

文章编号:1671-1637(2022)03-0115-11

基于 SVM-LSTM 的车辆跟驰行为 识别与信息可信甄别

史宇辰^{1,2}, 晏松^{1,2}, 姚丹亚^{1,2,3}, 张毅^{1,2,3,4}

(1. 清华大学 信息科学技术学院, 北京 100084; 2. 清华大学 北京信息科学与技术国家研究中心, 北京 100084;
3. 东南大学 现代城市交通技术江苏高校协同创新中心, 江苏 南京 210096;
4. 清华-伯克利深圳学院, 广东 深圳 518055;)

摘要:为利用智能车路协同系统内实时交互信息有效提升交通系统的安全性,提出了基于交通业务特征的交通信息可信甄别方法;重点构建了基于支持向量机(SVM)-长短时记忆(LSTM)神经网络的车辆跟驰行为识别与信息可信甄别模型,包括基于 SVM 的车辆跟驰行为识别模型和基于 LSTM 神经网络的车辆跟驰速度预测模型;设定了表征车辆行驶状态的特征向量,基于 SVM 的车辆跟驰行为识别模型将车辆行驶状态分为跟驰与非跟驰;对于跟驰车辆,基于 LSTM 神经网络的车辆跟驰速度预测模型根据其历史数据进行速度预测;SVM-LSTM 信息可信甄别模型通过检验跟驰车辆的预测速度与其实际速度的差是否在合理范围来判断车辆数据的可信性,实现信息可信甄别;采用公开数据集对提出的模型进行了训练与测试,并构建了不同异常类型和异常幅度的多个异常测试数据集,对基于 SVM-LSTM 神经网络的车辆跟驰行为识别与信息可信甄别模型进行了验证。研究表明:基于 SVM 的车辆跟驰行为识别模型对车辆行驶行为识别的准确率达到 99%,基于 LSTM 神经网络的车辆跟驰速度预测模型的跟驰速度预测精度达到了 $\text{cm} \cdot \text{s}^{-1}$ 数量级;基于 SVM-LSTM 神经网络的车辆跟驰行为识别与信息可信甄别模型在正常数据测试集与多个异常数据测试集上的甄别正确率达到了 97%。由此可见,提出的方法可用于路侧设备(RSUs)对车载单元(OBUs)实时信息和车载单元间实时信息的可信甄别。

关键词:智能交通;智能车路协同系统;SVM-LSTM;跟驰行为识别;车辆速度预测;可信甄别

中图分类号:U491.2 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2022.03.009

SVM-LSTM-based car-following behavior recognition and information credibility discrimination

SHI Yu-chen^{1,2}, YAN Song^{1,2}, YAO Dan-ya^{1,2,3}, ZHANG Yi^{1,2,3,4}

(1. School of Information Science and Technology, Tsinghua University, Beijing 100084, China; 2. Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China;
3. Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing 210096, Jiangsu, China; 4. Tsinghua-Berkeley Shenzhen Institute (TBSI), Shenzhen 518055, Guangdong, China)

Abstract: To effectively improve the traffic system security by using the real-time interaction

收稿日期:2021-12-31

基金项目:国家重点研发计划(2018YFB1600600)

作者简介:史宇辰(1999-),男,黑龙江哈尔滨人,清华大学工学博士研究生,从事智能交通研究。

导师简介:张毅(1964-),男,北京人,清华大学教授,工学博士。

引用格式:史宇辰,晏松,姚丹亚,等. 基于 SVM-LSTM 的车辆跟驰行为识别与信息可信甄别[J]. 交通运输工程学报, 2022, 22(3): 115-125.

Citation: SHI Yu-chen, YAN Song, YAO Dan-ya, et al. SVM-LSTM-based car-following behavior recognition and information credibility discrimination[J]. Journal of Traffic and Transportation Engineering, 2022, 22(3): 115-125.

information in intelligent vehicle-infrastructure cooperative systems (i-VICS), a credibility discrimination approach for traffic information based on the traffic business features was proposed. In particular, the model for the car-following behavior recognition and the information credibility discrimination was built based on the support vector machine (SVM) and long short-term memory (LSTM) neural network. It was composed of the SVM-based car-following behavior recognition model and the LSTM neural network-based car-following speed prediction model. The feature vector representing the vehicle driving states was set, and the vehicle driving states were divided into the following and non-following by the SVM-based car-following behavior recognition model. For following vehicles, their speeds were predicted by the LSTM neural network-based car-following speed prediction model according to the history data. With the SVM-LSTM-based information credibility discrimination model, the credibility of vehicle data was judged by checking whether the difference between the predicted speed and the actual speed of the following vehicles was within the reasonable range, and in this way, the information credibility discrimination was achieved. The public dataset was employed to train and test the proposed models, and several abnormal test datasets of various abnormality types and abnormality amplitude were built to verify the SVM-LSTM neural network-based model for the car-following behavior recognition and the information credibility discrimination. Research results show that the vehicle driving behavior recognition accuracy of the SVM-based car-following behavior recognition model is up to 99%, and the predicted car-following speed precision with an order of magnitude of $\text{cm} \cdot \text{s}^{-1}$ can be achieved by the LSTM neural network-based car-following speed prediction model. The discrimination accuracy of the SVM-LSTM neural network-based model for the car-following behavior recognition and information credibility discrimination is up to 97% on the normal test datasets and multiple abnormal test datasets. Thus, the proposed approach can be applied for the real-time information credibility discriminations of road side units (RSUs) to on-board units (OBUs) and between OBUs. 8 tabs, 9 figs, 30 refs.

Key words: intelligent transportation; intelligent vehicle-infrastructure cooperative system; SVM-LSTM; car-following behavior recognition; vehicle speed prediction; credibility discrimination

Author resumes: SHI Yu-chen(1999-), male, doctoral student, 1999031222@163.com; ZHANG Yi(1964-), male, professor, PhD, zhyi@mail.tsinghua.edu.cn.

Foundation item: National Key Research and Development Program of China (2018YFB1600600)

0 引 言

智能车路协同与智能驾驶技术的发展对交通信息交互的安全性提出了更高的要求,仅靠现有计算机信息安全技术和移动通信安全技术难以保障车路协同系统中交通信息交互的可信安全^[1]。当系统中的通信实体被劫持或遭攻击导致其计算机安全和通信安全技术失效时,存在实时交互数据被篡改、伪造等的风险,此时如果系统不能及时发现其异常并更新车辆信任列表,会造成严重危害。显然,如果交通系统能基于交通业务特征对数据本身的合理性进行分析和辨识,对不符合交通业务特征的异常信息进行甄别,及时在系统内对相应的可信标识进行传递

并更新系统的车辆可信列表,可在现有交通信息安全保障的基础上,进一步增强交通系统的信息安全防护能力。

交通业务数据指车辆基础数据如位置和速度、车辆总线提供的操控数据以及交通环境数据等。交通业务特征是指交通主体在具体的交通状态下其业务数据体现的规律和特点。不失一般性,交通业务特征可以归纳为基于交通物理边界、车辆运动状态和驾驶人驾驶行为 3 类特征。依据不同的交通业务特征,可以从不同层次对交互的车辆数据可信与否进行甄别。交通物理约束类特征主要关心车辆在物理世界中受到其他交通主体的限制,例如车辆不能超出道路边界,车辆的速度和加速度等不能超出其

