

DOI:10.11918/202205090

集成学习框架下的车辆跟驰行为建模

李立¹, 李仕琪¹, 徐志刚², 李光泽¹, 汪贵平¹

(1. 长安大学 电子与控制工程学院, 西安 710064; 2. 长安大学 信息工程学院, 西安 710064)

摘要: 为了提高复杂行驶环境下车辆跟驰行为预测精度, 提出了一种集成学习框架下融合理论驱动模型和数据驱动模型的车辆跟驰行为建模方法。基于 stacking 集成学习框架, 选择理论驱动的智能驾驶模型 (IDM)、考虑车辆队列和周围行驶条件因素的数据驱动的长短期记忆 (LSTM) 网络和门控循环单元 (GRU) 网络作为跟驰行为特征的一级学习算法, 选择 3 种线性和 8 种非线性回归方法作为备选二级学习算法来融合一级学习器的输出特征。通过对比使用实际车辆轨迹数据计算的模型预测精度, 确定了最优模型。研究结果表明: 包含车辆队列和周围行驶条件变量的数据驱动跟驰模型比 IDM 模型的预测精度更高; 多数情况下采用非线性二级学习算法的融合跟驰模型的预测精度高于 IDM 模型、数据驱动跟驰模型以及采用线性二级学习算法的融合跟驰模型; 分别采用 GBRT 回归和随机森林回归作为二级学习算法的 IDM-LSTM-stacking 模型和 IDM-GRU-stacking 模型具有最高的预测精度; 外界干扰下的融合跟驰模型稳定性优于单一的理论和数据驱动跟驰模型。集成学习为驾驶行为建模提供了新方法。

关键词: 交通工程; 跟驰模型; 集成学习; 理论驱动模型; 数据驱动模型

中图分类号: U491.252

文献标志码: A

文章编号: 0367-6234(2024)03-0046-10

Modeling of car-following behavior under an ensemble learning framework

LI Li¹, LI Shiqi¹, XU Zhigang², LI Guangze¹, WANG Guiping¹

(1. School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China;

2. School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: In order to improve prediction accuracy of vehicle's car-following behavior in the complex environment, a car-following behavior modeling method is proposed under an ensemble learning framework to integrate theory-driven model and data-driven model. Based on the stacking ensemble learning framework, a theory-driven model, intelligent driver model (IDM), and two data-driven models, long-short term memory (LSTM) network and gate recurrent unit (GRU) network considering the factors of vehicle platoon and surrounding driving condition, are selected as the first-level learning algorithms of car-following behavior features. Three linear and eight nonlinear regression methods are taken as the candidates of second-level learning algorithms to integrate the features of the first-level learner outputs. By comparing trajectory prediction accuracy calculated from real vehicle trajectory data, the optimal car-following model is determined. Results show that the data-driven car-following models, which take the variables of vehicle platoon and surrounding driving condition into account, have the higher trajectory prediction accuracy than the IDM. In most cases, the ensemble learning car-following models with the nonlinear second-level learning algorithm predicts vehicle trajectory in the higher accuracy than the IDM, the data-driven models and the ensemble learning car-following models with the linear second-level learning algorithm. The IDM-LSTM-stacking model, which uses the GBRT regression as second-level learning algorithm, and the IDM-GRU-stacking model, which uses the random forest regression as second-level learning algorithm, give the highest trajectory prediction accuracy. The stability of integrated car-following model under external disturbance is prior to single theory-driven model and data-driven model. The ensemble learning method provides a new approach for modeling driving behaviors.

Keywords: traffic engineering; car-following model; ensemble learning; theory-driven model; data-driven model

车辆跟驰模型描述了单车道上相邻两车之间的运动关系, 是微观交通流仿真、道路通行能力分析、自动驾驶算法等方面研究的基础模型之一, 在交通工程学理论体系中占据重要地位^[1]。现有跟驰模

收稿日期: 2022-05-24; 录用日期: 2022-09-01; 网络首发日期: 2024-03-21

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20240320.1819.006>

基金项目: 国家重点研发计划(2019YFB1600100); 国家自然科学基金(71901040); 陕西省自然科学基金(2021JC-28)

作者简介: 李立(1985—), 男, 博士, 副教授

通信作者: 汪贵平, gpwang@chd.edu.cn

型可以分为理论驱动模型和数据驱动模型两大类^[2]。理论驱动模型采用严格的数学表达式描述跟驰车辆的动力学机制,具有精确的物理意义,能够确保在合理的参数设定下具有明确稳定的输出,代表模型包括 GM (general motor) 模型、Gipps 模型和 IDM (intelligent driver model) 模型。实际交通流中,车辆跟驰行为受驾驶风格、交通环境、车辆类型等因素的影响呈现一定的随机性和非线性。理论驱动模型结构相对简单,模型参数较少,使其难以描述实际复杂交通流中的车辆跟驰行为,从而导致对跟驰行为的预测失准或仿真失真。针对这种情况,一种常见的解决方法是采用不同交通流状态下的实测数据标定跟驰模型,并为其设置分类应用条件,但是交通流状态划分方法、跟驰模型参数标定方法、预测结果评价指标等因素均可能影响跟驰模型的应用效果,提高模型应用的复杂度。

与理论驱动跟驰模型相比,近年来有较多研究采用数据驱动的方法构建跟驰模型。数据驱动跟驰模型利用数据科学与机器学习等理论和方法,从真实车辆行驶数据中挖掘跟驰行为特征,利用模型参数多、非线性规律拟合能力强的特点,学习数据中存在的复杂行为规律,使其经常获得比理论驱动模型更好的跟驰行为预测效果。文献[2]从模糊逻辑、人工神经网络、实例学习、支持向量回归、深度学习等5个方面对数据驱动跟驰模型的研究进行了综述。文献[3]将数据驱动跟驰模型分为基于神经网络、支持向量回归和 K 近邻等传统机器学习方法的模型和基于深度学习的模型。但是,数据驱动跟驰模型的参数连接关系缺乏明确的交通物理含义,且难以判断模型输入是否超出了安全驾驶阈值,这些缺陷可能影响模型预测结果可靠性并危及行车安全。通过近年来多次出现自动驾驶车辆安全事故^[4-5],也说明单纯依靠数据驱动方式进行车辆跟驰行为建模和应用存在潜在安全风险。

