异质混合交通中的网联自动驾驶汽车协同驾驶：一种博弈论方法

方诗宇、彭航、*IEEE*会员，魏崇锋、*IEEE*会员，邢阳、*IEEE*会员，和孙健

arXiv:2305.03563v1 [cs.MA] 5 May 2023

摘要—**高密度无信号交叉口一直是效率和安全性的瓶颈。网联自动驾驶汽车（CAV）的出现带来了混合交通状况，进一步增加了交通系统的复杂性。在此背景下，本文旨在研究高密度混合无信号交叉口中车辆之间复杂且异质的交互以及冲突消解问题。本文综合了关于网联自动驾驶汽车与人工驾驶车辆（HV）之间交互以及网联自动驾驶汽车协作的理论洞见，并据此提出一种适用于异质混合交通的新型协同决策框架。本文将归一化合作博弈与Level-k博弈（NCL博弈）结合，以生成系统最优解。随后，格点规划器为网联自动驾驶汽车生成最优的无碰撞轨迹。为再现混合交通中的人工驾驶车辆，本文从自然驾驶数据中提取交互行为作为先验知识，并将非合作博弈与逆向强化学习（IRL）相结合，以模拟异质人工驾驶车辆的决策过程。最后，通过三个案例验证所提算法的性能，包括与不同方法的比较分析、不同渗透率（ROP）下的案例研究，以及与异质人工驾驶车辆的交互分析。结果表明，所提出的协同决策框架有助于消解驾驶冲突并提升混合无信号交叉口的交通效率。此外，由于考虑了驾驶异质性，本文能够实现更好的人人机交互与协作。**

索引词—**网联自动驾驶汽车、异质混合交通、无信号交叉口、Level-k博弈、逆强化学习、协同驾驶**

I. 引言

THE 过去十年见证了网联自动驾驶汽车的快速发展。网联自动驾驶汽车具有更广的感知范围和更小的感知误差，因此能够更早地感知危险，甚至能够通过合理安排通行权（ROW）主动与其他车辆协作以解决冲突。然而，仍然存在一些挑战。由于现实世界情况的复杂性，网联自动驾驶汽车在与人类驾驶车辆互动时可能表现得不合理，从而导致碰撞或死锁。如何通过协调车辆之间的关系来提升系统性能，也是一个进一步需要探讨的问题。更不用说，包含网联自动驾驶汽车和异质人工驾驶车辆的混合交通中的网联自动驾驶汽车控制仍然是一个持续受到关注的问题。

几十年来，改善上述问题的一种思路是从人类交互中探寻潜在规则和模式。许多研究分析了驾驶员的决策逻辑、认知结构和内在特征，以设计出具有人类行为特征的网联自动驾驶汽车。博弈论描述了激励结构之间的交互，以及参与者策略之间的关系，已广泛用于复现社会决策 [1, 2]。除具有可解释性这一优势外，博弈论也适用于环形交叉口 [3]，、匝道汇入区域 [4, 5],和无信号交叉口[6, 7]等不同场景。博弈论方法不再将周围车辆视为移动障碍物，而是通过与周围车辆动态交互来模拟人类行为，从而实现自动变道 [8]。Level-k博弈分别与车辆控制器 [9]和蒙特卡洛树搜索（MTCS） [10] 相结合，以对网联自动驾驶汽车具有时间延展性、多步性和交互性的决策过程进行建模。同时，通过人类认知结构识别实时意图，对于开发以人为中心的智能车辆发挥着重要作用。因此，Chu等人 [11] 基于刺激-有机体-反应（SOR）认知理论，探索了分心驾驶意图背后的机制。此外，人工神经网络（ANN）通过模仿动物神经网络中的信息传递机制进行设计，并被用于预测匝道汇入 [12], 、变道 [13], 和左转 [14]时驾驶员的意图。Bi等人 [15] 提出了一种排队网络认知架构，用于推断有无分心任务时的意图；在驾驶模拟器上开展的实验表明，该架构在推断典型及快速变道行为意图方面表现良好。与多样且易变的意图一样，驾驶员自身也各具不同特征。Huang等人 [16] 采用逆向强化学习（IRL），对不同的驾驶行为进行建模

本研究部分获得国家自然科学基金（52232015 和 52125208）、中央高校基本科研业务费（编号：2022-5-ZD-02）以及之江实验室（2021NL0AB02）的资助。S. Fang、P. Hang 和 J. Sun 隶属同济大学交通工程系及教育部道路与交通工程重点实验室，地址：中国上海市 201804。（电子邮箱：{2111219, hangpeng, sunjian}@tongji.edu.cn）C. Wei 隶属贝尔法斯特女王大学机械与航空航天工程学院，地址：英国贝尔法斯特 BT7 1NN。（电子邮箱：c.wei@qub.ac.uk）Y. Xing 隶属克兰菲尔德大学航空航天、运输与制造学院，地址：英国 MK43 0AL。（电子邮箱：yang.x@cranfield.ac.uk）通讯作者：P. Hang。本研究已提交 IEEE，供其考虑发表。版权可能在未另行通知的情况下转让，届时此版本可能不再可供查阅。

其依据是自然驾驶数据。Schwarting 等人 [1] 研究了社会价值取向（SVO），该概念量化了智能体的自私或利他程度，并据此将智能体之间的交互建模为最佳响应博弈。尽管已有许多研究表明，网联自动驾驶汽车与人工驾驶车辆之间的交互可以通过最佳响应博弈来解决，但理性的集体决策可能与社会最优决策背道而驰 [17]。因此，解决这一现象的一种方法是建立制度安排，通过协同驾驶优化系统性能。

