

# 低渗透率智能网联环境下高风险事件预警方法

陈晓芸, 叶颖俊, 余荣杰, 孙 剑

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

**摘要:** 提出一种低渗透率智能网联环境下高风险事件预警方法。具体而言, 基于熵能表征系统状态的特点提出交通熵的概念, 将个体车辆的微观驾驶行为量化为交通熵, 以表征交通流状态; 再将交通熵作为长短时记忆网络模型(Long Short-term Memory, LSTM)的输入参数建立预警模型; 最后, 使用HighD轨迹数据集提取高风险事件, 并验证模型有效性。结果显示, 使用交通熵的模型误报率和漏报率大幅降低。以智能车渗透率 10% 为例, 误报率和漏报率分别从 6.18% 和 11.47% 下降到了 1.95% 和 3.12%; 在预测模式下, 对高风险事件误报率和漏报率为 2.28% 和 3.82%。

**关键词:** 交通安全; 智能网联交通; 高风险事件; 交通熵; 预警模型; 低渗透率

中图分类号: U491

文献标志码: A

## Early Warning Methods for Traffic High-risk Events Under Low Penetration of Connected and Autonomous Vehicles

CHEN Xiaoyun, YE Yingjun, YU Rongjie, SUN Jian

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

**Abstract:** We propose an early warning method for high-risk events of traffic operation under low penetration of connected and autonomous vehicles(CAVs). Specifically, we first define the concept of traffic entropy, and quantifies the micro driving behavior of individual vehicles as a parameter represented by traffic entropy, which is used to characterize the state of macroscopic traffic flow. And then the traffic entropy is used as the input parameter of the Long Short-Term Memory (LSTM) model to establish the early warning model of high-risk events. The HighD Dataset from

German highways was utilized for the empirical analyses. In order to compare the application results under CAVs environment, an autonomous-vehicles scenario and a connected-vehicles scenario were set for the high-risk events and non-risk events extracted from the HighD Dataset. and the effectiveness of the warning of high-risk events under different vehicle permeability was compared. Results show that, the false alarm and missed alarm rates of early warning model with traffic entropy parameters are both reduced. Taking the low-penetration CAVs of 10% as an example, the false alarm and missed alarm rates reduced from 6.18% and 11.47% to 1.95% and 3.12%, respectively. At the same time, the false alarm and missed alarm rates are only 2.28% and 3.82% under the prediction environment.

**Key words:** traffic safety; connected and autonomous vehicles(CAVs); high-risk event; traffic entropy; early-warning model; low penetration

在道路交通运行过程中, 驾驶人操控行为、交互行为不当等高风险事件极易引发道路交通事故<sup>[1]</sup>并导致交通拥堵等负面效应<sup>[2]</sup>。这些高风险事件的发生具有动态性、随机性、瞬时性的特点, 提前预警高风险事件可为驾驶员预留执行规避动作的反应时间, 有效降低交通运行风险<sup>[3-5]</sup>。传统交通系统状态分析多为基于路面检测线圈<sup>[68]</sup>采集流量、平均速度、密度等宏观交通流状态参数, 由于采集数据空间密度低、时间集计, 缺乏微观交通流特征与交通运行风险之间的理论支撑<sup>[4]</sup>, 导致高风险事件预警的漏报率和误报率较高, 亦不能精确定位异常行为发生的时空位置<sup>[4,9]</sup>。为此, 搭建连接宏观交通流状态与微

收稿日期: 2022-05-05

基金项目: 国家重点研发计划(2018YFB1600505); 国家自然科学基金重点项目(52125208); 浙江省重点研发计划(2021C01011)

第一作者: 陈晓芸(1993—), 女, 工学博士, 主要研究方向为交通运行建模与仿真、驾驶行为研究。

E-mail: 1610753@tongji.edu.cn

通信作者: 孙剑(1979—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为交通流理论与仿真, 智能网联汽车与车路协同。E-mail: sunjian@tongji.edu.cn



论文  
拓展  
介绍

观交通行为的桥梁,探究个体车辆的微观驾驶行为与交通流状态表征之间的关系,将成为精准预警高风险事件的破题之举。

随着智能网联汽车技术的发展,人类将逐步进入智能网联交通时代。不同于线圈采集的粗粒度状态参数,网联车辆可提供高分辨率的个体车辆轨迹数据<sup>[1011]</sup>,而搭载各类传感器设备的智能车辆,可以实时获取本车及周围车运动状态参数。但目前距离完全进入智能网联汽车时代尚远,低渗透率的智能网联车辆将是当前及未来交通的主要组成形态。

因此,针对智能网联车辆逐步渗入交通系统的现实,本文提出一种低渗透率智能网联环境下交通运行高风险事件预警方法。主要创新:提出了一种刻画微观个体行为与系统宏观状态的新指标——交通熵,通过交通熵量化微观个体车辆的驾驶行为,建立起个体车辆微观行为与交通流宏观状态的映射关系,进而有效表征交通流状态和紊乱程度。设计了基于长短期记忆网络(Long Short-Term Memory, LSTM)的深度学习模型作为高风险事件预警模型,通过滑动时空窗模式输入时序参数,利用有效的交通熵参数即可实现对高风险事件的提前预警。针对不同智能网联环境下数据采集和信息获取的差异,设置智能车场景和网联车场景对比分析,并利用从HighD轨迹数据集中提取的256个高风险事件和1024个常态交通流事件,分别测试使用交通熵参数与不使用交通熵参数的预警模型在低渗透率智能网联环境下的应用效果。

## 1 研究综述

目前,道路交通安全研究领域关注的高风险事件,通常包括接近碰撞事件(near-crashes)和碰撞事件(crashes,即事故)两大类。现有交通事故数据库(如GIDAS, German in-depth accident study)仅记录事故对象的有限轨迹,难以精确还原事故演化过程、实现准确预测<sup>[9,12]</sup>。而接近碰撞事件从车辆轨迹数据或者自然驾驶数据中获取,包含的前兆数据充足,对于风险演化过程有更强的可解释性和预测能力<sup>[1,3,12]</sup>。因此,本文所研究的高风险事件主要聚焦于接近碰撞事件及其发生前兆。

现有研究中,高风险事件的识别指标主要有两类:使用车辆动力学参数,设定阈值。该方法仅需本车的动力学参数(如横纵向加速度等),但误报率较高<sup>[3,13]</sup>。进而有研究通过识别加速度的变化模式,采