固有属性的边界等;车辆运动状态类特征主要用于分析车辆运动过程,包括不同来源与类型的传输数据和传感器获得的感知数据中的速度、加速度等信息是否一致,多车间车辆相对运动是否满足常规规律等;驾驶人驾驶行为类特征主要关注驾驶人在驾驶过程中所表现出的驾驶习惯和特点,包括驾驶人个性化驾驶特点,如踩踏油门踏板的习惯^[2]、驾驶激进程度^[3]以及驾驶人在特定场景下的共性驾驶特点,如跟驰时的速度滞后现象^[4]、非对称驾驶现象^[5-6]。3类特征的关注点由静态数据到数据动态关系再到数据统计规律,其特征提取的难度依次提高。目前,基于交通业务特征的信息可信甄别主要集中在依据交通物理边界和车辆运动状态2类特征上,依据驾驶人驾驶行为进行信息可信甄别的研究相对较少。Wong等^[7]提出了基于轨迹的分级防御(Trajectory-Based Hierarchical Defense, TBHD)框架,其3级框架分别对应车辆动力学的物理边界、车辆数据的连续性以及多车轨迹交叉的辨识检查;Huang等^[8]提出了一种数据驱动的方法来识别异常车辆提供的伪造轨迹,该方法实现了TBHD框架的第3级防御。这些甄别方法均可归纳为依据交通物理边界或车辆运动状态实施交通信息可信性甄别的方法。

考虑驾驶人在不同的车辆行驶状态时会存在不同的驾驶行为,祝俐菱等^[9]将车辆行驶状态划分为跟驰、换道准备和换道执行3类,并使用支持向量机(Support Vector Machine, SVM)识别了车辆的行驶状态;戢晓峰等^[10]使用SVM识别了货车遮挡条件下的小客车行驶状态,但其将车辆的行驶状态划分为跟驰、换道和超车3类。虽然研究人员对车辆行驶状态的划分存在差异,但必要的划分是车辆行驶状态分析、驾驶人驾驶行为特征提取或车辆行驶数据真实性甄别的前提条件。跟驰作为一种出现频率较高且较为典型的车辆行驶状态,是驾驶行为研究的基础,也是复杂、混合交通场景研究的前提。本文仅对跟驰状态时的驾驶行为进行展开,做深入分析,暂不讨论非跟驰状态。

通常可以将车辆跟驰模型分为传统模型和基于机器学习的模型2类。传统车辆跟驰模型包括Gazis-Herman-Rothery(GHR)模型、安全距离模型、线性模型、心理模型等^[11-15]。杨达等^[16-19]在分析传统车辆跟驰模型的基础上,提出了改进型车辆跟驰模型。随着深度学习的发展,将深度学习方法应用于驾驶行为分析取得了良好的效果^[20]。Wang等^[21-22]

引入深度学习中的长短时记忆(Long Short-Term Memory, LSTM)神经网络,在不区分驾驶人驾驶风格的情况下,实现了对车辆速度的预测,较好地展现了LSTM神经网络在车辆跟驰行为分析中的优势和效果。

考虑车辆跟驰行为模型可以作为车辆跟驰状态下信息可信甄别的基础,故可以在对车辆跟驰状态及其特征辨识的基础上,实现车辆行驶数据的可信甄别。于是,本文基于驾驶人驾驶行为分析,研究驾驶人在车辆跟驰状态下的驾驶行为特征,并设计相应的车辆行驶数据可信甄别模型,其基本思想是在车辆跟驰状态下驾驶人的驾驶行为应该满足固有规律或特征,由此可预测车辆近期行为如车辆速度,如实际数据与预测数据相符即表征车辆行驶数据在合理范围内,此时数据可视为真实可信。简言之,车路协同环境下当车辆行驶在跟驰状态时,可以基于一段时间内车辆的历史数据预测下一时刻的速度,若预测值与下一时刻系统实时采集到的速度接近,则认为采集的数据可信,否则认为不可信,在应用驾驶人驾驶行为特征对车辆数据进行可信甄别前,需先识别车辆的行驶状态;在此基础上为提高车辆跟驰状态下驾驶行为分析的准确性,引入深度学习中的LSTM神经网络,实现车辆跟驰状态下的行驶速度预测;最后,通过构建车辆行驶数据的可信甄别模型,完成车辆行驶数据的可信甄别计算。

基于上述分析,本文以构建基于交通业务特征的交通信息可信甄别方法为目标,研究基于SVM-LSTM的车辆跟驰行为识别与信息可信甄别方法,包括基于SVM的车辆跟驰行为识别模型和基于LSTM神经网络的车跟驰速度预测模型;提出可信甄别计算流程,采用公开数据集NGSIM(Next Generation Simulation)分别训练和测试基于SVM的车辆跟驰行为识别模型和基于LSTM神经网络的车跟驰速度预测模型,并验证基于SVM-LSTM的车辆跟驰行为识别与信息可信甄别模型在不同异常情况下的甄别效果。

1 问题描述

1.1 场景和参数

智能车路协同环境下任意车辆均可实时获取自车及周边车辆和道路环境等的实时信息,因此,其能够融合分析相关数据,进而可实现可信甄别。由于前车在跟驰场景中接近自由驾驶,后车会受跟驰影响,故以后车为研究对象,后文也将后车称为自车。

图 1 给出了车辆在 3 车道上行驶的典型场景,其中:着色车辆为自车; L 为车道编号; v_x 和 v_y 分别为自车的纵向(x)和横向(y)速度; v'_x 为前车的纵向速度; a_x 和 a'_x 分别为自车和前车的纵向加速度; Δv 和 Δx 分别为前后车的速度差和间距。

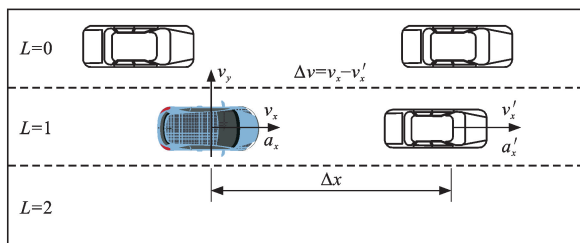


图 1 车辆驾驶场景与相关参数示意

Fig. 1 Sketch of vehicle driving scenario and related parameters

1.2 问题分析

基于 SVM-LSTM 的车辆跟驰行为识别与信息可信甄别问题可细分为 4 个子问题。

1.2.1 基于 SVM 的车辆跟驰行为识别问题

从车辆行驶的相关参数中选择合适的参数,可组成表征车辆行驶状态的特征向量 \mathbf{x} ;将车辆当前行驶状态记为 $s \in \{0, 1\}$,取 1 代表跟驰,取 0 代表非跟驰,可描述车辆处于非跟驰和跟驰状态,则该问题要求 SVM 分类器 $f_s(\cdot)$ 经训练后,可识别特征向量 \mathbf{x} 的状态 s 为

