计算机科学
COMPUTER SCIENCE

基于PPO算法的不同驾驶风格跟车模型研究

闫鑫, 黄志球, 石帆, 徐恒

引用本文

闫鑫, 黄志球, 石帆, 徐恒. 基于PPO算法的不同驾驶风格跟车模型研究[J]. 计算机科学, 2024, 51(9): 223-232.

YAN Xin, HUANG Zhiqiu, SHI Fan, XU Heng. Study on Following Car Model with Different Driving Styles Based on Proximal PolicyOptimization Algorithm [J]. Computer Science, 2024, 51(9): 223-232.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230700131

基于 PPO 算法的不同驾驶风格跟车模型研究

闫 鑫 黄志球 石 帆 徐 恒
南京航空航天大学计算机科学与技术学院 南京 210016
(yanxinsh@163.com)

摘 要 自动驾驶对于减少交通堵塞、提高驾驶舒适性具有非常重要的作用,如何提高人们对自动驾驶技术的接受程度仍具有重要的研究意义。针对不同需求的人群定制不同的驾驶风格,可以帮助驾驶人理解自动驾驶行为,提高驾驶人的乘车体验,在一定程度上消除驾驶人对使用自动驾驶系统的心理抵抗性。通过分析自动驾驶场景下的跟车行为,提出基于 PPO 算法的不同驾驶风格的深度强化学习模型设计方案。首先分析德国高速公路车辆行驶数据集(HDD)中大量驾驶行为轨迹,根据跟车时距(THW)、跟车距离(DHW)、行车加速度以及跟车速度特征进行归类,提取激进型的驾驶风格和稳健型的驾驶风格的特征数据,以此为基础编码能够反映驾驶人风格的奖励函数,经过迭代学习生成不同驾驶风格的深度强化学习模型,并在 highway env 平台上进行道路模拟。实验结果表明,基于 PPO 算法的不同风格驾驶模型具有完成任务目标的能力,且与传统的智能驾驶模型(IDM)相比,能够在驾驶行为中准确反映出不同的驾驶风格。

关键词:自动驾驶;智能驾驶模型;强化学习;PPO 算法;主成分分析;*K*-means
中图分类号 TP391

Study on Following Car Model with Different Driving Styles Based on Proximal Policy Optimization Algorithm

YAN Xin, HUANG Zhiqiu, SHI Fan and XU Heng
School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract Autonomous driving plays a crucial role in reducing traffic congestion and improving driving comfort. It remains of significant research importance to enhance public acceptance of autonomous driving technology. Customizing different driving styles for diverse user needs can aid drivers in understanding autonomous driving behavior, enhancing the overall driving experience, and reducing psychological resistance to using autonomous driving systems. This study proposes a design approach for deep reinforcement learning models based on the proximal policy optimization(PPO) algorithm, focusing on analyzing following behaviors in autonomous driving scenarios. Firstly, a large dataset of vehicle trajectories on German highways(HDD) is analyzed. The driving behaviors are classified based on features such as time headway(THW), distance headway(DHW), vehicle acceleration, and following speed. Characteristic data for aggressive and conservative driving styles are extracted. On this basis, an encoded reward function reflecting driver styles is developed. Through iterative learning, different driving style deep reinforcement learning models are generated using the PPO algorithm. Simulations are conducted on the highway environment platform. Experimental results demonstrate that the PPO-based driving models with different styles possess the capability to achieve task objectives. Moreover, when compared to traditional intelligent driver model(IDM), these models accurately reflect distinct driving styles in driving behaviors.

Keywords Autonomous driving, Intelligent driving model, Reinforcement learning, Proximal policy optimization, Principal component analysis, *K*-means

1 引言

自动驾驶技术能够很好地帮助人类解决在驾驶过程中所遇到的不稳定性问题,包括注意力不集中、动作反馈不及时等,同时也能够帮助人们减轻交通通行的压力。因此,如何提高人们对自动驾驶技术的接受度是一个值得关注的问题。

由于不同的驾驶人具有不同的驾驶经验、能力和习惯等,因此从驾驶人的驾驶行为中提取相关特征,针对不同需求的人群定制不同的驾驶风格,可以帮助驾驶人理解自动驾驶行为,提高驾驶人的乘车体验,在一定程度上消除驾驶人对使用

到稿日期:2023-07-19 返修日期:2024-01-19
基金项目:国家自然科学基金联合基金项目(U2241216)
This work was supported by the Joint Funds of the National Natural Science Foundation of China(U2241216).
通信作者:黄志球(zqhuang@nuaa.edu.cn)

自动驾驶系统的心理抵抗性[1]。

车道保持是驾驶过程中占比最多的驾驶行为，经过多年的发展，相关技术已经较为成熟。其主要可划分为基于数学的控制理论模型和基于数据驱动的机器学习模型，前者包括智能驾驶模型(Intelligent Driver Model，IDM)[2]、最优控制和自适应巡航[3]等；后者常用的机器学习方法有深度学习[4]、强化学习，以及深度强化学习[5]等。以深度Q网络算法(DQN)[6]、深度确定性策略梯度算法(DDPG)[7]和近端策略优化算法(PPO)[8]为代表的深度强化学习算法在训练数据中学习决策机制，具有很好的泛化能力[7]，适合环境复杂的道路交通场景。

尽管该领域的技术已经较为成熟，能够满足驾驶过程中的安全性和舒适性，但是自动驾驶的通用模型并没有根据具体人群分析不同驾驶人的驾驶风格。尽管自动驾驶系统能够采取最优的操作，但是对于具有不同驾驶风格的驾驶人来说，最优的驾驶操作却并不一定讨喜。因为传统的车辆驾驶模型的参数往往是固定的，以反映多数驾驶人的驾驶特征[9]，其并不适用于所有的驾驶场景和驾驶人，因此可以针对具有不同驾驶风格的驾驶人群提供对应的驾驶模型。深度强化学习作为一种基于数据驱动的方法，结合了深度学习和强化学习两个领域的技术，旨在通过从环境中获取大量数据并使用神经网络进行模式识别和决策，来训练智能体(Agent)进行学习和决策。

为了解决深度强化学习得到的模型仅具有一般驾驶人的驾驶特性问题，本文首先基于大规模的驾驶数据，获取车辆在行驶过程中的行为特征，利用主成分分析(PCA)方法进行降维，提取能体现驾驶人驾驶风格的数据特征。随后使用K-means方法进行聚类得到不同驾驶风格的驾驶人模型，并在此基础上设计奖励函数，以改进PPO算法进行学习，得到不同驾驶风格的参数模型，鼓励车辆采取更符合对应驾驶风格的行为动作。最后基于德国高速公路车辆行驶数据集(The highD Dataset)[10]进行实验，完整框架如图1所示。