用分类算法<sup>[14]</sup>、随机森林模型和支持向量机模型筛选<sup>[15]</sup>等改进算法。结合道路环境和车辆交互信息,使用替代指标。主要分为时间指标,如碰撞时间<sup>[16-17]</sup>(time to collision, TTC)、后侵入时间<sup>[18]</sup>和非时间指标,如避免碰撞减速率<sup>[18]</sup>、加加速度。这类指标考虑了周围车辆的交互信息,识别精度较高。其中,最常用的TTC假设车辆在发生碰撞之前保持恒定速度,当后车速度低于前车时便不能应用。为了解决该问题,修正的碰撞时间(modified time to collision, MTTC)<sup>[19]</sup>考虑了加速度对碰撞事件的影响,可识别纵向、横向等多种冲突类型<sup>[20]</sup>。与TTC类似,MTTC的数值越小,表明当前交通流状态中可能存在的风险程度越高。

高风险事件预警研究背后隐含的假设条件是,某些交通条件组合更容易导致交通运行中的高风险事件<sup>[4,9]</sup>,因此探究高风险事件发生前兆的交通流运行特征、分析高风险事件的影响因素是建立高风险事件预警模型的关键。常用的预警模型有广义线性回归模型<sup>[21-22]</sup>和机器学习模型等<sup>[23-24]</sup>两大类。其中,车辆速度是最常用的影响参数,车头时距、车头间距等也被使用<sup>[25]</sup>。现有研究大多使用传统线圈数据,受其精度不高、数据离散等限制,多侧重探究宏观交通特征的影响,缺少从车辆运行中微观层面开展影响因素分析方法。同时,统计学方法虽然具有理论基础,但无法处理相关变量<sup>[4]</sup>;基于数据挖掘的方法能够处理大量相关变量,但盲目输入大量参数可能会出现难以理解、甚至“虚假”的结果<sup>[9]</sup>。

## 2 交通熵定义与计算方法

交通熵的提出源于熵的基本概念,经过交叉熵的演变将其抽象为一个描述系统状态的指标,进而引入到交通系统中。熵最初是用以表征热力学中物质状态的物理量<sup>[26]</sup>,而后研究从分子运动论的角度,建立了系统宏观状态的熵与对应微观态数目的关系<sup>[27]</sup>,将熵视为体系混乱程度的度量。信息论之父香农认为<sup>[28]</sup>,通信过程中信息源的信号的不确定性称为信息熵,而消除了多少不确定性称为信息,其数学表达式为

$$S(p_i) = \sum_{i=1}^n p_i \log \frac{1}{p_i} \quad (1)$$

在信息论中,交叉熵主要用于度量两个概率分布间的差异性信息,计算公式为

$$S(p_i, q_i) = \sum_{i=1}^n p_i \cdot \log \frac{1}{q_i} \quad (2)$$

式中:  $p_i$  为真实分布概率;  $q_i$  为预测分布概率;  $n$  为信息中符号的数量。  $p_i$  表示真实标记的分布,  $q_i$  为训练后模型的预测标记分布。

在交通流系统内,不同车辆的微观驾驶行为、车辆之间的交互行为中都同样存在着大量的车辆行为和人类驾驶行为参数、以及概率分布特征,基于对于熵基本概念的认识,将熵作为交通流中概率分布特征的函数引入到交通系统中,用以量化表征个体车辆微观驾驶行为的异常程度,以及交通流系统状态的混乱或无序程度。对于某一具体的驾驶行为参数  $x$ ,假设其出现的概率越低行为异常程度越高,在交通流中造成系统的混乱度或无序性越高,表现为交通熵值越大。对  $x$  建立累积概率分布函数,  $b_i$  是在时刻  $i$  下的行为异常程度概率值,称为行为概率函数。其中,  $x_i$  是时刻  $i$  下的驾驶行为值。

$$b_i(x) = p(x \leq x_i) \quad (3)$$

图 1 为不同驾驶行为的行为概率函数示意图,无论累积分布形式如何变化,行为概率函数值的大小均可表示驾驶行为出现的异常程度。本研究中的交通熵,默认驾驶行为参数  $x$  的数值越小,其行为的异常程度越高(如速度、车头间距等);如果遇到某驾驶行为参数  $x$  的数值越大,其异常程度越高的情况,则将其行为概率函数的公式改为  $b_i = p(x \geq x_i)$ 。行为概率函数既适用于整体交通流中的驾驶行为,又适用于不同驾驶员异质性的驾驶行为。当行为概率函数是针对所有驾驶员建立的,则可表示交通流系统中某一驾驶行为出现的概率。而不同驾驶员之间驾驶行为的异质性则可通过分别统计不同驾驶员的累积概率分布实现。

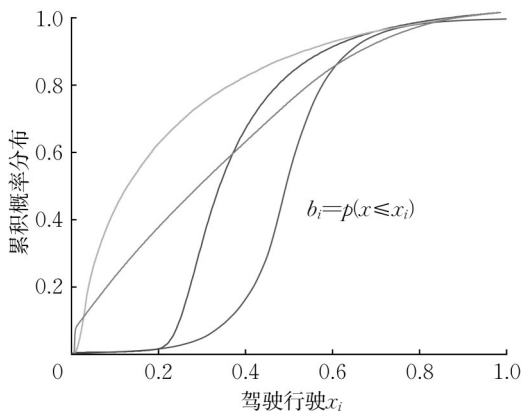


图 1 不同行为概率函数示意

Fig.1 Diagram of different behavior probability functions

通过比较行为概率函数  $b_i$  与整体分布中其他行为的概率  $(1 - b_i)$ ,即分别用  $(1 - b_i)$  和  $b_i$  替代交叉熵定义中的  $p_i$  和  $q_i$ ,量化驾驶行为的异常程度。其可能的表达式为

$$H(b_i, x) = (1 - b_i) \log \frac{1}{b_i} \quad (4)$$

$$H(b_i, x) = b_i \log \frac{1}{(1 - b_i)} \quad (5)$$

对比两式和示意图 2 可以看出,采用式(4)计算时  $b_i$  越低交通熵值越高,符合交通熵定义的内涵,即某行为参数出现的概率越低其行为的异常程度越高,进而在交通流中造成系统的混乱度或无序性越高,表现为交通熵值越大,对应“低熵有序,高熵无序”的概念,而“交通熵增”也表征着车辆行为异常程度增加与交通流系统紊乱程度增加。

交通熵通过比较个体车辆的微观行为与整体分布量化异常程度,表征宏观交通流状态。图 2a 中可看出交通熵具有以下特性:驾驶行为的异常程度始终为正值,即  $H \geq 0$ ;  $x$  越低  $b_i$  值越小,行为异常程度越高,交通流状态越混乱无序时,交通熵值越大,即交通熵随行为概率函数在  $(0, 1]$  区间单调递减;当  $b_i \rightarrow 0$  时,行为的异常程度最高,交通流状态的混乱程度为  $H \rightarrow +\infty$ ;当  $b_i = 1$  时,驾驶行为是无异常,此时交通流高度有序  $H = 0$ ;在连续驾驶过程中,每一时刻对应驾驶行为值,因此交通熵具有时间连续性。