网联自动驾驶汽车的协同驾驶已得到广泛研究。协同是指车辆通过决策过程协调彼此的行动，以实现单辆车无法独立实现的目标[18]。此外，根据网联自动驾驶汽车的自主程度，其协同方法可分为集中式、协商式、一致式和涌现式。在集中式控制中，协调器将充当交叉口管理器，为每辆驶近交叉口的车辆预留特定的时空资源 [19]，或为至少一项全局任务决定所有车辆的策略 [20]。然而，协调器的能力可能成为瓶颈。Zhou 等人 [21] 提出了一种具有情境感知能力的推理图（RG），并将其与一些机动兼容性和社会交互惯例规则相结合，从而快速搜索符合推理情境的速度剖面。Xu 等人 [22] 将蒙特卡洛树搜索与启发式规则相结合，为网联自动驾驶汽车寻找近似全局最优的通行顺序；结果表明，该方法能够在系统性能和计算灵活性之间保持良好的权衡。在集中式控制下，网联自动驾驶汽车只能被动遵循协调器的指令，因此其潜力远未得到充分发挥。协商式和一致式方法分别采用固定协议和动态协议，使网联自动驾驶汽车能够与其他车辆通信。Carlino 等人 [23] 提出了一种基于去中心化拍卖的自主交叉口管理方案，根据车辆的时间价值来决定其通行。Vu 等人[24] 进一步将拍卖机制移植到元胞自动机模型中。然而，由于人的行为无法控制，这些协同方法不适用于人工驾驶车辆。同时，在通信过程中，网联自动驾驶汽车遵循精心设计的协议还会进一步引发鲁棒性问题。涌现式方法不直接与其他车辆通信，也不遵循任何预定义的协同协议，而是由每辆车通过评估其他车辆行为的合理性，并基于当前状态推断其行动来做出决策。Wang 等人 [7] 通过博弈论为具有不同优先级的智能体建立了协同模型。此外，先前的研究还将目标—新颖性进化搜索相结合，以在仅网联自动驾驶汽车道路上合成网联自动驾驶汽车的协同行为 [25]。实验表明，多辆车交互时会涌现出理想的协同驾驶行为，但其结果无法推广到新的道路。总之，上述研究聚焦于仅网联自动驾驶汽车环境下的协同。然而，根据 [26],, 未来至少40年内，现实世界的交通流中都将混有人工驾驶车辆，因此这些方法在混合交通环境下的有效性仍有待验证。

无论是针对网联自动驾驶汽车与人工驾驶车辆之间的交互还是仅网联自动驾驶汽车环境中的协作方法，均只能造就尚不成熟的网联自动驾驶汽车，因此还应进一步评估其在复杂交互场景下的性能。迄今为止，关于网联自动驾驶汽车在混合交通中的验证与确认，相关讨论仍然很少。Wang 等人 [27]通过加入背景交通流展示了其控制器，但在其双车道合流场景中，最多仅涉及四辆车辆。此外，在驾驶模拟器上进行人在回路测试也是复现混合交通的一种实用方法。Sadigh 等人 [28] 率先探究了网联自动驾驶汽车主动获取周边人工驾驶车辆信息的能力，并在驾驶模拟器上开展了实验。美中不足的是，实验仅考虑了一对一交互。此外，由于设备限制，在驾驶模拟器上实现多重人机交互似乎并不切实际。因此，现有关于混合交通的研究主要针对低密度场景，这是一种极大的简化。然而，现实世界中的异质混合交通危险得多，因为驾驶员偏好会促使他们做出不同的决策。

为应对上述挑战，本文研究了异质混合交通条件下网联自动驾驶汽车的协同驾驶。本文的贡献如下：

1. 本文提出了一种新型协同驾驶框架，使网联自动驾驶汽车能够在高密度混合交通无信号交叉口中实现协同，并与异质人类驾驶员进行协调，从而有助于解决驾驶冲突、增强人机交互并提高交通效率。

2. 网联自动驾驶汽车的协同决策通过串联归一化合作博弈和 Level-k博弈（NCL 博弈）实现，并输出对应于系统整体效率最大值的 k 分配。随后结合格点规划器，生成最优且无碰撞的网联自动驾驶汽车轨迹。验证了不同渗透率（ROP）下网联自动驾驶汽车的协作性能。

3. 为模拟真实混合交通中的人机交互，对异质人工驾驶车辆决策进行建模。通过聚类获得驾驶员分类及其组成，并将其作为先验知识。采用非合作博弈来再现交互环境下驾驶员的理性决策，并利用逆向强化学习校准不同驾驶员的决策偏好。最后，将异质人工驾驶车辆引入仿真实验，以检验协同驾驶框架的有效性和鲁棒性。

本文其余部分组织如下。第二节介绍我们的研究所要解决的主要问题，以及相应的协同驾驶框架。第三节描述异质人工驾驶车辆决策的复现。第四节讨论网联自动驾驶汽车协同决策和轨迹规划的过程。第五节通过多项仿真实验验证我们的框架。最后，第六节给出结论。

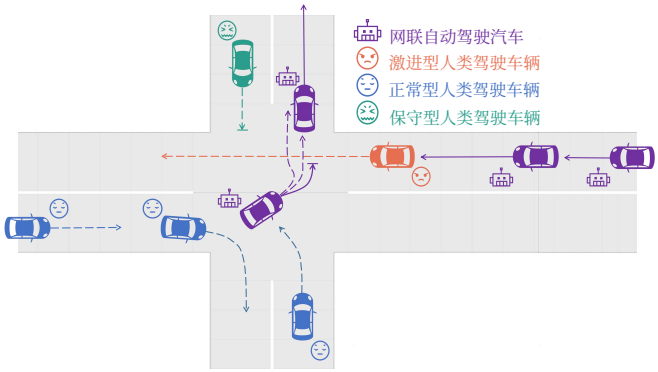

图1. 异质人工驾驶车辆参与的高密度无信号交叉口示意图。

II. 问题陈述与协同驾驶框架

网联自动驾驶汽车由于其复杂的机动行为和不完善的控制方法，仍然饱受诟病。当网联自动驾驶汽车在复杂且具有交互性的场景中与其他车辆交互时，可能会发生死锁和碰撞。在所有城市交通场景中，无信号交叉口由于具有众多冲突点以及多变的交互对象，最具挑战性。此外，人类驾驶员具有异质性，因此在相同情况下，可能会基于自身的决策偏好采取不同的机动行为。因此，网联自动驾驶汽车总是倾向于采取被动的防御性行动，例如过度减速，这会导致效率低下，甚至更糟的是，导致人们对网联自动驾驶汽车失去信任。

然而，很少有研究关注未来将无处不在的多重人机交互。因此，面向网联自动驾驶汽车的算法的有效性和鲁棒性仍不明确。我们认为，研究复杂场景下网联自动驾驶汽车与人工驾驶车辆之间的交互，对于验证和应用我们的网联自动驾驶汽车协同驾驶算法具有重要意义。因此，本文的主要问题是，使网联自动驾驶汽车能够在高密度无信号交叉口中恰当地响应异质人工驾驶车辆，或主动与其协调。图1展示了该交叉口的缩略图。