为了综合利用上述两类跟驰模型的优点,有学者提出了理论与数据混合驱动的跟驰模型。文献[6]采用线性函数融合了理论驱动与数据驱动跟驰模型,发现融合模型比非融合模型具有更高的跟驰行为预测精度。文献[7]采用自适应卡尔曼滤波方法组合了多种跟驰模型,找到了预测精度高的融合跟驰模型。文献[8]利用最优加权组合理论,建立 IDM 模型与径向基函数神经网络融合的跟驰模型。然而,现有理论与数据混合驱动跟驰模型所采用的融合方法通常直接对两类模型的输出进行线性加和,并未对两类模型的参数和权重进行深度融合,导致两类模型的优势未得到充分的发挥和利用。

除了与理论驱动跟驰模型融合之外,另一种提升数据驱动跟驰模型表现的途径在于合理增加模型变量。微观驾驶行为可能受到周边车辆行为、道路设施条件、车内外交通信息等因素的影响^[2],然而现有多数数据驱动跟驰模型中纳入的变量与理论驱动跟驰模型类似,这可能导致模型欠拟合,因此有必要进一步从车辆轨迹数据中提取更多能够反映真实驾驶过程中驾驶人观察信息的变量,并将其引入数据驱动跟驰模型中^[3]。

综上,本文提出了一种基于集成学习框架的理论和数据驱动跟驰模型融合方法。利用集成学习框架可以综合应用多种同质或异质基学习器的优势,通过构建多级学习器对理论和数据驱动跟驰模型进行组合、训练和应用,提高模型的预测精度和泛化能力。同时将跟驰车辆所在队列状态及其周边行驶条件因素引入数据驱动跟驰模型中,以进一步提升模型学习复杂交通条件下车辆跟驰行为的能力。使用真实车辆轨迹数据集对所构建模型进行训练和测试,来验证所提出方法的有效性。本文为复杂交通流条件下的车辆跟驰行为建模提供了新的思路,研究成果有助于提升跟驰行为预测和微观交通仿真精度。

1 数据与变量

1.1 数据

本文使用 highD 自然驾驶轨迹数据集^[9]作为基础数据。此数据集由无人机在德国高速公路6个路段上空拍摄的十余个小时的视频中提取的车辆轨迹数据组成,数据字段包括车辆经纬度、所在车道、行驶方向、外形尺寸、类型、速度、加速度、周边车辆编号等。数据采集地点为平直路段,上下游匝道距离较远,因此车辆换道现象较少,多数车辆处于跟驰或自由行驶状态。本文使用 IDM 模型作为理论驱动跟驰模型的代表,此模型具有明确的物理意义且参数相对较少,广泛用于多种交通状态下的跟驰行为建模,以及网联自动驾驶控制算法中,具有较强的适用性及可拓展性。IDM 模型变量包括跟驰车速度、与前导车的间距以及与前导车的速度差等,这些变量值可从数据集中直接计算获得,同时模型参数将参考数据集中特定字段的统计值设定。数据驱动模型中除了纳入上述变量外,其他纳入变量的计算相对复杂,详述如下。

1.2 车辆队列

在车流密度较高的条件下,同一车道上密集接续行驶的多辆车在运动方式上经常表现出一定的相似性,会从客观上形成车辆队列,如图1所示。过往有学者^[10-11]将跟驰车辆前后多车的运动状态变量

纳入跟驰模型中。实际上,跟驰驾驶员通常难以直接准确观察和衡量前后多车运动参数,更合理的行为模式是跟驰驾驶员根据非直接相邻的前后车辆运动状态估计其所在队列整体运动状态,并基于此调整驾驶行为。因此,本文将跟驰车辆所在队列的平均运动状态纳入数据驱动跟驰模型中。本文通过设置车辆跟驰状态判断条件划分车辆队列。筛选 highD 数据集中换道数为 0、有效跟驰时长 16 s 以上、处于稳定跟驰状态即跟驰车辆速度与所在车队的平均速度平均差值不超过 2 m/s 的车辆轨迹,然后根据数据字段中的跟驰车辆编号确定其是否归入车辆队列。一共得到 1 241 个车辆队列,共 7 306 条

车辆轨迹,所有车队中车辆的平均间距和平均车速分别为 34.304 m 和 21.479 m/s,平均车头时距为 1.6 s,最大加速度为 5.02 m/s²,最大减速度为 -4.46 m/s²。描述车辆队列平均运动状态的变量包括车辆队列中的平均间距、平均车头时距和平均碰撞时间(time to collision)。设定 t 时刻第 i 个队列平均运动状态变量 $P_i(t)$ 为

$$P_i(t) = [\bar{d}_i(t), \bar{h}_i(t), \bar{c}_i(t)] \quad (1)$$

式中: $\bar{d}_i(t)$ 、 $\bar{h}_i(t)$ 和 $\bar{c}_i(t)$ 分别代表 t 时刻队列 i 内部所有车辆的车辆间距、车头时距和碰撞时间的平均值。

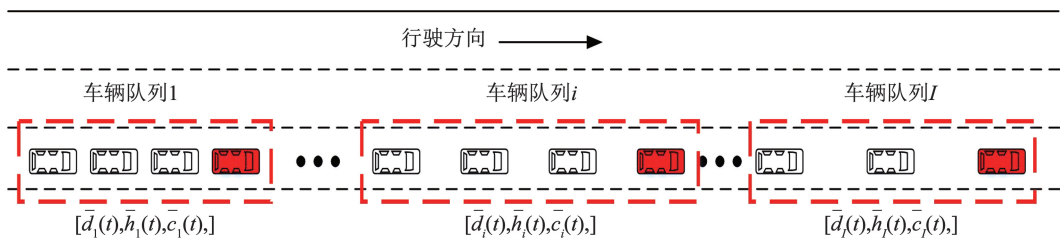


图 1 车辆队列及相关变量

Fig. 1 Vehicle platoon and related variables

1.3 周边行驶条件

除了本车道前后车辆的影响外,过往实证和仿真研究^[12-13]均发现跟驰车辆相邻车道的临近车辆也可能影响跟驰车辆的驾驶行为,影响因素包括车辆侧向间距、驾驶员换道超车意图等。为了描述车辆队列中跟驰车辆的周边行驶条件,定义 t 时刻队列 i 内部第 j 辆跟驰车周围行驶条件的变量 $F_{i,j}(t)$ 表达式为

$$F_{i,j}(t) = [\Delta l_{i,j}^n(t), \Delta v_{i,j}^n(t), \Delta a_{i,j}^n(t)] \quad (2)$$

