost)算法构建排队尾车驶离状态预测模型, 引入SHAP(SHapley Additive exPlanations)可解释机器学习框架解析所有特征的贡献度, 并确定最优特征组合及模型参数。研究表明: 本文基于XGBoost的尾车驶离时间预测模型平均绝对百分比误差(MAPE)为5.74%, 比运动学模型预测精度提升约10%; 尾车驶离速度预测模型MAPE为9.98%, 比运动学模型预测精度提升约6%, 且预测性能均优于随机森林、决策树和多层感知机神经网络3种常用机器学习方法。研究成果可为车路协同环境下交叉口信号相位最小绿灯时间调节与网联车辆生态驾驶提供技术支撑。

关键词: 交通运输系统工程; 排队消散特性; 排队尾车驶离时间; 排队尾车驶离速度; XGBoost; SHAP

中图分类号: U491.1 **文献标志码:** A **文章编号:** 1671-5497(2025)04-1275-12

DOI: 10.13229/j.cnki.jdxbgxb.20230784

Data-driven prediction of departure state for tail vehicles in queues at signalized intersections

LU Kai-ming^{1,2}, CHEN Yan-yan^{1,2}, TONG Yao^{1,2}, ZHANG Jian^{1,2}, LI Yong-xing^{1,2}, LUO Ying^{1,2}

(1. School of Urban Transportation, Urban Construction Department, Beijing University of Technology, Beijing 100124, China; 2. Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: Aiming at the problem that the traditional queue tail vehicle departure state prediction model is difficult to adapt to the uncertainty of queue dissipation, a queue tail vehicle departure state prediction model driven by trajectory data is proposed. By analyzing the shapes of queue dissipation trajectories and potential influencing factors, the uncertainty of departure state of tail vehicles is uncovered. Starting from the two stages of queue waiting and vehicle start-up, a feature set that influences the tail vehicle departure state is proposed. The extreme gradient boosting algorithm is employed to construct the prediction model,

收稿日期: 2013-07-26.

基金项目: 国家自然科学基金项目(72201010); 交通运输部交通运输行业重点科技项目(2022-ZD6-116); 北京市科技计划项目(Z221100005222021).

作者简介: 卢凯明(1992-), 男, 博士研究生. 研究方向: 智能交通控制. E-mail: luakaiming@emails.bjut.edu.cn

通信作者: 李永行(1990-), 男, 讲师, 博士. 研究方向: 行人流仿真, 交通系统分析. E-mail: liyx@bjut.edu.cn

incorporating the SHapley Additive exPlanations (SHAP) interpretable machine learning framework to dissect the contributions of features, and to determine the optimal feature combination and model parameters. The research results indicate that the proposed XGBoost-based departure time prediction model achieves an average mean absolute percentage error(MAPE) of 5.74%, which is improved by 10% approximately compared with the kinematic model. The MAPE for the queue departure speed is 9.98%, improved about 6% over the kinematic model. Furthermore, the performance of the proposed model surpasses three commonly used machine learning methods of random forest, decision trees, and multi-layer perceptron neural networks. The research outcomes provide technical support for adjusting the minimum green light time of intersection signals and eco-driving of connected vehicles in the vehicle-road cooperative environment.

Key words: engineering of communications and transportation system; queue dissipation characteristic; the departure time of tail vehicles in queues; the departure speed of tail vehicles in queues; XGBoost; SHAP

0 引言

信号控制交叉口车辆排队是一种常见的交通现象,排队尾车驶离状态是信号配时优化和生态驾驶速度轨迹优化的重要约束条件^[1-4]。准确预测排队尾车驶离状态,有利于及时调整相位最小绿灯时间,优化目标车辆(即在绿灯信号期间到达交叉口附近区域且有机会实现不停车/排队的车辆)速度轨迹,实现减少延误、降低能耗的目的。

近年来,有关信号交叉口车辆排队的研究主要聚焦于排队消散车头时距和排队长度。众多学者利用摄像机实地采集交叉口绿灯信号启亮后停车线断面的消散车头时距,证实了消散车头时距服从依赖车辆排队位置的对数正态分布^[5,6],改进了信号交叉口跟车模型及通行能力估算方法^[7,8]。利用交叉口进口道视频卡口数据,Zhan等^[9]构建了基于高斯过程的累积到达-消散曲线预测模型,实现了对车道级最大排队长度的间接预测。Mei等^[10]利用低渗透率的网联车轨迹数据分析了连续多个信号周期内车辆排队长度的变化趋势及关联关系,基于贝叶斯估计理论构建了周期级最大排队长度估计模型。该模型需要输入信号配时参数、交叉口上游车辆到达率信息。为了提高模型的实用性,谈超鹏等^[11]利用低采样浮动车轨迹数据,提出了一种数据驱动的排队长度估计方法,可不依赖任何外界信息输入和假设,实现时段级排队长度估计。

排队消散车头时距分布统计方法和排队长度估计方法主要用于信号交叉口通行能力估计和交叉口运行质量评估。然而,对信号配时优化和交

叉口上游车辆速度调控而言,不但需要估计得到排队长度,还需要预测排队尾车的驶离状态,即驶离时间和驶离速度。现阶段针对排队尾车驶离状态的预测方法可归纳为3种:①基于宏观交通流基本图模型,假设车辆到达服从泊松分布,结合交通波模型预测排队消散时间/排队尾车驶离时间^[12-14];②利用运动学模型,假设排队尾车由启动到驶离停车线过程中执行匀加速(或先匀加速,再匀速)的运动模式,通过设置恒定的停车间距、车辆启动反应时间来预测排队尾车驶离时间和速度^[15-17];③基于断面检测设备感知到的排队长度,通过设置恒定的平均饱和消散速率,实现对排队消散时间/排队尾车驶离时间的粗略预测^[2,18]。这3种理论计算方法均是在车辆均匀到达、驶离且为同质交通流的假设基础上提出的。然而,不同类型的车辆及驾驶员的停车间距、反应时间、启动速度均存在一定的差异^[15],基于稳态同质交通流假设建立的排队尾车驶离状态预测模型并不能适应迥异的车辆排队与消散形态,难以保证预测精度。

为此,本文基于无人机视频车辆轨迹数据,构建一种数据驱动的排队尾车驶离状态预测模型。该模型不依赖任何假设,仅从数据中挖掘影响排队尾车驶离状态的显著变量,并学习得到最优预测模型。

1 数据采集与处理

1.1 数据采集

本研究采用的数据来源于河北省廊坊市一个主-次干路相交(京榆大街-燕灵路)的信号交叉口。主路方向限速 60 km/h,交叉口无明显纵坡,

高峰时段交通流量较大,路口几何参数设计规范;主路方向进口道4车道,为“左转-直行-直行-右转”的车道功能划分方式。使用大疆无人机对交叉口进行垂直拍摄,拍摄帧率为30 Hz,飞行高度约为200 m,拍摄范围(沿道路方向)约为0~270 m,如图1所示。飞行时间包含早高峰(7:00~9:00)、晚高峰(17:30~19:00)和平峰(14:00~16:00)这3个时段。