为实现上述目标，本文提出了一个详细的框架，如图 2 所示。首先，收集并分析自然驾驶数据，以调查真实交叉路口中驾驶员决策类型及其分布。此外，需要指出的是，对异质驾驶员进行建模是一项庞大的工程。鉴于本文主要关注为网联自动驾驶汽车建立协同驾驶框架，我们选择通过复现异质决策来简化问题。随后，通过非合作博弈生成一系列可行轨迹，并将其与自然驾驶数据中的专家示范进行比较。通过最大熵逆强化学习，对不同的决策偏好 $\boldsymbol{\theta}^T$ 进行校准。接着，利用纳什均衡解复现异质人工驾驶车辆的决策，并通过车辆动力学生成相应的下一状态。

此外，对于网联自动驾驶汽车的协同决策和轨迹规划，本文采用了大多数自主机器人控制所使用的三层层级结构，以生成安全且连续的状态 [29, 30]。在层级顶端，引入Level-k博弈来模拟人类推理深度，并将其作为下一层的自变量。然后，基于合作博弈，通过枚举可实现导向系统最优的 k 分配。在底层，格点规划器将生成满足车辆动力学约束的最优无碰撞轨迹。

仿真中的混合交通将通过车辆之间动态地相互竞争与协调，每 Δt =0.1s更新一次。最后，我们通过实验验证了所提出的框架，包括与不同方法的对比分析、不同渗透率（ROP）下的案例研究，以及与异质人工驾驶车辆的交互分析。

III. 混合交通中异质人工驾驶车辆决策的再现

尽管没有交通信号灯会带来不确定性，但人工驾驶车辆不可预测且无法控制的决策潜在意图，可能会给网联自动驾驶汽车带来更大的困境。为验证网联自动驾驶汽车在异质混合交通中的协同驾驶算法，本节提取了真实世界中经验丰富的驾驶员演示数据，以再现异质人工驾驶车辆决策。

主要流程包括四个步骤。首先，收集真实世界交叉口的交互数据。其次，通过K均值聚类将驾驶员的决策分为三组。然后，选用非合作博弈来指导人工驾驶车辆决策的生成；非合作博弈被认为是一种有望再现人类决策过程的模型。最后，通过逆向强化学习校准各组的决策偏好。

A. 数据收集

如前所述，研究人工驾驶车辆的交互，对于构建人工驾驶车辆模型或类人网联自动驾驶汽车都大有裨益。本文依据Ni等人以往的研究 [31]，在中国上海的仙霞路—剑河路（XXJH）交叉口采集数据。随后，提取侵入后时间（PET）小于3秒的交互事件。

筛选后，131 个交互事件中保留了 79 个。同时，通过半自动化视频图像处理工具提取人工驾驶车辆的轨迹，并将其作为先验知识。此外，还计算了其速度和加速度，用于聚类。

B. 基于 *K*均值聚类的驾驶员决策分类

众所周知，驾驶员的决策可分为激进型、普通型和保守型等几种类型。为了再现具有不同决策偏好的驾驶员，首要任务是确定 XXJH 交叉口的驾驶员可以划分为多少个决策簇。

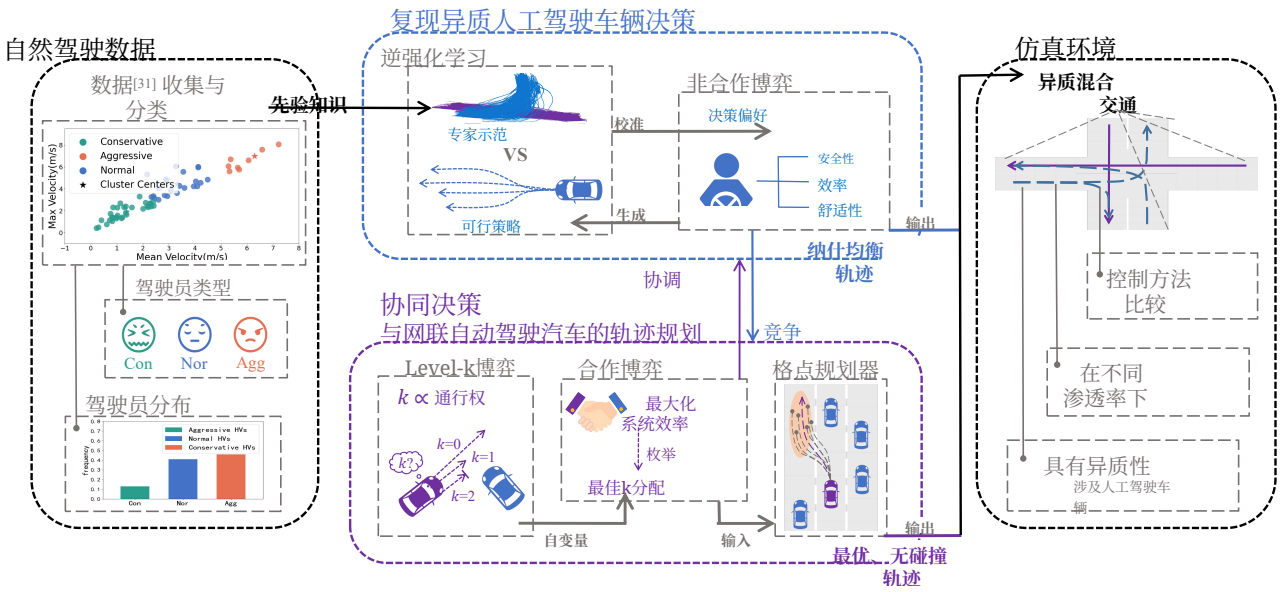

图 2. 混合交通中的协同驾驶框架。

基于肘部法的结果——这是一种使用误差平方和（SSE）确定簇数量的简单而通用的理论——三个簇对我们的数据效果最佳。考虑到速度和加速度是驾驶过程中决策最直观的体现，根据 Chen 等人 [32] 的研究，车辆的速度和加速度的均值、最大值、最小值和标准差被选作聚类参数。随后，使用 K均值聚类对 XXJH 中的自然驾驶数据样本进行分类。

表 I XXJH 数据的聚类中心

参数	$V(m/s)$			
	mean	max	max	min
簇-1	3.31	4.42	2.36	0.66
簇-2	1.34	1.60	0.99	0.27
簇-3	6.29	6.98	5.40	0.50
参数	$A\ (m/s^2)$			
	mean	max	max	min
簇-1	-0.07	0.49	-0.72	0.37
簇-2	0.08	0.55	-0.34	0.28
簇-3	-0.77	-0.30	-1.14	0.28