由于微观行为具有随机性,且多个个体车辆之间的行为交互也易导致交通流紊乱引发质变,因此交通熵的提出恰好可作为联系个体车辆微观驾驶行为、多车交互行为与交通流整体宏观现象的统一量纲。根据研究对象的不同,其对于行为异常程度和状态的量化表征可以涵盖车辆动力学参数、驾驶行为参数、驾驶员姿态动作参数等。就其应用范围而言,交通熵既可以量化微观驾驶行为,识别个体车辆的异常驾驶状况,也可以通过量化交通流内车辆微观行为,进一步表征整体交通流的状态。

### 3 高风险事件预警方法

#### 3.1 高风险事件预警框架

低渗透率智能网联环境下的交通运行高风险事件预警模型的方法框架如图 3 所示,分为信息采集、交通熵值计算、高风险事件预警模型 3 部分。

在信息采集部分,结合智能网联环境,采用滑动时空窗的方式采集交通流中相关车辆的信息。滑动

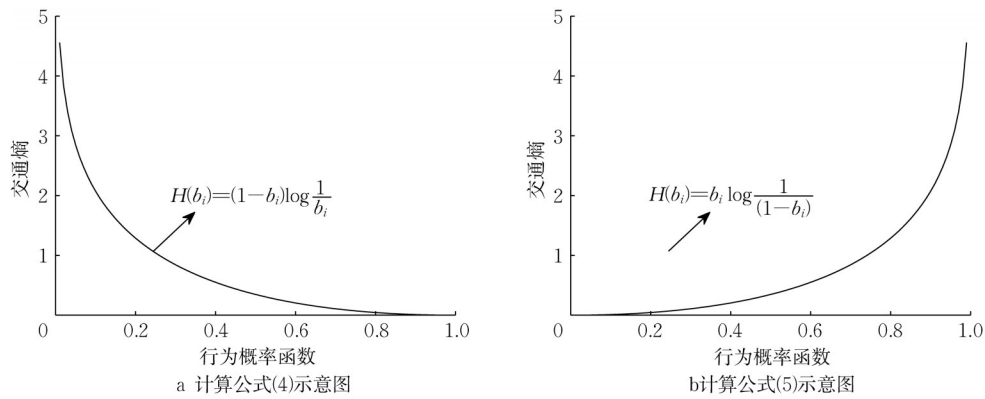


图2 交通熵函数示意图

Fig.2 Diagram of Traffic entropy

时空窗的时间长度为 $s_t$ ,空间宽度为所在路段宽度 $s_l$  (一般取左右相邻车道),信息采集范围是滑动时空窗内的智能车辆或网联车辆,滑动时空窗的滑动步长为 $w$ 。以时刻 $t$ 为例,采集目标车辆从时刻 $(t - s_t)$

到时刻 $t$ 时间内行驶过的路段范围内的所有智能车辆或网联车辆的信息。由于实际应用中智能网联车辆的车辆渗透率较低,本文设定的低渗透率环境为5%和10%。

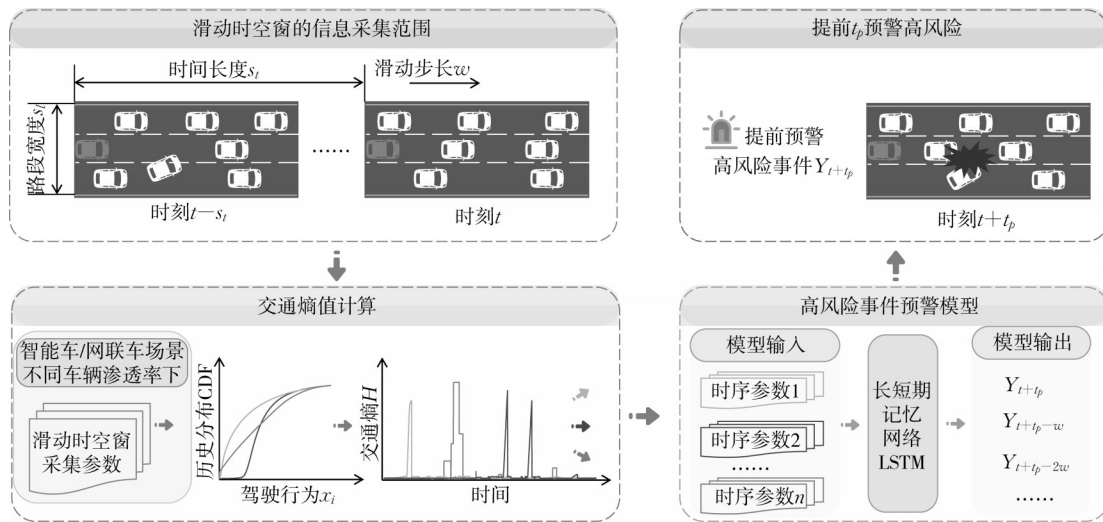


图3 高风险事件预警的方法框架

Fig.3 The framework of high-risk events early warning

在交通熵值计算部分,根据滑动时空窗内采集到的参数和预先建立起的驾驶行为历史分布,计算得到各个参数的对应的交通熵值。

在高风险事件预警模型部分,选用长短期记忆网络LSTM<sup>[29]</sup>对高风险事件进行时间序列预测,模型输入为基于滑动时间窗的计算得到的时序参数,模型的输出为是否是高风险事件。由于LSTM对有价值的信息进行长期记忆,因此在许多交通预测类问题中得到应用<sup>[30-33]</sup>。在本研究中,模型参数经过交通熵量化后,已可以表征微观驾驶行为与交通流的状态,因此采用基础的LSTM模型结构即可快速有效地满足高风险事件的预警需求。

### 3.2 数据来源与高风险事件提取

本文采用2019年德国高速公路无人机摄影技术采集的自然车辆轨迹数据——HighD数据集(highway drone dataset)<sup>[34]</sup>作为研究数据。该数据集是免费、非商业目的的开放数据集,已广泛应用于交通仿真建模和驾驶行为分析<sup>[21]</sup>。记录了德国六条不同高速公路上的交通运行情况,采集110 500辆车44 500km行驶里程内的相关交通参数。数据采集时间为工作日早8:00至晚19:00之间,并划分为60段时长17min,路段长度约为420m的数据。采集的数据精度为25 Hz的帧频,包含车辆的轨迹数据(横纵向位置信息、速度和加速度等)及其物理特征(车辆