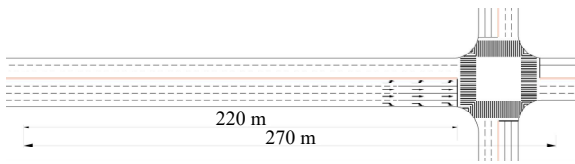


图1 视频采集区域

Fig. 1 Video acquisition area

1.2 数据处理

基于OpenCV开发的目标识别及轨迹跟踪算法,以10帧/s的速率提取车辆行驶轨迹。在无人机视频采集过程中路侧架设摄像机同步采集信号配时方案。采用Savitzky-Golay滤波器对提取的车辆轨迹数据进行分段平滑处理,滑动窗口长度为51帧。图2展示了一个轨迹平滑样例,可以发现,车辆消散过程中并非始终保持匀加速行驶,而是首先经历了短暂加速,达到一定车速后又发生了微小减速,最后再二次加速通过停车线,体现了车辆在消散过程中的不稳定性。

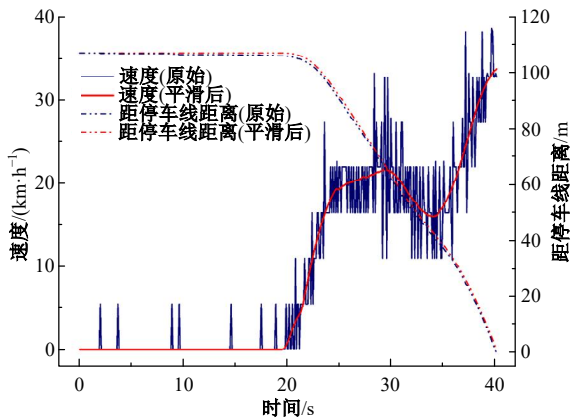


图2 排队消散过程车辆轨迹平滑样例

Fig. 2 Vehicle trajectory smoothing sample during queue dissipation process

2 排队尾车驶离状态分析

2.1 排队尾车驶离状态界定

依据德国仿真软件VISSIM用户手册和美国通行能力手册^[19]中对排队车辆的定义,本文所述

排队尾车是指绿灯启亮时刻由同一车道内所有速度低于10 km/h的车辆构成的车队中最后一辆车。

本文所述排队尾车驶离时间,表示排队尾车由绿灯启亮时刻所在位置到驶离停车线所经历的时间。就感应式信号控制而言,已有研究指出,相位最小绿灯时间需保证感知范围内一定数量的排队车辆完全通过停车线,其取值可随每个信号周期排队状况的变化而变化^[1,18,20],因此,本文所述排队尾车驶离时间可等同于相位最小绿灯时间。就生态速度轨迹规划而言,排队尾车驶离时间是确定车辆速度轨迹规划时域的重要依据。

排队尾车驶离速度,表示排队尾车驶过停车线时的瞬时车速。该状态变量主要应用于交叉口车辆生态速度轨迹规划,预测排队尾车驶离速度有利于确定上游目标车辆过停车线时的终端速度,使目标车辆平稳驶离交叉口。

2.2 排队消散轨迹特征分析

为了探究排队尾车驶离过程潜在的影响因素,支撑数据驱动的排队尾车驶离状态预测特征集构建,本文对信号交叉口车队消散轨迹进行对比分析。

如图3(a)所示,当信号交叉口绿灯启亮后,处于同一车道内的排队车辆将在经历均匀且较短的反应时间后实现启动和快速加速,所有车辆的停车及启动过程可表现为桃红色的三角区域,该三角形较为窄长,呈现出一种“紧凑”的消散跟驰状态。

与图3(a)情形相比,图3(b)则反映出了现实中的另一种普遍现象:由于车队存在大型车辆(车辆5),出于安全起见,其与前车的停车间距较大;同时由于车队内存在启动反应较慢的车辆(车辆3),这些车辆的延迟启动行为可能会传导至后方车队,使消散波速(桃红色三角形斜边的倾斜程度)呈现明显差异。即使这两种情形具有相同排队车辆数,但最终图3(b)情形下的排队尾车驶离时间(36 s)明显大于图3(a)所示的驶离时间(28 s)。

同样,“紧凑”的消散跟驰状态虽然可能有利于车辆同步加速,提升整体效率,但其容易受到启动阶段车辆随机行为的干扰。如图3(c)中车辆7在停车等待时期由于停车间隙较大提前向前滑行

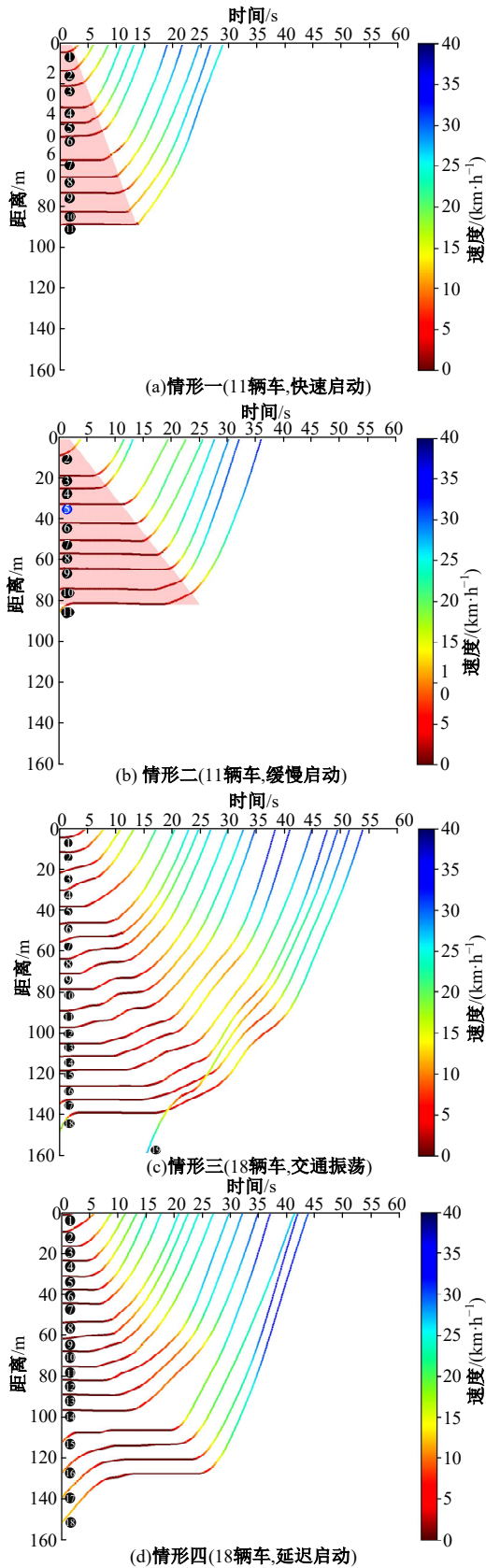


图3 4种典型排队消散轨迹时空图

Fig. 3 Time-spatial diagrams of four typical queue dissipation trajectories

微小距离,但诱发了后方车辆的二次停车波动现象。此外,前方排队车辆低速消散过程中车辆13的一次微小减速又引发了后续车辆的再一次振荡,从而延长了排队整体的启动和加速时间,导致消散时间明显增加。