式中: $\Delta l_{i,j}^n(t)$ 为 t 时刻队列 i 内部第 j 辆跟驰车与其第 n 辆临近车辆二者几何中心的距离($n = 1 \sim 6$), $\Delta v_{i,j}^n(t)$ 为 t 时刻第 n 辆临近车辆与跟驰车辆 j 的速度差, $\Delta a_{i,j}^n(t)$ 为第 n 辆临近车辆与跟驰车辆 j 的加速度差。图 2 所示为跟驰车辆周边 6 个方向上临近车辆的空间位置。

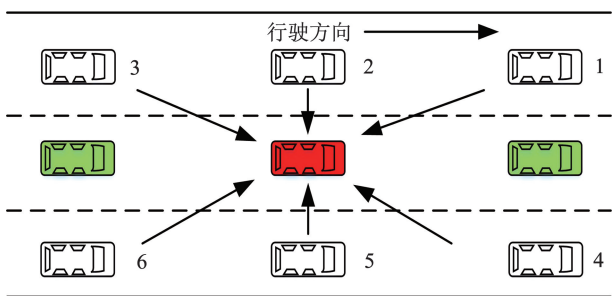


图 2 周边车辆及相关变量

Fig. 2 Surrounding vehicles and related variables

2 集成学习框架下的跟驰行为建模

2.1 stacking 集成学习方法

集成学习是一种通过构建和结合多个学习器,降低单一学习器灵敏度不同带来的误差,提高模型泛化能力、鲁棒性和预测精度的机器学习方法。常见的集成学习方法包括自助聚合(bagging)法、提升(boosting)法、堆叠(stackings)法。由于 bagging 法和 boosting 法主要用于同质学习器的集成,而理论和数据驱动跟驰模型的建模机制区别明显,因此本文引入面向异质学习器的 stacking 法,将跟驰行为建模划分成两个阶段,针对不同阶段的建模需求分别构建一级学习器和二级学习器,对理论和数据驱动跟驰模型进行集成与融合。首先,将数据集划分成 3 个子集,分别是训练集、验证集和测试集。接着,将理论驱动模型和数据驱动模型作为两个一级学习算法,使用训练集对其进行训练,即标定理论驱动模型的参数和训练数据驱动模型的网络权重,构成一级学习器。然后,使用验证集作为一级学习器的输入,将获得的输出与验证集中的数据标签组合,生成一个新的数据集。接着,使用新数据集训练二级学习算法,获得二级学习器的网络权重和模型参数。最后,将测试集作为训练后的一级和二级学习器的输入,对比测试集成跟驰模型的预测效果。此方法的伪代码如下所示。流程如图 3 所示。

定义:数据集 $D_r = \{(x_{r,m}, y_{r,m})\} (r = 1 \sim 3)$,其中 D_1 、 D_2 和 D_3 分别为训练集、验证集和测试集,样本编号 $m = 1 \sim M_r$,其中 M_1 、 M_2 和 M_3 分别为子集 D_1 、 D_2 和 D_3 的样本数量, $x_{r,m}$ 和 $y_{r,m}$ 分别为不同子集中样本的数值与标签;新的训练集 $D'_2 = \varnothing$;一级学习算法 ζ_1 、 ζ_2 ,一级学习器 θ_1 、 θ_2 ;二级学习算法 μ ,二级学习器 β 。

步骤:
for $q = 1, 2$ do:
 训练 $\theta_q: \theta_q = \zeta_q(D_1)$

end for
for $m = 1, \dots, M_2$ do:
 for $q = 1, 2$ do:
 θ_q 的输出为: $z_{2,m}^q = \theta_q(x_{2,m})$
 end for
 生成新训练集 $D'_2: D'_2 = D'_2 \cup \{(z_{2,m}^1, z_{2,m}^2), y_{2m}\}$
end for
训练 $\beta: \beta = \mu(D'_2)$
输出:集成模型 $H(x_{3m}): H(x_{3m}) = \beta(\theta_1(x_{3,m}), \theta_2(x_{3,m}))$