的宽度和长度)。

通过对现有高风险事件识别方法与指标的研究综述,结合车辆运行轨迹数据特征,本研究使用修正的碰撞时间(modified time to collision, MTTC)<sup>[27]</sup>识别潜在冲突概率,如图4所示,计算如下:

$$t_1 = \frac{-\Delta v - \sqrt{\Delta v^2 + 2\Delta a D}}{\Delta a} \quad (6)$$

$$t_2 = \frac{-\Delta v + \sqrt{\Delta v^2 + 2\Delta a D}}{\Delta a} \quad (7)$$

$$m_{tc} = \begin{cases} \min(t_1, t_2), & \text{若 } t_1 > 0, t_2 > 0 \\ \max(t_1, t_2), & \text{若 } t_1 t_2 > 0 \\ D/\Delta v, & \text{若 } \Delta a = 0 \end{cases} \quad (8)$$

$$\Delta v = v_f - v_p, \Delta a = a_f - a_p, D = x_p - x_f - l_p \quad (9)$$

式中: $v_f$ 指后车速度; $a_f$ 指后车加速度; $x_f$ 指后车位置坐标; $v_p$ 指前车速度; $a_p$ 指前车加速度; $x_p$ 指前车位置坐标; $l_p$ 指前车长度。

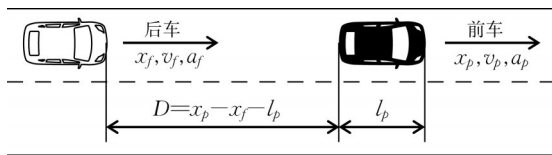


图4 MTTC 计算示意  
Fig.4 Diagram of MTTC

目前不同研究和适用场景中 MTTC 识别高风险事件的阈值范围尚未达成一致的标准。为合理选取阈值,将相关研究中所用阈值汇总分析于表1。考虑到高速公路高速行驶场景下驾驶员反应差异,在本团队针对 HighD 轨迹数据集的研究<sup>[21]</sup>中,以 MTTC 小于 2 s 作为高风险事件判定标准<sup>[34]</sup>可有效支撑高风险事件相关研究,因此本研究中沿用该阈值作为交通运行高风险事件的判定标准。

高风险事件提取的时空范围以高风险事件发生作为零时刻,向前提取 5 s 作为前兆时间;以高风险车辆尾部为基准,向前提取其前方直至路段尽头的各车辆运行轨迹。对于在时空范围内有重叠的高风险事件,选取空间位置最靠近道路起点、发生时间最早的车辆作为该高风险事件的提取合并样本。通过 MTTC 的阈值判断和时空提取范围的初步合并,

表1 各类研究中TTC及MTTC 阈值汇总表

Tab.1 TTC and MTTC thresholds in various studies

文献研究	最小值/s	期望值/s	适用场景
[16]	1	1.5	交叉口
[35]	1.6	2.0	低密度交通流
	0	0.9	高密度交通流
[36]	—	3.5	未安装安全警告系统的驾驶员
	—	2.6	安装安全警告系统的驾驶员
[17]	1	2	交叉口
[37]	—	1.6	交叉口
[38]	2	4	城市隧道(追尾事故)
[20]	—	4	高速公路(追尾、侧碰事故)
[21]	2	4	高速公路

从 HighD 轨迹数据集的 60 个路段中筛选出 439 个高风险事件。而后,基于高风险事件提取的时空范围,剔除持续时间(即高风险零时刻减去车辆开始出现时刻)低于 5 s 的 147 个事件。本文高速公路限速为  $120 \text{ km}\cdot\text{h}^{-1}$ ,安全车距为 50 m,因此需将前向距离不足 50 m 的高风险事件合并,剔除前向距离低于 50 m 的 37 个事件。经过上述处理后,最终得到符合条件的 256 个高风险事件。

为对比常态交通流与高风险事件下的交通状态,还需要提取常态交通流事件作为对照组,为便于区分将其命名为非风险事件。将是流行病学研究中经典的病例对照法,引入交通领域的事故风险预测建模研究<sup>[39]</sup>,按照 1:4 的常用比例,以分层随机抽样的方式确定非风险事件,提取流程如下:

(1) 设定对照组中包含的非风险事件个数。

(2) 设定非高风险事件提取范围。为消除已经提取出的高风险事件的影响,假定在正常交通流中高风险事件的持续影响为 20 s,因此非风险事件的提取首先要剔除已有高风险事件前后 20 s 的数据。

(3) 分层随机抽样选取对照组。为保证对照组能够代表整体交通流中 MTTC 的分布情况,根据 MTTC 分布选取非风险事件。基于 MTTC 分布特征将对照组车辆分为 4 层,并根据每层的车辆数占比进行随机抽取对照组车辆样本,分层随机抽样情况如表 2 所示。同样,基于对照组中的 1 024 个非高风险事件,以非风险事件发生的零时刻向前提取 5 s 前兆时间,以车辆尾部为基准提取其前方至路段尽头的车辆运行轨迹。

表2 对照组中非风险事件的分层随机抽样数

Tab.2 Stratified random sampling of non- risk events

分层区间	$2 < m_{tc} \leq 4$	$4 < m_{tc} \leq 10$	$10 < m_{tc} \leq 16$	$16 < m_{tc} \leq 200$	总计
车辆数	2 389	20 717	27 182	51 769	102 057
占比/%	2.34	20.30	26.63	50.73	100
随机抽取样本数	23	207	272	519	1 024

## 4 预警模型的应用

### 4.1 模型输入与场景设置

本文HighD轨迹数据集的数据精度为25 Hz,将滑动时空窗的时间长度 $s_t$ 设定为25帧,即目标车辆在1 s时间内行驶过的路段长度;空间宽度为路段宽度。滑动时空窗滑动步长为 $w = 1$ 帧。因此,在高风险事件或非风险事件发生前的5 s内共有125个数据采集点,可以得到101个滑动时空窗。提取全部1 024个非风险事件的3个微观参数(速度、加/减速度、车头间距),建立概率密度函数(probability density function, PDF)和累积分布函数(cumulative distribution function, CDF)如图5所示,并将非风险事件3个参数的累积分布函数作为后续熵值计算中的历史分布。

鉴于智能网联环境的实际应用,设定5%和10%两种低渗透率场景,根据不同车辆渗透率随机选取滑动时空窗内对应数量的车辆,作为可被感知的智能车辆或网联车辆。根据256个高风险事件和1 024个非风险事件中,每个滑动时空窗中的对应微观参数在历史分布中的占位值得到式(3)中行为概率值 $b_i$ ,再带入式(4)得到该滑动时空窗中的交通熵值。此外,分别统计计算3个微观参数和其交通熵值的最大/小值、均值、标准差、变异系数这4项统计值,共计28个参数,参数名与其对应含义如表3所示。

为验证评价交通熵参数在不同智能网联环境下数据采集和信息获取的差异、及其应用效果,分别设计智能车场景和网联车场景。其中,智能车场景是针对可以主动发出雷达等检测信号,从交通流中获取其他车辆信息的情况;网联车场景针对现有普通车辆,