然而,较大的车间距和延迟启动行为并不一定会对排队消散过程造成负面影响。如图3(d)中的车辆15,这些换道和延迟启动行为虽然短时间内增加了后方车队的启动延误,但是其前方预留的较大停车间距也为车辆提供了有利的加速空间,使其能够快速达到稳态车速,并带动后方车辆快速驶离。图3(c)(d)虽然排队长度相同,但是启动消散过程中的不确定性,使排队尾车驶离时间存在明显差异。

分析上述4种典型排队消散形态可以看出,车辆在启动阶段的行为可能对排队整体的消散具有较大影响,驾驶员反应时间、车辆类型、停车间距、驾驶风格等随机因素都可能间接影响到排队尾车的驶离时间和驶离速度。

2.3 不同排队长度下排队尾车驶离状态特征分析

图4、图5展示了排队尾车驶离时间和驶离速度与排队长度的关系。

由图4、图5可知,随着排队长度的增加,排队尾车驶离时间总体呈线性增长趋势,但排队长度相同时排队尾车驶离时间分布在一个区间,具有不确定性。相比驶离时间,排队尾车驶离速度分布更加离散化,且与排队长度的关联性并不显著。

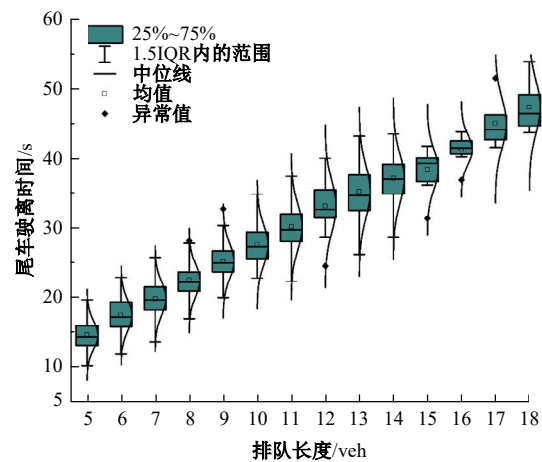


图4 排队尾车驶离时间与排队长度的关系图 (样本量:583)

Fig. 4 Relationship diagram of tail vehicle departure time with queue length (sample size: 583)

因此,单纯利用排队长度预测排队尾车驶离状态无法达到理想的预测效果。

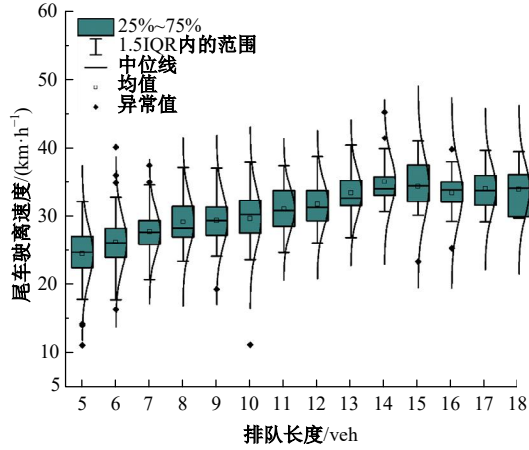


图5 排队尾车驶离速度与排队长度关系图
(样本量: 583)

Fig. 5 Relationship diagram of tail vehicle departure speed with queue length (sample size: 583)

3 基于XGBoost的排队尾车驶离状态预测模型

3.1 XGBoost算法

由3.2、3.3节分析可知,排队尾车驶离状态将受到多重因素的综合影响。为了能够从实际数据中学习一种排队尾车驶离状态与队列停车等待变量和启动行为变量的多元非线性函数关系,使其能够应对现实交通环境中排队消散过程的不确定性,提高排队尾车驶离状态预测精度,本文选取了被广泛应用于交通预测和机理挖掘^[21-24]的XGboost(eXtreme gradient boosting)算法。XGBoost是一个优化的分布式梯度增强库,是对梯度提升算法的改进,具有分布式计算、计算速度快等优势。在求解损失函数极值时将损失函数泰勒展开到了二阶导数,同时损失函数中将树模型复杂度作为目标函数中的正则化项,以防止过拟合^[25]。损失函数定义为:

$$O^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + \Omega(f_i) \quad (1)$$

$$\Omega(f_i) = \gamma\Gamma + \frac{1}{2}\lambda\|w\|^2 \quad (2)$$

式中: $O^{(t)}$ 为经过 t 次迭代后的目标函数; $l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i))$ 为选定的训练损失函数; y_i 为第 i 个样本的真实值; $\hat{y}_i^{(t-1)}$ 为经过 $t-1$ 次迭代后

第 i 个样本的预测值; $f_i(x_i)$ 为经过 t 次迭代训练后的样本 x_i 的决策树函数; $\Omega(f_i)$ 为第 t 次迭代的复杂性函数; γ 为复杂度参数; Γ 为叶节点数; λ 为正则项惩罚系数; w 为叶子节点上的权值。

本研究中的排队尾车驶离时间和速度不存在零值或接近零的极小值,因此,选用平均绝对百分比误差(MAPE)和均方根误差(RMSE)两项指标来评价模型的预测性能,计算方法如下所示:

$$\text{MAPE} = 100\% \times \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (4)$$

式中: N 为测试集样本数量; y_i 为测试样本(真实值); \hat{y}_i 为预测样本(预测值)。

3.2 SHAP可解释机器学习框架

XGBoost算法虽然能够输出特征的重要度排序,但是无法展现特征与预测结果的关系。在完成建模后,有必要量化各特征变量对排队尾车驶离状态的影响,提高预测模型的可解释性。同时通过模型重要特征的分析,有利于优化模型特征结构,提高模型预测结果的可信度。本文选取SHAP(SHapley additive exPlanation)作为模型解释工具。

