arXiv:2304.00397v1 [cs.LG] 1 Apr 2023

混合交通中的网联自动驾驶车辆：学习人类驾驶员行为以实现有效的匝道汇入

Nishanth Venkatesh, Viet-Anh Le, Aditya Dave,*IEEE* 学生会员，
Andreas A. Malikopoulos，*IEEE* 高级会员

摘要—混合交通条件下的高速公路汇入场景给网联自动驾驶车辆（CAVs）与从匝道驶入的人类驾驶车辆（HDV）之间的交互建模和控制带来了重大挑战。本文提出一种方法，为网联自动驾驶车辆（CAV）学习 CAV-HDV交互的近似信息状态模型，使其能够在高速公路汇入过程中安全行驶。该方法中，CAV先利用近似信息状态学习驶入匝道的HDV的行为，再生成控制策略以促成汇入。首先，我们利用从下一代仿真数据库中提取的混合交通场景，预测HDV的行为，并据此在真实世界数据上验证该框架的有效性。然后，我们采用标准的逆强化学习方法，为高速公路汇入场景中的有人类驾驶车辆与网联自动驾驶车辆的交互生成仿真数据。在不假设已知生成模型的前提下，我们展示了该近似信息状态模型仅凭观测即可学习预测HDV的未来轨迹。随后，我们为CAV在与HDV共同汇入时生成安全控制策略，展示从激进到保守的一系列驾驶行为。我们通过数值仿真展示了所提方法的有效性。

I. 引言

网联自动驾驶车辆（联网自动驾驶汽车）有望显著提升能源效率、安全性和舒适度 [1]。大量研究致力于在网联自动驾驶车辆渗透率达到 100% 的情况下，确保能源效率 [2], 安全性 [3], [4],和乘客舒适度 [5] 。我们预计道路上的网联自动驾驶车辆数量将逐步增加，因此，网联自动驾驶车辆必须能够与有人类驾驶的车辆安全地协同运行。由于人类驾驶行为在变道、交叉路口、环岛和高速公路匝道汇入等各种情况下具有不可预测性，有人类驾驶的车辆的存在给网联自动驾驶车辆的控制带来了挑战。通过开发在线预测人类驾驶行为的方法，并动态调整网联自动驾驶车辆的轨迹，可以应对这一挑战。

近年来，车辆交互与控制的博弈论模型受到了广泛关注。Liu等人。 [6] 提出了一种基于领导者—跟随者博弈的决策算法，用于处理合流场景中的车辆两两交互。Chandra 和 Manocha[7] 提出了一种基于机器人学的拍卖框架，用于车辆在无信号交叉路口、环岛和合流场景中的导航。文献 [8], [9]提出了一种博弈论模型预测控制方法，并将移动时域逆强化学习与权重自适应策略相结合，用于网联自动驾驶车辆与人类驾驶员交互。

其他方法则侧重于利用深度学习方法，在不同情境下预测人类驾驶行为，并通过循环层学习数据中的时间关系。例如，Altch´e 和 Fortelle [10] 提出了一种基于长短期记忆（LSTM）的神经网络架构，用于预测在直路上行驶的车辆的未来轨迹。Park 等. [11] 使用基于 LSTM 的模型生成 HDV 周围所有车辆的未来轨迹。类似地，Deo 和 Trivedi 使用 LSTM 对高速公路上周围车辆进行交互感知运动预测 [12]，并稳健地学习车辆运动之间的相互依赖关系 [13]。LSTM 还与 [14]基于占用栅格的框架结合使用，以考虑预测轨迹中的潜在因素，例如道路结构、交通规则和驾驶员意图。

虽然大多数深度学习方法的主要目标是预测 HDV 的未来行为，但人们对利用此类预测来控制与 HDV 交互过程中的 CAV 越来越感兴趣。Kherroubi 等. [15]训练了一个神经网络模型，用于预测人类驾驶员的意图，并通过深度策略梯度方法学习驾驶策略。Guo 等. [16] 采用 Q学习控制混合交通中的 CAV 变道，其中 HDV 遵循 Gipps 跟驰模型。尽管 HDV 行为预测已取得显著进展，但如何在保留对不同 HDV 行为模型泛化能力的同时优先保障安全性，仍是一个有待解决的研究问题。建立一个能够指导设计和实现学习赋能型[17]CAV 的框架，并确保其具备高置信度的安全性，正当其时。

本文聚焦于这一综合问题：学习有人类驾驶的车辆的行为，并利用学得的模型为网联自动驾驶车辆生成安全控制策略。为此，我们借鉴了部分可观测强化学习领域的相关研究思路。Subramanian 等. [18], [19]提出了一个系统化框架，用于学习状态空间表示，并为隐藏状态不可观测的系统计算控制策略。这种称为近似信息状态模型的模型，已用于机器人学中的系统辨识及后续控制 [20]，也用于强化学习中的医疗死路识别 [21]。类似地，Dave 等. [22], [23] 提出了一种非随机

本研究得到美国国家科学基金会（NSF）资助，资助编号为 CNS-2149520 和 CMMI-2219761。作者来自特拉华大学机械工程系，美国特拉华州纽瓦克市，邮编 DE19716（电子邮箱： {nish; vietale；adidave；andreas}@udel.edu）。

框架，用于学习近似模型并推导极小极大控制策略。近似信息状态这一概念构成了我们方法的基础。

本文的主要贡献如下：我们提出了一种方法，用于学习高速公路汇入场景中匝道有人类驾驶的车辆的行为，并利用学习得到的模型，为高速公路上的网联自动驾驶车辆生成优先保障安全的策略。为此，我们根据一组预期目标（包括安全性和能源效率）制定网联自动驾驶车辆的控制问题，并实施以下步骤：(1) 构建并训练一种神经网络架构，在无需预先了解有人类驾驶的车辆动力学的情况下预测其轨迹。我们将观测历史视为能够代表有人类驾驶的车辆驾驶行为，并将其压缩为受 [19] 启发的近似状态表示，从而实现这一目标。(2) 开发一种迭代式模型预测控制（MPC）算法，利用学习得到的近似信息状态模型，在线为网联自动驾驶车辆生成安全且节能的控制策略。(3) 通过成功预测下一代仿真（NGSIM）数据集中的车辆轨迹，验证所提出学习框架的有效性。(4) 在基于有人类驾驶车辆与网联自动驾驶车辆交互生成的数据上进行 5000 次仿真，涵盖从激进到保守的多种有人类驾驶的车辆行为，验证结合学习与控制的方法生成网联自动驾驶车辆动作的有效性。