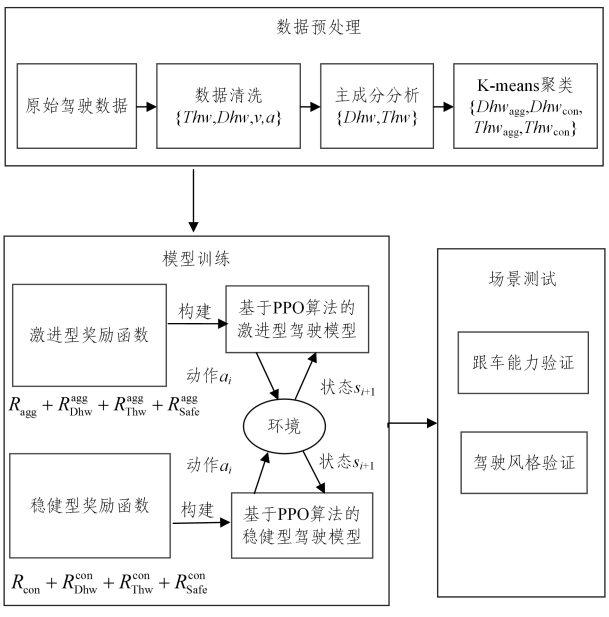

图1 基于PPO算法的风格驾驶模型框架

Fig. 1 Style driving model framework based on PPO algorithm

本文的主要贡献如下：

(1)根据The highD Dataset数据集提取激进型和稳健型的驾驶风格特征，在此基础上设计能够体现不同风格的奖励函数。

(2)将不同风格的驾驶人模型编码为奖励函数，改进PPO算法，生成具有不同驾驶风格的深度强化学习模型。

(3)在highway env测试环境中与传统的IDM控制方法进行对比，以验证基于PPO算法的不同驾驶风格的跟车模型具有良好的完成任务目标的能力和风格表达能力。

2 相关理论

2.1 高速公路车道保持

高速公路的车道保持过程可以划分为横向控制和纵向控制。横向控制的主要决策包括车道跟踪和车道变换，其中车道跟踪的目标是在驾驶过程中进行转向，使车辆以较小的误差跟踪车道线。车道变换的任务是选择更加合适的车道以满足驾驶需求，例如转向需求和超车需求等[11]，主要的控制策略是MOBIL(Minimize Overall Braking Induced by Lane Change)，其数学约束为：

$$a_c^- - a_c + p(a_n^- - a_n + a_o^- - a_o) > \Delta a_{th} \tag{1}$$

其中，a_c和a_c^-分别表示受控车辆变道前加速度和变道后的加速度；a_n和a_n^-分别表示受控车辆变道前和变道后新的跟随车辆加速度；a_o和a_o^-分别表示受控车辆变道前和变道后旧的跟随车辆加速度；p为礼让系数；Δa_{th}为设定阈值，当变道后的加速度收益大于该阈值后才可以变道。

Gao等[12]基于车辆运动学模型和Taylor展开式，提出了一个新的有界等价函数，用于解决横向运动的轨迹跟随控制，以确保强大的鲁棒性和控制精度。Xie等[13]提出基于物理和机动的轨迹预测模型，在结合无迹卡尔曼滤波器非线性估计和基于机动的长期预测能力优点后，能够在横向变道场景进行精确的预测。在某些倒车场景的横向控制场景中，Gao等[14]基于反步和切换控制理论解决了全自动停车难以实现的问题。针对自动驾驶车辆如何换道至专用车道的场景，Zhang等[15]提出了使用深度强化学习选择控制信号实现自动驾驶车辆的换道汇入。

纵向控制的主要任务是在车辆与跟随点和跟随车辆之间进行选择跟随，要求车辆之间保持一定的车距以紧急避险[16]，并收缩一定车距来提高道路的车辆容量。半个世纪以来，车道控制领域的技术已走向成熟，主要的控制策略有基于数学控制的IDM模型以及交互多模型(IMMS)[17]，还有基于机器学习的模仿学习[18]和深度强化学习。

智能驾驶模型(IDM)是基于驾驶人的行为和反应，以及车辆的动力学特征来建立数学模型，从而预测车辆的加速度、速度和位置等行驶状态，其需要满足动力学约束公式：

$$v_\alpha = a\left(1-\left(\frac{v_\alpha}{v_0}\right)^\delta - \left(\frac{s^*(v_\alpha,\Delta v_\alpha)}{s_\alpha}\right)^2\right) \tag{2}$$

$$s^*(v_\alpha,\Delta v_\alpha) = s_0 + v_\alpha T + \frac{v_\alpha \Delta v_\alpha}{2\sqrt{ab}} \tag{3}$$

其中，对于受控车辆α，v_α为车辆此时的速度，Δv为该车辆与前车的速度差，v_0代表期望速度，s_0为最小跟车距离，T为

期望跟车时距，a 为加速度，$b>0$ 为安全系数，参数 δ 通常设置为 4。

本文的实验环境为，在多车道直线行驶的高速公路中，存在以下 3 种基本跟车场景：侧道车辆变道插入受控车辆车道、受控车辆的前方车辆变道离开、受控车辆的前方车辆保持直行且没有其他车辆插入，如图 2－图 4 所示。

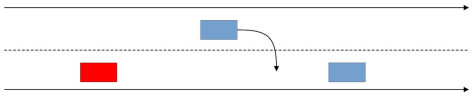

图 2　侧道车辆变道插入受控车辆车道

Fig. 2　Side lane vehicle lane change and insertion into controlled vehicle lane

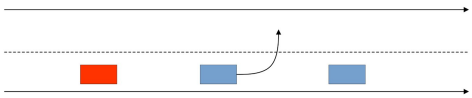

图 3　受控车辆的前方车辆变道离开

Fig. 3　Front vehicle of controlled vehicle changing lanes and exiting

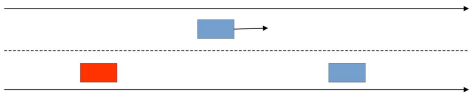

图 4　受控车辆的前方车辆保持直行且没有其他车辆插入

Fig. 4　Vehicle in front of the controlled vehicle continues straight and no other vehicles insert

本文的研究问题是如何让受控车辆在 3 种基本场景中进行自动跟车，受控车辆 FV 的驾驶目标是尽可能保持指定的驾驶风格，完成行驶任务。特定的驾驶风格下，受控车辆将保持理想的行车效率并与前方车辆保持期望的理想距离。除了受控车辆 FV 采用本文算法控制的纵向行驶外，其他车辆在纵向控制上均采用 IDM 控制策略，横向控制采用 MOBIL 方法。跟车系统描述如下：

$$action=f(V_{FV}(t),\Delta V(t),\Delta S(t)) \tag{4}$$

$$\Delta S(t)=S_{LV}(t)-S_{FV}(t) \tag{5}$$

$$\Delta V(t)=V_{LV}(t)-V_{FV}(t) \tag{6}$$

其中，S_{FV} 和 V_{FV} 分别为 t 时刻的受控车辆 FV 的位置和速度信息；而 S_{LV} 和 V_{LV} 是 t 时刻前方车辆的位置信息和速度信息。f 是一个关系函数，它将受控车辆的行驶速度以及与前方车辆之间的相对速度 ΔV 和相对纵向距离 ΔS 映射到自动驾驶车辆的执行动作上。本文中考虑的执行动作仅为纵向动作、油门和刹车。

不同驾驶人在跟车的过程中，受到生理因素、交通路况、驾驶经验等的影响，在跟车过程中有自己偏好的跟车特征，包括行车过程与前车之间的跟车距离、跟车速率、加速度、与前车的跟车时距等。当受控车辆的跟车特征偏离驾驶人期望时，驾驶人会适当调整油门和刹车来满足偏好需求。本文基于文献[19]，将驾驶人的风格主要分为激进型和稳健型。