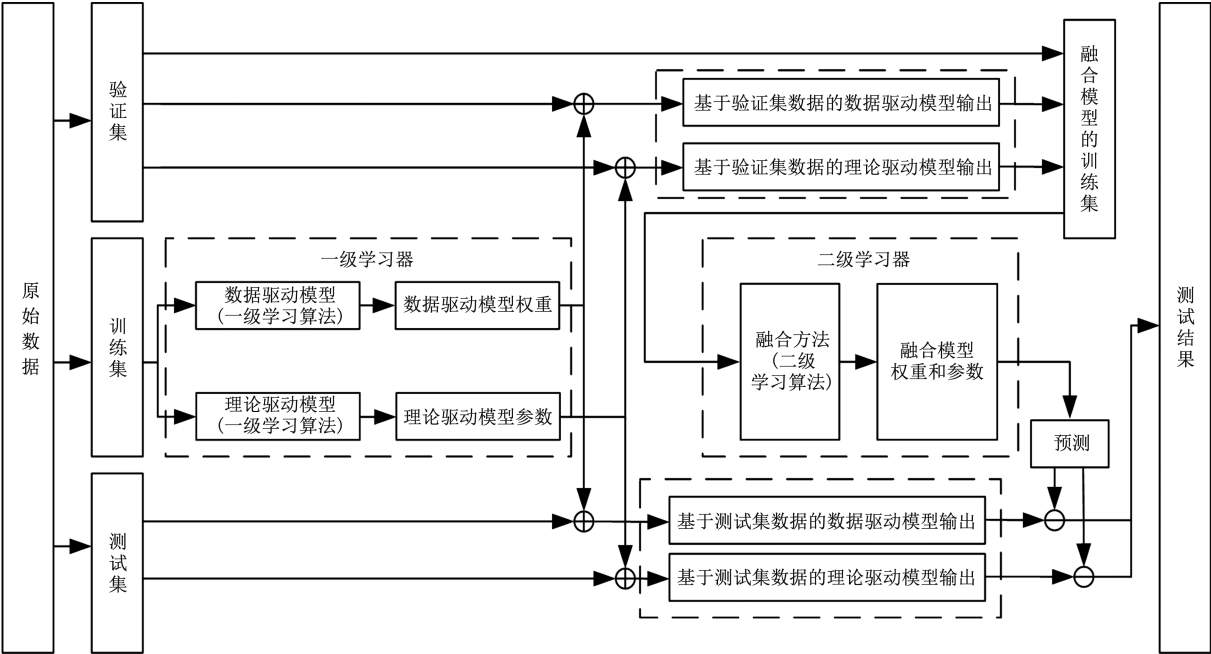


图3 集成学习框架下的跟驰行为建模方法

Fig. 3 Modelling method of car-following behavior under ensemble learning framework

其中,一级学习算法的设置详见2.2节和2.3节。为了对一级学习器的输出特征进行再学习,在二级学习算法中可以采用线性或非线性方法对一级学习器的输出进行再学习。本文选择了表1所示的11种线性 and 非线性方法作为二级学习算法,将通过对比测试选择性能最佳的二级学习算法。

表1 二级学习算法分类

类型	序号	名称
线性方法	1	算数平均值法
	2	泰尔森估算 ^[14]
	3	随机抽样一致性算法 ^[15]
非线性方法	1	决策树回归 ^[16]
	2	SVM 回归 ^[17]
	3	KNN 回归 ^[18]
	4	随机森林回归 ^[19]
	5	Adaboost 回归 ^[20]
	6	GBRT 回归 ^[21]
	7	Bagging 回归 ^[22]
	8	极端随机树回归 ^[23]

2.2 理论驱动跟驰模型

本文选择 IDM 模型作为 stacking 集成学习框架下的一级学习算法之一。IDM 模型描述了跟驰车的加速度与速度、与前导车的间距以及速度差的关系。将 IDM 模型应用于本文研究场景中,可得 t 时刻队列 i 内部第 j 辆跟驰车的加速度为

$$a_{i,j}(t) = a_{\max} \left[1 - \left(\frac{v_{i,j}(t)}{v_e} \right)^\delta \right] - a_{\max} \left(\frac{d^*(t)}{d_{i,j,j-1}(t)} \right)^2$$

(3)

式(3)的前半部分为第 j 辆跟驰车的自由加速度,其与最大加速度 $a_{\max}$ 、期望速度 v_e 以及 t 时刻车速 $v_{i,j}(t)$ 有关,其中 δ 为加速度指数。式(3)的后半部分为跟驰车的减速策略,当跟驰车与前导车间距过小时,该策略被触发,其中 $d_{i,j,j-1}(t)$ 为 t 时刻第 j 辆跟驰车与其前导车之间的距离, $d^*(t)$ 为 t 时刻的期望最小间距,可表示为

$$d^*(t) = d_{\min} + v_{i,j}(t) h_\sigma + \frac{v_{i,j}(t) \Delta v_{i,j,j-1}(t)}{2 \sqrt{a_{\max} \Delta b}}$$

(4)

式中: h_σ 为安全车头时距, $d_{\min}$ 为拥堵间距, b 为驾驶员期望减速度, $\Delta v_{i,j-1}(t)$ 为 t 时刻第 j 辆跟驰车与其前导车之间的速度差。参考 1.2 节所述所纳入车辆运动状态信息。其中, $a_{\max}$ 为 5 m/s^2 , v_e 为 30 m/s , δ 为 4 , h_σ 为 1.5 s , $d_{\min}$ 为 2 m , b 为 4.5 m/s^2 。

2.3 数据驱动跟驰模型

由于跟驰驾驶员往往依赖历史驾驶行为和过去一段时间内的交通状态进行行驶决策, 即其决策行为存在记忆效应, 因此本文选择长短时记忆 (long-short term memory, LSTM) 网络^[24] 和门控循环单元

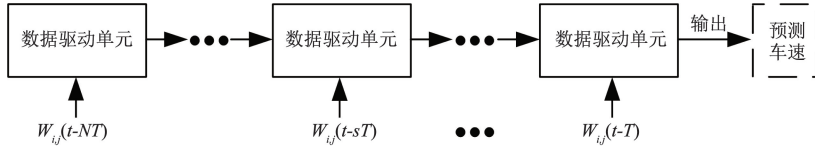


图 4 数据驱动跟驰模型的结构

Fig. 4 Structure of data-driven car-following model

定义 t 时刻影响第 i 个车辆队列中第 j 辆跟驰车辆行为的因素组合 $W_{i,j}(t)$ 为

$$W_{i,j}(t) = [P_i(t), F_{i,j}(t), G_{i,j-1,j}(t)] \quad (5)$$

式中: $P_i(t)$ 为车辆队列影响因素 (见式 (1)), $F_{i,j}(t)$ 为周围驾驶环境影响因素 (见式 (2)), $G_{i,j-1,j}(t)$ 为第 j 辆跟驰车与其前导车之间的时空关系, 定义为

$$G_{i,j-1,j}(t) = [d_{i,j-1,j}(t), \Delta v_{i,j-1,j}(t), v_{i,j}(t), a_{i,j}(t)] \quad (6)$$

综上, 作为数据驱动跟驰模型的输入, t 时刻之前历史时段内不同时间节点上跟驰车辆影响因素组合 $W_{i,j}^h(t)$ 为

$$W_{i,j}^h(t) = [W_{i,j}(t-T), \dots, W_{i,j}(t-sT), \dots, W_{i,j}(t-NT)] \quad (7)$$