本文其余部分安排如下。II节中，我们针对有人类驾驶的车辆从匝道汇入时网联自动驾驶车辆的汇入问题进行建模。III节中，我们介绍基于近似信息状态模型的预备知识，随后介绍用于学习和预测网联自动驾驶车辆与有人类驾驶的车辆之间的交互的编码器-解码器模型，以及如何训练神经网络架构。IV节中，我们分析如何利用所学习的模型求解问题建模，并给出各种合流场景的数值仿真结果。最后，在V节中，我们总结全文并讨论正在开展的工作。

II. 问题建模

本节针对混合交通高速公路汇入场景，建立模型预测控制问题，以计算网联自动驾驶车辆的控制输入。如图1所示，编号为1的网联自动驾驶车辆和编号为2的有人类驾驶的车辆分别行驶在主路和匝道上。我们将图1中绿色区域所示的范围定义为控制区域。在控制区域内，CAV–1的控制输入由我们提出的方法确定；在控制区域外，则可由跟驰模型（例如智能驾驶员模型 [24]）确定。我们设定控制区域从每条道路上距离冲突点上游 $L_c \in \mathbb{R}_{>0}$处开始，即从CAV–1和HDV–2的行驶路径交汇的位置开始。冲突点是可能发生侧向碰撞的区域，在图1中以红圈标出。在此合流场景中，考虑到HDV–2的存在，我们的目标是控制CAV–1安全且有效地汇入主路。

在我们的表述中，CAV–1 和 HDV–2 的纵向位置从控制区域起点开始测量

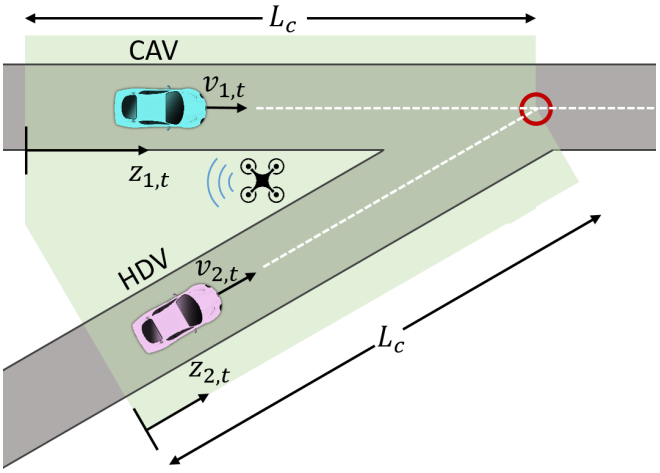


图 1：混合交通合流场景中的控制区域（绿色区域）和冲突点（红色圆圈）。

在任意 $t \in \mathbb{N}$ 时，我们分别用 $z_{1,t} \in \mathbb{R}_{\geq 0}$ 和$z_{2,t} \in \mathbb{R}_{\geq 0}$ 表示。令$v_{1,t} \in \mathbb{R}_{\geq 0}$ 和 $u_{1,t} \in \mathbb{R}$分别表示 CAV–1 在时刻 t 的速度和加速度，$v_{2,t} \in \mathbb{R}_{\geq 0}$ 和$u_{2,t} \in \mathbb{R}$ 分别表示 HDV–2 在时刻 t 的速度和加速度。我们将每辆车辆$i = 1, 2$ 的位置和速度统称为其在时刻 t 的状态 $x_{i,t} = (z_{i,t}, v_{i,t})$。然后，从初始状态 $x_{1,0}$ 和 $x_{2,0}$ 开始，在 $t = 0$ 时，每辆车的状态按照以下规律演变

$$x_{1,t+1} = f_{1,t}(x_{1,t}, u_{1,t}), \tag{1}$$
$$x_{2,t+1} = f_{2,t}(x_{2,t}, u_{2,t}). \tag{2}$$

本文假设，对于所有 $t \in \mathbb{N}$，式 (1) 中 CAV–1 的动力学采用以下双积分器模型

$$\begin{aligned} z_{1,t+1} &= z_{1,t} + \Delta T v_{1,t} + \frac{1}{2}\Delta T^2 u_{1,t}, \\ v_{1,t+1} &= v_{1,t} + \Delta T u_{1,t}, \end{aligned} \tag{3}$$

其中，$\Delta T \in \mathbb{R}_{>0}$ 是两个离散时间步之间的采样时间。我们考虑 CAV–1 的以下状态和控制约束

$$0 \leq v_{\min} \leq v_{1,t} \leq v_{\max}, \quad \text{and} \quad u_{\min} \leq u_{1,t} \leq u_{\max}, \tag{4}$$

其中，$u_{\min}$、$u_{\max} \in \mathbb{R}$ 分别为最小减速度和最大加速度，$v_{\min}$、$v_{\max} \in \mathbb{R}_{>0}$ 分别为最低和最高速度限制。CAV–1 和 HDV–2 均只能沿前进方向行驶，因此，对于 $i = 1$，其速度 $v_{i,t}$ 2 始终非负。然而，我们不对 HDV–2 的速度施加已知的先验上界，因为人类驾驶行为可能导致其超出该上界。下一小节将对 CAV–1 的控制问题进行建模。

A. 模型预测控制表述

在本小节中，给定系统动力学 (3) 以及状态和控制约束 (4)，我们针对合流场景构建一个用于控制 CAV–1 的模型预测控制问题。令

$H \in \mathbb{N}$ 为控制时域的长度， $t \in \mathbb{N}$为当前时间步，$\mathcal{I}_t = \{t, \dots, t+H-1\}$ 为以 t 为起点的控制时域内所有时间步的集合。模型预测控制问题的目标可由以下三个特征线性组合而成：（*i*）最小化控制输入，（*ii*）最小化与最大允许速度之间的偏差，以缩短行驶时间，以及（*iii*）与碰撞避免约束 [9] 相对应的对数惩罚函数。因此，CAV–1 在任意时刻 k 的目标函数由以下成本给出

$$\begin{aligned} l(x_{1,k+1}, u_{1,k}) = \omega_1 u_{1,k}^2 + \omega_2 (v_{1,k+1} - v_{\max})^2 \\ - \omega_3 \log\big((z_{1,k+1} - z_c + \rho v_{1,k+1})^2 \\ + (\hat{z}_{2,k+1} - z_c + \rho \hat{v}_{2,k+1})^2\big), \end{aligned} \tag{5}$$