聚类算法收敛后，可以得到每个簇的中心，如表 I 所示。由于这并非本研究的主要关注点，我们仅根据 Huang等人 [33] 的研究，假设驾驶过程中的速度是影响决策激进程度的最重要变量。因此，79 个交互事件中驾驶员的决策被分为三个簇：普通型、保守型和激进。此外，这些轨迹被用作专家示范，并与人工驾驶车辆的可行轨迹进行比较，以校准不同的决策偏好。本文选择非合作博弈来估计人类驾驶员面对其他驾驶员时可能采取的决策，因为在交叉口执行冲突机动时，它能够实现更符合实际的人类行为 [34, 35]。结合车辆动力学，可以得出每一帧的可行轨迹。

C. 通过非合作博弈生成人工驾驶车辆的可行轨迹

如前所述，博弈论将人类描述为理性的决策者，他们会将行动之间的依赖关系考虑在内。因此，博弈论在建模人类决策领域的应用日益广泛。本文将人类驾驶员视为理性决策者，并认为他们意识到自身行为会对在时间和空间上与其存在冲突的其他驾驶员产生影响。因此，我们建立了非合作博弈，以模拟驾驶员在共享空间中驾驶时如何推测彼此的行为并相互竞争。

我们采用了采样间隔 $\Delta t = 0.1s$，并设置了一个包含六种策略的动作集，这些策略基于 Tian 等人提出的方法，代表了城市交通中的常见驾驶操作， [3],见表 II。

表 II 动作集
U

动作 u	$A(m/s^2)$	$\omega(rad/s)$
保持 u_1	0	0
加速 u_2	2	0
减速 u_3	-2	0
制动 u_4	-4	0
左转 u_5	0	$\pi/4$
右转 u_6	0	$-\pi/4$

其中， U 为动作集， $u_t^i = [a_t^i, \omega_t^i]^T$ 为车辆 i 在第 t 帧的动作， a_t^i 为加速度， ω_t^i 为航向角速度。因此，可以生成六条可行轨迹

在每一帧中，将动作 u_t^i 与式1中的车辆动力学相结合即可生成这些轨迹。

$$f(s_t^i, u_t^i) = \begin{bmatrix} x_t^i + v_t^i \cos(\gamma_t^i)\Delta t \\ y_t^i + v_t^i \sin(\gamma_t^i)\Delta t \\ v_t^i + a_t^i \Delta t \\ \gamma_t^i + \omega_t^i \Delta t \end{bmatrix} \tag{1}$$

其中，f表示单轮模型，$s_t^i = [x_t^i, y_t^i, v_t^i, \gamma_t^i]^T$表示车辆的状态（$x_t^i, y_t^i$），$v_t^i$ 和 γ_t^i 分别表示车辆 i 在 t 帧的位置、速度和偏航角。

此外，在非合作博弈中，每个 $\Delta t = 0.1s$ 都需要在每个动作的后果与对手可能采取的相应反应之间进行谨慎权衡。最后，在考虑表 II 中的每个动作后，可以得出应对对手每个动作的最佳响应。当每个参与者都寻求能够带来最低成本或最高奖励的最佳响应时，纳什均衡将得以实现。在纳什均衡状态下，任何参与者都无法通过单方面调整其动作来获得更低的成本。因此，纳什均衡的数学定义为

$$R(u_t^{i^*}, u_t^{-i^*}) \geq R(u_t^i, u_t^{-i^*}) \tag{2}$$

其中，R 是奖励函数，$u_t^{i^*}$ 和 $u_t^{-i^*}$ 分别是在游戏于 t 帧达到纳什均衡时由玩家 i 和除玩家 i 以外的其他玩家执行的动作，u_t^i 是玩家 i 在动作集 U 中的任意动作。

在某些情况下，纳什均衡并不总是最优解 [17]。不过，我们不认为这是一个缺点，而是将其视为人类推理能力局限的替代指标，以及绝对理性的结果。正如现实世界中，并非所有游戏都有圆满结局。

这里，我们考虑由效率、舒适性和安全性构成的奖励，分别以到目的地的距离、相对于预期路径的偏移量以及碰撞时间（TTC）表示。然后，可以通过式 3 中的 s_t^i、u_t^i 和 u_t^{-i} 计算每个可能动作的奖励。

$$R(s_t^i, u_t^i, u_t^{-i}) = [-d_t^i - o_t^i + \frac{1}{TTC_t^i}]\boldsymbol{\theta}^T \tag{3}$$

其中，d_t^i 是车辆 i 在第 t 帧的当前位置与其目的地之间的距离，o_t^i 是相对于预期路径的偏移量，TTC_t^i 是碰撞时间，$\boldsymbol{\theta}^T$ 是驾驶员决策偏好的权重。

因此，结合式 1 和表 II，可以在每一帧生成六条可行轨迹。随后引入最大熵逆向强化学习，以量化可行轨迹与专家示范之间的差异，从而校准人工驾驶车辆的不同决策偏好。

D. 基于逆向强化学习的决策偏好校准

将驾驶员划分为若干组，并通过非合作博弈生成可行轨迹后，引入逆向强化学习（IRL），以挖掘影响专家决策的内在特征。逆向强化学习（IRL）提出的时间晚于行为克隆。尽管两者有许多相似之处，但与单纯模仿专家操作不同，IRL 试图推断专家做出决策的原因，并据此优化策略。换句话说，IRL 不仅直接学习状态—动作映射，还会推断奖励权重的形式，并通过奖励权重优化操作。

鉴于上述优势，逆向强化学习已广泛应用于许多领域。在众多逆向强化学习算法中，最大熵方法因能够解决歧义问题而脱颖而出。人类驾驶员遵循随机策略，这可能导致候选决策呈现不同的分布 [16]。引入最大熵原理，是为了选择一种不体现任何额外偏好的分布。因此，歧义得以消除。通常，轨迹的概率与公式 4 中奖励的指数成正比。

$$P(\zeta|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{R(\zeta)} = \frac{1}{Z(\boldsymbol{\theta})} e^{\boldsymbol{\theta}^T f_\zeta} \tag{4}$$

其中，$P(\zeta|\boldsymbol{\theta})$ 是轨迹的概率，ζ，$R(\zeta)$ 是专家轨迹 ζ 的对应奖励函数，它由奖励权重向量 $\boldsymbol{\theta}^T$ 与轨迹特征向量 f_ζ 相乘得到，而 $Z(\boldsymbol{\theta})$ 是配分函数。

面对连续或高维空间问题时，配分函数可能无法收敛。然而，在有限时域问题中，最大熵方法中的奖励权重必然收敛 [36]。因此，生成了有限的可行轨迹，公式 4 可改写如下。

$$P(\zeta|\boldsymbol{\theta}) = \frac{e^{\boldsymbol{\theta}^T f_\zeta}}{\sum_{i=1}^{M} e^{\boldsymbol{\theta}^T f_{\tilde{\zeta}}}} \tag{5}$$