式中: N 为所分析历史时段内时间节点数量, T 为此时段内相邻时间节点的间隔, s 为此时段内第 s 个历史时刻。

2.4 评价指标

使用 3 种评价指标测试所构建跟驰模型的性能, 分别是对称平均绝对百分比误差 (symmetric mean absolute percentage error, SMAPE)、平均绝对误差 (mean absolute error, MAE) 和平均绝对相对误差 (mean absolute relative error, MARE), 这些指标常用与评价跟驰模型对实际车辆轨迹的预测精度^[3,6,26], 分别为

$$\text{SMAPE} = \frac{1}{M} \sum_{m=1}^M \frac{2 |v_j(t) - \hat{v}_j(t)|}{(|v_j(t)| + |\hat{v}_j(t)|)} \quad (8)$$

$$\text{MAE} = \frac{1}{M} \sum_{m=1}^M |v_j(t) - \hat{v}_j(t)| \quad (9)$$

(gate recurrent unit, GRU) 网络^[25] 对跟驰行为进行建模, 并将其作为 stacking 集成学习框架下的另一种一级学习算法。LSTM 网络在其单元之间传递信息时, 分别使用单元状态和隐藏状态存储长期记忆和短期记忆。GRU 网络可看作是 LSTM 网络的一种变体, 其结构比 LSTM 网络简单, 不包含单元状态, 使用隐藏状态传递信息。图 4 所示为数据驱动跟驰模型的结构。模型训练中采用均方误差 (mean-square error, MSE) 作为回归损失函数, Sigmoid 作为激活函数, Adam 作为优化算法。

$$\text{MARE} = \frac{1}{M} \sum_{m=1}^M \frac{|v_j(t) - \hat{v}_j(t)|}{v_j(t)} \quad (10)$$

式中: M 表示测试数据的样本个数, $v_j(t)$ 为 t 时刻第 j 辆跟驰车的真实车速, $\hat{v}_j(t)$ 为 t 时刻 d_j 的预测车速。

3 计算结果与分析

3.1 一级学习器

通过两组试验确定作为一级学习器的数据驱动跟驰模型的网络结构和训练过程的学习率, 并采用一组试验对比理论驱动模型和数据驱动模型的性能。

1) 网络结构。LSTM 网络和 GRU 网络包含有输入层、隐藏层和输出层, 其中输入层和输出层的维数分别根据跟驰模型中所纳入的影响因素数量和模型输出的预测值数量确定。本文为这两种网络设置了单一隐藏层, 以避免传统递归式结构网络可能发生的梯度弥散问题。表 2 报告了采用不同隐藏层神经元数量的数据驱动跟驰模型性能, 可以发现, 采用结构 1、2、3 的模型隐藏层神经元个数过少, 导致学习效果不如采用结构 4、5 的模型。由于增加隐藏层神经元个数会延长模型解算时间, 综合对比采用结构 4、5、6 的模型性能, 最终选定结构 5 为最优模型结构。

2) 学习率。学习率是训练数据驱动模型的重要参数。表 3 中报告了不同学习率下具有最优结构的模型的性能。可以发现, 由于学习率过高使得算法很难收敛到最低点, 采用学习率 1、2 的模型误差较大。与之相比, 采用学习率 3、4 的模型误差明显

降低。由于采用学习率 4 的模型训练时间较长,且学习效果与采用学习率 3 的模型接近,因此选择学习率 3 为最优学习率。

表 2 不同结构数据驱动模型性能
Tab. 2 Performances of data-driven model with different structures

模型结构 编号	隐藏层神经 元数量	模型	SMAPE/%	MAE	MARE
1	16	LSTM	0.211	0.194	0.008
		GRU	0.202	0.187	0.008
2	32	LSTM	0.411	0.380	0.016
		GRU	0.197	0.181	0.008
3	64	LSTM	0.188	0.173	0.008
		GRU	0.191	0.177	0.008
4	128	LSTM	0.173	0.159	0.007
		GRU	0.164	0.149	0.007
5	256	LSTM	0.155	0.142	0.006
		GRU	0.151	0.138	0.006
6	512	LSTM	0.206	0.189	0.008
		GRU	0.130	0.115	0.005

3)对比模型。上述试验结果表明,使用网络结构 5 且设置学习率为 0.001 时,数据驱动模型能够获得最佳跟驰行为预测精度。采用相同的数据对 IDM 模型进行标定和测试,发现 IDM 模型的跟驰行为预测精度为 SMAPE = 0.740%, MAE = 0.427, MARE = 0.020。对比二者模型性能可以发现,具有最优网络结构和学习率的数据驱动跟驰模型的预测

精度优于 IDM 模型。

表 3 不同学习率下数据驱动模型性能
Tab. 3 Performances of data-driven models with different learning rates

学习率编号	学习率数值	模型	SMAPE/%	MAE	MARE
1	0.1	LSTM	0.844	0.784	0.033
		GRU	0.805	0.749	0.031
2	0.01	LSTM	0.776	0.723	0.030
		GRU	0.779	0.725	0.030
3	0.001	LSTM	0.150	0.132	0.006
		GRU	0.130	0.115	0.005
4	0.000 1	LSTM	0.160	0.147	0.006
		GRU	0.161	0.147	0.006

3.2 融合模型

对于在 stacking 集成学习框架下获得的两个融合跟驰模型 IDM-LSTM-stacking 和 IDM-GRU-stacking,表 4 给出了采用均值法、泰尔森估算和随机采样一致性线性回归 3 种线性二级学习算法后的模型预测结果。为了便于对比,表中同时给出了单独使用 IDM 模型和数据驱动模型的预测结果。可以发现,IDM 模型的预测误差大于 GRU 和 LSTM 模型的预测误差,而 IDM-LSTM-stacking 和 IDM-GRU-stacking 模型的误差在理论驱动模型和数据驱动模型之间。这说明采用表中 3 种线性二级学习算法的模型并未能很好地对一级学习器的输出特征进行再学习,而其效果类似于对理论和数据驱动模型的输出进行了折中。

表 4 3 类跟驰模型性能
Tab. 4 Performances of three types of car-following models

模型类型	模型名称	二级学习算法	SMAPE/%	MAE	MARE
理论驱动模型	IDM		0.740	0.427	0.021
数据驱动模型	GRU		0.131	0.116	0.005
	LSTM		0.150	0.133	0.006
融合模型	IDM-LSTM-stacking	均值法	0.294	0.250	0.012
		泰尔森估算	0.438	0.369	0.018
		随机抽样一致性算法	0.436	0.299	0.014
	IDM-GRU-stacking	均值法	0.293	0.249	0.012
		泰尔森估算	0.387	0.327	0.016
		随机抽样一致性算法	0.365	0.260	0.013