其中，ω_1、ω_2和 ω_3 为正权重，z_c 为冲突点的位置，$\rho \in \mathbb{R}_{>0}$ 为一个用于考虑 HDV–2 反应延迟的参数。增大 ρ表示通过假设 HDV–2 的反应延迟更长，来提高目标函数的保守性。在（5）中，$\hat{}z_{2,\ k+1}$ 和 $\hat{}v_{2,\ k+1}$ 表示 HDV–2 在时间步 $k+1$的预测未来位置和速度。于是，针对每个 $t \in \mathbb{N}$，CAV–1 的模型预测控制问题可表述为

$$\underset{\{u_{1,k}\}_{k \in \mathcal{I}_t}}{\text{minimize}} \sum_{k \in \mathcal{I}_t} l(x_{1,k+1}, u_{1,k}), \tag{6a}$$

$$\text{subject to: } (3), (4). \tag{6b}$$

我们希望在每个 $t \in \mathbb{N}$，求得控制输入（$u_{1,\ t},\ \dots,\ u_{1,\ t+H-1}$），使（6）最小化，而无需预先了解（2）中给出的 HDV–2 动力学。因此，CAV–1 必须学习一个模型，以预测目标函数（5）所需的 HDV–2 未来轨迹。为确保 CAV–1 能够获取 HDV–2 的实时数据，我们在框架中作出如下假设。

假设1. 存在一个协调器（可将其视为图1所示的无人机），用于收集 HDV–2 的状态和控制动作的实时观测，并立即将其传输给 CAV–1。协调器能够在观测和通信过程中均不受任何噪声影响地完成上述操作。

备注1. 在该时域内，HDV–2 未来的控制动作和轨迹并非预先确定，因为：(i) HDV–2 的动力学未知；(ii) CAV–1 的状态和动作对 HDV–2 动作的影响也未知。鉴于 HDV–1 会对 CAV–1 [8], 的控制动作作出反应，我们将 CAV–1 与 HDV–2 的联合动力学建模为一个未知的部分可观测马尔可夫决策过程(POMDP)。接下来，我们采用 POMDP 的近似信息状态框架，学习一个模型，以预测 HDV–2 与 CAV–1 动作相耦合的轨迹。

III. 学习框架

A. 预备知识

本小节介绍部分可观测马尔可夫决策过程以及近似来自 [18] 的信息状态框架。这一表述有助于我们奠定基础，以理解如何在对底层动态了解不完全的情况下分析系统。随后，我们将在合流场景中运用这一框架，学习一个用于预测 HDV–2 未来轨迹的模型。

部分可观测马尔可夫决策过程基础：我们将部分可观测马尔可夫决策过程定义为一个元组 $(\mathcal{X}, \mathcal{U}, \mathcal{Y}, T, O, c, \gamma)$，其中，可行状态集合为 $\mathcal{X}$，可行控制动作集合为$\mathcal{U}$，可行观测集合为 $\mathcal{Y}$；函数T：$\mathcal{X} \times \mathcal{U} \times \mathcal{X} \to [0,1]$ 给出转移概率$T(x_t, u_{t-1}, x_{t+1}) = p(x_{t+1}|x_t, u_{t-1})$，对所有 x_t、$x_{t+1} \in \mathcal{X}$和 $u_t \in \mathcal{U}$均成立；函数 O：$\mathcal{X} \times \mathcal{Y} \to [0,1]$ 给出观测概率 $O(x_{t+1}, y_{t+1}) = p(y_{t+1}|x_{t+1})$，对所有 $x_{t+1} \in \mathcal{X}$ 和 $y_{t+1} \in \mathcal{Y}$均成立；函数 c：$\mathcal{X} \times \mathcal{U} \to \mathbb{R}$给出成本 $c(x_t, u_t)$，对所有 $x_t \in \mathcal{X}$ 和 $u_t \in \mathcal{U}$均成立；而$\gamma \in [0,1)$为折扣因子。在给定时刻 t 之前的观测和控制动作历史构成记忆，定义为 $m_t = (m_{t-1}, y_t, u_{t-1}) \in \mathcal{M}_t$。

近似信息状态：近似信息状态是对系统数据的压缩，这些数据存储在记忆中，可用于通过动态规划分解，为部分可观测马尔可夫决策过程计算近似最优控制策略。在我们的问题中，它具有一项重要性质：无需完整了解底层动态，仅凭观测数据即可学习近似信息状态模型。考虑一个可测空间 $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$，其中 $\mathcal{B}(\mathcal{X})$ 是 $\mathcal{X}$ 上的 σ-代数。我们用 $d(\cdot,\cdot)$ 表示任意两个概率分布 $\mu, \nu \in \Delta(\mathcal{X})$ 之间的积分概率度量。这类度量的例子包括全变差度量、Wasserstein度量和最大均值差异度量。于是，对于时域为 $T \in \mathbb{N}$ 的部分可观测马尔可夫决策过程，时不变的近似信息状态模型定义为一个元组 $(\mathcal{S}, \hat{}p, \hat{c}, \sigma_t : t = 0, \dots, T)$，其中 $\mathcal{S}$ 是可行信息状态构成的巴拿赫空间，$\hat{}p$：$\mathcal{S} \times \mathcal{U} \to \Delta(\mathcal{S})$ 是马尔可夫概率核，$\hat{c}$：$\mathcal{S} \times \mathcal{U} \to \mathbb{R}$ 是近似代价函数，σ_t：$\mathcal{M}_t \to \mathcal{S}$ 是记忆压缩函数。如果存在 $\varepsilon, \delta \in \mathbb{R}_{>0}$，使得对于记忆 $m_t \in \mathcal{M}_t$ 的所有取值以及控制动作 $u_t \in \mathcal{U}$ 的所有取值，以下性质均成立，则该模型构成此部分可观测马尔可夫决策过程的一个 (ε, δ)-近似模型：

(AP1) 足以近似预测成本 $c_t = c(x_t,\ u_t)$

$$|\mathbb{E}[c_t(x_t, u_t) \,|\, m_t, u_t] - \hat{c}(\sigma_t(m_t), u_t)| \le \varepsilon. \tag{7}$$

(AP2a) 近似信息状态以确定性方式演变，即存在一个可测的更新函数 ψ：$\mathcal{S} \times \mathcal{Y} \times \mathcal{U} \to \mathcal{S}$，使得

$$\sigma_t(m_t) = \psi(\sigma_t(m_{t-1}), y_t, u_{t-1}). \tag{8}$$

(AP2b) 足以近似预测未来观测，即存在一个概率核 $\hat{}p^{\mathcal{Y}}$：$\mathcal{S} \times \mathcal{U} \to \Delta(\mathcal{Y})$，使得对于 $\mathcal{Y}$ 的任意博雷尔子集 B，基于记忆的分布 $\mu_t(B) = p(y_{t+1} \in B | m_t, u_t)$ 和