2.2　主成分分析

本文提取驾驶人风格时涉及多个特征，使用主成分分析（PCA）[20-21]能够很好地提取驾驶数据中和驾驶人风格有关的主要特征，有效地进行降维。PCA 将原始数据转换为一组新的线性无关变量，这些变量具有原始数据中最大的信息量，称为主成分。利用主成分取代原始数据进行学习计算，可以减少计算的存储和计算成本。PCA 是现代数据分析中的标准工具，其目标是识别给定数据集最有影响力的因子，提取的因子能揭示数据集中的隐藏结构并过滤了噪声，目前在数据处理方面已经有很多应用，如数据降维、数据压缩、特征提取和数据可视化。

在降维过程中，本文对 m 行 n 列的数据矩阵 $\boldsymbol{X}$ 进行降维，其中每一行表示一个样本，每一列为一个特征。本文将矩阵 $\boldsymbol{X}$ 从 N 维降维为 K 维，并保证不损失太多的原始数据信息。具体步骤如下：

(1)中心均值化。计算 $\boldsymbol{X}$ 每个特征的均值，并对每一行的特征值减去均值。

(2)主成分挑选。计算协方差矩阵 $\boldsymbol{C}$，根据协方差矩阵的特征值大小将其特征向量进行排序，选取前 K 行得到新矩阵 $\boldsymbol{P}$。

(3)降维表示。$\boldsymbol{Y}=\boldsymbol{PX}$，得到降维后 K 维的数据。

2.3　*K*-means 聚类

聚类利用数据分布的潜在结构，并定义规则来将具有相似特征的数据进行分组[22]。这个过程是根据聚类标准将给定数据集重新划分成若干部分，划分依据不需要任何关于此数据集的先验知识，并且划分的簇群实例具有较大差异性。为了得到不同驾驶风格的模型参数，本文需要对 PCA 降维过后的相关数据进行聚类，得到想要的风格信息。采用 K-means 方法将驾驶数据分割成能代表不同驾驶风格的簇群。K-means 聚类方法具有适用于大规模数据集、结果具有可解释性等优点，本文采用 K-means[23]方法将数据点与最近的质心相关联从而实现聚类，而质心由每个簇群中所有的数据点的平均值计算得到。具体步骤如下：

(1)随机选取 K 个初始质心点；

(2)对于每个数据点，计算它与 K 个中心点之间的距离，并将其归纳到距离它最近的中心点所对应的簇群中；

(3)根据每个簇群中所有的数据点的平均值重新计算中心点；

(4)重复步骤(2)和(3)，直到中心点的变化收敛。

K-means 方法能够快速算簇心和分簇后的数据，但具有一定的不稳定性，当选取不同的初始点时，收敛后的结果可能略有差别。

2.4　近端策略优化

2.4.1　强化学习

在标准的强化学习模型中，智能体感知环境状态信息 s_t，在行为策略 $\pi(a_t|s_t)$ 指示下做出行动 a_t，与环境进行交互，取得奖励 r_t，达到下一个状态 s_{t+1}，形成轨迹 τ。强化学习的任务是帮助智能体在与环境交互中学习最优的行为策略 π^*，使智能体取得最大的累计期望奖励 $R(\tau)=\sum_{t=0}^{T}\gamma r_t$[24]。由于未来奖励具有不确定性，因此需要采用折扣因子 $\gamma\in(0,1]$ 来对

未来奖励进行折现。在强化学习中存在两种学习方式：基于值函数的强化学习和基于策略梯度的强化学习。

(1)基于值函数的强化学习

基于值函数的强化学习目的是学习一个价值函数，用于预测在给定状态下某个动作的长期奖励。值函数通常可分为基于动作价值函数的方法和基于状态价值函数的方法，两种方法均通过迭代价值函数的方式进行学习。基于动作价值函数的方法使用$Q^{\pi}(s,a)$代表当前状态 s 采取动作 a 预计取得的回报奖励，常用的算法为 Q-learning，Saras 等算法。基于状态价值函数的方法使用$V^{\pi}(s)$来表示在当前状态 s 预计取得的回报奖励[24]。

(2)基于策略梯度的强化学习

基于策略梯度的强化学习直接学习最优策略而不是价值函数。策略函数 $\pi(a|s;\theta)$是一个映射，它将状态映射在相应的动作概率分布。计算 $E[R(\tau)]$的导数$\nabla_{\theta}\log\pi(a|s;\theta)$，通过梯度上升来更新策略 $\pi(a|s;\theta)$的参数 θ，使获得高奖励的动作概率增大，而获得低奖励的动作概率减小，从而最大化累计奖励的预期值。常用的算法有 REINFORCE、Actor-Critic 方法，以及近端策略优化等。

2.4.2 近端策略优化

当环境过于复杂时，需要采用神经网络对环境进行建模，以此来解决维度爆炸的问题。近端策略优化(Proximal Policy Optimization，PPO)算法是一种基于策略梯度的深度强化学习方法，具有在离散动作空间中表现优秀并且收敛速度快的特点，适合本文实验环境(简化的高速公路行驶场景)。PPO 算法由 OpenAI 于 2017 年提出[25]，其基于 Actor-Critic 框架。算法结构如图 5 所示。

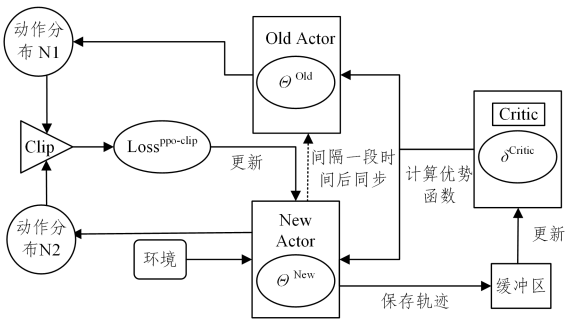

图 5 PPO 算法结构图

Fig. 5 Structure diagram of PPO algorithm

PPO 算法使用优势函数 $A(s,a)$衡量在状态 s 下采取动作 a 带来高回报的相对优越性。优势函数的定义如下：

$$A(s,a)=Q(s,a)-V(s) \tag{7}$$

其中，$Q(s,a)$是在状态 s 下采取动作 a 的动作价值函数，表示采取该动作 a 后带来的预期收益；而 $V(s)$是在状态 s 下的状态价值函数，代表在当前状态 s 下的预期收益。优势函数表示当前状态下采用当前动作 a 与其他动作相对的收益，如果 $A(s,a)>0$，说明采用当前动作更具有优势，反之，采用当前动作效果更差。本文使用时序差分残差对优势函数进行估计，估计函数如下：

$$A(s_t,a_t)=r_t+\gamma V(s_{t+1})-V(s_t) \tag{8}$$

PPO 算法中的价值评价模块 Critic 和策略执行模块 Actor 由神经网络构成。其中 Critic 单元的作用是对当前状态 s 的价值函数 $V(s)$进行评价，便于后续计算优势函数，价值损失函数为：

$$L(\phi)=-\sum_{t=1}^{T}\left(\sum_{t'>t}\gamma^{t'-t}R_{t'}-V_{\phi}(s_t)\right)^2 \tag{9}$$