SHAP框架是来源于博弈论的一种方法,能够定量描述每个样本特征值对模型输出值的贡献值,即特征的SHAP值,等价于各个特征对样本预测结果的影响。将所有特征的SHAP值进行累加即得到模型对样本的预测结果。设第 i 个样本 x_i 的第 j 个特征为 x_{ij} ,模型对 x_i 的预测结果为 y_i , x_{ij} 的SHAP值为 $f(x_{ij})$,则 y_i 与 $f(x_{ij})$ 的关系为:

$$y_i = y_0 + \sum_{j=1}^k f(x_{ij}) \quad (5)$$

式中: y_0 为模型对样本预测的基准值,通常为所有样本预测均值。

由式(5)可知, $f(x_{ij})$ 为特征 j 对样本预测结果的贡献。若 $f(x_{ij})$ 大于0,表示该特征对模型有正向作用,增大了模型的预测值;否则,该特征将起到负向作用,减小了模型的预测值。因此,使用SHAP可解释框架进行排队尾车驶离状态的特征重要度分析,辅助快速识别重要影响因素及作用机理,并为精简模型特征组合提供依据。

3.3 模型构建

3.3.1 特征初选

排队消散过程中主要包括 3 个阶段:排队等待阶段(S_1)、启动阶段(S_2)、加速阶段(S_3)。

这 3 个阶段通过以下两个时刻进行划分: T_0 时刻,绿灯启亮,排队车辆由排队等待阶段进入启动阶段。 T_1 时刻,车队启动波传播到初始排队(T_0 时刻对应的排队)尾车位置,此时所有初始排队车辆均完成启动。值得注意的是,本文研究重点面向相位最小绿灯时间设置和排队情形下的目标车辆生态速度轨迹规划,虽然绿灯启亮后前方队列消散过程中仍可能有新的车辆加入排队,但这些车辆不包含在本文所述排队车辆内。

参考已有研究中信号交叉口排队消散影响因素^[8,26-28],并结合对排队消散特征的分析,初步构建了 17 个特征变量来描述排队尾车驶离状态影响特征集,具体如表 1 所示。基础交通参数包括排队长度、启动车速、车间净距、行驶速度及车型,其余特征均为由基础交通参数得到的统计变量。对本文所提出的启动车速定义如下:表示车辆从

表 1 排队尾车驶离状态影响特征集

Table 1 Feature set influencing the tail vehicle departure state

符号	时刻/阶段	特征变量	变量类型
Y_1	—	驶离时间/s	连续
Y_2	—	驶离速度/($\text{km}\cdot\text{h}^{-1}$)	连续
x_1	S_1	尾车车型	离散
x_2	S_1	尾车相邻前车车型	离散
x_3	T_0	排队长度/veh	连续
x_4	T_0	尾车前方车队大车比	连续
x_5	T_0	尾车前方车辆间距平均值/m	连续
x_6	T_0	尾车前方车辆间距标准差/m	连续
x_7	S_2	尾车启动车速/($\text{km}\cdot\text{h}^{-1}$)	连续
x_8	S_2	尾车相邻前车启动车速/($\text{km}\cdot\text{h}^{-1}$)	连续
x_9	S_2	尾车前方车辆启动车速平均值/($\text{km}\cdot\text{h}^{-1}$)	连续
x_{10}	S_2	尾车前方车辆启动车速标准差/($\text{km}\cdot\text{h}^{-1}$)	连续
x_{11}	S_2	头车启动车速/($\text{km}\cdot\text{h}^{-1}$)	连续
x_{12}	T_1	尾车前方车辆速度平均值/($\text{km}\cdot\text{h}^{-1}$)	连续
x_{13}	T_1	尾车前方车辆速度标准差/($\text{km}\cdot\text{h}^{-1}$)	连续
x_{14}	T_1	尾车前方车辆间距平均值/m	连续
x_{15}	T_1	尾车前方车辆间距标准差/m	连续
x_{16}	T_1	尾车前方车队大车比	连续
x_{17}	T_1	尾车前方车队长度/m	连续

启动到行驶至前车初始所在位置的平均行驶速度,即车头间距与车头时距的比值。如图 6 所示, Δd 为排队等待状态下自车与前车的车头间距, Δt 为启动阶段的车头时距, $\bar{v}=\Delta d/\Delta t$,为车辆的启动车速。车辆的启动车速可以综合反映驾驶员反应时间、车型、驾驶风格等多因素对排队消散的组合影响。由于排队队列的头车没有前车,因此,本文定义头车启动车速为头车由启动时刻到驶过停车线 10 m 的平均车速。该指标能够综合反映绿灯启亮后头车驾驶员反应强度,以及相位切换期间交叉口内部交通冲突对头车启动行为的影响。

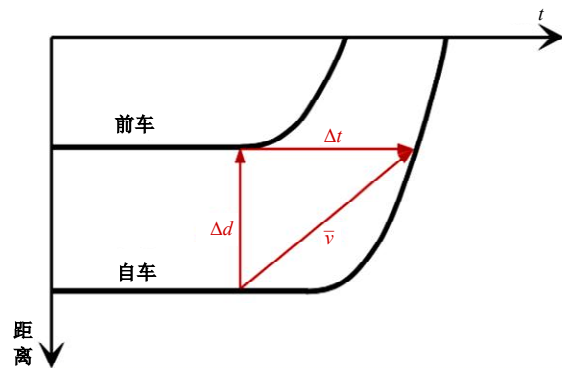


图 6 启动车速示意图

Fig. 6 Schematic diagram of start-up speed

3.3.2 模型训练与特征解析

(1)模型训练。本文共提取了 583 个车队样本,随机切分为训练集(占比 80%)与测试集(占比 20%)。基于贝叶斯优化算法和 5 折交叉检验法求解模型最优超参数。贝叶斯优化算法迭代次数为 100 次。采用平均绝对误差作为模型拟合效果评价函数。

(2)特征解析。引入 SHAP 进行特征分析。图 7、图 8 显示了初步选定的 17 个特征对排队尾车驶离时间和驶离速度的影响程度排序,由上到下重要度依次减小。横坐标为各特征的 SHAP 值,SHAP 越大表示对样本预测值的贡献越大。每个点代表一个样本,颜色越红代表特征本身数值越大,颜色越蓝则越小。

可见,对排队尾车驶离时间影响最为显著的是 T_0 时刻的排队长度、头车启动车速、 T_0 时刻排队尾车前方车辆间距标准差 3 个特征。具体而言,随着排队长度增加,尾车驶离时间趋于延长,呈现明显正向作用。而头车启动车速和车间距标准差均呈现负向作用,即头车启动车速越低、车间距标准差越小,尾车驶离时间反而越长。其余特