预测分布 $\phi(B) = \hat{p}^{y}(y_{t+1} \in B|\sigma_t(m_t)\ u_t)$ 满足

$$d(\mu_t, \phi) \leq \delta. \tag{9}$$

一个（ε, δ）近似信息状态模型构成了一个完全可观测的马尔可夫决策过程，其每个 t 的状态由 $s_t \in \mathcal{S}$给出。因此，我们可以通过对问题进行动态规划分解，利用该模型计算近似最优控制策略。已有研究证明，这种近似策略的性能损失有界，并且性能会随着 ε 和 δ [18, 的降低而线性提升（定理9]）。这一特性意味着，近似信息状态构成了一种有理论依据的记忆压缩，可以通过尽可能满足（**AP1**）和（**AP2b**），同时确保满足（**AP2a**），来学习这种压缩。

鉴于这些优势，我们在CAV-HDV汇入场景中采用这一框架来预测HDV–1的轨迹，进而为CAV–1计算控制策略。在该框架中，在每个 $t \in \mathbb{N}$，CAV–1的状态为$x_{1,t} = (z_{1,t}, v_{1,t})$，控制动作为$u_{1,t}$，而CAV–1接收到的观测为$y_t = (z_{1,t}, v_{1,t}, u_{1,t-1}, z_{2,t}, v_{2,t}, u_{2,t-1})$。我们采用编码器-解码器神经网络架构来表示一个近似信息状态模型，该模型学习表征CAV–1的动作对HDV–2动作的未知影响。如图2所示，我们将编码器记为 ψ：$\mathcal{S}\times\mathcal{Y}\times\mathcal{U} \to \mathcal{S}$，将解码器记为 ϕ：$\mathcal{S} \times \mathcal{U} \to \mathcal{Y}^3$，其中 $\mathcal{Y} \subset \mathbb{R}^6$。我们旨在完全通过观测数据学习这一近似信息状态模型。为此，我们将解码器预测的分布与构成真实值采样点的数据进行比较。接下来，我们将介绍神经网络架构的各个组成部分。

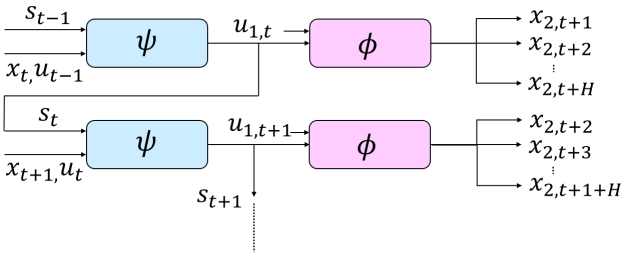


图 2：编码器-解码器架构的示意图。

B. 编码器

编码器 ψ 旨在学习一种函数映射，以强制满足（**AP2a**），即其目标是将观测历史压缩为能够表示 *CAV-HDV*交互 的近似信息状态。因此，在每个 $t \in \mathbb{N}$，编码器的输出为近似信息状态 $s_t = \sigma_t(m_t)$，其中 m_t 是记忆，其输入包括之前的近似信息状态、最新观测以及之前给定的控制动作，并以元组（s_{t-1}，y_t，u_{t-1}）表示。我们在图 2 中对此进行了说明。编码器由两个全连接层和一个循环神经网络（RNN）组成。RNN 在每个 $t \in \mathbb{N}$ 的隐藏状态都被视为编码器的输出s_t。线性层负责接收输入（y_t，u_{t-1}），而 RNN 确保编码器在时间 t也将之前的近似信息状态 s_{t-1} 作为输入。因此，RNN 的结构确保近似信息状态在每个 $t \in \mathbb{N}$ 更新为

$$s_t = \psi(s_{t-1}, y_t, u_{t-1}), \tag{10}$$

其中，ψ 是一个确定性函数，表示由编码器中的神经网络层生成的映射。

C. 解码器

解码器 ϕ 旨在学习一种函数映射，以满足 **(AP1)** 和 **(AP2b)**，即在任意时刻，根据近似信息状态和控制动作预测未来观测的分布及期望成本。在我们的实现中，解码器 ϕ 由三个全连接层组成。在每个 $t \in \mathbb{N}$，解码器以 CAV–1 的近似信息状态 s_t 和控制动作 $u_{1,\,t}$ 作为输入。解码器输出 HDV–2 在时域 $\mathcal{I}_t$内的未来状态分布 $(x_{2,\,t+1},\ x_{2,\,t+2},\ \ldots,\ x_{2,\,t+H})$，而非仅输出下一个观测 y_{t+1}。因此，解码器的输出为

$$\phi(s_t, u_{1,t}) = \hat{p}^{y}(x_{2,t+1}, \ldots, x_{2,t+H} \mid s_t, u_{1,t}), \tag{11}$$

其中 $\hat{p}^{y}$ 是观测的预测概率分布。因此，在我们的架构中，解码器旨在对整个控制时域 $\mathcal{I}_t$ 强制满足（**AP2b**），而不仅仅是在式（9）所述的时刻 $t+1$ 满足。这有两个作用：（i）在实现模型预测控制器时，可在控制时域 $\mathcal{I}_t$ 内进行深度规划；（ii）使解码器的输出专注于预测 HDV–2 的未来状态。我们没有在解码器的输出中加入期望成本项，因为每个时刻的成本（5）都是观测和 CAV–1 预测的确定性函数。注意，我们只需关注 HDV–2，因为 CAV–1 的动力学已由式（3）给出，无需学习。因此，如果解码器能够学会尽可能准确地预测轨迹，就会自动确保（**AP1**）和（**AP2b**）均成立。

D. 训练

在本小节中，我们将（**AP2b**）中式（9）左侧的内容视为训练损失，并据此说明如何训练编码器 ψ 和解码器 ϕ。因此，训练损失通过积分概率度量，衡量观测到的未来状态的预测概率分布与真实值的底层分布之间的距离。为此，我们采用 [19, Proposition32] 中定义的基于距离的最大均值差异（MMD）度量，以刻画两个概率分布之间的距离。

将真实值的底层分布记为 p，将解码器的预测分布记为$\hat{p}^{y}$。因此，训练损失为 $d(\hat{p}^{y},\ p)^2$,，其中 $d(\cdot,\ \cdot)$是 MMD 指标。由式（11）可知，该分布