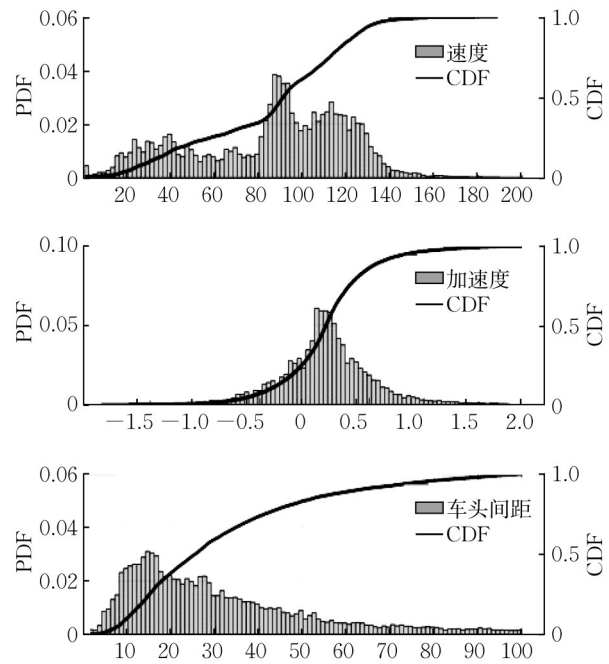


图5 非风险事件概率密度函数与累积分布函数  
Fig.5 The PDF and CDF of non-risk events

通过车载网联设备、手机移动端网联设备上传本车相关信息的情况。即智能车场景中可以使用车头间距等信息,网联车场景中只能使用网联车速度、加速度信息。在每类场景中选取3种类型的参数组合作为模型的输入参数:仅使用交通熵参数;混合使用交通熵参数和非交通熵参数;仅使用非交通熵的参数。

关键变量筛选过程如图6所示,通过Pearson相关系数检验和随机森林算法的重要度排序,初步筛选28个参数与高风险事件的关系,依照结果综合确定各类场景下的具体使用参数,优先选取重要度排序高的参

表3 参数名与含义

Tab.3 Nomenclature of the variables

参数含义	参数的统计值		参数转化为交通熵后统计值	
滑动时空窗内 速度	最大值	MaxV	最大值	TeMaxV
	均值	MeanV	均值	TeMeanV
	标准差	SdV	标准差	TeSdV
	变异系数	CvV	变异系数	TeCvV
滑动时空窗内 加/减速度	加速度最大值	MaxACC	加/减速度最大值	TeMaxAD
	减速度最大值	MaxDEC		
	加速度均值	MeanACC	加/减速度均值	TeMeanAD
	减速度均值	MeanDEC		
	加速度标准差	SdACC	加/减速度标准差	TeSdAD
	减速度标准差	SdDEC		
加速度变异系数	CvACC	加/减速度变异系数	TeCvAD	
减速度变异系数	CvDEC			
滑动时空窗内 车头时距	最大值	MaxDHW	最大值	TeMaxDHW
	均值	MeanV	均值	TeMeanDHW
	标准差	SdDHW	标准差	TeSdDHW
	变异系数	CvDHW	变异系数	TeCvDHW

数,并对于相关性较强参数进行合理剔除。具体场景中选取的参数如表 4 所示,其中网联车场景仅使用交通熵参数的情况时参数个数不足 6 个,因此只测试混合使用和仅使用非交通熵参数两种情况。

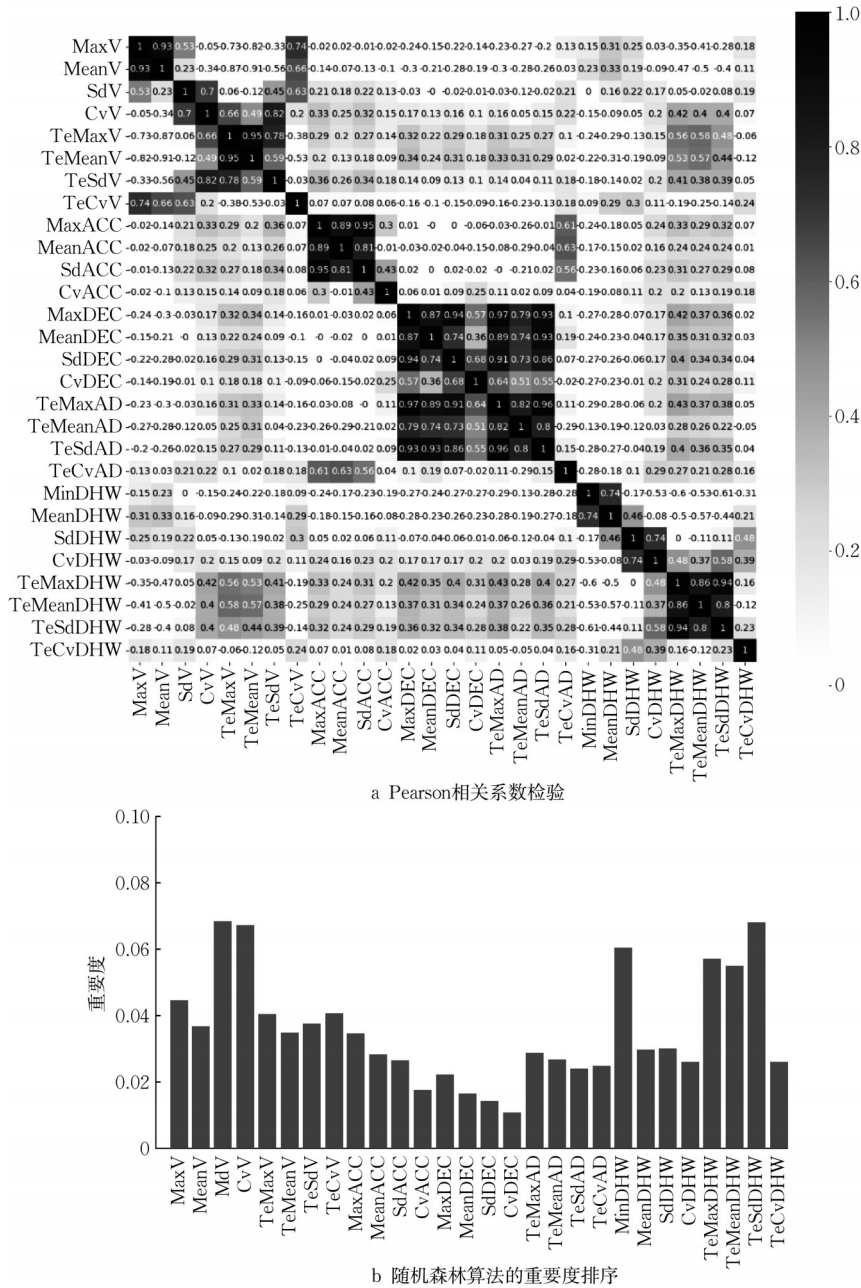


图 6 关键变量筛选

Fig.6 Key parameters selection

4.2 提前预警时间

为测试高风险事件预警模型在不同输入数据量和不同预警时间下的效果,分别选用不同的时间窗长度(1-5s 时间内)和不同的预警时间(提前 0-4s 预警)进行测试,共 15 组数据量与预警时间的组合形式,如表 5 所示。其中,高风险事件发生时刻为 0,以 [-5, -3) 为例,其表示输入模型的是高风险事件发生前 5 s 到前 3 s 内共计 2 s 时间窗内的数据。