其中，$f_{\tilde{\zeta}}$ 是可行轨迹 ζ的特征，M 是有限的可行轨迹。通过这种近似，概率更容易计算。

逆向强化学习的核心是调整奖励函数的权重，以生成与专家示范相匹配的机动行为。为实现这一目的，应通过调整 $\boldsymbol{\theta}^T$来最大化专家示范的似然。

$$\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \sum_{\zeta \in D} logP(\zeta|\boldsymbol{\theta}) \tag{6}$$

其中，$L(\boldsymbol{\theta})$是似然函数，同时也是目标函数；$D = \{\zeta_i\}_{i=1}^N$ 是轨迹集，包含 N 个专家示范，这些示范是我们之前筛选并聚类的真实世界的人类轨迹。

将式 6 中的 $P(\zeta|\boldsymbol{\theta})$替换为式 5 后，$L(\boldsymbol{\theta})$可以重写为

$$L(\boldsymbol{\theta}) = \sum_{\zeta \in D} [\boldsymbol{\theta}^T f_\zeta + log \sum_{i=1}^{M} e^{\boldsymbol{\theta}^T f_{\tilde{\zeta}}}] \tag{7}$$

该函数是凸函数，并采用基于梯度的优化来求解最优奖励权重。梯度可以表示为式 8 中专家示范 f_ζ 与可行轨迹 $f_{\tilde{\zeta}}$ 之间的差值。

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \sum_{\zeta \in D} [f(\zeta) - \sum_{i=1}^{M} \frac{e^{\boldsymbol{\theta}^T f_\zeta}}{\sum_{i=1}^{M} e^{\boldsymbol{\theta}^T f_{\tilde{\zeta}}}}] \tag{8}$$

在似然函数中引入 L2 正则化，以防止过拟合，其中 $\lambda > 0$ 为正则化参数。因此，式 7 和式 8 可改写为

$$L(\boldsymbol{\theta}) = \sum_{\zeta \in D} [\boldsymbol{\theta}^T f_\zeta + log \sum_{i=1}^{M} e^{\boldsymbol{\theta}^T f_{\tilde{\zeta}}}] - \lambda \boldsymbol{\theta}^2 \tag{9}$$

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \sum_{\zeta \in D} [f(\zeta) - \sum_{i=1}^{M} \frac{e^{\boldsymbol{\theta}^T f_\zeta}}{\sum_{i=1}^{M} e^{\boldsymbol{\theta}^T f_{\tilde{\zeta}}}}] - 2\lambda \boldsymbol{\theta} \tag{10}$$

接下来，算法1总结了最大熵逆向强化学习的伪代码。基于聚类结果，分别使用逆向强化学习对三组专家示范进行训练，以校准不同驾驶员的决策偏好。通过非合作博弈和公式3，分别可以获得可行轨迹及其对应的特征。

算法1： 最大熵逆向强化学习。

数据： 三组专家示范数据集，学习率 α，正则化参数 λ，训练轮次 E

结果： 优化后的奖励权重 $\boldsymbol{\theta}^T$，包括效率、舒适性和安全性1 初始化 $\boldsymbol{\theta}^T \leftarrow \mathcal{N}$(0，0。05)；

2 初始化缓冲区 $B \leftarrow []$；

3 计算人类特征 $f \leftarrow \sum_{i=1}^{M} f_\zeta$；

4 **当** $\zeta_i \in D$ **时执行**

5 生成可行轨迹 $f_{\tilde{\zeta}}$；

6 计算以下对象的特征 $\tilde{\zeta}$；

7 将特征添加到缓冲区 $B \xleftarrow{+} \tilde{\zeta}$；

8 $i \leftarrow i + 1$；

9 **结束**

10 **当** $epoch < E$ **时**

11 计算每个样本的似然函数，其中缓冲区 B；

12 $L(\boldsymbol{\theta}) = \sum_{\zeta \in D} [\boldsymbol{\theta}^T f_\zeta + log \sum_{i=1}^{M} e^{\boldsymbol{\theta}^T f_{\tilde{\zeta}}}]$；

13 计算梯度 $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$；

14 更新奖励权重 $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$】【。

15 $\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}$；

16 **结束**

训练完成后，可通过最大化熵并进行迭代，校准不同驾驶员群体的奖励权重，如表 III 所示。

如表 III 所示，激进型驾驶员群体的效率值最高，而安全性和舒适性值最低，表明他们偏好以高速超车，并且更愿意绕行，而不是停车等待。与此同时，保守型驾驶员更加关注

表 III 逆向强化学习校准结果

Type	奖励权重		
	效率	舒适性	安全性
激进型人工驾驶车辆	8.33	1.56	3.69
普通人工驾驶车辆	8.2	1.72	5.7
保守型人工驾驶车辆	7.79	2.1	8.44

驾驶舒适性和安全性，因此会做出保守的决策。普通驾驶员的奖励权重处于中间范围，这些驾驶员没有过度偏好，而是在驾驶过程中努力平衡效率、舒适性和安全性。

总之，上述结果符合现实预期。另一方面，这些结果也证实了我们在 A 小节中对聚类中心的定义是合理的。在对不同驾驶员的决策偏好进行校准后，可以推断出与纳什均衡相对应的动作，并将该动作代入公式 1，即可计算出车辆的下一状态。

到目前为止，我们已对异质人工驾驶车辆进行了建模，以构建混合环境。接下来，我们设计了一种协同驾驶框架，使网联自动驾驶汽车不仅能够与其他车辆互动，还能主动与其协调，从而提升交通效率和安全性。

IV. 网联自动驾驶汽车的协同决策与轨迹规划

与人工驾驶车辆的轨迹不可预测且无法控制不同，网联自动驾驶汽车能够与其他车辆协作，甚至协调行动，从而提升系统效率。如图 2 所示，网联自动驾驶汽车的协同驾驶算法分为三个小节介绍。A 小节和 B 小节分别介绍 Level-k博弈和合作博弈，以寻找使系统效率最高的最优 k 分配方案。在 C 小节中，使用格点规划器根据 k 分配结果规划最优无碰撞轨迹。

A. 运用 *Level-k*博弈解决多股交通流冲突

Level-k博弈假设参与者根据自身推理层级生成策略。此外，k 也可以定义为通行权 [7],驾驶风格 [3, 6]。这里，考虑到没有交通信号灯会导致多股交通流在时空上相交，我们将 k 定义为通行权以解决冲突。