$$s = f_s(\mathbf{x}) \quad (1)$$

1.2.2 基于 LSTM 神经网络的车辆速度预测问题

选择跟驰场景下前车与后车运动的相关参数,可组成描述车辆跟驰行为的特征矩阵 \mathbf{X} ,其元素需包含所选参数 t 时刻以前一段时间内的数值,则该问题要求神经网络 $f_L(\mathbf{X})$ 经训练后,可预测后车在时刻 t 的纵向速度 $v_{p,x}(t)$ 为

$$v_{p,x}(t) = f_L(\mathbf{X}) \quad (2)$$

1.2.3 可信甄别计算问题

选择参量 $e(t)$ 表示时刻 t 时基于 LSTM 神经网络算法获得的预测速度 $v_{p,x}(t)$ 与系统采集到的速度 $v_{s,x}(t)$ 的偏差,则将其与设定的偏差阈值进行比较,即可计算获得车辆行驶数据可信甄别的结果 $R(t) \in \{0, 1\}$,分别表示 t 时刻数据不可信($R(t)=0$)与可信($R(t)=1$),因此,该问题即要求设计一个偏差计算算法 $f_e(\cdot)$,使其可以表达预测值与采集值之间的偏差

$$e(t) = f_e[v_{p,x}(t), v_{s,x}(t)] \quad (3)$$

1.2.4 甄别结果计算流程

显然,可信甄别结果的计算需要综合以上 3 个子问题的结果,但考虑实际应用中车辆的计算资源有限以及车辆速度预测需要一定时长的历史数据等

因素,系统不宜在每个时刻都重复上述全部计算过程,即不能将以上 3 个子问题简单地进行组合,因此,该问题要求可信甄别过程启动后,需建立可稳定执行的可信甄别计算流程,以形成完整的可信甄别模型,其最终输出可表示为

$$f_R(t) = R(t) \quad (4)$$

式中: $f_R(t) \in \{0, 1\}$,为可信甄别模型最终的输出结果,分别表示 t 时刻数据不可信($f_R(t)=0$)与可信($f_R(t)=1$)。

2 交通信息可信甄别

2.1 车辆跟驰行为识别

2.1.1 SVM 分类

鉴于 SVM 是一种二分类器,能够支持线性和非线性核函数对复杂样本进行分类,因此,将其用于车辆跟驰行为识别。考虑车辆行驶特征间具有较为复杂的相互作用关系,可选用径向基核函数 $K(\mathbf{x}_1, \mathbf{x}_2)$ 为

$$K(\mathbf{x}_1, \mathbf{x}_2) = e^{-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2} \quad (5)$$

式中: \mathbf{x}_1 和 \mathbf{x}_2 分别为 2 个不同样本的特征向量; γ 为径向基核函数的超参数。

于是,SVM 二分类器可将车辆行驶状态划分为跟驰与非跟驰,以实现车辆跟驰行为识别。

2.1.2 车辆状态特征量设定

根据自车和前车的实时运动参数,自车的状态特征向量、样本输入和标签可分别设定为

$$\mathbf{V}(t) = (v_x(t), v_y(t), a_x(t), \Delta L(t), \Delta v(t), \Delta x(t), a'_x(t)) \quad (6)$$

$$\mathbf{x} = (\mathbf{V}(t-0.4), \mathbf{V}(t-0.2), \mathbf{V}(t)) \quad s \in \{0, 1\} \quad (7)$$

式中: $\Delta L(t)$ 为 t 时刻车道编号变化; $\mathbf{V}(t)$ 为 t 时刻的 7 维状态特征向量,其对应的参数如表 1 所示。

表 1 SVM 输入数据参数

Table 1 Parameters of input data for SVM

参数名称	符号
t 时刻纵向速度/($\text{m} \cdot \text{s}^{-1}$)	$v_x(t)$
t 时刻横向速度/($\text{m} \cdot \text{s}^{-1}$)	$v_y(t)$
t 时刻纵向加速度/($\text{m} \cdot \text{s}^{-2}$)	$a_x(t)$
t 时刻车道编号变化	$\Delta L(t)$
t 时刻与前车速度差/($\text{m} \cdot \text{s}^{-1}$)	$\Delta v(t)$
t 时刻与前车车头间距/m	$\Delta x(t)$
t 时刻前车纵向加速度/($\text{m} \cdot \text{s}^{-2}$)	$a'_x(t)$

为使样本输入能够反映更大时间范围的变化规律,设定样本输入包括 $t-0.4$ 、 $t-0.2$ 和 t 时刻的状态特征,即为 21 维特征向量。

2.2 车辆跟驰速度预测

2.2.1 LSTM 神经网络

LSTM 神经网络^[23]是一种特殊的循环神经网络,其核心细胞状态 C 受控于遗忘门、输入门和输出门,其神经元结构如图2所示,其中: σ 为 Sigmoid 函数; x_k 为第 k 个单元的输入; C_k 为第 k 个单元的细胞状态; h_k 为第 k 个单元的隐状态。该网络可以处理较长时间序列的历史数据,设有遗忘门的 LSTM 神经网络具有衰落记忆功能,可以按需要弱化过时数据。考虑到 LSTM 神经网络具有的特性与人为驾驶过程较为一致,因此,本文采用 LSTM 神经网络描述车辆跟驰行驶状态下的驾驶行为。

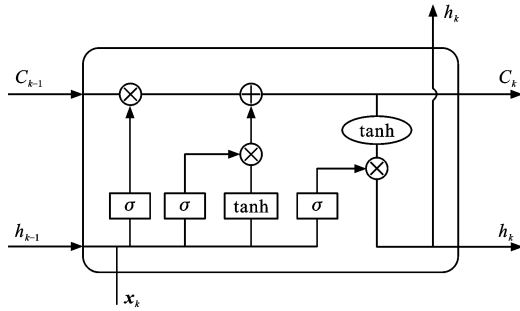


图2 LSTM 神经元结构

Fig. 2 Structure of LSTM neuron

2.2.2 参数设定

在实际车辆跟驰场景中,后车驾驶人通常会根据一段时间内自车与前车的速度以及两车间距来调整自车的速度。于是,选用一段时间内后车纵向速度、两车速度差和两车位置差分别构成的3个序列作为 LSTM 神经网络输入,可预测下一时刻的后车速度,则 LSTM 神经网络的输入、输出关系可以表达为

$$v_{p,x}(t) = f_L[v_x(t-\tau:t-\Delta t), \Delta v(t-\tau:t-\Delta t), \Delta x(t-\tau:t-\Delta t)] \quad (8)$$

式中: τ 为所选序列的时间长度; Δt 为采样间隔; $f_L(\cdot)$ 为输入和输出变量之间的映射关系; $v_x(t-\tau:t-\Delta t)$ 为时刻 $t-\tau \sim t-\Delta t$ 的自车纵向速度序列; $\Delta v(t-\tau:t-\Delta t)$ 为时刻 $t-\tau \sim t-\Delta t$ 的两车速度差序列; $\Delta x(t-\tau:t-\Delta t)$ 为时刻 $t-\tau \sim t-\Delta t$ 的两车位置差序列。