5 预警模型结果分析

高风险事件预警模型选用 python3.7 中 Keras<sup>[40]</sup>工具箱的长短期记忆网络 LSTM 对高风险事件进行时间序列预测,模型输入为基于滑动时间窗的计算得到的时序参数,模型的输出为是否是高风险事件。由于高风险事件和非风险事件样本数差异较大,使用 Focal Loss 函数<sup>[41]</sup>修正损失函数,使其在样本比例

表 4 不同场景中选用的模型输入参数

Tab.4 Model input parameters selected in different scenarios

场景类型	参数类型	具体参数
智能车场景 6 参数	仅使用交通熵参数	TeMaxV, TeCvV, TeMaxAD, TeCvAD, TeSdDHW, TeCvDHW
	混合使用交通熵和非交通熵参数	SdV, TeMaxV, MaxACC, TeMaxAD, MinDHW, TeSdDHW
	仅使用非交通熵的参数	SdV, MaxV, MaxACC, MaxDEC, MinDHW, SdDHW
网联车场景 6 参数	仅使用交通熵参数	参数个数不足
	混合使用交通熵和非交通熵参数	MaxV, SdV, TeMaxV, MaxACC, MaxDEC, TeMaxAD
	仅使用非交通熵的参数	MaxV, SdV, CvV, MaxACC, CvACC, MaxDEC
网联车场景 2 参数	仅使用交通熵参数	TeMaxV, TeCvV
	混合使用交通熵和非交通熵参数	SdV, TeMaxV
	仅使用非交通熵的参数	SdV, MaxV

表 5 不同数据量与提前预警时间的组合形式

Tab.5 Combinations of used data and early warning time

输入长度 预警时间	1 s 时间窗	2 s 时间窗	3 s 时间窗	4 s 时间窗	5 s 时间窗
提前 0 s	[-1, 0)	[-2, 0)	[-3, 0)	[-4, 0)	[-5, 0)
提前 1 s	[-2, -1)	[-3, -1)	[-4, -1)	[-5, -1)	-
提前 2 s	[-3, -2)	[-4, -2)	[-5, -2)	-	-
提前 3 s	[-4, -3)	[-5, -3)	-	-	-
提前 4 s	[-5, -4)	-	-	-	-

不均等的情况下适配二分类问题的训练过程,并保证训练结果的有效性。其余的 LSTM 模型参数设置

表 6 LSTM 参数设置

Tab.6 Parameters in LSTM

参数含义	输入层 维度	输出层 维度	批次大小	隐藏层 层数	隐藏层中的 神经元个数	随机失活 层比例	损失函数	激活函数	优化器	训练轮数
取值	6/2	1	32	2	100	0.2	$\gamma=2, \alpha=0.25$	Sigmoid	Adam	300

由于本文提出的高风险事件预警模型在整体准确率上已经能够达到较高的精度,为更直观展示预警模型误报率和漏报率,模型结果采用混淆矩阵的方式表示,其中,  $T_p$  (true positive) 表示真阳性;  $F_p$  (false positive) 表示假阳性;  $F_n$  (false negative) 表示假阴性;  $T_n$  (True Negative) 表示真阴性。模型的精度即阳性预测值(positive predictive value, PPV) 计算如下:

$$\epsilon_{ppv} = \frac{T_p}{T_p + F_p} \quad (10)$$

模型的敏感即真阳性率(true positive rate, TPR)的计算公式为

$$\epsilon_{tpr} = \frac{T_p}{T_p + F_n} \quad (11)$$

对于高风险事件预警的误报率为

$$\text{误报率} = 1 - \epsilon_{ppv} \quad (12)$$

对于高风险事件预警的漏报率为

$$\text{漏报率} = 1 - \epsilon_{tpr} \quad (13)$$

分别统计车辆渗透率 10 % 和 5 % 下,预警模型的误报率和漏报率,对于 15 种不同输入时间窗长度与提前预警时间的组合形式都进行了结果统计。考虑到不同随机选择下模型预测结果的差异性,所展示的结果均为 10 次预警结果的均值。由于组别繁多,为便于读者直观对比,选取“全部”与“预测”两种模式展示结果,如表 7~9 所示。其中,“全部”模式表示全部 15 种组合形式预警结果的均值,用于表征具有普适性的结论;“预测”模式表示,所有提前预警时间为提前 2 s 和 3 s 的 5 种组合形式预警结果的均值,重点关注于在实际需求场景下的预警效果。

### 5.1 交通熵参数与非交通熵参数结果对比

以表 7 智能车场景-6 参数为例,在 10 % 渗透率下,与“仅非交通熵参数”相比,在“仅交通熵参数”作为模型输入后,高风险事件预警模型的误报率从 6.18 % 下降到 1.95 %,漏报率更是从 11.47 % 下降到 3.12 %。在 5 % 渗透率下,误报率和漏报率分别从 7.33 %、14.62 % 下降到 4.09 %、7.05 %,误报率和



表 7 智能车场景—6 参数的模型结果

Tab.7 Model results for autonomous -vehicles scenario-6 variables

渗透率		仅交通熵参数		混合使用		仅非交通熵参数	
		误报率/%	漏报率/%	误报率/%	漏报率/%	误报率/%	漏报率/%
10%	全部	1.95	3.12	3.68	7.26	6.18	11.47
	预测	2.28	3.82	3.69	6.30	6.48	9.40
5%	全部	4.09	7.05	4.62	10.44	7.33	14.62
	预测	4.24	8.88	3.69	6.30	9.39	12.44

表 8 网联车场景—6 参数的模型结果

Tab.8 Model results for connected-vehicles scenario-6 variables

渗透率		仅交通熵参数		混合使用		仅非交通熵参数	
		误报率/%	漏报率/%	误报率/%	漏报率/%	误报率/%	漏报率/%
10%	全部	—	—	4.00	7.69	6.65	14.03
	预测	—	—	3.74	7.23	9.39	15.97
5%	全部	—	—	6.01	10.52	8.88	14.63
	预测	—	—	5.72	10.81	9.39	15.97