在轨迹规划中，零级驾驶员将其他车辆视为静态障碍物，并基于这一基本假设规划自己的轨迹。接着，一级驾驶员会推断零级驾驶员的规划轨迹，并在避免与零级轨迹发生碰撞的同时规划自己的轨迹。以此类推，所有驾驶员都可以根据各自的推理深度获得预期路径。因此，车辆 i 的预期路径记为 ζ_i，可由下式计算：

$$\zeta_i(k) = \max R(k, s_u^i, \zeta_j(k-1)) \tag{11}$$

其中，$\zeta_i(k)$ 是 i 车辆在 k 级别下的轨迹，轨迹由最大化奖励函数R确定。

在本文中，我们的网联自动驾驶汽车可以自由生成自身的推理深度，以最大化系统的整体效率。此外，当多股交通流相交时，将为每股交通流分配一个唯一的 k 值，以防止发生碰撞或死锁。研究人员还指出，根据[6, 37]，人类的推理深度通常小于或等于 2。因此，为防止网联自动驾驶汽车出现不合理的行为，基于人类推理深度的有限性，将 k 大于 2 的交通流修正为 k 等于 2。

Level-k博弈为车辆有序通行提供了理论基础。尽管无信号交叉口无法从时间维度解决冲突，但可计数的车流意味着 k 分配方案的数量有限。通过优化各车流的 k 分配，可以实现系统最优并解决冲突。因此，本文采用合作博弈来量化每种 k 分配方案的系统性能。

B. 通过合作博弈搜索最佳*k*分配

先到先服务（FCFS）和批处理策略等基于预约的控制方法历史悠久，且因形式简洁而得到广泛应用 [19]。FCFS 控制简单地假设通行权与车辆进入顺序成正比，并最初由 Dresner和Stone 将其作为自动驾驶车辆的预约机制 [38]。批处理策略认为，从通行能力的角度来看，交叉口采用批次通行会更有效。因此，各车流中的车辆会被划分为若干批次，并按先到先服务原则依次通行。然而，这些方法没有考虑系统效率，且仅依据车辆进入顺序分配通行权。

因此，将合作博弈与上述 Level-k博弈相结合，以优化系统效率。与将其他参与者视为对手并强调最大化个人收益的非合作博弈不同，合作博弈是指参与者组成若干联盟，并通过与其他联盟合作来实现系统目标的最大化。同时，合作博弈应满足超可加性、个体理性和群体理性的要求 [39, 40]。

超可加性是指联盟的收益与各个独立车辆的总收益之间满足的不等式。它可以表示为

$$R(V_i) + R(V_j) \leq R(V_i \cup V_j) \tag{12}$$

其中， $R(V_i)$ 表示车辆 i 的奖励函数， $V_i \cup V_j$ 表示由车辆 i 和车辆 j 组成的联盟。此外，个体理性应满足以下不等式。

$$R(V_i) \leq R_i(V_i \cup V_j ... \cup V_n) \tag{13}$$

其中， $R_i(V_i \cup V_j ... \cup V_n)$表示在 n 辆车辆联合组成联盟进行合作博弈时，车辆 i 的个体收益。该不等式规定，每辆车辆在采用合作博弈进行控制时，都能获得比之前更高的收益。最后，群体理性意味着系统目标是个体收益之和，即，

$$O = \sum_{i \in N} R(V_i) \tag{14}$$

其中， O 既是合作博弈的系统目标，也是个体奖励之和。系统目标通常表现为最小化总体延误或最大化效率，例如行进得更快或更远。

通过实验，我们最终选择目的地距离倒数来表示系统效率，因为它的表现优于其他替代变量。因此，目标函数 O 如下。

$$O(k) = \max \sum_{i \in lane} \sum_{j \in group} \frac{1}{\bar{d}_{i,j}} \tag{15}$$

其中，O(k) 表示在某一 k 分配方案下目标函数的值，$d_{i,j}$ 表示属于 i 车道和 j 组中车辆的预期路径到目的地的平均距离。每一种 k 分配方案都会产生不同的$d_{i,\ j}$。

需要注意的是，此距离并非到目的地的当前距离，而是 k 级规划轨迹到目的地的平均距离。将特定的 k 提供给格点规划器，即可获得 $d_{i,j}$ 的值，下一小节将对此进行讨论。

与此同时，我们意识到直行车辆和转弯车辆对系统的影响不同。因此，目的地距离的系数应有所不同，公式 15 可以改写为

$$\max(\sum_{j \in group} \frac{1}{\bar{d}_{gs,j}} + \sum_{j' \in group'} \frac{\alpha}{\bar{d}_{turn,j'}}) \tag{16}$$

其中， α 表示转弯车辆相对于直行车辆的权重， d_{gs} 和 d_{turn} 分别表示直行车辆和转弯车辆到目的地的距离。随后，研究旨在统一 d_{gs} 和 d_{turn}的量纲。

为研究直行车辆与转弯车辆之间的现实世界关系，基于 XXJH 数据进一步提取了直行车辆与转弯车辆之间的交互。接下来，引入逻辑回归模型，探究初始目的地距离与实际通行顺序之间的关系。利用交互数据进行校准后，结果如下。

$$ln\frac{p}{1-p} = 0.0925 d_{turn} - 0.1332 d_{gs} + 2.35 \tag{17}$$

其中， p 表示直行车辆优先通行的概率。 p 与 d_{turn}、d_{gs} 之间的关系如图 3 所示。

将 $p = 0.5$ 代入式 17，得

$$d_{turn} = 1.44 d_{gs} - 25.4 \tag{18}$$

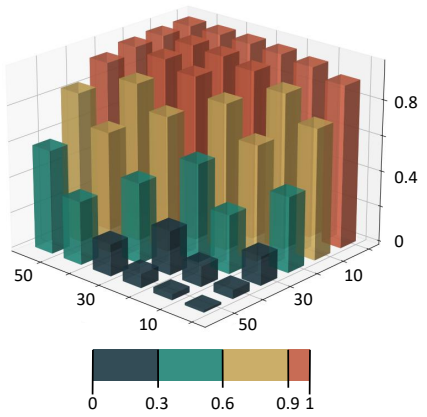

图 3. 目的地距离与直行车辆优先通行概率之间的逻辑关系。

从 d_{turn} 和 d_{gs}的系数可以看出，转弯车辆需要付出更高代价，才能与直行车辆保持平等。这一结论与相关法律规定一致：在交叉口发生冲突时，转弯车辆应让直行车辆先行。此外，将式 16 代入式 18，可得目标函数如下。

$$\max\Big(\sum_{j\in group}\frac{1}{\bar{d}_{gs,j}}+\sum_{j'\in group'}\frac{1}{1.44\bar{d}_{gs,j'}-25.4}\Big) \tag{19}$$