为保证所选时间序列能够覆盖车辆某种行驶行为的完整过程,同时避免因时间序列过长所造成的计算资源消耗过大的问题,应用中可选择 $\tau=5\text{ s}$ ^[24];采用 NGSIM 数据集进行验证,其采样间隔选为 0.1 s ,即 $\Delta t=0.1\text{ s}$,因此,采用 LSTM 神经网络预测车辆行驶速度时, $t-5 \sim t-0.1$ 内的 v_x 、 Δv 和 Δx

序列为网络输入, $v_{p,x}(t)$ 为网络输出。完成上述计算过程的 LSTM 神经网络结构如图3所示,单元结构如图2所示,隐藏层的输出 h_{50} 即为 LSTM 神经网络的最终输出 $v_{p,x}(t)$ 。

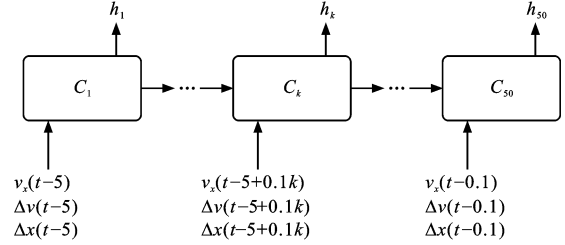


图3 跟驰场景下车辆速度预测的 LSTM 神经网络结构

Fig. 3 LSTM neural network structure for speed prediction of vehicle under car-following scenario

2.3 可信甄别计算

在车辆速度预测问题中,驾驶的主观性和随机性对预测结果的准确性有较大影响,即表现为一定概率上驾驶人会不按常规行为进行驾驶。

为保证可信甄别算法的稳定性和准确性,减少驾驶人主观因素和其他意外情况的干扰,选择当前时刻以及回溯 N 个连续采样周期的车辆速度预测值及其对应的实采值,可对当前时刻车辆行驶数据的可信进行甄别计算,即

$$e(t) = \frac{1}{N} \sum_{i=0}^{N-1} \frac{|v_{p,x}(t-i\Delta t) - v_{s,x}(t-i\Delta t)|}{v_{s,x}(t-i\Delta t)} \quad (9)$$

式中: i 为不同时时刻的计数变量。

于是,可获得甄别结果为

$$R(t) = \begin{cases} 0 & e(t) \geq \alpha \\ 1 & e(t) < \alpha \end{cases} \quad (10)$$

式中: α 为可信程度阈值,可由数据可信程度的具体要求设定。

使用该算法,即使驾驶人在 $N\Delta t$ 时长内某个时刻做出了特殊动作,导致该时刻 LSTM 神经网络预测值与实采值偏差较大,但是由于其他 $N-1$ 个时刻的偏差相对较小,也可保证平均误差在合理范围内,不会导致判断异常,有效减少了随机性对可信甄别结果的影响。

实际应用时,需要根据相关情况确定参数值。例如,在本文验证试验中采用的 NGSIM 数据集采样间隔为 0.1 s ,故选择 $\Delta t=0.1\text{ s}$;为在提高算法稳定性的同时保证甄别延迟不至过高,可令计算周期为 0.5 s ,故选择 $N=5$;经测试,选择 $\alpha=2.5\%$ 时,异常数据的检测准确率较高,正常数据的误报率较低。

2.4 可信甄别计算流程

图4给出了可信甄别计算流程的架构,其任务

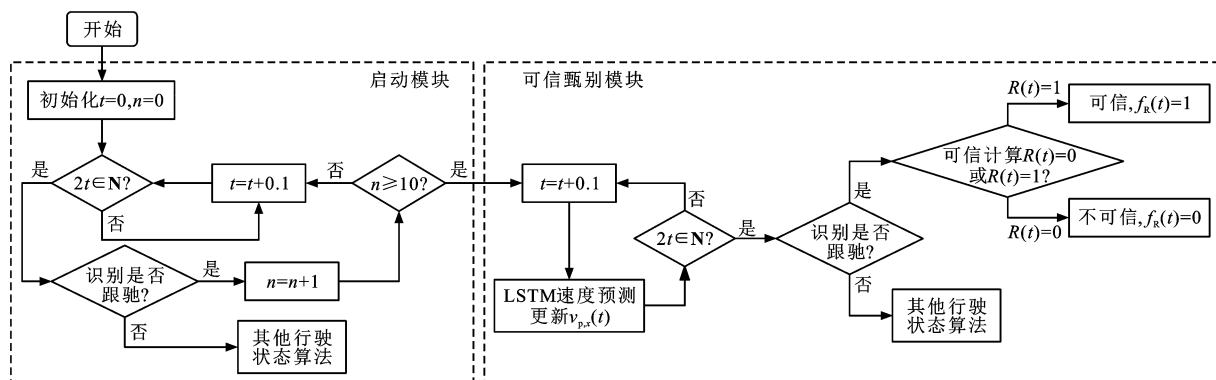


图4 可信甄别计算流程架构

Fig. 4 Framework of calculation process for credibility discrimination

可划分为启动模块与可信甄别模块两部分,其中, n 为当前已连续识别为跟驰状态的次数。

2.4.1 启动模块

启动模块是可信甄别的计算基础,需按设计要求为可信甄别模块准备必须的数据。如前所述,信息采样周期选为 0.1 s ;基于SVM的车辆跟驰状态识别算法需要使用当前时刻以及回溯 0.5 s 范围内参数构成的车辆状态特征向量作为输入,故每 0.5 s 即满足 $2t \in \mathbf{N}$ (\mathbf{N} 为正整数集)时进行一次跟驰状态识别;基于LSTM神经网络的车辆速度预测算法所需的历史行驶数据时段长度选为 $\tau=5\text{ s}$ 。于是,当系统或车辆间保持稳定通信超过 5 s ,且车辆行驶状态连续 10 次的识别结果为跟驰后,说明车辆已经持续 5 s 处于跟驰状态,数据准备已达到可信甄别模块的运行前置条件,此后即可进入可信甄别模块。若过程中某次车辆行驶状态识别结果为非跟驰状态,则启动其他行驶状态所对应的计算。

2.4.2 可信甄别模块

可信甄别模块负责完成车辆行驶数据可信甄别的完整计算过程。当满足其运行前置条件后即可进入模块的计算过程。本模块启动后,每个时刻均使用LSTM神经网络算法计算该时刻车辆速度预测值 $v_{p,x}(t)$ 。当满足 $2t \in \mathbf{N}$ 时则执行跟驰状态识别,如识别结果为跟驰状态,则由 5 个连续时刻的车辆速度预测值及其对应的实采值进行比对以获得可信甄别计算结果 $R(t)$,并由可信甄别模型输出甄别结果 $f_R(t)=R(t)$ 。

3 测试验证

使用美国联邦公路局提供的NGSIM数据集,选取其中的US101高速公路相关数据,对基于SVM的车辆跟驰行为识别模型、基于LSTM神经

网络的车辆跟驰速度预测模型以及车辆行驶数据可信甄别模型进行试验验证。NGSIM数据集使用摄像机进行采集,并经视频图像分析方法获得,包括车辆位置、速度和加速度等,其采样频率为 10 Hz ,即数据更新时间间隔为 0.1 s [24-26]。