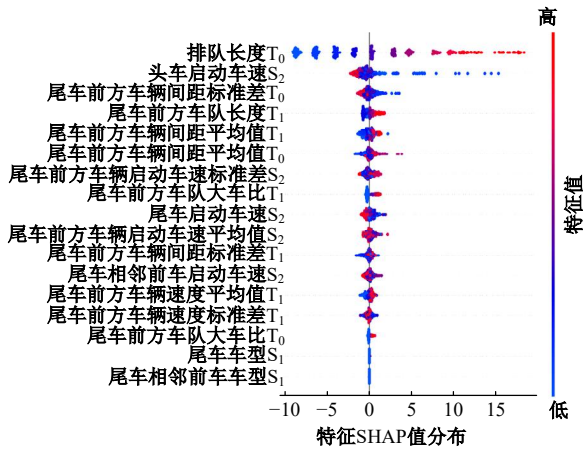


图 7 尾车驶离时间预测特征重要度 SHAP 图

Fig. 7 SHAP diagram of feature importance to departure time prediction of tail vehicle

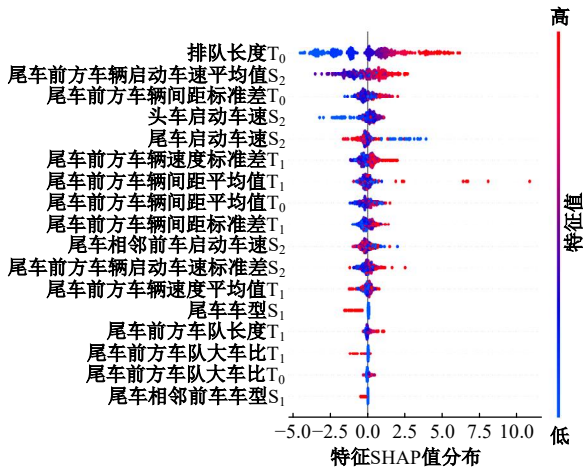


图 8 尾车驶离车速预测特征重要度 SHAP 图

Fig. 8 SHAP diagram of feature importance to departure speed prediction of tail vehicle

征的 SHAP 值均小于 5,相比前 3 个特征贡献度并不显著。

相比之下,尾车驶离车速的影响因素错综复杂,其中排在前三位的分别是 T_0 时刻排队长度、启动阶段尾车前方车辆启动车速平均值、 T_0 时刻排队尾车前方车辆间距标准差,头车启动车速居第 4 位。具体而言,排队长度越长,则排队尾车驶离速度越快,但排队长度对排队尾车驶离速度的贡献度(SHAP 值接近 5)远小于对排队尾车驶离时间的贡献度(SHAP 值高于 15),因此,排队长度对驶离速度的影响程度远不及尾车驶离时间,原因是当消散车队车速达到交叉口期望车速时,速度将趋于稳定,排队长度与尾车驶离车速并不是简单的线性关系。尾车前方车辆平均启动车速越快,尾车驶离速度将越快,表明排队车辆一致且均

匀地启动行为有利于排队尾车获得较快的驶离速度。绿灯启亮时尾车前方车间距标准差越大(车间距分布越离散),越有利于提高排队尾车驶离速度,原因是较大的停车间距有利于消解排队消散初期不稳定交通状态下车队内部的交通波动。

3.3.3 最优特征组合比选

为保证模型预测精度和时效性,同时尽可能简化特征结构,降低模型复杂度,本研究在 4.3.2 节特征贡献度分析基础上,选取对排队尾车驶离状态贡献度较大的特征进行组合试验,制订了如表 2 和表 3 所示的用于排队尾车驶离时间和驶离速度预测的多种特征组合试验方案,并分别进行了模型训练。

表 2 排队尾车驶离时间预测不同特征组合

Table 2 Feature consists for departure time prediction of tail vehicle

组合编号	特征组成
1	所有特征
2	T_0 时刻所有特征+头车启动车速
3	T_0 时刻特征
4	初始排队长度+所有车间距相关特征+头车启动车速
5	T_0 时刻排队长度+所有车间距相关特征
6	排队长度
7	Based_3
8	Based_5
9	Based_6
10	Based_8

注:Based_3表示上述图 7 特征重要度排前 3 位的特征。

表 3 排队尾车驶离速度预测不同特征组合

Table 3 Feature consists for departure speed prediction of tail vehicle

组合编号	特征组成
1	所有特征
2	T_0 时刻所有特征
3	T_0 时刻排队长度+所有车间距相关特征
4	T_0 时刻所有特征+头车启动车速
5	Based_8
6	Based_6
7	Based_5
8	Based_4
9	Based_1

注:Based_1表示上述图 8 特征重要度排第 1 位的特征。

图9展示了不同特征组合方式下的排队尾车驶离时间预测结果,可见特征组合不同会引起预测性能评价指标的差异。其中,组合7预测精度和稳定性最好,MAPE为5.74%,RMSE为1.53,与组合6单纯采用排队长度预测相比,精度有明显提升。此外,组合8、组合9添加冗余特征反而会降低模型预测精度。因此,最终将组合7(包含T₀时刻排队长度、尾车前方车间距标准差和头车启动车速3个特征)作为排队尾车驶离时间预测的最优特征组合。

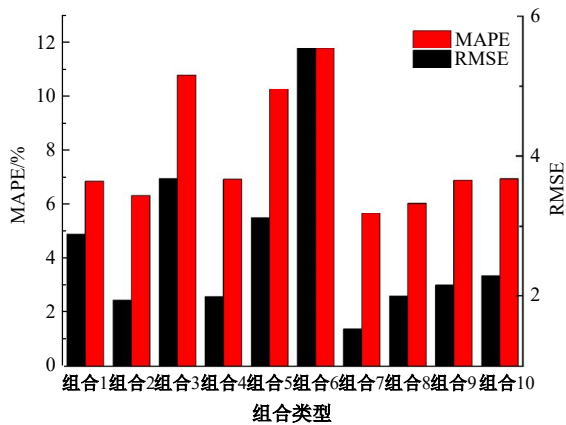


图9 不同特征组合方式下尾车驶离时间预测对比

Fig. 9 Prediction comparison of tail vehicle departure time under different feature consists

图10展示了不同特征组合方式下的尾车驶离速度预测结果,可以发现,MAPE稳定在8.5%左右,RMSE稳定在3.6左右,表明改变特征组合对尾车驶离速度的预测精度影响并不显著。因此,为了保证预测的时效性,选定组合2(包含T₀时刻对应的所有特征)作为排队尾车驶离速度预

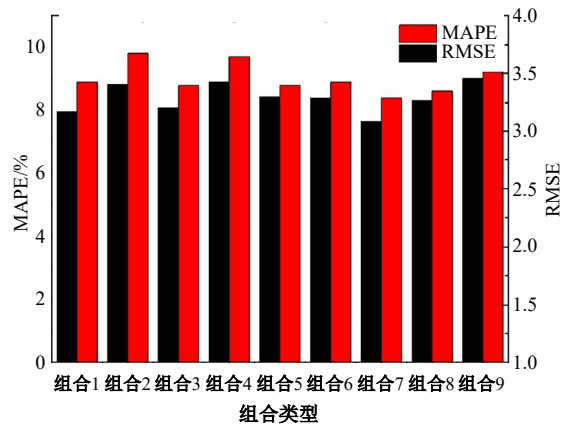


图10 不同特征组合方式下尾车驶离车速预测对比

Fig. 10 Prediction comparison of tail vehicle departure speed under different feature consists