表 5 中给出了采用表 1 中 8 种非线性方法作为二级学习算法的融合跟驰模型的预测精度。对比表 4 中的结果可以发现,对于同样的输入数据,采用

非线性二级学习算法的模型预测精度明显比 IDM 模型和采用线性二级学习算法的模型更高,也高于单纯的数据驱动模型。对于 IDM-LSTM-stacking 模

型来说,采用非线性方法 6 即 GBRT 回归法的模型预测精度最高;对于 IDM-GRU-stacking 模型来说,采用非线性方法 4 即随机森林回归法的模型预测精度最高。上述结果说明,两种一级学习器即理论驱

动跟驰模型和数据驱动跟驰模型的输出特征存在着复杂的非线性关系,需要采用非线性二级学习算法才能更好地对它们的输出特征进行再学习。

表 5 采用非线性二级学习算法的跟驰模型性能

Tab. 5 Performances of car-following model with non-linear second-level learner

模型	二级学习算法	SMAPE/%	MAE	MARE
IDM-LSTM-stacking	非线性方法 1	0.163	0.143	0.007
	非线性方法 2	0.216	0.186	0.009
	非线性方法 3	0.122	0.106	0.005
	非线性方法 4	0.117	0.103	0.005
	非线性方法 5	0.145	0.128	0.006
	非线性方法 6	0.115	0.101	0.005
	非线性方法 7	0.117	0.102	0.005
	非线性方法 8	0.149	0.131	0.006
IDM-GRU-stacking	非线性方法 1	0.135	0.119	0.005
	非线性方法 2	0.215	0.186	0.009
	非线性方法 3	0.121	0.105	0.005
	非线性方法 4	0.115	0.101	0.005
	非线性方法 5	0.127	0.111	0.005
	非线性方法 6	0.118	0.104	0.005
	非线性方法 7	0.117	0.103	0.005
	非线性方法 8	0.135	0.119	0.005

图 5 对比了 IDM 模型、GRU 模型、LSTM 模型、采用随机森林回归方法的 IDM-GRU-stacking 模型和采用 GBRT 回归方法的 IDM-LSTM-stacking 模型预测的车辆轨迹和真实车辆的轨迹。可以发现,采

用随机森林回归方法的 IDM-GRU-stacking 模型和采用 GBRT 回归方法的 IDM-LSTM-stacking 模型比 IDM 模型、GRU 模型和 LSTM 模型的轨迹预测精度高。

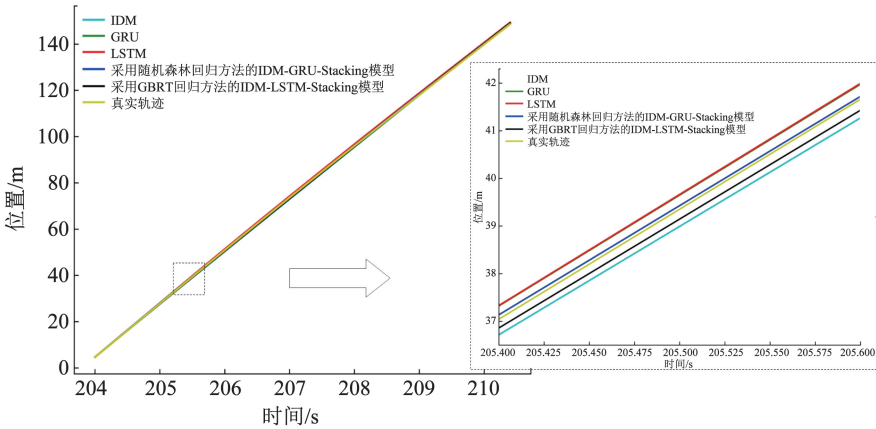


图 5 真实轨迹与 3 类跟驰模型预测轨迹对比

Fig. 5 Comparison of real trajectory and predicted trajectories of three types of car-following models

为了对比验证所构建集成学习跟驰模型稳定性及其对宏观交通流状态的影响,设计了如下数值模拟实验。假定由 100 辆跟驰车组成的车队在长度为 2 000 m 的三车道环路的中间车道上行驶,三车道环

路的左右两侧车道交通流随机生成。跟驰车队的初始状态为稳定均衡状态,即车辆等间距分布,任意两车初始车头间距为 20 m,所有车辆的车身长度为 5 m,初始速度 21.466 m/s,初始加速度为 0 m/s²,

其他仿真条件参考文中理论和数据驱动跟驰模型设置。实验中,在 $t=300\text{ s}$ 时向第一辆车施加扰动,令该车速度降为初始速度的一半即 10.733 m/s ,同时