表 9 网联车场景—2 参数的模型结果

Tab.9 Model results for connected-vehicles scenario-2 variables

渗透率		仅交通熵参数		混合使用		仅非交通熵参数	
		误报率/%	漏报率/%	误报率/%	漏报率/%	误报率/%	漏报率/%
10%	全部	6.12	9.89	7.75	12.31	9.50	16.45
	预测	6.33	11.08	7.16	16.06	12.42	17.22
5%	全部	7.75	13.54	10.93	15.51	11.48	20.27
	预测	7.93	13.37	20.91	20.91	16.26	23.72

漏报率的下降幅度都达到 50% 以上。“混合使用”是介于上述两种情况之间的过渡状态,在智能车场景-6 参数中混合使用了 3 个交通熵参数和 3 个非交通熵参数。其模型结果在 10% 渗透率下,误报率和漏报率的下降幅度超过 50%;在 5% 渗透率下,误报率和漏报率下降幅度超过 30%。由此可知,在模型输入参数中加入交通熵参数后,高风险事件预警效果更好,可以有效降低预警模型的误报率和漏报率。

在表 8 和表 9 的网联车场景中,即使缺乏车辆之间的交互信息,只使用车辆速度与加速度等参数对高风险事件预警,交通熵参数仍有较好表现。在网联车场景-2 参数和 10% 渗透率情况下,只使用两个交通熵参数,误报率和漏报率分别为 6.12% 和 9.89%;即使车辆渗透率降低到 5%,误报率和漏报率也在 7.75% 和 13.54% 以内。“仅交通熵参数”与“仅非交通熵参数”的对比,也同样验证了交通熵参数对交通运行高风险事件识别中的优势,能够有效降低高风险事件的误报率和漏报率。

## 5.2 智能车场景与网联车场景结果对比

从模型结果看,智能车场景下对高风险事件的预

警结果整体优于网联车场景。表 7 智能车场景—6 参数相较表 8 网联车场景—6 参数,高风险事件预警模型误报率的平均值从 6.72% 下降到 4.80%,漏报率的平均值从 12.11% 下降到 8.43%。但也要充分认识到,在智能车场景推广应用的成本相比网联车场景更高,而本研究提出的高风险事件预警模型,在智能网联时代的发展初期能够仅基于网联车场景就得到不错的高风险事件预警效果,也能在后期发展的智能车场景中快速拓展应用,整体适用范围更广。

表 8 和表 9 均为网联车场景但模型中输入的参数个数不同,表 9 中使用 2 参数的模型误报率、漏报率较高,分别为 10.38% 和 15.86%。因此在网联车场景下可通过适当增加输入参数的个数提升模型效果。

## 5.3 提前预警时间结果对比

在实际应用场景中,提前预留足够的反应时间对于高风险事件预警十分必要。为此专门在模型结果中设置“预测”一栏,取所有可以提前 2 s 和 3 s 预警的结果平均值。表 7 中的智能车场景 10% 渗透率下“仅交通熵参数”,在预测模式下,其对高风险事件误报率和漏报率为 2.28% 和 3.82%,其中提前 2 s 预警的误报率和漏报率分别为 1.31% 和 2.52%;提前 3 s 预警的误报率和漏报率分别为 3.57% 和 5.77%。表 9 中的网联车场景,10% 渗透率下“仅交通熵参数”的预测环境中,误报率和漏报率分别为 6.12% 和 9.89%。均可以证明该高风险事件预警模型具有良好的预警能力,能够为驾驶员预留出反应时间,从而采取具体动作规避驾驶风险,提高道路交通运行的安全性。

## 6 结论

针对现有高风险事件预警模型误报率高、预测时空精准度低等问题,结合智能网联时代的发展趋势,本文首先定义了交通熵及其计算方法,用以量化表征微观个体行为与交通系统的宏观状态;而后基于 LSTM 深度学习模型,提出一种低渗透率智能网联环境下高风险事件的预警方法。主要结论如下:

(1)交通熵作为交通流中的通用性指标,建立起了个体车辆微观行为与交通流宏观状态的映射关系,相比于传统的统计性指标(极值、标准差、离散系数等)能够更好地表征交通流状态和紊乱程度。在宏观层面,利用交通熵表征的交通流状态,能够有效识别与预警交通运行高风险事件。同时在微观行为层面,交通熵可以定量刻画驾驶行为,应用于异常驾驶行为识别等研究中。

(2)利用HighD轨迹数据集中提取的256个高风险事件和1024个非风险事件,分别测试低渗透率水平(5%和10%)智能车场景和网联车场景下的模型预警效果。结果表明,使用交通熵参数后高风险事件预警模型误报率和漏报率大幅降低。在预测模式下,对高风险事件误报率和漏报率为2.28%和3.82%,其中提前2s预警的误报率和漏报率分别为1.31%和2.52%;提前3s预警的误报率和漏报率分别为3.57%和5.77%。可为驾驶员预留充分反应时间、采取具体动作规避驾驶风险,有助于提高道路交通运行安全。

(3)在智能网联环境中,智能车场景对高风险事件的预警结果整体优于网联车场景。智能车辆的应用推广有助于更精准预警道路交通运行风险,减少甚至防止事故的发生,提高道路交通安全水平。在网联车场景下可通过适当增加输入参数的个数提升模型效果。

本文所提出的高风险事件预警模型针对智能网联时代低渗透率场景,对交通流中可能存在的运行风险提前进行预警。随着智能网联车渗透率的不断提升,还可对个体车辆的微观行为识别进一步优化,如监测个体车辆的疲劳驾驶、分心驾驶、甚至路怒行为等。这不仅有助于交通熵在微观行为层面的拓展应用,也需要进一步结合其他交通要素的感知,这部分工作正在进行中。

#### 作者贡献声明:

陈晓芸:交通熵理论及算法,高风险预警模型构建;  
叶颖俊:数据预处理,模型结果分析;  
余荣杰:数据预处理,高风险事件识别方法与提取;  
孙剑:方法论构建,交通熵理论及算法。