$\hat{p}^y$ 由编码器-解码器网络的权重参数化。我们将其视为一个多元正态分布，该分布描述了时域 $\mathcal{I}_t$内 HDV–2 各状态未来可能实现值($x_{2,\ t+1}$、……、 $x_{2,\ t+H}$)的分布。随后，我们将$\hat{p}^y$ 的标准差固定为 1，并认为解码器的输出表示该分布的均值向量。在每个 $t \in \mathbb{N}$，我们都可以从底层分布 p中获得一个采样观测($x_2^{\text{ref}}{}_{,\ t+1}$、……、 $x_2^{\text{ref}}{}_{,\ t+H}$)。为了最小化损失 $d(\hat{p}^y,\ p)^2$,，我们需要仅利用采样点构造该训练损失梯度的无偏估计量。根据 [19, 命题 33], ，可以通过对$\hat{p}^y$ 的均值和来自 p 的采样点求梯度，以无偏方式估计梯度 $\nabla d(\hat{p}^y,\ p)^2$：

$$(x_{2,t+i} - 2 \cdot x_{2,t+i}^{ref})^\top x_{2,t+i}. \tag{12}$$

我们使用 (12) 作为训练的替代损失，该损失根据预测均值（$x_{2,\ t+1}$，……，$x_{2,\ t+H}$）和真实值样本（$x_2^{\text{ref}}{}_{,\ t+1}$，……，$x_2^{\text{ref}}{}_{,\ t+H}$）计算得出。

IV. 仿真与结果

A. 数据集

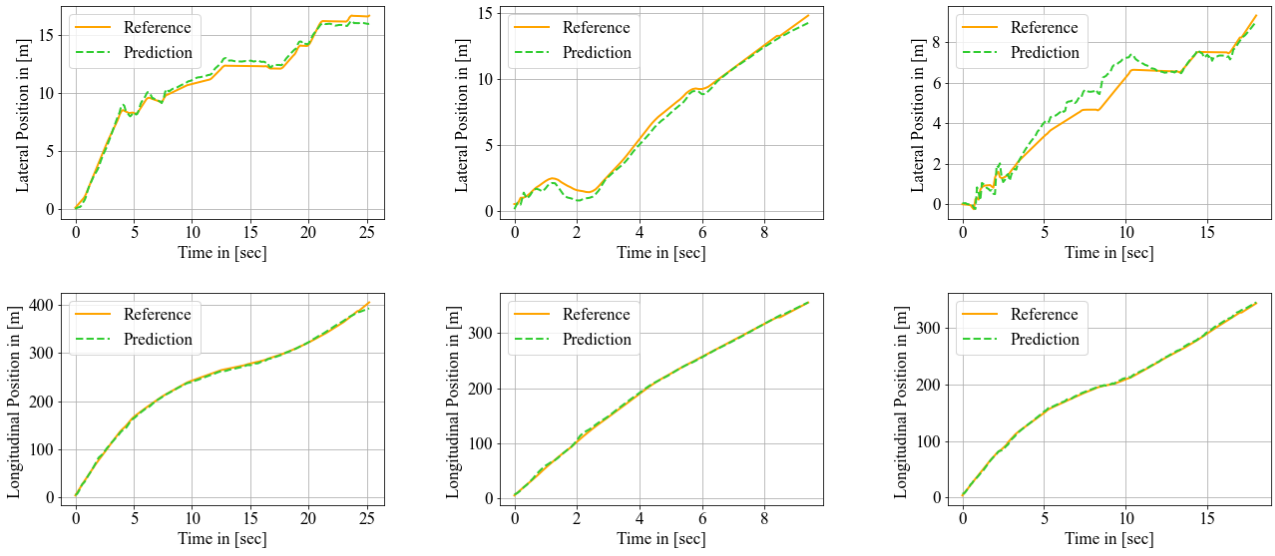
我们使用三个数据集训练并验证近似信息状态模型。这些数据集包含高速公路匝道汇入场景中有人类驾驶的车辆的轨迹。对于第一个数据集，我们从 I-80 [25] 的 NGSIM 存储库中收集轨迹数据。具体而言，我们提取匝道上车辆的位置、速度和加速度，以及它在汇入过程中与之交互的前后两辆公路车辆的位置、速度和加速度。我们使用这些数据验证近似信息状态模型在真实汇入场景中预测人类驾驶车辆轨迹的能力。针对这一任务对第三节中的架构进行专门训练的细节及结果见 IV-B小节。

接下来，我们采用 [26] 中介绍的逆强化学习 (IRL) 技术，在仿真中复现人类驾驶员的行为，并为 II 节所述场景生成两个数据集。使用该 IRL 技术时，我们假设，有人类驾驶的车辆的控制动作可以表示为某个基于特征的目标函数的最小化器，其中不同的驾驶风格由特征权重刻画。通过使用有人类驾驶的车辆模型的不同参数开展仿真并收集数据，我们可以捕捉人类驾驶员可能采取的各种动作，范围从保守驾驶到激进驾驶。这种方法生成了第二和第三个数据集，供后续内容中近似信息状态模型的训练和验证使用。在第二个数据集中，我们假设 CAV–1 针对不同的有人类驾驶的车辆采用 [9] 提出的安全控制策略。这确保数据集中仅有 1% 的情形呈现不安全汇入场景。在第三个数据集中，我们假设 HDV–2 和 CAV–1 的控制动作均通过权重随机化的 IRL 技术生成。这模拟了 CAV 的一种探索性控制策略，且在数据生成过程中不考虑任何安全性。我们使用第二和第三个数据集验证 IV-D 小节所述模型预测控制策略的性能。该模型预测控制方法利用分别在每个数据集上训练的近似信息状态模型的预测结果，为 CAV–1 针对 HDV–2 的各种行为生成控制动作。

在接下来的小节中，我们将介绍在各个数据集上训练近似信息状态模型的具体细节，以及对训练后模型的验证。请注意，编码器-解码器神经网络架构并不假设事先了解用于生成数据的策略或模型。