位置向前移动 14 m 。
图 6 为数值模拟实验结果,图中第一列为各模型输出的车队行驶轨迹,此列各子图中竖线标记出

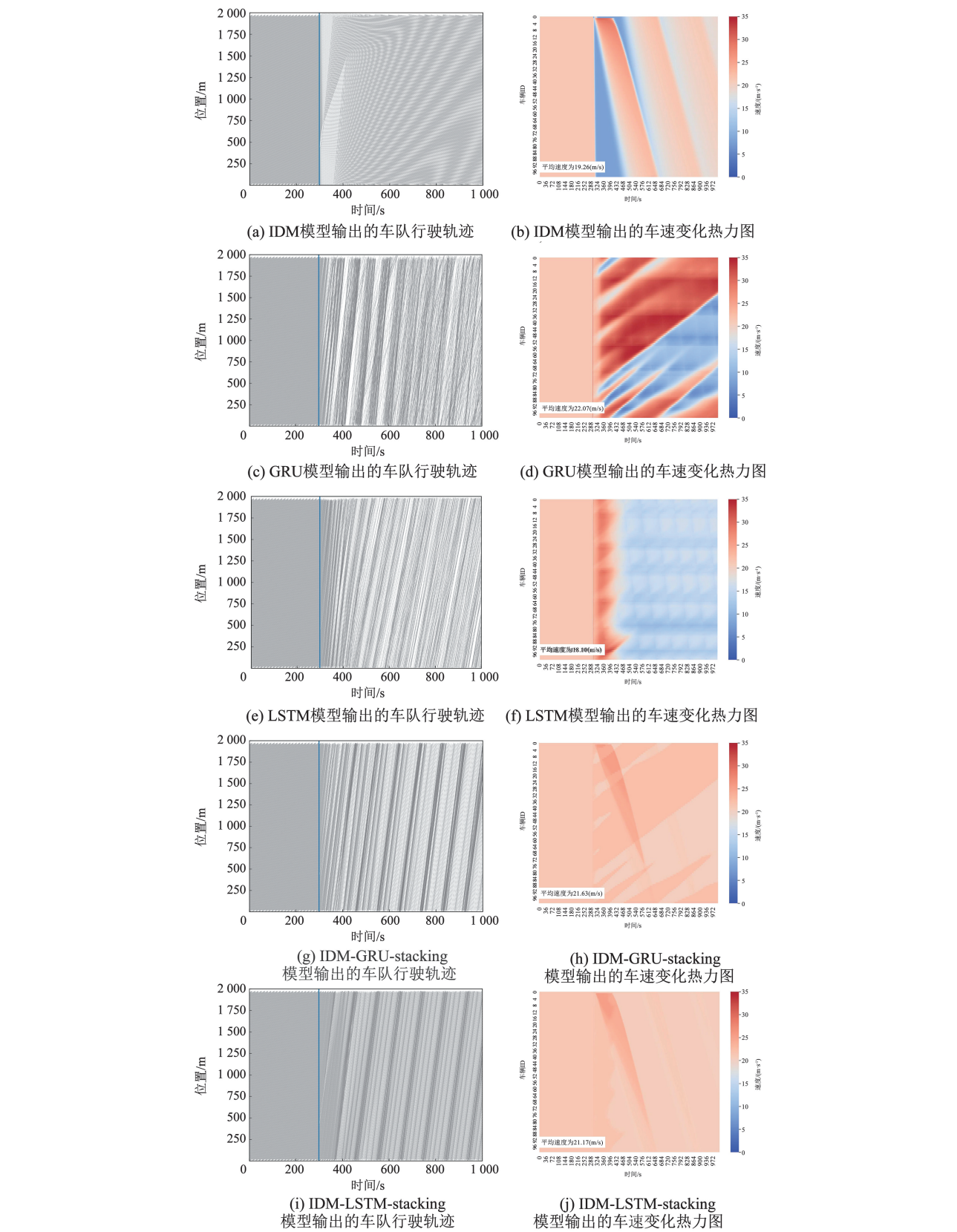


图 6 环路仿真实验结果
Fig.6 Results of ring road test

扰动的施加时刻,第二列为各模型输出的车速变化热力图。从图 6 中可以看出,施加扰动后 GRU 模型输出的跟驰车队行驶轨迹出现了明显紊乱,每辆车的速度随时间变化出现了显著波动,说明此跟驰模型的稳定性不佳;与 GRU 模型相比,IDM 模型和 LSTM 模型输出的速度值虽然存在周期性波动,但是各车速度变化幅度相对较小,说明其稳定性优于 GRU 模型;与上述 3 个模型相比,施加扰动后 IDM-LSTM-stacking 模型和 IDM-GRU-stacking 输出的跟驰车队行驶轨迹更为平滑,且每辆车的速度并未呈现出周期性波动,其中 IDM-GRU-stacking 模型输出的各车速度虽然在个别时间点上会发生较大变化,但是速度波动能够很快平抑,而 IDM-LSTM-stacking 模型输出的速度波动随着时间变化逐步消失,车队恢复到稳定跟驰状态,即车队内车辆速度与车队平均速度的平均差值维持在 0.6 m/s 以内。

4 结 论

1) 为了提高考虑车辆队列和周边车辆干扰条件下车辆跟驰行为预测精度,提出了一种 stacking 集成学习框架下的跟驰模型融合方法,将代表理论驱动 IDM 跟驰模型和数据驱动的 LSTM 和 GRU 跟驰模型作为一级学习算法,再使用特定的二级学习算法对一级学习器的输出特征进行再学习,构建出融合跟驰模型 IDM-LSTM-stacking 模型和 IDM-GRU-stacking 模型。

2) 基于实测数据的模型性能测试结果表明,考虑了车辆队列和周围驾驶条件因素的 LSTM 和 GRU 跟驰模型比 IDM 模型的预测精度高;采用非线性二级学习算法的融合跟驰模型的预测精度高于采用线性二级学习算法的融合模型以及数据驱动模型。其中,采用 GBRT 回归的 IDM-LSTM-stacking 模型和采用随机森林回归的 IDM-GRU-stacking 模型能够获得最高的预测精度。基于数值仿真的模型稳定性测试结果表明,施加扰动后 IDM 模型和 LSTM 模型的稳定性优于 GRU 模型,IDM-LSTM-stacking 模型和 IDM-GRU-stacking 模型的稳定性优于 IDM 模型、LSTM 模型和 GRU 模型。

3) 除了本文所考虑的因素以外,车辆跟驰行为还可能受到包括驾驶员心理在内的其他因素的影响,后续研究中可进一步将其纳入跟驰模型中,另外本文所提出的集成学习方法也可应用于其他微观驾