3.1 基于SVM的车辆跟驰行为识别模型训练与测试

3.1.1 数据准备

在对基于SVM的车辆跟驰行为识别模型进行训练和测试前,需提取和标注NGSIM数据集的所有数据。首先,根据变换车道前后的横向位置变化提取出换道过程的数据,将其标注为非跟驰状态 0 。然后,对余下数据进行分析,若 t 时刻后车存在有前车,则可结合前车行驶数据计算 $\Delta v(t)$ 和 $\Delta x(t)$,记录 $a_x(t)$,并计算 t 时刻车头时距 $t_H(t)$ 为

$$t_H(t) = \frac{\Delta x(t)}{v_x(t)} \quad (11)$$

于是,将 $t_H(t) < 2.5\text{ s}$ 的数据标注为跟驰状态 1 ,将 $t_H(t) \geq 2.5\text{ s}$ 的数据标注为非跟驰状态 0 。若 t 时刻不存在前车,则将 $\Delta v(t)$ 和 $a'_x(t)$ 分别记为 $-100\text{ m} \cdot \text{s}^{-1}$ 和 $-100\text{ m} \cdot \text{s}^{-2}$, $\Delta x(t)$ 记为 0 以作区分,同时将该时刻的数据标注为非跟驰状态 0 。最后,将 $t-0.4$ 、 $t-0.2$ 和 t 时刻的状态特征向量拼接为 21 维向量 \mathbf{x} 并进行归一化,以 t 时刻的标注结果作为该特征向量的标签 s 。

从NGSIM数据集包含的 $1\,000$ 辆车的行驶数据共获得 $701\,728$ 条特征状态数据,每隔 1 s 进行一次拼接即可得到特征向量样本及相应的标签,由此最终获得数据集样本总数为 $69\,525$ 。其中跟驰状态下的样本数为 $35\,667$,非跟驰状态的样本数为 $33\,858$ 。为便于说明,特选车辆编号为 37 的行驶轨迹及其行驶状态数据绘制行驶过程,如图5所示。

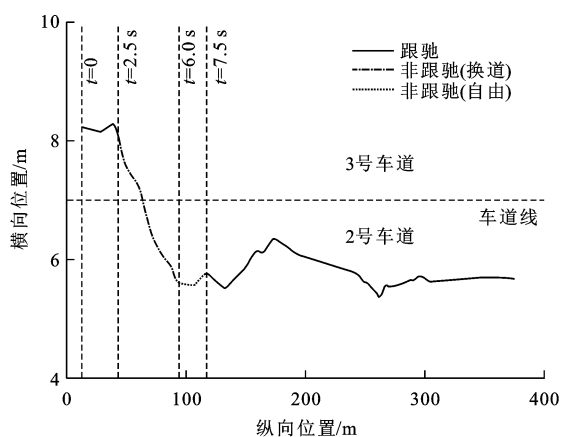


图 5 37 号车辆行驶轨迹与行驶状态分类
Fig. 5 Classifications of vehicle trajectory and driving states for vehicle No. 37

3.1.2 训练与测试

在模型训练前,需要对 SVM 二分类器的惩罚参数 c 、核函数的超参数 γ 进行超参数寻优。可以使用网格搜索法对超参数组合进行遍历,选择可获得最高准确率的超参数组合为寻优结果。为了减少寻优时间,可以选择 70% 的数据作为 SVM 二分类器的输入。如图 6 所示,寻优结果为 $c=100, \gamma=0.01$, 此时取得最高准确率为 99.11%。

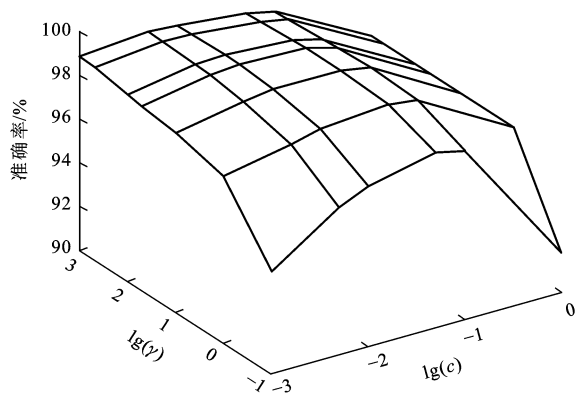


图 6 SVM 二分类器超参数寻优
Fig. 6 Hyper-parameter optimization of SVM binary classifier

选用该超参数组合作为 SVM 的超参数并进行 SVM 的训练与测试。按照 7 : 3 的比例将提取和标注后的数据划分为训练集和测试集。于是,列出 SVM 在训练集与测试集上的混淆矩阵,分别如表 2、3 所示,混淆矩阵给出了 SVM 二分类器对不同类别数据正确与错误分类的情况,可以使用混淆矩阵的数据计算准确率、精准率和召回率等多项指标。

在训练集上,跟驰状态识别的总体准确率为 99.38%,精准率为 99.25%,召回率为 99.54%;在测试集上,跟驰状态识别的总体准确率为 99.10%,

精准率为 99.03%,召回率为 99.21%。由此可见, SVM 二分类器对跟驰状态的识别效果较好。

表 2 SVM 二分类器在训练集上的混淆矩阵

训练集		预测表现		总计
		1	0	
实际表现	1	24 853	114	24 967
	0	188	23 512	23 700
总计		25 041	23 626	48 667

表 3 SVM 二分类器在测试集上的混淆矩阵

测试集		预测表现		总计
		1	0	
实际表现	1	10 616	84	10 700
	0	104	10 054	10 158
总计		10 720	10 138	20 858

3.2 基于 LSTM 神经网络的车辆跟驰速度预测模型训练与测试

3.2.1 数据准备

由 3.1.1 节中标注的跟驰状态下的车辆行驶数据可以筛选出用于 LSTM 神经网络的车辆跟驰速度预测模型训练与测试所需的数据。考虑 5 s 内连续 51 个采样点均为跟驰状态的片段,称为跟驰片段,则每个跟驰片段的前 50 个采样点 $v_x, \Delta v$ 和 Δx 组合成维度为 50×3 的矩阵作为网络的输入,第 51 个采样点的 v_x 作为网络的目标输出,由此可获得共 37 806 组样本数据。按照 3 : 1 : 1 将上述数据划分成训练集、验证集和测试集;同时,设定 LSTM 神经网络所需的结构参数与训练参数如表 4 所示。

表 4 LSTM 神经网络结构参数与训练参数

参数	数值	参数	数值
神经网络层数	2	批大小	32
输入维度	3	损失函数	均方误差
隐藏层维度	1	学习率	0.002
序列长度	50	优化器	Adam
训练轮次	60		

3.2.2 模型训练

对输入和输出数据进行归一化,以满足 LSTM 神经网络对输入输出的要求。模型训练过程中,每采用 100 组数据进行一次训练后,即使用验证集的数据进行一次验证,该过程称为一次迭代,网络训练中损失变化过程如图 7。选择所有迭代轮次中验证