其中，R_t为 t 时刻的即时奖励，V_{ϕ}表示在s_t状态下的状态价值函数，γ 表示奖励折扣因子。策略执行网络由两个 Actor 网络构成，分别为 $actor_{old}$和 $actor_{new}$。$actor_{new}$网络接受当前的状态作为输入，输出一个代表智能体将会采取的动作概率分布，并计算策略梯度，对当前策略进行评估和更新。$actor_{old}$网络提供与当前 $actor_{new}$网络的重要性采样比，用于控制更新幅度，从而限制学习策略和之前一轮的差距。其优化策略为：

$$\theta_{k+1}=\arg\max_{\theta}E_{s\sim S}E_{a\sim\pi_{\theta_k}}(\cdot|s)\left[\min\left(\frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)}A^{\pi_{\theta_k}}(s_t,a_t),clip\left(\frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)},1-\epsilon,1+\epsilon\right)A^{\pi_{\theta_k}}(s_t,a_t)\right)\right] \tag{10}$$

其中，π_{θ_k}表示 $actor_{old}$的策略网络；π_{θ}表示 $actor_{new}$的策略网络；$A(s,a)$表示状态 s 动作 a 的优势函数；ϵ是一个超参数，表示截断的范围。截取函数 $clip$ 的计算方式为：

$$clip(x,1-\epsilon,1+\epsilon)=\begin{cases}1-\epsilon, & x<1-\epsilon\\ 1+\epsilon, & x>1+\epsilon\\ x, & \text{else}\end{cases} \tag{11}$$

3 个性化的车道保持过程

3.1 个性化驾驶风格

Sagberg 等[26]认为有意识的偏好行为和自动化的习惯都可以定义为驾驶风格，并概括了激进型或者风险型和防御型、稳健型或者专注型的特点。激进型的驾驶风格以实现目标为导向，而稳健型的驾驶风格以避免风险为导向。具体来说，不同的驾驶人对跟车过程中的跟车间距、速度、相对速度以及加速度的敏感度不同，在不同的行车场景中又由于交通规则等的限制，同一个驾驶人在不同的跟车任务中也可能有不同的驾驶偏好。例如在行驶过程中保持较快的车速，会造成部分老年人群的恐慌，而保持较慢的车速又会使部分年轻驾驶人认为驾驶体验平淡无趣；过于频繁的加速减速会让部分老年驾驶人抱有不适和安全疑虑，但也会给部分年轻驾驶人带来较为刺激的驾驶体验。

因此，针对不同类型的驾驶人群的不同驾驶风格建立模型是非常有必要的，这样能够很好地还原本人在驾驶过程中的驾驶行为，减轻驾驶人对自动驾驶的抵抗性。本文主要对高速公路的跟车场景下的驾驶人行为进行分析[10]，对不同的驾驶风格进行建模，并根据前人的调查[19]将驾驶风格分为激进型和稳健型两种类型。识别不同的驾驶风格需要分析驾驶过程中大量的驾驶行为，选用不同的指标得到的风格分析也往往不同。

在对驾驶风格进行评估的过程中，选择关键的驾驶特征

进行量化、正确地评估驾驶风格是非常重要的步骤。在前人的研究中，驾驶风格可以体现在不同的规范上，例如超速和急加速动作等单一动作，再到具有攻击性的驾驶行为等由若干动作的组合[26]。Murphey 等[27]提出了一种基于颠簸曲线的驾驶风格分类方法，通过测量司机加速和减速的速度，将驾驶风格分为 4 种类型，分别为激进驾驶、平稳驾驶、正常驾驶和静止。在 Ahmad 等的研究中，综合加速度和速度作为驾驶风格的分类依据，采用模糊推理来推断是否为激进的驾驶风格。但是加速度和速度并不能在所有场景中作为分类依据，De Waard 等调研了高速公路入口匝道上司机的汇入速度，发现较慢车速完成汇入动作反而更具有危险性[28]。

在 Liu 等[29]的研究中，采用平均车头时距（THW）和制动时的平均 THW 作为特征，将驾驶人分为两种不同的驾驶风格。车头时距 THW 定义为前后两辆车的车头通过同一地点的时间差。根据测量的车头时距信息，MacAdam 等制定了一个“驾驶侵略性指数”来衡量车辆快速靠近的程度[30]。越小的跟车 THW 可以视为车辆更为靠近，是更具有驾驶侵略性的行为，可以作为激进型驾驶风格的分类依据。

本文认为，在驾驶过程中只考虑车辆驾驶过程中的某一特征作为驾驶风格的分类依据过于单一，并不能全面地反映驾驶过程中不同行驶风格的驾驶人心理，驾驶人在跟车过程中也会考虑环境因素，如天气、自身车辆的行驶速度、与前车的跟车距离（DHW）、与前车的跟车时距（THW）、自身车辆的加速度大小等。本文选用高速公路驾驶环境下的自身车辆行驶速度、THW、DHW 和自身车辆的加速度大小等作为特征向量，经过剔除离群值、归一化等操作，使用 PCA 降维技术选取主要特征后，通过 K-means 方法将驾驶行为分为两种驾驶风格：激进型和稳健型。其中，本文采用的跟车距离和跟车时距计算式如下：

$$DHW=S_{\mathrm{LV}}(t)-S_{\mathrm{FV}}(t) \tag{12}$$

$$THW=\frac{DHW}{v_{\mathrm{FV}}(t)} \tag{13}$$

3.2　高速公路跟车过程的数据集

本文使用了德国亚琛工业大学汽车工程研究所的一个用于研究自动驾驶的数据集 The highD Dataset（HDD）[10]，该数据集来源于德国科隆附近的 6 个不同地点的高速公路大型自然车辆轨迹数据，记录的数据包括轿车和卡车。图 6 展示了高速公路录像示例。

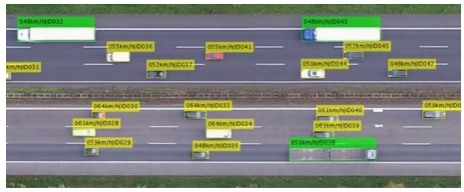

图 6　德国科隆高速公路录像示例

Fig. 6　Example of German cologne highway video

该数据集保留了行驶车辆的高保真多维特征信息。本文选用 HDD 数据记录中的 DHW（Distance Headway）、THW（Time Headway）、车辆行驶速度以及车辆加速度等特征值来确定驾驶人的驾驶风格。首先计算行驶过程中每个驾驶人的 DHW、THW、当前行驶速度和行驶加速度的平均值。

本文通过 PCA 技术对原始特征进行降维来提取关键特征。为了更好地提取关联信息，去除其不同维度上的计量单位差别，对平均化后的 DHW、THW、车辆当前行驶速度以及加速度进行 Z-score 标准化，使其收敛在[0，1]区间内，再使用 PCA 技术将其降维。通过计算前 K 个主成分方差占总方差的比重，确定当参数 K 为 2 时，方差比重为 91.31。因此本文挑选权重最大的 DHW 和 THW 作为分类依据。根据降维后的数据，使用 K-means 分类方法，将降维后的数据进行分类，分别得到激进型和稳健型驾驶风格分类结果。表 1 列出了两种不同驾驶风格下跟车行为的统计数据。