B. NGSIM 的网络架构与结果

在本小节中，我们训练编码器和解码器，以累积学习从NGSIM数据库收集的第一个数据集的近似信息状态模型。回顾IV-A小节，我们的数据集包含匝道车辆以及与其交互的两辆连续公路车辆的轨迹信息。因此，在每个时间点，编码器 ψ 将每辆车辆状态和动作的最新观测作为输入，即一个9维输入，其中每辆车辆分别包含2维的位置和速度观测以及1维动作。编码器由两个维度分别为（9,8）和（8,16）的全连接层组成，每层均采用ReLU激活，随后连接一个带门控循环单元（GRU） [27]层，其隐藏状态大小为24。我们将隐藏状态视为编码器的输出，并将其作为任意时间点的近似信息状态。近似信息状态和每辆车辆的最新动作被作为解码器 ϕ的输入。解码器由三个全连接层组成，其维度分别为（$24+3$, 32），(32, 64）和（64, 2），前两层采用ReLU激活。解码器经过训练，用于预测匝道车辆在 $H=1$ 个时间步时域内的轨迹。图3绘制了三辆匝道车辆的预测轨迹与其实际轨迹的对比结果。该图的每一列对应一辆车辆，其中第一行的图表示垂直于公路方向的位置（标记为横向位置）的对比，第二行则绘制沿公路方向的位置（标记为纵向位置）。近似信息状态模型在预测纵向位置方面表现良好。同时，我们的模型能够沿横向和纵向两个轴 closely 预测匝道车辆的轨迹。参考轨迹与预测轨迹之间横向位置的均方根误差分别为3(a)中的0.35米、3(b)中的0.39米和3(c)中的0.54米；而纵向位置的均方根误差分别为3(a)中的2.62米、3(b)中的1.92米和3(c)中的1.73米。这些结果验证了我们的近似信息状态模型在涉及人工驾驶车辆的现实汇入场景中的预测能力。



(a) (b) (c)

图 3：展示三个从匝道汇入的不同汽车 (a)、(b) 和 (c) 的 NGSIM 数据集轨迹预测结果与参考轨迹对比的一组图表。

C. 逆强化学习的网络架构

在本小节中，我们介绍用于学习通过IRL生成的两个数据集的近似信息状态模型的编码器-解码器架构的详细信息。

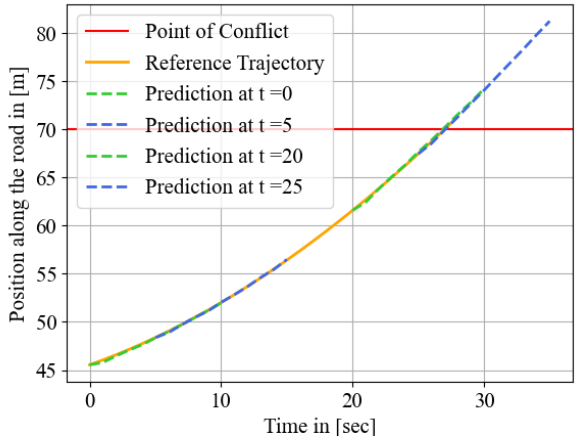


图 4：轨迹预测 v/s 是生成数据集中的实际轨迹。

编码器和解码器遵循与 IV-B 小节中所使用网络相同的基本结构。由于我们考虑的是 II 节中介绍的场景，因此与之前的网络相比，我们减少了每一层的输入和输出数量。请注意，在生成的数据集之间，我们使用完全相同的网络。在每个时刻，编码器 ψ 以 CAV–1 和 HDV–2 的状态与动作的最新观测作为输入，这是一个 6 维输入，其中每辆车辆分别包含 2 维的位置和速度观测，以及 1 维动作。编码器结构为（6，8）、（8，16），并具有 ReLU 激活，后接一个隐藏状态大小为 4 的门控循环单元。再次强调，我们将隐藏状态视为编码器的输出，并将其作为任意时刻的近似信息状态。近似信息状态和 CAV–1 的最新动作作为解码器 ϕ的输入。解码器由三个全连接层组成，其维度分别为（$4+1$，2）、（2，4）和（4，H），其中 H 是时域的长度，前两层使用 ReLU 激活。我们使用（12）中所述的基于距离的最大均值差异来训练编码器-解码器神经网络架构。我们在图 4 中展示轨迹预测结果，将不同时间点 $t=0$、5、20、25 秒的预测轨迹与参考轨迹进行比较。所有预测轨迹与参考轨迹之间的均方根误差为 0．17 米。对于这些数据集，我们将控制区域的长度设为 $L_c=70$ 米，并使用一条红色水平线表示冲突点。

D. 迭代式模型预测控制实现

为求解模型预测控制问题并获得当前时间步 t下 CAV–1 的控制输入$u_{1,\,t}$ 我们利用编码器-解码器神经网络预测控制时域内的人类驾驶轨迹。同时，神经网络的解码器还需要将 $u_{1,\,t}$ 作为输入来进行预测。可以将编码器-解码器神经网络直接视为模型预测控制问题中的约束。然而，由于神经网络模型的复杂性，由此产生的模型预测控制问题在计算上变得难以处理。因此，我们提出了一种迭代式模型预测控制实现，依次计算神经网络预测结果并求解模型预测控制问题。迭代式模型预测控制实现详见算法1。在每个时间步，我们初始化 $u_{1}^{(0)}{}_{,\,t}=0$ 并计算近似信息

状态的预测，表示为 $s_t = \psi(s_{t-1},\ x_t,\ u_{t-1})$，其中输入包括先前的近似信息状态 s_{t-1}、当前状态 x_t和先前的控制输入 u_{t-1}。接下来，在算法的每次迭代 j 中，我们利用 s_t 近似信息状态和上一次迭代中模型预测控制的控制输入$u_{1\,t}^{(j-1)}$，，通过 $y_{t+1:\,t+H}^{(j)} = \phi(s_t,\ u_{1\,t}^{(j-1)},)$预测 HDV–2 的状态和动作。然后，根据 $y_{t+1:\,t+H}^{(j)}$ 的预测结果求解模型预测控制问题，以获得新的控制输入 $u_{1\,t}^{(j)},$。该过程重复进行，直至达到最大迭代次数 $j_{\max}$。

算法1 迭代式模型预测控制实现

要求： $t,\ H \in \mathbb{N},\ j_{\max} \in \mathbb{N},\ u_{1,t}^{(0)} = 0$

1: 通过 s_t 预测近似信息状态 $= \psi(s_{t-1}, x_t, u_{t-1})$

2: **对于** $j = 1,\ 2,\ \ldots\ldots,\ j_{\max}$ **执行**

3: 通过以下方式预测 HDV–2 的状态和控制动作 $y_{t+1:t+H}^{(j)} = \phi(s_t,\ u_{1,t}^{(j-1)})$

4: 求解 (6) 得到 $u_{1,t:t+H-1}^{(j)}$

5: **返回** $u_{1,t}^{(j_{\max})}$

E. 仿真结果

我们使用 Python 编程语言进行了仿真，其中分别使用 CasADi [28] 和内置的 IPOPT 求解器[29] 来构建和求解模型预测控制问题。模型预测控制的参数设定为：$\Delta T = 0.2$s，$H = 10$，$v_{\min} = 0.0$m/s，$v_{\max} = 14.0$ m/s，$u_{\min} = -3.0$ m/s^2，$u_{\max} = 2.0$m/s^2，$\omega_1 = 1.0$，$\omega_2 = 10.0$，$\omega_3 = 10^3$，$\rho = 1.0$s。回顾一下，我们通过采用具有不同权重的逆强化学习模型来模拟人类驾驶行为，以验证控制方法对不同驾驶风格的适用性。