测的最优特征组合。

采用最优特征组合,对排队尾车驶离状态预测模型参数进行迭代优化,最终得到表4所示的最优超参数取值。

表4 模型超参数最优取值

Table 4 Optimal values of main parameters of model

模型主要参数	取值范围	模型参数最优取值	
		排队尾车驶离时间	排队尾车驶离速度
基本学习器的深度	[2,10]	3	8
基本学习器的数量	[100,800]	700	400
学习率	[0.01,0.5]	0.2	0.03
损失减少阈值	[0,1]	0.7	0.1
L ₁ 正则化	[0.0001,100]	1	0.000 1
L ₂ 正则化	[0.0001,100]	10	10
子集最小权重	[2,200]	6	10
样本子采样	[0.6,0.9]	0.7	0.8
列子采样	[0.6,1]	0.8	0.9

4 模型预测性能评估

4.1 对比方法

4.1.1 常用机器学习方法

为了验证基于XGBoost算法的排队尾车驶离状态预测模型性能,选取了随机森林、决策树和多层感知机(MLP)神经网络3种常用的回归预测算法进行对比。采用与4.3.2节所述相同的数据集和最优特征组合分别对3种机器学习算法驱动的排队尾车驶离状态预测模型进行训练。

4.1.2 理论模型

除了其他3种常用机器学习算法以外,重点对比了本文提出的模型与理论模型的预测性能差异。选取了被广泛运用的运动学模型^[15,17]。该模型将排队尾车驶离时间分为两个组成部分:一是绿灯启亮后消散波由头车传播到队尾车辆所需的时间;二是尾车从启动到驶过停车线所需的时间,计算方法如式(6)所示。排队尾车驶离速度理论预测方法如式(7)所示,若尾车以期望加速度匀加速行驶且过停车线时未达到最大期望速度,则尾车驶离速度为 $a_d T$;否则,尾车驶离速度为最大期望速度。

$$T = \begin{cases} \frac{L}{\bar{w}} + \frac{\sqrt{2La_d}}{a_d}, & L \leq \frac{v_{\max}^2}{2a_d} \\ \frac{L}{\bar{w}} + \frac{v_{\max}}{a_d} + \frac{L - v_{\max}^2/(2a_d)}{v_{\max}}, & L > \frac{v_{\max}^2}{2a_d} \end{cases} \quad (6)$$

$$v_T = \begin{cases} a_d T, & a_d T \leq v_{\max} \\ v_{\max}, & a_d T > v_{\max} \end{cases} \quad (7)$$

式中: T 为排队尾车驶离时间; v_T 为排队尾车驶离速度; L 为尾车排队位置距停车线的距离, m; v_{\max} 为排队尾车通过停车线时的最大期望速度, km/h, 实际数据显示排队尾车消散过程中往往无法达到最大限速值(60 km/h), 因此, 本文选取所有样本车辆(583个样本)过停车线速度累积分布曲线中的第 85% 分位速度作为最大期望速度, 取值为 48 km/h; a_d 为期望加速度, 通常取最大加速度的一半, 为更贴近真实情况, 本文采用式(8)计算获得, 其中, q 为排队长度, veh; $a_{d,q}$ 为排队长度为 q 时尾车的平均期望加速度; N 为相同排队长度的样本数; m 为每个样本(尾车)从启动到驶过停车线期间的加速度采样数; a_{ij} 为第 i 个样本在第 j 个采样时刻的瞬时加速度。

$$a_{d,q} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{m} \sum_{j=1}^m a_{ij} \right) \quad (8)$$

除此之外, 式(6)中 \bar{w} 为队列平均消散速率, s/veh, 计算方法如式(9)所示:

$$\bar{w}_q = \frac{1}{N} \sum_{k=1}^N \bar{w}_k \quad (9)$$

式中: \bar{w}_q 为排队长度为 q 时的平均消散速率; \bar{w}_k 为相同排队长度下第 k 个排队样本的平均消散速率。

采用 4.3.2 节所述模型训练数据集, 获得式(6)所述排队尾车消散时间预测理论模型中 a_d 和 \bar{w} 的标定结果, 如表 5 所示。

表 5 理论模型参数标定结果

Table 5 Parameter calibration results of theory model

平均期望加速度/(m·s ⁻²)	标定取值	平均消散速率/(s·veh ⁻¹)	标定取值
$a_{d,5}$	0.691	\bar{w}_5	1.277
$a_{d,6}$	0.647	\bar{w}_6	1.277
$a_{d,7}$	0.607	\bar{w}_7	1.308
$a_{d,8}$	0.594	\bar{w}_8	1.318
$a_{d,9}$	0.538	\bar{w}_9	1.346
$a_{d,10}$	0.496	\bar{w}_{10}	1.356
$a_{d,11}$	0.492	\bar{w}_{11}	1.356
$a_{d,12}$	0.465	\bar{w}_{12}	1.363
$a_{d,13}$	0.464	\bar{w}_{13}	1.388
$a_{d,14}$	0.454	\bar{w}_{14}	1.401
$a_{d,15}$	0.438	\bar{w}_{15}	1.379
$a_{d,16}$	0.434	\bar{w}_{16}	1.342
$a_{d,17}$	0.421	\bar{w}_{17}	1.315
$a_{d,18}$	0.389	\bar{w}_{18}	1.313

4.2 结果分析

由表 6 可以发现, 基于 XGBoost 算法的尾车驶离时间预测模型精度均优于其他 3 种常用的机器学习预测方法(随机森林、决策树和多层感知机), 相比理论模型, MAPE 也有明显降低(接近 10%), RMSE 降低约 4 km·h⁻¹。

此外, 图 11 折线图绘制了本文提出的模型和理论模型的预测值与真实值的分布详情, 横坐标表示测试样本编号, 样本编号越大, 表示排队长度越大。可以看出: 在排队长度较短时, 本文提出的预测模型和理论模型预测误差均比较小, 但随着排队长度的增加, 理论模型预测值与真实值的误差明显增大; 理论模型预测值总体走势较为平缓, 且针对大多数测试样本, 采用理论模型的预测值小于真实值, 表明基于稳态同质交通流的假设建立的运动学模型无法满足变化的车辆排队消散特性。相反, 利用本文提出的模型得到的预测值与真实值变化趋势更加吻合, 重要的是当实际尾车驶离时间出现异常陡增时, 模型同样可以得到较好的预测结果(如样本编号为 58、79 和 82 对应的尾车驶离时间)。