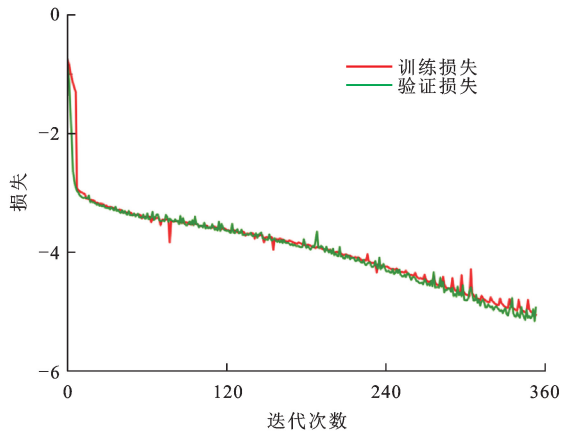


图7 网络训练中损失变化

Fig. 7 Loss changes in network training

损失最小的迭代模型参数作为最终模型参数,该轮迭代中模型在训练集和验证集的均方差损失分别为 5.00×10^{-6} 和 3.28×10^{-6} 。

3.2.3 模型评估

对输数据进行去归一化得到预测速度,选用平均绝对误差 L_a 和均方误差 L_s 来评价模型的表现,其计算公式分别为

$$L_a = \frac{1}{N_j} \sum_{j=1}^{N_j} |v_{p,j} - v_{s,j}| \quad (12)$$

$$L_s = \frac{1}{N_j} \sum_{j=1}^{N_j} (v_{p,j} - v_{s,j})^2 \quad (13)$$

式中: $v_{p,j}$ 为第 j 个样本中第 51 个采样点的预测速度; $v_{s,j}$ 为第 j 个样本中第 51 个采样点的实际速度; N_j 为数据集的样本个数。

于是,模型在训练集、验证集和测试集上的表现可归纳如表 5 所示。

表5 模型评估指标

Table 5 Indices for model evaluation

评估指标	$L_a / (\text{m} \cdot \text{s}^{-1})$	$L_s / (\text{m}^2 \cdot \text{s}^{-2})$
训练集	5.86×10^{-2}	3.25×10^{-3}
验证集	5.91×10^{-2}	2.96×10^{-3}
测试集	5.89×10^{-2}	3.98×10^{-3}

模型在测试集上的评价指标 L_a 和 L_s 分别为 5.89×10^{-2} 和 3.98×10^{-3} ,说明基于 LSTM 神经网络的车辆跟驰速度预测值误差在 $\text{cm} \cdot \text{s}^{-1}$ 数量级,由此表明模型取得了良好效果。

选择测试集中的一个跟驰片段,对比该片段车辆速度预测值与实采值,如图 8 所示,可以看到,预测速度与实际速度曲线的末端重合,该片段中的速度预测结果准确。

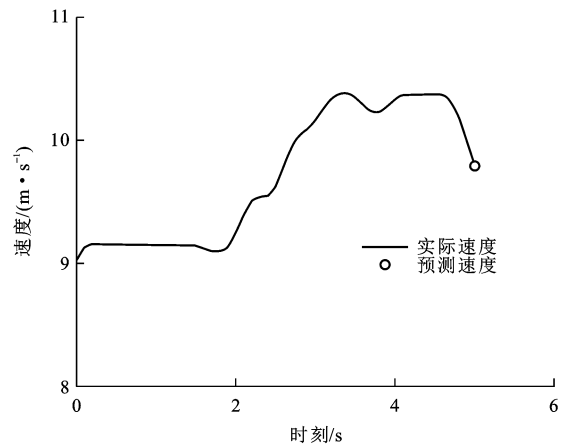


图8 预测速度与实际速度对比

Fig. 8 Comparison between predicted and real speeds

3.3 可信甄别模型试验验证

3.3.1 数据准备

在原有数据集的基础上,可按照数据异常类型和异常幅度的特点,通过修改正常数据获得所需异常数据测试集。然后,再分别使用正常数据测试集和异常数据测试集对可信甄别模型进行试验,以验证本文方法的有效性。

试验验证时从数据集中选择 700 辆车,提取每辆车时间长度为 5.5 s 的行驶数据,构成由 55 个采样点组成的数据片段,按照以下 5 种异常类型和 2 种异常幅度,对每辆车数据判断中的第 51~55 个采样点进行修改,共生成 10 个异常数据集。于是,考虑已有原始正常数据集,共获得 11 个测试数据集用于试验验证,如表 6 所示。异常幅度和异常类型的设置方法如下。

表6 测试数据集划分

Table 6 Classification of testing dataset

异常类型	低偏移幅度	高偏移幅度
类型 1	数据集 1	数据集 2
类型 2	数据集 3	数据集 4
类型 3	数据集 5	数据集 6
类型 4	数据集 7	数据集 8
类型 5	数据集 9	数据集 10
正常数据	数据集 11	

(1)数据异常幅度:低偏移幅度为 10%~20%;高偏移幅度为 20%~40%

(2)数据异常类型:类型 1 为随机单采样点异常;类型 2 为随机 3 采样点异常,且不同采样点的异常间相互独立;类型 3 为全部采样点异常,且不同采样点的异常间相互独立;类型 4 为全部采样点异常,第 5 个采样点服从异常幅度设定,同时前 4 个采样

点也异常,且 5 个采样点的异常幅度比例为 1/5 : 1/4 : 1/3 : 1/2 : 1,其异常数据曲线相对连续;类型 5 为全部采样点异常,第 5 个采样点服从异常幅度设定,同时设定前 4 个采样点也异常,且 5 个采样点的异常幅度比例为 1 : 2 : 3 : 4 : 5,其异常数据曲线形成直线关系。

3.3.2 试验分析

(1)启动模块试验分析

考虑长度为 5 s 的数据片段,若存在某个时刻车辆处于非跟驰状态,则该片段为非跟驰片段;若 10 次车辆的行驶状态识别均为跟驰,则为跟驰片段。

表 7 给出了启动模块对车辆行驶状态识别结果的混淆矩阵,可见:启动模块识别的总体准确率为 99.4%,精准率和召回率均为 99.6%。

表 7 启动模块对车辆行驶状态识别结果的混淆矩阵

Table 7 Confusion matrix of vehicle driving state recognition results in starting module

识别结果		SVM 识别		总计
		跟驰	非跟驰	
实际表现	跟驰	470	2	472
	非跟驰	2	226	228
总计		472	228	700

(2)可信甄别模块试验分析

采用可信甄别模块对正确识别的 470 个跟驰片段进行检验,结果表明:470 个跟驰片段中第 51~55 个采样点所组成的样本的 SVM 识别结果均为跟驰状态,且识别正确。采用可信甄别模块对 11 个测试数据集进行检验,检验出的异常数据数据如表 8 所示,其中 10 个异常数据集对应的检验结果表示的是甄别为异常的片段数及其准确率,而正常数据集对应的检验结果表示的是甄别为异常的片段数及其误报率,可见:除数据集 9 外,其他 10 个数据集甄别效果良好。