驶行为建模中,以提高复杂交通条件下的车辆轨迹预测精度和微观交通流仿真精度。

参考文献

- [1] 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): 116
WANG Dianhai, JIN Sheng. Review and outlook of modeling of car following behavior[J]. China Journal of Highway and Transport, 2012, 25(1): 116. DOI:10.19721/j.cnki.1001-7372.2012.01.018
- [2] 杨龙海,张春,仇晓赞,等. 车辆跟驰模型研究进展[J]. 交通运输工程学报, 2019, 19(5): 127
YANG Longhai, ZHANG Chun, QIU Xiaoyun, et al. Research progress on car-following models[J]. Journal of Traffic and Transportation Engineering, 2019, 19(5): 127. DOI:10.19818/j.cnki.1671-1637.2019.05.013
- [3] 贺正冰,徐瑞康,谢东繁,等. 数据驱动跟驰模型综述[J]. 交通运输系统工程与信息, 2021, 21(5): 103
HE Zhengbing, XU Ruikang, XIE Dongfan, et al. A review of data-driven car-following models[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(5): 103. DOI: 10.16097/j.cnki.1009-6744.2021.05.010
- [4] SONG Y, CHITTURI V, NOYCE A. Automated vehicle crash sequences: patterns and potential uses in safety testing[J]. Accident Analysis & Prevention, 2021, 153: 106017. DOI:10.1016/j.aap.2021.106017
- [5] GOODALL N J. Can you program ethics into a self-driving car?[J]. IEEE Spectrum, 2016, 53(6): 32. DOI:10.1109/MSPEC.2016.7473149
- [6] YANG D, ZHU L, LIU Y, et al. A novel car-following control model combining machine learning and kinematics models for automated vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(6): 1991. DOI:10.1109/TITS.2018.2854827
- [7] LI Y, LU X, REN C, et al. Fusion modeling method of car-following characteristics[J]. IEEE Access, 2019, 7: 162778. DOI: 10.1109/ACCESS.2019.2949305
- [8] 罗颖,秦文虎. 基于 IDM 与 RBFNN 的组合型车辆低速跟驰模型[J]. 计算机应用研究, 2020, 37(8): 2354
LUO Ying, QIN Wenhui. Combination low-speed car-following model based on IDM and RBFNN[J]. Application Research of Computers, 2020, 37(8): 2354. DOI:10.19734/j.issn.1001-3695.2019.03.0052
- [9] 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. Maui: IEEE Press, 2018: 2122. DOI:10.1109/ITSC.2018.8569552
- [10] 唐毅,刘卫宁,孙棣华,等. 考虑前后车辆综合效应的跟驰模

型及其稳定性分析[J]. 哈尔滨工业大学学报, 2014, 46(2): 115

TANG Yi, LIU Weining, SUN Dihua, et al. A new car following model with considering the synergy effect of ahead-backward vehicles and its stability analysis[J]. Journal of Harbin Institute of Technology, 2014, 46(2): 115. DOI:10.11918/j.issn.0367-6234.2014.02.021

[11] 彭勇, 刘世洁, YU D Z. 前后多车影响的跟驰模型及稳定性分析[J]. 交通运输系统工程与信息, 2018, 18(6): 41

PENG Yong, LIU Shijie, YU D Z. Stability analysis of a car-following model with consideration of preceding and following vehicles[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(6): 41. DOI: 10.16097/j.cnki.1009-6744.2018.06.007

[12] HE Z, WANG Y, SUN W, et al. Modelling car-following behaviour with lateral separation and overtaking expectation [J]. Transportmetrica B: Transport Dynamics, 2016, 4(3): 223. DOI: 10.1080/21680566.2015.1083911

[13] CHOUDHURY F, ISLAM M. Modelling acceleration decisions in traffic streams with weak lane discipline: a latent leader approach [J]. Transportation Research Part C: Emerging Technologies, 2016, 67: 214. DOI:10.1016/j.trc.2016.02.010

[14] OHLSON J A, KIM S. Linear valuation without OLS: the Theil-Sen estimation approach[J]. Review of Accounting Studies, 2015, 20(1): 395. DOI:10.2139/ssrn.2276927

[15] ÇETINER H, ÇETINER İ. Analysis of different regression algorithms for the estimate of energy consumption[J]. European Journal of Science and Technology, 2021, 31: 23

[16] PEKEL E. Estimation of soil moisture using decision tree regression [J]. Theoretical and Applied Climatology, 2020, 139(3): 1111. DOI:10.1007/s00704-019-03048-8

[17] LAREF R, LOSSON E, SAVA A, et al. On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications [J]. Chemometrics and Intelligent Laboratory Systems, 2019, 184: 22. DOI:10.1016/j.chemolab.2018.11.011

[18] MODARESI F, ARAGHINEJAD S, EBRAHIMI K. A comparative assessment of artificial neural network, generalized regression neural network, least-square support vector regression, and K -nearest neighbor regression for monthly streamflow forecasting in linear and nonlinear conditions[J]. Water Resources Management, 2018, 32(1): 243. DOI:10.1007/s11269-017-1807-2

[19] ZHANG W, WU C, LI Y, et al. Assessment of pile drivability using random forest regression and multivariate adaptive regression splines [J]. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 2021, 15(1): 27. DOI:10.1080/17499518.2019.1674340

[20] ZHU X, ZHANG P, XIE M. A joint long short-term memory and AdaBoost regression approach with application to remaining useful life estimation[J]. Measurement, 2021, 170: 108707. DOI:10.1016/j.measurement.2020.108707

[21] HE X, LUO J, LI P, et al. A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting [J]. Water Resources Management, 2020, 34(2): 865. DOI:10.1007/s11269-020-02483-x

[22] SHRIVASTAVA S, JEYANTHI M, SINGH S. Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting [J]. Cogent Economics & Finance, 2020, 8(1): 1729569. DOI: 10.1080/23322039.2020.1729569

[23] JOHN V, KARUNAKARAN M, GUO C, et al. Free space, visible and missing lane marker estimation using the PsiNet and extra trees regression [C]//2018 24th International Conference on Pattern Recognition. Beijing: IEEE Press, 2018: 189. DOI:10.1109/ICPR.2018.8546108

[24] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735. DOI:10.1162/neco.1997.9.8.1735

[25] CHUNG J, GULCEHRE C, CHO H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J/OL]. arXiv: 1412.3555. (2014-12-11). <https://arxiv.org/abs/1412.3555>

[26] 李腾龙, 惠飞. 考虑后视和最优速度记忆的跟驰模型及仿真 [J]. 计算机工程与应用, 2017, 53(12): 253

LI Tenglong, HUI Fei. Numerical simulation of car-following model considering optimal velocity changes with memory and backward looking effect[J]. Computer Engineering and Applications, 2017, 53(12): 253. DOI:10.3778/j.issn.1002-8331.1611-0152