表 6 尾车驶离时间预测模型性能对比

Table 6 Performance comparison of tail vehicle departure time prediction models

模型	MAPE/%	RMSE (km·h ⁻¹)
XGBoost	5.74	1.74
随机森林	10.58	3.75
决策树	12.84	5.01
MLP 神经网络	30.39	7.33
理论模型	14.64	5.78

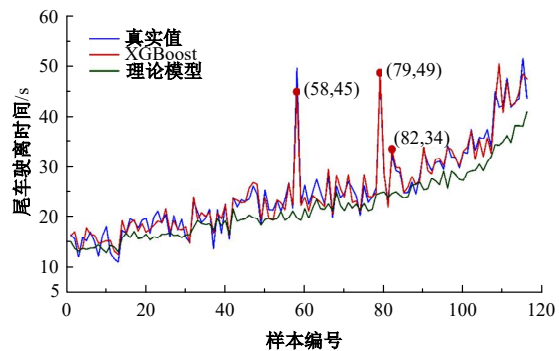


图 11 基于 XGBoost 的模型与理论模型预测的尾车驶离时间对比

Fig. 11 Comparison of departure time of tail vehicle predicted by XGBoost-based model and theoretical model

由表 7 可以发现,基于 XGBoost 的尾车驶离速度预测模型 MAPE 为 9.98%,均优于其他 3 种机器学习方法,但基于决策树与基于 XGBoost 的预测模型精度非常接近。相比理论模型,基于 XGBoost 的预测模型 MAPE 约降低了 6%。各种模型之间 RMSE 差异不大。此外,由图 12 所示模型预测值与真实值的对比可见,理论模型预测值随着排队长度的增加而平缓增大,模型无法适应驶离速度的非线性变化特性。相反,基于 XGBoost 的预测模型却能很好地适应驶离速度的非线性变化特性,且预测值与真实值的速度振荡趋势基本吻合,只有个别样本(如样本编号为 81 的样本)预测值明显高于真实值。

表 7 尾车驶离速度预测模型性能对比

Table 7 Performance comparison of tail vehicle departure speed prediction models

模型	MAPE/%	RMSE (km·h ⁻¹)
XGBoost	9.98	3.48
随机森林	10.06	3.44
决策树	13.34	4.73
MLP 神经网络	13.27	4.45
理论模型	15.78	3.50

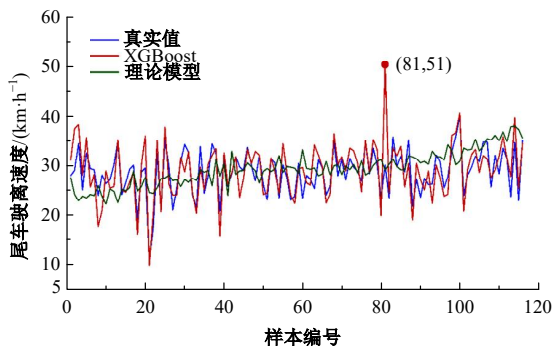


图 12 基于 XGBoost 的模型与理论模型预测的尾车驶离速度对比

Fig. 12 Comparison of departure speed of tail vehicle predicted by XGBoost-based model and theoretical model

分析表明:基于 XGBoost 的排队尾车驶离状态预测模型综合性能最优。相比之下,即使对理论模型中消散波速和期望加速度进行精细化标定,仍会产生较大的预测误差,无法适应排队消散的不确定性。

5 结 论

(1) 构建了由绿灯启亮时刻的排队长度、绿灯启亮时刻排队尾车前方车间距标准差和绿灯启亮

后头车启动车速 3 个特征变量构成的排队尾车驶离时间预测最优特征组合;构建了由绿灯启亮时刻排队长度、尾车前方车队大车占比、尾车前方平均车间距、尾车前方车辆间距标准差 4 个特征变量构成的排队尾车驶离速度预测最优特征组合。

(2) 提出的排队尾车驶离时间预测模型,预测结果绝对百分比误差可达 5.74%,相比于运动学模型预测精度提升约 10%。排队尾车驶离速度预测模型预测结果平均绝对百分比误差可达 9.98%,比理论模型精度提升 6%。本文模型可为相位最小绿灯时间设置和网联车辆生态速度轨迹规划提供技术支撑。

(3) 排队长度对排队尾车驶离时间具有正向作用;头车启动车速和车间距标准差对排队尾车驶离时间具有负效应。排队长度对驶离速度的影响并不显著;前方车辆一致且均匀地启动行为有利于排队尾车获得较快的驶离速度;绿灯启亮时尾车前方车间距分布越离散,越有利于提高排队尾车驶离速度。基于 SHAP 的特征贡献度分析结果有利于深入认识排队消散特性及影响机理,从而辅助制订科学的交通管控策略。本研究数据主要来源于平直道路,未来将采集更多不同类型信号交叉口(譬如有明显纵坡的信号交叉口)的车辆轨迹数据对模型加以验证和改进,以提高模型的普适性和鲁棒性。

参考文献:

[1] 徐洪峰,王殿海. BRT 优先控制交叉口的机动车相位固定最小绿灯时间计算方法[J]. 吉林大学学报:工学版, 2009, 39(增刊 1): 92-97.
 Xu Hong-feng, Wang Dian-hai. Absolute minimum green calculation for vehicle phase at signalized intersections with bus rapid transit signal priority[J]. Journal of Jilin University(Engineering and Technology Edition), 2009, 39(Sup. 1): 92-97.

[2] 刘东波,沈莉潇,代磊磊,等. 基于多目标雷达数据的单点交通信号控制方法[J]. 吉林大学学报:工学版, 2022, 52(10): 2456-2465.
 Liu Dong-bo, Shen Li-xiao, Dai Lei-lei, et al. Traffic signal control method at isolated intersections based on multi-target radar data[J]. Journal of Jilin University(Engineering and Technology Edition), 2022, 52(10): 2456-2465.