表 1　高速公路跟车行为的统计性描述

Table 1　Statistical description of highway following behavior

类　型	DHW/m			THW/s			行驶速度/(m/s)			行驶加速度/(m/s²)		
	Mean	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*
全部	57.93	13.25	123.91	2.01	0.45	4.34	30.97	21.01	49.42	0.07	−0.42	0.49
激进型	45.35	13.25	81.01	1.57	0.45	2.83	32.88	23.01	49.42	0.06	−0.42	0.49
稳健型	72.31	42.47	123.91	2.51	1.48	4.34	28.88	21.01	47.57	0.08	−0.23	0.41

从表 1 可以看出，激进的驾驶风格具有保持较短的跟车距离、较快的车速、较小的跟车时距，以及较大的加速度等特点。稳健型的驾驶风格具有较长的跟车距离、较慢的车速、较大的跟车时距，以及较小的加速度等特点。

3.3　奖励函数设计

强化学习的目标是找到最优策略，该策略能够使未来的预期奖励最大。奖励函数指导策略的学习方向，通过自定义奖励函数，能够指导行为策略朝着预期方向收敛。在本节中，对高速公路上跟车行为数据提取了不同风格的跟车特征，因为本文所提出的深度强化学习模型需要反映不同驾驶风格的跟车距离、跟车效率特征，因此将其编码进奖励函数，如图 7 所示。

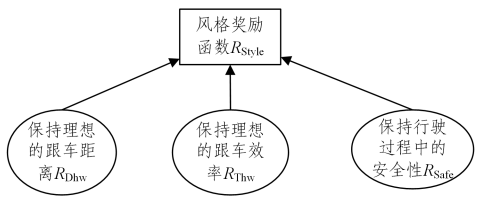

图 7　驾驶风格的奖励函数构成

Fig. 7　Composition of reward functions for driving styles

3.3.1　保持理想的跟车距离

Helly[31]提出每个驾驶人都有一个期望的驾驶距离，并有在行驶过程中与前车保持期望驾驶距离的习惯。图 8 和图 9 为不同驾驶风格的跟车距离的分布图。

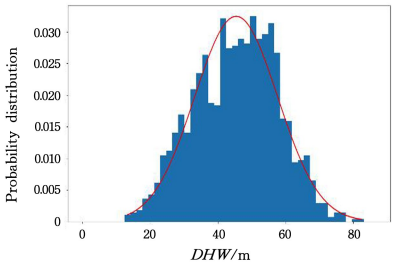

图 8 激进型驾驶风格的跟车距离分布图

Fig. 8 Distance headway distribution graph for aggressive driving style

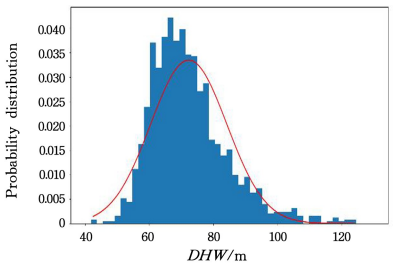

图 9 稳健型驾驶风格的跟车距离分布图

Fig. 9 Distance headway distribution graph for conservative driving style

为了拟合不同驾驶风格的跟车距离分布，需要拟合得到其近似曲线。从分布特点来看，其分布片段和正态分布契合，因此本文采用正态分布拟合的方式对两种不同风格的分布函数进行拟合。对服从正态分布的概率分布，可以得到其相应的概率函数为：

$$f(x|u,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-(\ln x-u)^2}{2\sigma^2}},x>0 \tag{14}$$

为了使驾驶风格反映在奖励函数上，鼓励智能体保持不同风格的跟车距离，本文根据图 8 和图 9 中拟合的分布函数设计奖励函数R_{Dhw}。

$$R_{Dhw_{agg}}=30.65f(DHW|u=45.35,\sigma=12.23) \tag{15}$$

$$R_{Dhw_{con}}=29.84f(DHW|u=72.31,\sigma=11.91) \tag{16}$$

从分布上可以看出，在跟车距离越接近 0 时，越具有危险驾驶的倾向，奖励函数越小；而在行驶距离接近期望的跟车距离时，奖励函数逐渐增大，直到跟车距离远离期望的跟车距离时再次减小，符合人类驾驶习惯。

3.3.2 保持理想的跟车效率

不同驾驶风格的驾驶人在驾驶过程中除了保持不同的行驶距离外，还需要保持较高的行驶效率。激进型驾驶人往往采取高速并有效率的跟车方式，而稳健型驾驶人的跟车效率相比之下更为稳健。稳健型和激进型的驾驶人在驾驶过程中使用车头时距 THW 来描述不同人群对效率的敏感性。图 10 和图 11 分别为激进型驾驶人和稳健型驾驶人在高速公路行驶过程中 THW 的分布图。

与跟车距离的处理类似，保持理想的跟车效率需要对跟车时距的分布进行拟合，提取驾驶风格的特点，编码在奖励函数中。由于在跟车时距中，不同驾驶风格类型的跟车时距分布片段与正态分布相似，因此依旧采用正态分布的方式进行拟合，拟合函数如图 10 和图 11 所示。为了使跟车时距的驾驶风格反映在奖励函数上，鼓励智能体保持不同风格的跟车时距，本文设计了用于保持不同风格的跟车时距奖励函数R_{Thw}。

$$R_{Thw_{agg}}=1.06f(THW|u=1.57,\sigma=0.423) \tag{17}$$

$$R_{Thw_{con}}=1.11f(THW|u=2.51,\sigma=0.440) \tag{18}$$

从两种不同驾驶风格的分布函数上可以看出，当跟车时距越接近于 0 时，反应时间越短，越具有危险驾驶的倾向，因此奖励函数也越接近于 0；而当跟车时距越靠近期望的跟车时距时，奖励函数逐渐增大，在达到期望的跟车时距时达到最大；当跟车时距逐渐偏离期望跟车时距时，跟车的收益逐渐下滑，符合人类的驾驶行为习惯。

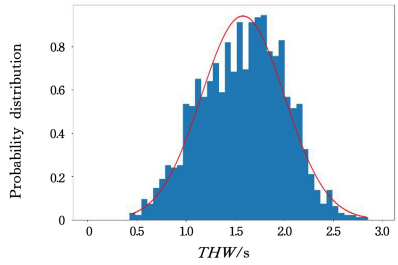

图 10 激进型驾驶风格的跟车时距分布图

Fig. 10 Time headway distribution graph for aggressive driving style

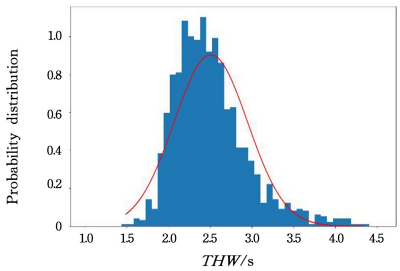

图 11 稳健型驾驶风格的跟车时距分布图

Fig. 11 Time headway distribution graph for conservative driving style

3.3.3 保持行驶的安全性

在驾驶过程中，除了保证车辆行驶符合驾驶人员的行驶风格外，也不能忽略驾驶过程中的安全性。为了避免在行驶过程中车辆过分追求符合驾驶风格的行驶轨迹，造成过度的危险驾驶行为，需要对其行驶过程中的安全性加以限制。本文采用碰撞时间(Time to Collision，TTC)指标来辅助车辆在行驶过程中保持安全性。TTC 经常被用来评估紧急避让情况下的安全性，通过计算 TTC，驾驶人可以预测和障碍物之间发生碰撞的时间，从而采取必要的操作，比如刹车和转向[32]。TTC 的计算式为：

$$TTC=\frac{\Delta s}{\Delta v} \tag{19}$$

其中，Δs 为前车与当前行驶车辆的相对距离，Δv 是当前车辆与前方车辆的相对速度差。通常情况下，TTC 小于 1.5 时会认为具有危险性驾驶的风险。本文使用如下仿射函数将安全性编码到奖励函数中：

$$R_{\text{Safe}}=\min(0,2/3(TTC-1.5)) \tag{20}$$

为了避免安全性的约束过多影响到风格模型的决策，限制模型对不同驾驶风格的表现，本文设计安全奖励函数在 TTC 大于 1.5 时不发挥作用，在小于 1.5 时线性增强。TTC 值越小，危险性越高，对驾驶风格的奖励函数的干扰越大。