#### 参考文献:

- [1] GUO F, KLAUER S, HANKEY J, *et al.* Near crashes as crash surrogate for naturalistic driving studies [J]. *Transportation Research Record Journal of the Transportation Research Board*, 2010, 2147(2147):66.
- [2] PARK H, HAGHANI A. Real-time prediction of secondary incident occurrences using vehicle probe data[J]. *Transportation Research Part C Emerging Technologies*, 2016, 70(9):69.
- [3] DINGUS T, KLAUER S, NEALE V L, *et al.* The 100-car naturalistic driving study, phase II-results of the 100-car field experiment[R]. [S.l.]:Department of Transportation, National Highway Traffic Safety Administration, 2006.
- [4] HOSSAIN M, ABDEL-ATY M, QUDDUS M A, *et al.* Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements [J]. *Accident Analysis & Prevention*, 2019, 124(3):66.
- [5] FEDERAL HIGHWAY ADMINISTRATION. Highway Safety Improvement Program (HSIP) [R]. Federal Highway Administration, 2020.
- [6] KWAK H C, KHO S. Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data [J]. *Accident Analysis and Prevention*, 2016, 88(3):9.
- [7] 孙剑,孙杰.城市快速路实时交通流运行安全主动风险评估[J].*同济大学学报(自然科学版)*,2014,42(6):873.  
SUN Jian, SUN Jie. Proactive assessment of real time traffic flow accident risk on urban expressway [J]. *Journal of Tongji University (Natural Science)*, 2014,42(6):873.
- [8] YU R, WANG X, YANG K, *et al.* Crash risk analysis for Shanghai urban expressways: A Bayesian semi-parametric modeling approach[J]. *Accident Analysis & Prevention*, 2016, 95(Pt B):495.
- [9] ROSHANDEL S, ZHENG Z, WASHINGTON S. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis [J]. *Accident Analysis & Prevention*, 2015, 79: 198.
- [10] KATRAKAZAS C, QUDDUS M, CHEN W, *et al.* Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions [J]. *Transportation research, Part C. Emerging technologies*, 2015, 60: 416.
- [11] WANG L, ABDEL-ATY M, MA W, *et al.* Quasi-vehicle-trajectory-based real-time safety analysis for expressways [J]. *Transportation Research Part C: Emerging Technologies*, 2019, 103(6): 30.
- [12] YAN K, QU X. A review of crash surrogate events[C/CD]//Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM) and the Sixth International Symposium on Uncertainty, Modeling, and Analysis (ISUMA). 2014.
- [13] PEREZ M, SUDWEEKS J, SEARS E, *et al.* Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data [J]. *Accident Analysis & Prevention*, 2017, 103: 10.
- [14] SUDWEEKS J D. Using functional classification to enhance naturalistic driving data crash/near crash algorithms [R]. Blacksburg: Virginia Tech Transportation Institute, 2015.
- [15] 王雪松,徐晓妍.基于自然驾驶数据的危险事件识别方法[J].*同济大学学报(自然科学版)*,2020,48(1):51.

- WANG Xuesong, XU Xiaoyan. Detection of safety-critical events based on naturalistic driving data [J]. Journal of Tongji University (Natural Science), 2020, 48(1):51.
- [16] VAN A. A time-based analysis of road user behaviour in normal and critical encounters [D]. Delft; TNO, 1990.
- [17] VOGEL K. A comparison of headway and time to collision as safety indicators [J]. Accident Analysis and Prevention, 2003, 35(3):427.
- [18] ARCHER J. Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections [D]. Stockholm: Royal Institute of Technology, 2005.
- [19] OZBAY K, YANG H, BARTIN B, *et al.* Derivation and validation of new simulation-based surrogate safety measure [J]. Transportation Research record, 2008, 2083(1): 105.
- [20] YANG H. Simulation-based evaluation of traffic safety performance using surrogate safety measures [D]. New Brunswick: Rutgers University Graduate School, 2012.
- [21] YU R, HAN L, ZHANG H. Trajectory data based freeway high-risk events prediction and its influencing factors analyses [J]. Accident Analysis & Prevention, 2021, 154(1):106085.
- [22] YU R, ABDEL-ATY M. Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes [J]. Accident Analysis & Prevention, 2013, 58(9):97.
- [23] YU R, ABDEL-ATY M. Utilizing support vector machine in real-time crash risk evaluation [J]. Accident Analysis & Prevention, 2013, 51(3):252.
- [24] SUN J, SUN J. Real-time crash prediction on urban expressways: identification of key variables and a hybrid support vector machine model [J]. IET Intelligent Transport Systems, 2016, 10(5): 331.
- [25] ABDEL-ATY M, SHI Q, PANDE A, *et al.* Real time traffic operations and safety [M]. [S.l.]: Emerald Group Publishing Limited, 2018.
- [26] CLAUSIUS R. Über die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen [J]. Annalen der Physik, 1850, 155(3): 368.
- [27] BOLTZMANN L. Über die Beziehung zwischen dem zweiten Hauptsatze des mechanischen Wärmethorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht [M]. [S.l.]: Kk Hof- und Staatsdruckerei, 1877.
- [28] SHANNON C E. A mathematical theory of communication [J]. The Bell System Technical Journal, 1948, 27(3): 379.
- [29] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8):1735.
- [30] VLAHOGIANNI E I, KARLAFTIS M G, GOLIAS J C. Short-term traffic forecasting: Where we are and where we're going [J]. Transportation Research Part C Emerging Technologies, 2014, 43(1):1
- [31] MA X, TAO Z, WANG Y, *et al.* Long short-term memory neural network for traffic speed prediction using remote microwave sensor data [J]. Transportation Research Part C: Emerging Technologies, 2015, 54:187.
- [32] HUANG X, SUN J, SUN J. A car-following model considering asymmetric driving behavior based on long short-term memory neural networks [J]. Transportation Research Part C: Emerging Technologies, 2018, 95(10):346.
- [33] SUN J, QI X, XU Y, *et al.* Vehicle turning behavior modeling at conflicting areas of mixed-flow intersections based on deep learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9): 3674.
- [34] KRAJEWSKI R, BOCK J, KLOEKER L, *et al.* The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems [C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC). [S.l.]: IEEE, 2018: 2118-2125.
- [35] SAYED T, BROWN G, NAVIN F. Simulation of traffic conflicts at unsignalized intersections with TSC-Sim [J]. Accident Analysis & Prevention, 1994, 26(5):593.
- [36] HOGEMA J H, JANSSEN W H. Effects of intelligent cruise control on driving behaviour : a simulator study [C/CD]// Intelligent Transportation: Realizing the Future Abstracts of the Third World Congress on Intelligent Transport Systems. [S.l.]: TNO, 1996.
- [37] HUANG F, LIU P, YU H, *et al.* Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections [J]. Accident Analysis & Prevention, 2013, 50(1):1014.
- [38] QIANG MENG, QU X. Estimation of rear-end vehicle crash frequencies in urban road tunnels [J]. Accident Analysis & Prevention, 2012, 48:254.
- [39] 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 Journal of the Transportation Research Board, 2004, 1897:88.
- [40] GRAZIOTIN D, ABRAHAMSSON P. A web-based modeling tool for the SEMAT essence theory of software engineering [J]. Journal of Open Research Software, 2013, 1(1): 4.
- [41] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99:2999.