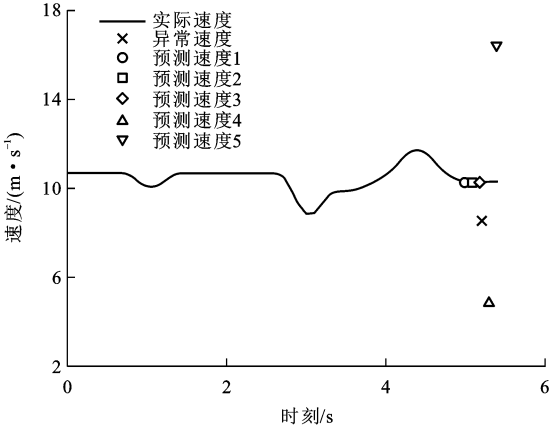
表 8 可信甄别模块异常数据甄别结果

Table 8 Discrimination results of abnormal data in credibility discrimination module

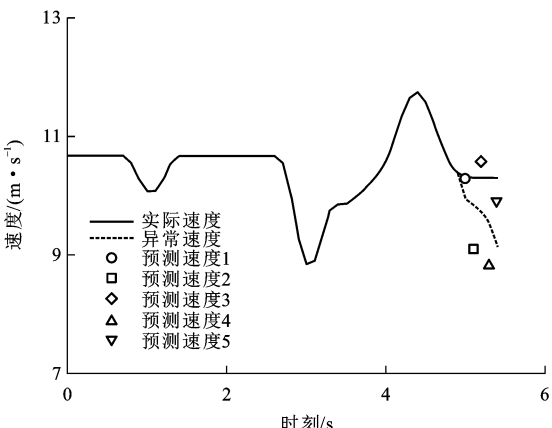
甄别结果	低偏移幅度		高偏移幅度	
	异常数	准确率/%	异常数	准确率/%
类型 1	456	97.0	470	100.0
类型 2	470	100.0	470	100.0
类型 3	470	100.0	470	100.0
类型 4	459	97.7	470	100.0
类型 5	227	48.3	468	99.6
正常数据	异常数	5	误报率/%	1.1

显然,异常类型 5 条件下出现低偏移幅度异常时,由于异常的车辆速度曲线与实际车辆可能的速度曲线较为相似,因此,容易出现漏检的情况。

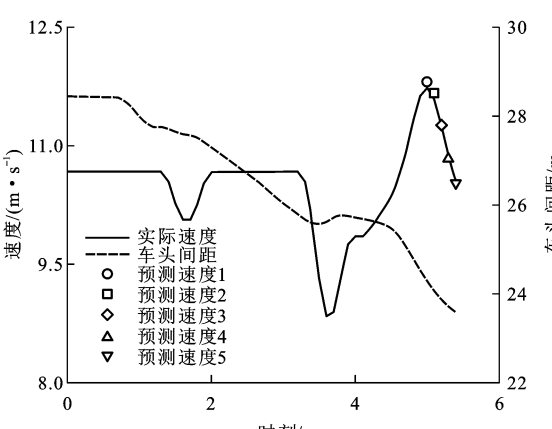
图 9(a)、(b)分别给出了编号为 100 的车辆在类型 1 和 4 条件下低偏移幅度异常时的 2 个实例,可以看出:预测速度较大幅度地偏离系统所采集的速度数据,于是均被识别为异常。图 9(c)给出了该车辆在由加速转为减速时的正常数据的识别实例,可



(a) 类型1:低偏移幅度异常



(b) 类型4:低偏移幅度异常



(c) 类型11:正常数据无异常

图 9 100 号车甄别结果

Fig. 9 Discrimination results for vehicle No. 100

以看到:预测速度与系统所采集的速度数据较为接近,因此,未出现误报;同时其解释了急加速、急减速等情况下如何避免短时数据预测错误的问题,一方面,LSTM神经网络可以通过输入数据中车头间距的快速减小来预判后车在短暂延时后会做出的减速反应,另一方面,甄别模型综合了 N 个连续采样时刻的数据进行甄别,弱化了某时刻出现的特殊动作所带来的偏差的影响。

3.3.3 整体分析

以数据集 1、9 和 11 为例,结合启动模块和可信甄别模块的试验结果,可分析可信甄别模型的整体效果:使用数据集 1 的 700 组异常数据对方法进行检验时,模型对其中 4 组测试数据对应的跟驰状态识别错误,对 14 组测试数据对应的速度误判为正常,误判次数合计为 18 次,准确率即为 97.4%;使用数据集 9 的 700 组异常数据对方法进行检验时,模型对其中 4 组测试数据对应的跟驰状态识别错误,对 11 组测试数据对应的速度误判为正常,误判次数合计为 15 次,准确率为 97.8%;使用数据集 11 的 700 组正常数据对方法进行检验时,模型对其中 4 组测试数据对应的跟驰状态识别错误,对 5 组测试数据对应的速度误判为异常,误判次数合计为 9 次,准确率为 98.7%。可见,使用多个测试数据集对方法进行检验,其准确率均可达 97% 以上,说明提出的可信甄别模型在大多数异常情况下可取得良好的效果。

4 结 语

(1)构建了基于 SVM-LSTM 的车辆跟驰行为状态识别与信息可信甄别模型,包括基于 SVM 的车辆跟驰行为识别模型和基于 LSTM 神经网络的车辆跟驰速度预测模型。基于 SVM 的车辆跟驰行驶状态识别算法准确率达到 99% 以上,基于 LSTM 神经网络的车辆跟驰速度预测算法平均误差达到 $\text{cm} \cdot \text{s}^{-1}$ 的数量级,算法效果较好。

(2)将基于 SVM 的车辆跟驰行为识别和基于 LSTM 神经网络的车辆跟驰速度预测两部分算法相结合,可以通过对交通业务特征的分析实现车辆行驶数据的可信甄别。经正常数据集和不同类型异常数据集的测试,其可信甄别结果准确率达到 97%,表明提出的可信甄别模型及其流程可在多种场景下使用,包括可用于路侧设备(Road Side Units, RSUs)对车载单元(On-Board Units, OBUs)实时信息的可信甄别,也可用于在缺少 RSUs 的场景中

车载单元 OBUs 间对实时信息的可信甄别。

(3)本文讨论的跟驰状态下车辆行驶数据的可信甄别方法也可拓展到换道^[27-28]和超车^[29]状态下车辆行驶行为识别及其信息可信甄别;同时,还可与车联网直连通信系统中的公钥基础设施机制^[30]相结合,构建起道路交通应用系统的可信甄别模型与方法,以实现车辆信息可信交互的分层次安全验证和甄别。

参 考 文 献 :

References :