图 5 和图 6 展示了在两种仿真中，CAV–1 和 HDV–2 的轨迹、速度和控制输入；这两种仿真分别对应两种人类驾驶风格，即激进型和保守型人类驾驶员。如图所示，在两种仿真中都能保证安全距离，但 CAV–1 会根据 HDV–2 的行为采取不同的汇入操作。如果人类驾驶员加速并抢先通过，CAV–1 就会减速让行；如果人类驾驶员较为保守，CAV–1 则会判断在 HDV–2 之前通过是安全的。

表 I：使用安全和探索性策略数据训练的模型比较。

反应时间延迟（ρ）	**Safe**	**探索性**
0.6	4984 (99.6%)	4578 (91.5%)
0.8	4996 (99.9%)	4793（95.8%）
1.0	5000（100%）	4971（99.4%）

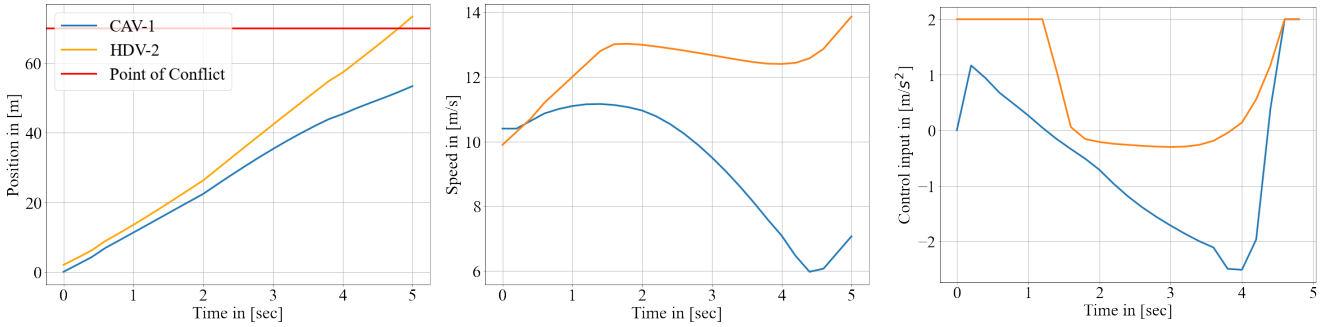
为评估采用两个学习得到的预测模型的MPC性能，我们针对车辆的不同初始条件以及HDV–2的不同逆强化学习模型进行了5000次仿真，并在 ρ的三个不同取值下比较安全性水平。回顾II-A小节，CAV–1的保守性会随 ρ 增加。因此，我们观察到不安全情况的数量会随着 ρ的增加而减少。结果如表I所示，采用基于安全策略数据集训练的模型的MPC，相较于采用基于探索策略数据集训练的模型的MPC，能够实现更高的安全性水平。我们认为，这可能是因为针对探索策略数据集的模型需要进一步训练。

V. 结语

本文提出了一种方法，用于学习网联自动驾驶车辆与有人类驾驶的车辆之间交互的近似信息状态模型，使网联自动驾驶车辆能够在高速公路汇入时安全地进行机动。我们通过使用 NGSIM数据库中的现实汇入场景对模型进行训练并作出预测，验证了该近似信息状态模型的能力。随后，我们展示了利用针对包含一辆网联自动驾驶车辆和一辆有人类驾驶的车辆的混合交通场景、由逆强化学习模型生成的数据，可以学习近似信息状态模型，以预测有人类驾驶的车辆的未来轨迹。我们提出了一种迭代模型预测控制算法，利用这些预测为网联自动驾驶车辆生成控制动作，并通过数值仿真表明，这些控制动作能够应对有人类驾驶的车辆的一系列驾驶行为，确保安全。未来的研究应着重于：（i）将该框架扩展至连续时间控制；（ii）考虑更多样化的场景，包括多条车道、多辆汽车、环岛和交叉路口；（iii）利用实验测试平台，研究所提出方法的实际有效性[30]。