3.3.4　奖励函数

对于自动跟车任务，基于激进型和稳健型驾驶风格的奖励函数可以表示为驾驶风格奖励函数以及安全奖励函数的线性组合，而驾驶风格奖励函数由跟车距离和跟车时距共同决定：

$$R_{\text{Style}}=w_1R_{Dhw}+w_2R_{Thw}+w_3R_{\text{Safe}} \tag{21}$$

其中，R_{style}对应两种行驶风格的总奖励函数；w_1，w_2和w_3分别是风格奖励函数和安全奖励函数的系数，在本文中设定为 1。

3.4　网络结构的设计

本文采用上述 PPO 算法进行深度强化学习，PPO 算法的网络架构是基于深度强化学习中的 Actor-Critic 架构进行设计。其中 Actor 网络将状态$s_t=(DHW,THW)$作为输入，输出离散动作 $a(t)$。本实验中，受控车辆的动作空间仅限于油门和刹车，受控车辆的加速度为[−1,1]的离散化变量。Critic 的输入是状态s_t，输出一个标量值 $V(s_t)$。

Actor 和 Critic 网络均有 3 层结构：输入层、输出层，以及两个含有 256 个神经元的隐藏层。在隐藏层中，采用校正线性单元 RELU 激活函数来加快网络参数的收敛；对于输出层采用 Sigmod 激活函数，将函数值映射到[0,1]。表 2 列出了训练过程中所设置的网络超参数。

表 2　训练模型的超参数设置

Table 2　Hyperparameter settings of training models

超参数	值
Replay memory size	64
Discount factor	0.8
Batch size	128
Learning rate	0.0005
Number of steps	128

本文提出的基于 PPO 算法的风格驾驶模型完整设计如算法 1 所示。在进行初始化后，受控车辆与环境进行交互，在每一个步骤中，由 PPO 算法中的 Actor 网络根据当前状态信息s_t计算受控车辆此时应采取的加速度，受控车辆与环境交互得到回报奖励r_t和新状态s_{t+1}，以此更新 Actor 网络和 Critic 网络。

算法 1　PPO-style

输入：初始化策略参数 θ 和值函数参数 ϕ，驾驶行为数据集 D

输出：激进型驾驶风格策略参数 θ_{agg}和值函数 ϕ_{agg}，稳健型驾驶风格策略参数 θ_{con}和值函数 ϕ_{con}

1. $D_{PCA}\leftarrow PCA(D)$；/ * 对数据集 D 采用 PCA 降维 * /
2. $\{Dis_{agg},Dis_{con}\}\leftarrow PCA(D_{PCA})$；/ * 对数据集$D_{PCA}$采用 K-means 方法聚类，得到激进型驾驶风格分布Dis_{agg}以及稳健型驾驶风格分布Dis_{con}} * /
3. 根据不同驾驶风格的分布特征编码奖励函数R_{style}；
4. for k = 1,2,…,n do
5. 通过在环境中运行策略 $\pi_{\theta_k}=\pi(\theta_k)$保存轨迹 $\tau=\{(s_0,a_0,s_1,r)\cdots\}$；
6. 计算轨迹奖励R_{style}；
7. 基于当前值函数$V_{\phi k}$，使用 TD 方法计算优势估计值$A^{\pi_{\theta_k}}$
8. 基于以下规则更新策略网络θ_{k+1}，

$$\theta_{k+1}=\arg\max_{\theta}\frac{1}{T}\sum_{t=0}^{T}\min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)}A^{\pi_{\theta_k}}(s_t,a_t),\text{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)},1-\epsilon,1+\epsilon\right)A^{\pi_{\theta_k}}(s_t,a_t)\right)$$

其中，clip 函数为：

$$\text{clip}(x,1-\epsilon,1+\epsilon)=\begin{cases}1-\epsilon, & x<1-\epsilon\\ 1+\epsilon, & x>1+\epsilon\\ x, & \text{Else}\end{cases}$$

9. 基于以下规则进行更新价值网络，其中R_t为 t 时刻后的累积折扣奖励和，

$$\varphi_{k+1}=\arg\min_{\phi}\frac{1}{T}\sum_{t=0}^{T}(V_\phi(s_t)-R_t)^2$$

10. end For

4　实验设置和结果分析

4.1　实验环境设置

本节搭建了一个总长度为 10 km 的高速公路仿真环境。领先车辆的速度变化是随机的，平均行驶速度保持在 30～50 km/h，以和数据集的环境保持一致。受控车辆由深度强化学习算法控制采用自动驾驶，初始车辆间距随机在 2～200 m 之间，受控车辆的初始速度随机在 30～60 km/h。本研究中，所有车辆的加速度控制在[−1,1]之间。在采样时刻，受控车辆和仿真环境进行交互得到一个奖励值作为反馈。跟车任务的终止条件为：

(1)跟车过程中发生碰撞；

(2)跟车时间超过 40 s。

4.2　实验及结果比较

为了验证提出的自动跟车策略具有目标可达性和自适应性，能够完成跟车任务，本节计算两种不同的跟车风格中所有时间步积累的平均奖励，以及在跟车过程中的平均时间步。

从图 12 和图 13 中可以看出，随着迭代次数的增加，不论是激进型驾驶风格的训练模型还是稳健型驾驶风格的训练模型，平均行驶时间均在增加，且在 50000 次后均得到收敛。稳健型模型在训练过程中平均行驶时间可以保持接近 40 s，激进型由于鼓励采用较为激进的行驶行为，更容易与前方车辆碰撞，对换道车辆躲避不及时，训练后平均驾驶的时间保持接近在 35 s。两个模型均具有完成跟车任务的能力。

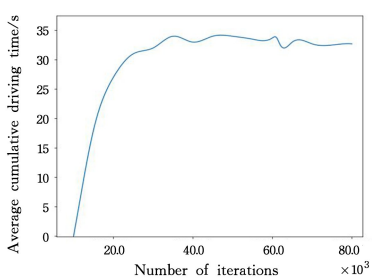

图 12　激进型驾驶风格训练过程的平均行驶时间长度

Fig. 12　Average driving duration during the training process of aggressive driving style

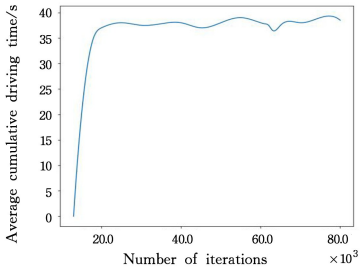

图 13 稳健型驾驶风格训练过程的平均行驶时间长度

Fig. 13 Average driving duration during the training process of conservative driving style