- [1] 张 毅,姚丹亚,李 力,等.智能车路协同系统关键技术与应用[J].交通运输系统工程与信息,2021,21(5):40-51.
ZHANG Yi, YAO Dan-ya, LI Li, et al. Technologies and applications for intelligent vehicle-infrastructure cooperation systems[J]. Transportation Systems Engineering and Information Technology, 2021, 21(5): 40-51. (in Chinese)
- [2] MIYAJIMA C, NISHIWAKI Y, OZAWA K, et al. Driver modeling based on driving behavior and its evaluation in driver identification[J]. Proceedings of the IEEE, 2007, 95(2): 427-437.
- [3] WANG Da-jun, PEI Xin, LI Li, et al. Risky driver recognition based on vehicle speed time series[J]. IEEE Transactions on Human-Machine Systems, 2018, 48(1): 63-71.
- [4] NEWELL G F. A simplified car-following theory: a lower order model[J]. Transportation Research Part B: Methodological, 2002, 36(3): 195-205.
- [5] YEO H. Asymmetric microscopic driving behavior theory[D]. Berkeley: University of California, 2008.
- [6] WEI Da-li, LIU Hong-chao. Analysis of asymmetric driving behavior using a self-learning approach[J]. Transportation Research Part B: Methodological, 2013, 47: 1-14.
- [7] WONG W, HUANG Shi-hong, FENG Yi-heng, et al. Trajectory-based hierarchical defense model to detect cyber-attacks on transportation infrastructure[R]. Washington DC: Transportation Research Board, 2019.
- [8] HUANG S E, FENG Yi-heng, LIU H X. A data-driven method for falsified vehicle trajectory identification by anomaly detection[J]. Transportation Research Part C: Emerging Technologies, 2021, 128: 103196.
- [9] 祝俐菱,刘 澜,赵新朋,等.基于支持向量机的车辆驾驶行为识别研究[J].交通运输系统工程与信息,2017,17(1):91-97.
ZHU Li-ling, LIU Lan, ZHAO Xin-peng, et al. Driver behavior recognition based on support vector machine[J]. Journal of Transportation Systems Engineering and Information Technology, 2017, 17(1): 91-97. (in Chinese)
- [10] 戴晓峰,卢梦媛,覃文文.货车移动遮断影响下的小客车驾驶行为识别[J].交通运输系统工程与信息,2021,21(5):174-182.
JI Xiao-feng, LU Meng-yuan, QIN Wen-wen, et al. Passenger cars driving behaviors recognition under truck movement interruption[J]. Journal of Transportation Systems Engineering

- and Information Technology, 2021, 21(5): 174-182. (in Chinese)
- [11] BRACKSTONE M, MCDONALD M. Car-following: a historical review[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 1999, 2(4): 181-196.
- [12] 王殿海,金 盛. 车辆跟驰行为建模的回顾与展望[J]. 中国公路学报, 2012, 25(1): 115-127.
- WANG Dian-hai, JIN Sheng. Review and outlook of modeling of car following behavior [J]. China Journal of Highway and Transport, 2012, 25(1): 115-127. (in Chinese)
- [13] CHANDLER R E, HERMAN R, MONTROLL E W. Traffic dynamics; studies in car following[J]. Operations Research, 1958, 6(2): 165-184.
- [14] GIPPS P G. A behavioural car-following model for computer simulation[J]. Transportation Research Part B: Methodological, 1981, 15(2): 105-111.
- [15] KIKUCHI S, CHAKROBORTY P. Car-following model based on fuzzy inference system[J]. Transportation Research Record, 1992(1365): 82-91.
- [16] 杨 达,蒲 云,祝俐菱,等. 改进的基于安全距离的车辆跟驰模型[J]. 北京工业大学学报, 2013, 39(9): 1335-1342.
- YANG Da, PU Yun, ZHU Li-ling, et al. Improved safe distance car-following model [J]. Journal of Beijing University of Technology, 2013, 39(9): 1335-1342. (in Chinese)
- [17] 杨龙海,赵 顺,罗 沂. 基于期望跟驰间距的跟驰模型研究[J]. 武汉理工大学学报(交通科学与工程版), 2016, 40(4): 580-584, 591.
- YANG Long-hai, ZHAO Shun, LUO Yi. Car-following model based on expected headway[J]. Journal of Wuhan University of Technology (Transportation Science and Engineering), 2016, 40(4): 580-584, 591. (in Chinese)
- [18] 杨龙海,王 晖,李 帅,等. 改进分子动力学的车辆跟驰模型[J]. 重庆大学学报, 2021, 44(7): 26-33.
- YANG Long-hai, WANG Hui, LI Shuai, et al. Car-following model with improved molecular dynamics [J]. Journal of Chongqing University, 2021, 44(7): 26-33. (in Chinese)
- [19] 邱小平,孙若晓,杨 达,等. 考虑速度对期望间距影响的车辆跟驰模型研究[J]. 计算机应用研究, 2015, 32(6): 1649-1652.
- QIU Xiao-ping, SUN Ruo-xiao, YANG Da, et al. Consider impact of speed on desired distance car-following model[J]. Application Research of Computers, 2015, 32(6): 1649-1652. (in Chinese)
- [20] ELAMRANI ABOU ELASSAD Z, MOUSANNIF H, AL MOATASSIME H, et al. The application of machine learning techniques for driving behavior analysis: a conceptual framework and a systematic literature review[J]. Engineering Applications of Artificial Intelligence, 2020, 87: 103312.
- [21] WANG Xiao, JIANG Rui, LI Li, et al. Capturing car-following behaviors by deep learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(3): 910-920.
- [22] HUANG Xiu-ling, SUN Jie, SUN Jian. 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: 346-362.
- [23] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [24] HE Zheng-bing. Research based on high-fidelity NGSIM vehicle trajectory datasets: a review [R]. Beijing: Beijing University of Technology, 2017.
- [25] THIEMANN C, CHRISTIAN M, KESTING A, et al. Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data [J]. Transportation Research Record, 2008(2088): 90-101.
- [26] 李 依,赵建有,叶云涛. NGSIM轨迹重构研究综述[J]. 汽车实用技术, 2021, 46(4): 204-208.
- LI Yi, ZHAO Jian-you, YE Yun-tao. A review of NGSIM trajectory reconstruction [J]. Automobile Applied Technology, 2021, 46(4): 204-208. (in Chinese)
- [27] DING Chen-xi, WANG Wu-hong, WANG Xiao, et al. A neural network model for driver's lane-changing trajectory prediction in urban traffic flow[J]. Mathematical Problems in Engineering, 2013, 2013: 967358.
- [28] TANG Jun-jun, LIU Fang, ZHANG Wen-hui, et al. Lane-changes prediction based on adaptive fuzzy neural network[J]. Expert Systems with Applications, 2018, 91: 452-463.
- [29] 许伦辉,胡三根,伍 帅,等. 考虑车辆运行特性的双车道超车模型[J]. 华南理工大学学报(自然科学版), 2015, 43(4): 7-13, 27.
- XU Lun-hui, HU San-gen, WU Shuai, et al. Overtaking model for two-lane highway considering vehicle running characteristics[J]. Journal of South China University of Technology (Natural Science Edition), 2015, 43(4): 7-13, 27. (in Chinese)
- [30] 李 峰,陈 新. 基于LTE-V2X技术的PKI系统架构探讨[J]. 信息技术与网络安全, 2020, 39(7): 41-47.
- LI Feng, CHEN Xin. PKI infrastructure architecture based on LTE-V2X[J]. Information Technology and Network Security, 2020, 39(7): 41-47. (in Chinese)