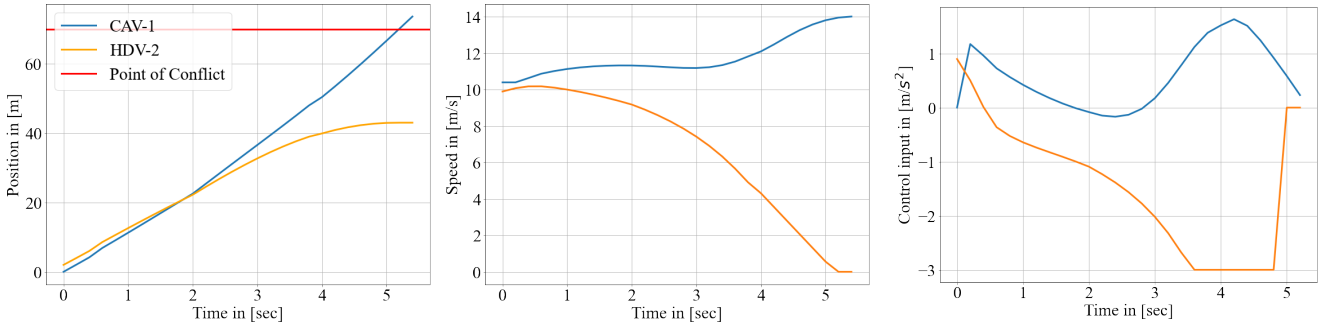
参考文献

[1] L. Zhao 和 A. A. Malikopoulos，“通过网联与自动化提升出行效率：共享自动驾驶车辆系统综述”，*IEEE*智能交通系统杂志，第14卷，第1期，第87–102页，2022年。[2] A. A. Malikopoulos、L. E. Beaver 和 I. V. Chremos，“网联自动驾驶车辆的最优时间轨迹与协调”，*Automatica*，第125卷，编号109469，2021年。[3] W. Xiao 和 C. G. Cassandras，“具有安全约束保证的网联自动驾驶车辆分散式最优汇入控制”，*Automatica*，第123卷，文章109333，2021年。[4] B. Chalaki 和 A. A. Malikopoulos，“网联自动驾驶车辆的障碍认证最优协调框架”，载于第*61*届*IEEE*决策与控制会议（*CDC*）论文集，第2264–2269页，2022年。[5] W. Gao、A. Odekunle、Y. Chen 和 Z. Jiang，“通过强化学习实现网联自动驾驶车辆的预测巡航控制”，*IET*控制理论与应用，第13卷，第2849–2855页，2019年11月。[6] K. Liu、N. Li、H. E. Tseng、I. Kolmanovsky、A. Girard 和 D. Filev，“面向协作的自动驾驶车辆汇入场景决策”，载于*2021*年第*60*届*IEEE*决策与控制会议（*CDC*），第5006–5012页，IEEE，2021年。[7] R. Chandra 和 D. Manocha，“Gameplan：在交叉路口、环岛和汇入场景中与人类驾驶员进行博弈论多智能体规划”，*IEEE*机器人与自动化快报，第7卷，第2期，第2676–2683页，2022年。[8] V.-A. Le 和 A. A. Malikopoulos，“利用社会价值取向实现混合交通中网联自动驾驶车辆的协作最优控制框架”，载于*2022*年第*61*届*IEEE*决策与控制会议（*CDC*），第6272–6277页，2022年。[9] V.-A. Le 和 A. A. Malikopoulos，“利用贝叶斯优化实现混合交通中网联自动驾驶车辆模型预测控制的最优权重自适应”，载于*2023*年美国控制会议（*ACC*）论文集，2023年（待发表）。[10] F. Altch´e 和 A. de La Fortelle，“用于高速公路轨迹预测的长短期记忆网络”，载于*2017*年第*20*届*IEEE*智能交通系统国际会议（*ITSC*），第353–359页，IEEE，2017年。



(a) (b) (c)

图5：轨迹 、速度和距离：激进型人类驾驶员参与仿真的车辆 r.



(a) (b) (c)

图6：轨迹 、速度和距离：保守型人类驾驶员参与仿真的车辆 r.

[11] S. H. Park、B. Kim、C. M. Kang、C. C. Chung 和 J. W. Choi，“通过LSTM编码器-解码器架构进行车辆轨迹的序列到序列预测”，载于*2018*年*IEEE*智能车辆研讨会（*IV*），第1672–1678页，IEEE，2018年。[12] N. Deo 和 M. M. Trivedi，“利用基于机动行为的LSTM对周边车辆进行多模态轨迹预测”，载于*2018*年*IEEE*智能车辆研讨会（*IV*），第1179–1184页，IEEE，2018年。[13] N. Deo 和 M. M. Trivedi，“用于车辆轨迹预测的卷积社会池化”，载于*IEEE*计算机视觉与模式识别会议研讨会论文集，第1468–1476页，2018年。[14] B. Kim、C. M. Kang、J. Kim、S. H. Lee、C. C. Chung 和 J. W. Choi，“通过循环神经网络在占据栅格地图上进行概率车辆轨迹预测”，载于*2017*年第*20*届*IEEE*智能交通系统国际会议（*ITSC*），第399–404页，IEEE，2017年。[15] Z. el abidine Kherroubi、S. Aknine 和 R. Bacha，“网联自动驾驶车辆在高速公路匝道汇入中的新型决策策略”，*IEEE*智能交通系统汇刊，第23卷，第8期，第12490–12502页，2021年。[16] J. Guo、S. Cheng 和 Y. Liu，“常规车辆与自动驾驶车辆的汇入与分流对混合交通的影响”，*IEEE*智能交通系统汇刊，第22卷，第3期，第1639–1649页，2020年。[17] A. A. Malikopoulos，“信息物理系统中的学习与控制分离”，*Automatica*，第151卷，编号110912，2023年。[18] J. Subramanian 和 A. Mahajan，“部分观测系统的近似信息状态”，载于*2019*年第*58*届*IEEE*决策与控制会议（*CDC*），第1629–1636页，IEEE，2019年。[19] J. Subramanian、A. Sinha、R. Seraj 和 A. Mahajan，“部分观测系统中用于近似规划和强化学习的近似信息状态”，机器学习研究杂志，第23卷，第12期，第1–83页，2022年。[20] L. Yang、K. Zhang、A. Amice、Y. Li 和 R. Tedrake，“部分可观测环境中的离散近似信息状态”，载于*2022*年美国控制会议（*ACC*），第1406–1413页，IEEE，2022年。

[21] M. Fatemi、T. W. Killian、J. Subramanian 和 M. Ghassemi，《医疗死胡同：学习识别高风险状态和治疗措施》，神经信息处理系统进展，第34卷，第4856–4870页，2021年。[22] A. Dave、N. Venkatesh 和 A. A. Malikopoulos，《用于不确定系统最坏情况控制的近似信息状态》，载于第*61*届*IEEE*决策与控制会议（*CDC*）论文集，第4945–4950页，2022年。[23] A. Dave、N. Venkatesh 和 A. A. Malikopoulos，《用于不确定系统最坏情况控制与学习的近似信息状态》，*arXiv:2301.05089*（审稿中），2023年。[24] M. Treiber 和 A. Kesting，《交通流动力学》，《交通流动力学：数据、模型与仿真》，*Springer-Verlag Berlin Heidelberg*，第983–1000页，2013年。[25] 美国交通部联邦公路管理局，《下一代仿真（NGSIM）车辆轨迹及辅助数据》，2016年。[26] M. Kuderer、S. Gulati 和 W. Burgard，《通过示范学习自动驾驶车辆的驾驶风格》，载于*2015*年*IEEE*国际机器人与自动化会议（*ICRA*），第2641–2646页，IEEE，2015年。[27] K. Cho、B. Van Merri¨enboer、D. Bahdanau 和 Y. Bengio，《神经机器翻译的特性：编码器—解码器方法》，*arXiv*预印本*arXiv:1409.1259*，2014年。[28] J. A. Andersson、J. Gillis、G. Horn、J. B. Rawlings 和 M. Diehl，《CasADi：用于非线性优化和最优控制的软件框架》，数学规划计算，第11卷，第1期，第1–36页，2019年。[29] A. W¨achter 和 L. T. Biegler，《大型非线性规划中内点滤波线搜索算法的实现》，数学规划，第106卷，第1期，第25–57页，2006年。[30] B. Chalaki、L. E. Beaver、A. I. Mahbub、H. Bang 和 A. A. Malikopoulos，《用于新兴交通系统实时控制的研究与教育机器人试验平台：从理论到缩比实验 [控制应用],》，*IEEE* 控制系统杂志，第42卷，第6期，第20–34页，2022年。