通过逻辑回归模型，对直行车辆和转弯车辆对系统的影响进行了归一化。在归一化过程中，也微妙地考虑了交叉口处直行车辆优先级高于转弯车辆这一社会规范。在此，我们将其称为归一化合作博弈。

如上所述，在合作博弈中，车辆将联合形成若干联盟，而最大化系统整体效率是所有联盟共同追求的目标。这一过程可以通过调整自身策略来影响其他联盟而实现。不同于上述非合作博弈中将影响视为离散的动作-动作映射，合作博弈需要采用轨迹-轨迹策略映射，这是由目标函数的形式决定的。为了计算式（19），应根据 k 层级博弈的概念和格点规划器，推断每辆车的规划轨迹。

因此，通过将归一化合作博弈与 Level-k 博弈相结合，可以通过枚举有限个 k 分配方案，轻松获得使系统效率达到最优的 k 分配方案。此外，还需要一个轨迹规划器，根据 Level-k 博弈的概念生成轨迹，以评估每个 k 分配方案的表现。因此，我们引入了格点规划器。

C. 使用格点规划器规划最优轨迹

格点规划器是一种广受认可的运动规划器，可根据预期生成最优且无碰撞的轨迹路径和障碍物。它主要由坐标变换、采样和曲率多项式拟合组成。

在坐标转换步骤中，根据笛卡尔坐标系中的初始状态，将相应状态转换为 Frenet 坐标系中的状态。随后，对规划时域和最大转向角进行采样。为提高计算速度，初始采样空间相对较小。如果没有可用解，则扩大采样空间。若仍然没有可用解，则停止仿真并输出无解信息。如果存在多个解，则根据奖励确定最优轨迹。最后，采样完成后，使用曲率多项式生成起点与终点之间的完整轨迹。

为控制变量，网联自动驾驶汽车的奖励构成和奖励权重设计为与普通人工驾驶车辆保持一致。区别在于，人工驾驶车辆通过在非合作博弈中搜索纳什均衡来做出决策，随后根据车辆动力学生成轨迹；而网联自动驾驶汽车则分别用我们提出的 NCL 博弈和格点规划器替代这些步骤。算法 2 描述了网联自动驾驶汽车的协同决策和轨迹规划流程。

算法 2： 网联自动驾驶汽车的协同决策与轨迹规划。

输入： 车辆状态 s_t^i，交通流数量 $lane$，每条交通流中的车辆数量 $group$
输出： 车辆决策深度 k 和下一状态 s_{t+1}^i（依据规划轨迹）

计算所有可能的 k 分配方案 $A \leftarrow A_{lane}^2$;
初始化合作博弈 O 的目标函数列表;
对 $a \in A$ **执行遍历**
  评估每个 a 的性能;
  对每个 $stream \in lane$ **执行**
    对 $i \in group$ **中的每个项执行**
      计算到目的地的平均距离 d_j 通过 $Lattice(k)$;
      归一化 $\bar{d}_{gs,j} \leftarrow \bar{d}_j$;
      计算此解的目标函数 $O_s += \bar{d}_{gs,j}$;
    end
  end
  添加到 $O \xleftarrow{+} O_s$
结束
求解最佳 k 分配方案最大值 O;
获取每辆车辆 k，并规划其下一个状态 s_{t+1}^i，通过 $Lattice(k)$ 和最佳 k 分配方案;

算法 2 中的步骤展示了网联自动驾驶汽车协同决策与轨迹规划的完整过程。该过程将每隔 $\Delta t = 0.1s$ 重复一次，直至达到仿真时长。在完成对人工驾驶车辆和网联自动驾驶汽车的建模后，我们开展了若干实验，以验证所提出的协同驾驶框架。

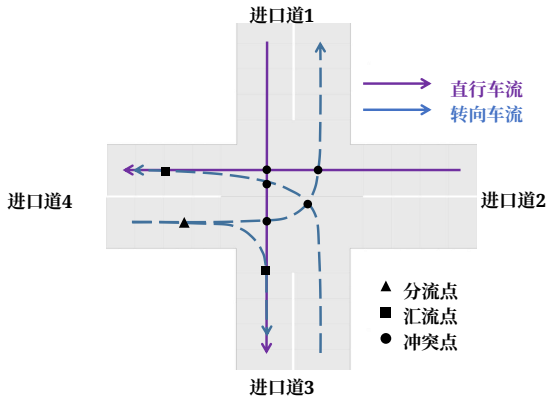

图 4. 仿真环境。

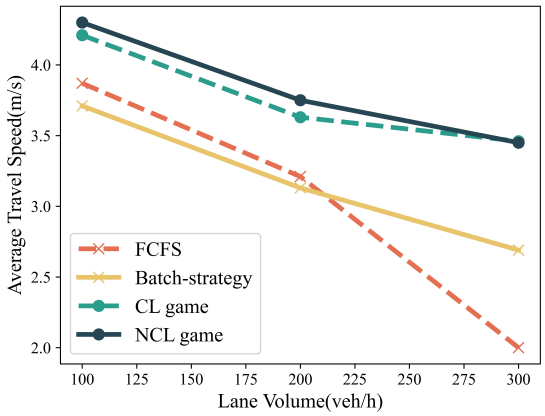

图5. 不同车道流量下控制方法的平均行驶速度比较。

V. 仿真验证

A. 仿真设计

为评估所提出算法框架的性能，采用图 4 所示的孤立四岔无信号控制交叉口作为仿真环境。所有进口道长 40 米，每个进口道包含一条或两条交通流，如图 4 所示。

尽管我们简化了交通流的数量，但根据拓扑关系，仍然存在近十个冲突。除了各种冲突点外，仿真环境还包括汇流点和分流点。这导致了合流和分流行为，进而对混合交通中的车辆行为产生了显著影响 [41]。结合高密度交通流和异质人工驾驶车辆，我们认为，该无信号交叉口足够复杂，能够用于检验我们的算法框架。

此外，偏好不仅影响驾驶员的决策，也会影响其初始速度和目标速度。因此，根据基于 XXJH 数据的先验知识，人工驾驶车辆的初始速度设定为平均速度，目标速度设定为最高速度。例如，根据表 I，激进型驾驶员将以 6.3m/s 的速度进入仿真环境，并期望在可能的情况下以 7m/s 的速度行驶。为了检验我们框架的有效性和鲁棒性，本文进行了三种仿真案例。