在一个典型的深度强化学习训练过程中，每一个 episode 的奖励值都可能会出现波动。本节采用移动平均法展示平均奖励的变化趋势，计算式为：

$$R_t = 0.8R_{t-1} + 0.2R_t \tag{22}$$

其中，R_t 表示第 t 次训练的平均奖励。

图 14 和 15 展示了两种不同风格的驾驶模型在行驶过程中的平均奖励相对于训练次数的变化情况。可以看到，无论是激进型风格模型的训练过程还是稳健型风格模型的训练过程，随着迭代的次数增多，奖励值不断增大，并均在迭代 50000 次后，模型收敛。

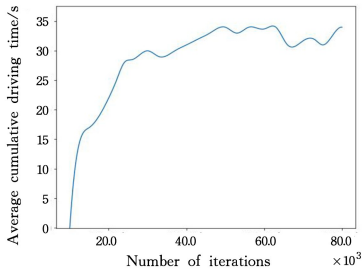

图 14 激进型驾驶风格训练过程平均奖励变化情况

Fig. 14 Changes in average reward during training process of aggressive driving style

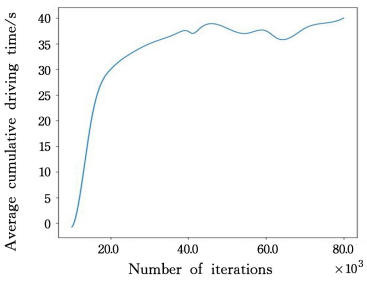

图 15 稳健型驾驶风格训练过程平均奖励变化情况

Fig. 15 Changes in average reward during training process of conservative driving style

平均行驶时长和平均行驶奖励的收敛表明了基于 PPO 算法的风格模型的有效性，并且两种行驶风格的跟车时长都接近于规定的 40 s，表明两种模型均具有完成跟车任务的能力。

为了验证所提出的自动驾驶跟车策略确实能反映驾驶人的行驶风格，本文在相同跟车环境下将两种不同的驾驶风格模型表现以及传统的跟车 IDM 模型[33]、基于 HDD 统计所得到的数据分布进行对比，以验证基于 PPO 算法的风格驾驶模型能够表现出不同的风格。

如图 16—图 19 所示，在选取的 DHW 和 THW 的数据对比中，本文训练所得的激进型模型和稳健型模型在仿真环境中的结果均与 HDD 数据集中的激进型和稳健型驾驶行为特征分布相似。相比之下，IDM 模型在仿真环境中的结果与 HDD 中激进型和稳健型的驾驶行为特征分布差异更大。

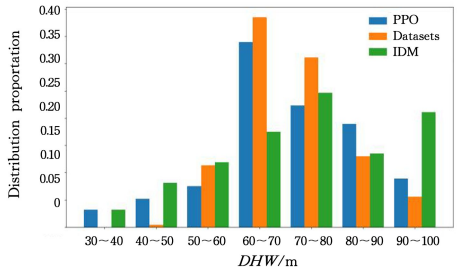

图 16 稳健型驾驶风格在 DHW 上的对比

Fig. 16 Comparison of conservative driving style on DHW

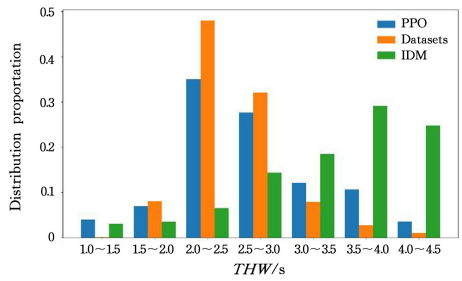

图 17 稳健型驾驶风格在 THW 上的对比

Fig. 17 Comparison of conservative driving style on THW

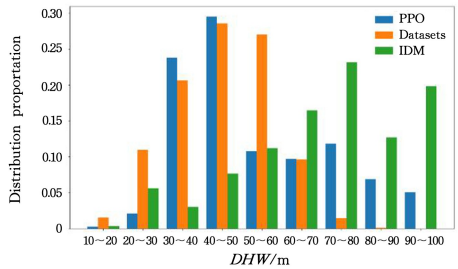

图 18 激进型驾驶风格在 DHW 上的对比

Fig. 18 Comparison of aggressive driving style on DHW

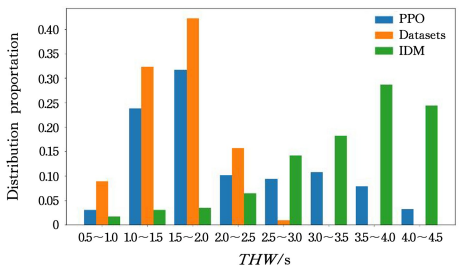

图 19 激进型驾驶风格在 THW 上的对比

Fig. 19 Comparison of aggressive driving style on THW

为了量化基于 PPO 算法的不同风格驾驶模型与 IDM，HDD 数据集原始驾驶数据之间的差异，本文选取海林格距离(Hellinger Distance，HD)[34]和平均绝对误差(Mean Absolute Error，MAE)进行量化。HD 可以量化两个概率分布 $P=\{p_i\}_{i\in N}$ 和 $Q=\{q_i\}_{i\in N}$ 之间的整体相似性；MAE 是两个概率

分布在每一个取值区间上概率差绝对值的平均值，反映的是分布中每一个对应区间的平均相似性。HD 和 MAE 均为取值范围为[0，1]的无因次量，取值越靠近 0，则两个分布相似的程度越高。两个参数的计算方法如下：

$$HD(P,Q)=1/\sqrt{2}\ \|\ \sqrt{P}-\sqrt{Q}\ \|^2 \tag{23}$$

$$MAE=\frac{1}{n\Sigma|\ p_i-q_i\ |} \tag{24}$$

表 3 列出了基于 PPO 算法的不同驾驶风格的驾驶模型与 IDM 模型在 HDD 数据集上的 HD 和 MAE 的量化分析，量化数值越接近 0，证明仿真模型的实验分布与 HDD 数据分布越相似，表现的风格性越突出。

表 3　HDD 数据集上不同风格驾驶行为的分布对比后的量化结果

Table 3　Quantified results of distribution comparison of different driving behaviors styles on HDD dataset

驾驶风格	指标类型	*HD*		*MAE*	
		PPO	IDM	PPO	IDM
激进型	DHW	**0.292**	0.491	**0.067**	0.120
激进型	THW	**0.280**	0.779	**0.039**	0.187
稳健型	DHW	**0.285**	0.338	**0.049**	0.058
稳健型	THW	**0.245**	0.590	**0.038**	0.123

从表 3 中可以看出，基于 PPO 算法的激进型模型和稳健型模型在 HD 和 MAE 上比 IDM 模型得到更小的值，即基于 PPO 算法的不同驾驶风格模型比 IDM 模型更接近 HDD 数据集不同驾驶风格的特征分布，表达出了不同驾驶风格的行为特征，表明基于 PPO 算法的风格驾驶模型具有表达不同驾驶风格的能力。