[3] 刘显贵,王晖年,洪经纬,等. 网联环境下信号交叉口车速控制策略及优化[J]. 交通运输系统工程与信

- 息. 2021, 21(2): 82-90.
- Liu Xian-gui, Wang Hui-nian, Hong Jing-wei, et al. Traffic speed control strategy and optimization at signalized intersections under connected environment[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2021, 21(2): 82-90.
- [4] 杨澜, 赵祥模, 吴国垣, 等. 智能网联汽车协同生态驾驶策略综述[J]. *交通运输工程学报*, 2020, 20(5): 58-72.
- Yang Lan, Zhao Xiang-mo, Wu Guo-yuan, et al. A review of collaborative eco-driving strategies for intelligent connected vehicles[J]. *Journal of Traffic and Transportation Engineering*, 2020, 20(5): 58-72.
- [5] Luo Q, Yuan J, Chen X, et al. Analyzing start-up time headway distribution characteristics at signalized intersections[J]. *Physica A: Statistical Mechanics and its Applications*, 2019, 535: No. 122348.
- [6] Mondal S, Gupta A. Discharge characteristics analysis of queued-up vehicles at signal-controlled intersections under heterogeneous traffic conditions[J]. *International Journal of Civil Engineering*, 2019, 17(5): 619-628.
- [7] Jin X, Zhang Y, Wang F, et al. Departure headways at signalized intersections: a log-normal distribution model approach[J]. *Transportation Research Part C: Emerging Technologies*, 2009, 17(3): 318-327.
- [8] 唐克双, 董可然, 黄志荣, 等. 信号交叉口排队消散特性实证对比[J]. *同济大学学报: 自然科学版*, 2015, 43(11): 1689-1695.
- Tang Ke-shuang, Dong Ke-ran, Huang Zhi-rong, et al. Empirical comparison of queuing dissipation characteristics at signalized intersections[J]. *Journal of Tongji University(Natural Science)*, 2015, 43(11): 1689-1695.
- [9] Zhan X, Li R, Ukkusuri S V. Lane-based real-time queue length estimation using license plate recognition data[J]. *Transportation Research Part C: Emerging Technologies*, 2015, 57: 85-102.
- [10] Mei Y, Gu W, Chung E C S, et al. A Bayesian approach for estimating vehicle queue lengths at signalized intersections using probe vehicle data[J]. *Transportation Research Part C: Emerging Technologies*, 2019, 109: 233-249.
- [11] 谈超鹏, 姚佳蓉, 唐克双. 基于抽样车辆轨迹数据的信号控制交叉口排队长度分布估计[J]. *中国公路学报*. 2021, 34(11): 282-295.
- Tan Chao-peng, Yao Jia-rong, Tang Ke-shuang. Queue length distribution estimation at signalized intersections based on sampled vehicle trajectory data[J]. *China Journal of Highway and Transport*, 2021, 34(11): 282-295.
- [12] He X, Liu H X, Liu X. Optimal vehicle speed trajectory on a signalized arterial with consideration of queue[J]. *Transportation Research Part C: Emerging Technologies*, 2015, 61: 106-120.
- [13] Wu L, Ci Y, Wang Y, et al. Fuel consumption at the oversaturated signalized intersection considering queue effects: a case study in Harbin, China[J]. *Energy*, 2020, 192: No. 116654.
- [14] Yang H, Rakha H, Ala M V. Eco-cooperative adaptive cruise control at signalized intersections considering queue effects[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(6): 1575-1585.
- [15] Dong H, Zhuang W, Chen B, et al. Enhanced eco-approach control of connected electric vehicles at signalized intersection with queue discharge prediction[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(6): 5457-5469.
- [16] Dong H, Zhuang W, Yin G, et al. Energy-optimal velocity planning for connected electric vehicles at signalized intersection with queue prediction[C]//*IEEE ASME International Conference on Advanced Intelligent Mechatronics(AIM)*, Boston, USA, 2020: 238-243.
- [17] Ye F, Hao P, Qi X, et al. Prediction-based eco-approach and departure at signalized intersections with speed forecasting on preceding vehicles[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(4): 1378-1389.
- [18] 徐洪峰, 郑明明, 李克平. 机动车相位固定最小绿灯时间计算方法[J]. *公路交通科技*, 2008(5): 105-110.
- Xu Hong-Feng, Zheng Ming-ming, Li Ke-ping. Absolute minimum green time calculation for vehicle phase[J]. *Journal of Highway and Transportation Research and Development*, 2008(5): 105-110.
- [19] Elefteriadou L. The highway capacity manual 6th edition: a guide for multimodal mobility analysis[J]. *Institute of Transportation Engineers*. 2016, 86(4): 14-18.
- [20] 罗小芹, 王殿海, 金盛. 面向混合交通的感应式交通信号控制方法[J]. *吉林大学学报: 工学版*, 2019, 49(3): 695-704.
- Luo Xiao-qin, Wang Dian-hai, Jin Sheng. Traffic signal actuated control at isolated intersections for het-

- erogeneous traffic[J]. Journal of Jilin University (Engineering and Technology Edition), 2019, 49(3): 695-704.
- [21] Shi X P, Wong Y D, Li M Z F, et al. A feature learning approach based on XGBoost for driving assessment and risk prediction[J]. Accident Analysis & Prevention, 2019, 129: 170-179.
- [22] Shi R, Xu X, Li J, et al. Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization[J]. Applied Soft Computing, 2021, 109: No. 107538.
- [23] 魏田正, 魏雯, 李海梅, 等. 基于 XGBoost 算法的危险场景驾驶行为模式分析及安全评估[J]. 交通信息与安全, 2022, 40(5): 53-60.
- Wei Tian-zheng, Wei Wen, Li Hai-mei, et al. An analysis of driving behavior model and safety assessment under risky scenarios based on an XGBoost algorithm[J]. Journal of Transport Information and Safety, 2022, 40(5): 53-60.
- [24] 赵晓华, 亓航, 姚莹, 等. 基于可解释机器学习框架的快速路立交出口风险预测及致因解析[J]. 东南大学学报:自然科学版, 2022, 52(1): 152-161.
- Zhao Xiao-hua, Qi Hang, Yao Ying, et al. Risk prediction and causation analysis of expressway interchange exits based on interpretable machine learning framework[J]. Journal of Southeast University(Natural Science Edition), 2022, 52(1): 152-161.
- [25] Roger E, Torlay L, Gardette J, et al. A machine learning approach to explore cognitive signatures in patients with temporo-mesial epilepsy[J]. Neuropsychologia, 2020, 142: No. 107455.
- [26] 陈秀锋, 田家斌, 石英杰, 等. 基于排队消散模型的干线协调控制[J]. 科学技术与工程, 2018, 18(10): 279-283.
- Chen Xiu-feng, Tian Jia-bin, Shi Ying-jie, et al. Arterial coordination control based on queuing dissipation model[J]. Science Technology and Engineering, 2018, 18(10): 279-283.
- [27] 赵巍, 徐汉清. 长周期倒计时信号控制对排队消散特性的影响[J]. 城市交通, 2016, 14(6): 67-74.
- Zhao Wei, Xu Han-qing. Impacts of signal with long cycle and duration countdown on queue discharge at signalized intersections[J]. Urban Transport of China, 2016, 14(6): 67-74.
- [28] 王殿海, 郭佳林, 蔡正义. 基于自动车牌识别数据的混合交通流饱和和流率实时估计[J]. 交通运输系统工程与信息, 2021, 21(2): 37-43.
- Wang Dian-hai, Guo Jia-lin, Cai Zheng-yi. Real-time estimation of saturated flow rate of mixed traffic flow based on automatic license plate recognition data [J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(2): 37-43.