结束语　本文通过分析大规模驾驶行为数据中提取激进型和稳健型驾驶风格数据特征并用于编码深度强化学习的奖励函数，使得所提模型在跟车场景下尽量还原不同风格的驾驶特征。相关实验结果表明：

(1)经过学习后的 PPO 模型具有目标可达性，能够完成跟车任务，证明了其有效性。

(2)PPO 模型在 HD 和 MAE 两个评价指标上与 HDD 数据集特征分布进行对比，其相似性明显高于 IDM 和 HDD 的对比结果，证明了基于 PPO 的不同驾驶风格模型能够表达驾驶人的不同风格特征。

在稳健型的驾驶模型训练过程中，训练车辆收敛时的平均驾驶时长恒定接近于 40 s，而激进型的驾驶模型训练过程中，训练车辆收敛时的平均驾驶时长恒定接近于 35 s。由此可见，在激进型驾驶模型行驶的过程中更容易发生追尾事故而提前结束驾驶任务。在尽量保持驾驶风格的行驶过程中，如何约束并保证车辆行驶的安全性需要进一步的研究。

参 考 文 献

[1] WEI J，DOLAN J M，LITKOUHI B. A learning-based autonomous driver：emulate human driver's intelligence in low-speed car following[C]// Unattended Ground，Sea，and Air Sensor Technologies and Applications XII. SPIE，2010，7693：93-104.

[2] KESTING A，TREIBER M，HELBING D. Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity[J]. Philosophical Transactions of the Royal Society A：Mathematical，Physical and Engineering Sciences，2010，368(1928)：4585-4605.

[3] CAO W，LIU S，LI J，et al. Analysis and design of adaptive cruise control for smart electric vehicle with domain-based poly-service loop delay[J]. IEEE Transactions on Industrial Electronics，2022，70(1)：866-877.

[4] DARAPANENI N，RAJ P，PADURI A R，et al. Autonomous car driving using deep learning[C]// 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). IEEE，2021：29-33.

[5] YI L M. Lane change of vehicles based on dqn[C]// 2020 5th International Conference on Information Science，Computer Technology and Transportation(ISCTT). IEEE，2020：593-597.

[6] GIPPS P G. A behavioural car-following model for computer simulation[J]. Transportation Research(Part B)：Methodological，1981，15(2)：105-111.

[7] MNIH V，KAVUKCUOGLU K，SILVER D，et al. Human-level control through deep reinforcement learning[J]. Nature，2015，518(7540)：529-533.

[8] SCHULMAN J，WOLSKI F，DHARIWAL P，et al. Proximal policy optimization algorithms[J]. arXiv：1707.06347，2017.

[9] WANG J，ZHANG L，ZHANG D，et al. An adaptive longitudinal driving assistance system based on driver characteristics[J]. IEEE Transactions on Intelligent Transportation Systems，2012，14(1)：1-12.

[10] 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). IEEE，2018：2118-2125.

[11] HEDRICK J K，TOMIZUKA M，VARAIYA P. Control issues in automated highway systems[J]. IEEE Control Systems Magazine，1994，14(6)：21-32.

[12] GAO H，KAN Z，LI K. Robust lateral trajectory following control of unmanned vehicle based on model predictive control[J]. IEEE/ASME Transactions on Mechatronics，2021，27(3)：1278-1287.

[13] XIE G，GAO H，QIAN L，et al. Vehicle trajectory prediction by integrating physics-and maneuver-based approaches using interactive multiple models[J]. IEEE Transactions on Industrial Electronics，2017，65(7)：5999-6008.

[14] GAO H，ZHU J，LI X，et al. Automatic parking control of unmanned vehicle based on switching control algorithm and backstepping[J]. IEEE/ASME Transactions on Mechatronics，2020，27(3)：1233-1243.

[15] ZHANG J，LI Q Y，LI D，et al. Merging guidance of exclusive lanes for connected and autonomous vehicles based on deep reinforcement learning[J]. Journal of Jilin University(Engineering and Technology Edition)，2023，53(9)：2508-2518.

[16] VARAIYA P. Smart cars on smart roads：problems of control[J]. IEEE Transactions on Automatic Control，1993，38(2)：195-207.

[17] GAO H，QIN Y，HU C，et al. An interacting multiple model for

trajectory prediction of intelligent vehicles in typical road traffic scenario[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021.

[18] CODEVILLA F, MÜLLER M, LÓPEZ A, et al. End-to-end driving via conditional imitation learning[C]// 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 4693-4700.

[19] WANG W, XI J, CHEN H. Modeling and recognizing driver behavior based on driving data: A survey[J]. Mathematical Problems in Engineering, 2014, 2014: 245611.

[20] KURITA T. Principal component analysis(PCA)[J/OL]. HTTPS://DOI.ORG/10.1007/978-3-030-03243-2_649-1.

[21] SHLENS J. A tutorial on principal component analysis [J]. arXiv: 1404.1100, 2014.

[22] JAIN A K, MURTY M N, FLYNN P J. Data clustering: a review[J]. ACM Computing Surveys(CSUR), 1999, 31(3): 264-323.

[23] AHMED M, SERAJ R, ISLAM S M S. The k-means algorithm: A comprehensive survey and performance evaluation[J]. Electronics, 2020, 9(8): 1295.

[24] KAELBLING L P, LITTMAN M L, MOORE A W. Reinforcement learning: A survey[J]. Journal of Artificial Intelligence Research, 1996, 4: 237-285.

[25] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[J]. arXiv: 1707.06347, 2017.

[26] SAGBERG F, SELPI, BIANCHI PICCININI G F, et al. A review of research on driving styles and road safety[J]. Human Factors, 2015, 57(7): 1248-1275.

[27] MURPHEY Y L, MILTON R, KILIARIS L. Driver's style classification using jerk analysis[C]// 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems. IEEE, 2009: 23-28.

[28] DE WAARD D, DIJKSTERHUIS C, BROOKHUIS K A. Merging into heavy motorway traffic by young and elderly drivers [J]. Accident Analysis & Prevention, 2009, 41(3): 588-597.

[29] LIU L, LIN J, YAO J, et al. Path planning for smart car based on Dijkstra algorithm and dynamic window approach[J]. Wireless Communications and Mobile Computing, 2021, 2021(1): 8881684.

[30] MACADAM C, BAREKET Z, FANCHER P, et al. Using neural networks to identify driving style and headway control behavior of drivers[J]. Vehicle System Dynamics, 1998, 29(S1): 143-160.

[31] HELLY W. Simulation of bottlenecks in single-lane traffic flow [J]. Theory of Traffic Flow, 1959, 6(2): 207-238.

[32] VAN DER HORST A R A, HOGEMA J H. Time-to-collision and collision avoidance systems[C]// Proceeding of the 6th ICTCT Workshop. 1994: 59-66.

[33] TREIBER M, HENNECKE A, HELBING D. Congested traffic states in empirical observations and microscopic simulations[J]. Physical review E, 2000, 62(2): 1805.

[34] RAO C R. A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance [J]. Qüestiió: Quaderns Destadística i Investigació Operativa, 1995, 19: 23-63.

YAN Xin, born in 1999, postgraduate. His main research interests include reinforcement learning, autonomous driving and so on.

HUANG Zhiqiu, born in 1965, Ph. D, professor, is a distinguished member of CCF(No. 09028D). His main research interests include software quality assurance, system safety, and formal methods.

(责任编辑:何杨)