首先，我们在不同交通密度下，将 NCL 博弈和 CL 博弈（未经归一化的 NCL 博弈）与基于预约的控制方法进行比较，以评估我们提出的基于底层控制的方法。基于 [19]，我们采用 FCFS 和批处理策略作为基于预约的方法的代表。其次，为了验证所提出的协同驾驶算法的有效性，我们在不同网联自动驾驶汽车渗透率下进行了仿真实验。最后，与将所有人类驾驶员均视为做出普通型决策的情况不同，我们根据表 III 中的逆向强化学习校准结果，将具有不同个人偏好的驾驶员引入仿真环境中。

所有仿真持续了 2 分钟，即 1200 帧。仿真结束后，记录了所有车辆的轨迹，以供进一步评估。

B. 案例*1*：不同控制方法的比较 方法

为评估不同控制方法的结果，我们首先在全网联自动驾驶汽车环境下进行了仿真。采用平均行驶速度来评估系统效率。

图5展示了不同车道流量下的平均行驶速度。显然，平均行驶速度与车道流量呈负相关，这与Greenshields等人关于速度与流量关系的基本认识一致 [42]。此外，在先到先服务控制下，随着车道流量增加，行驶速度大幅下降。但当车道流量相对较低（低于200辆/小时）时，先到先服务控制甚至优于批处理策略，因为在低密度下将车辆分组处理可能会增加不必要的步骤，从而降低效率。

如图5所示，在所有车道流量下，CL博弈和NCL 博弈的行驶速度均高于基于预约的控制方法。这表明，从系统角度看，合作博弈与Level-k博弈的结合是有效的。此外，随着车道流量上升，CL博弈与NCL 博弈之间的差距逐渐缩小。这一现象是因为冲突交通流中的车辆数量变得比交通流类型（转向车流或直行车流）更为重要。因此，随着车道流量增加，归一化的作用减弱。

此外，图 6 展示了在批处理策略控制和 NCL 博弈下，仿真环境中同时存在 8 辆车辆的情况。图 6 (a) 描绘了它们的初始位置，车辆旁边的数字表示车辆的进入顺序。图 6 (b)-(c) 展示了车辆位置随时间迭代变化的过程。由于批处理策略以批次为单位服务车辆，V_7（在图中代表车辆 7）获得了与 V_3 相同的通行权，并以队列形式通过。然而，批处理策略仍然遵循先到先服务原则，因此即使 V_1 所在的车流中只有一辆车辆，V_1 仍然先行通过。这一现象导致其他车流.....

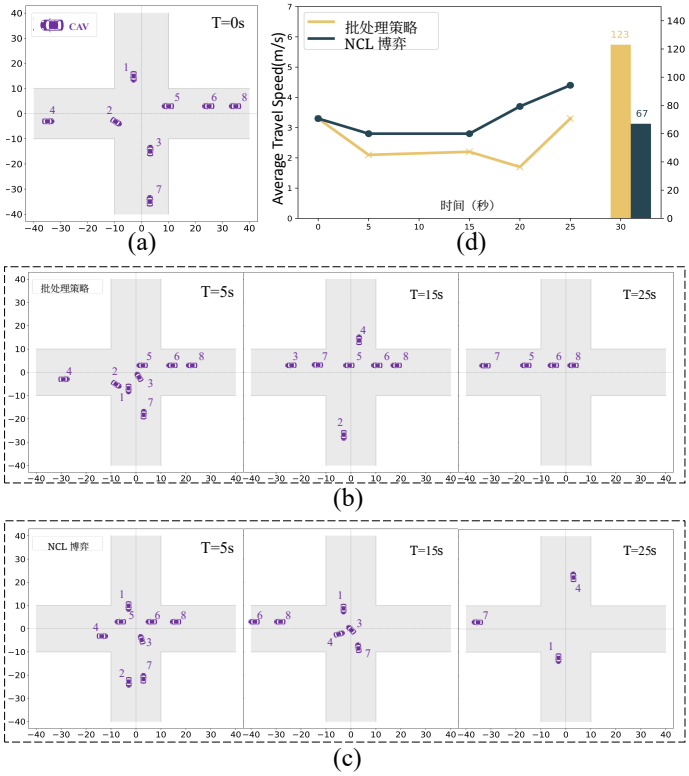

图 6. 交互案例：NCL 博弈与批处理策略的比较：(a) 显示车辆的初始位置；(b)–(c) 分别显示采用批处理策略和 NCL 博弈进行网联自动驾驶汽车控制时后续的三个步骤。(d) 总结了该交互案例的平均行驶速度和总延误；车辆旁的数字表示其进入仿真环境的顺序。

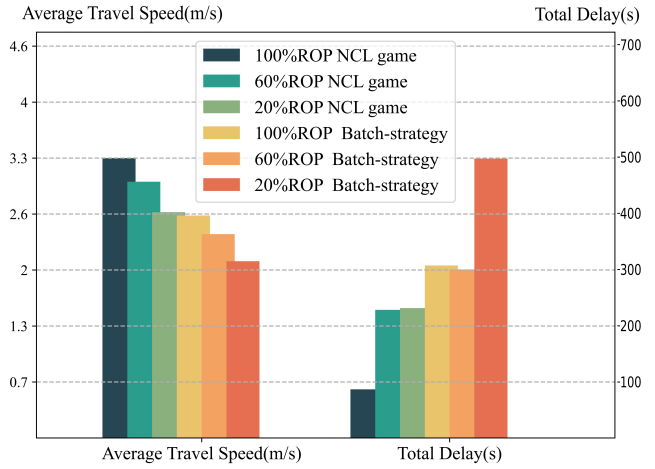

图 7. 不同 ROP 下的平均行驶速度和延误。

与 V_1 发生冲突的车辆必须退让，无论该组中有多少辆车辆，导致效率低下。

可以明显看出，在采用 NCL 博弈进行控制时，在筛选所有可能的 k 分配方案并找出效率最高的方案后，进口道2的车辆拥有最高的通行权，进口道3的车辆次之。图 6 (d) 显示了整个系统在不同仿真时间下的平均行驶速度。显然，我们提出的 NCL 博弈比批处理策略带来了更高的效率。此外，整个交叉口的总延误几乎只有采用批处理策略控制时的一半。需要注意的是，如果车辆持续不断地进入仿真环境，这一差距还将进一步扩大。

因此，考虑到高密度连续交通流会给网联自动驾驶汽车轨迹规划带来更多困难，也会给整个系统带来更大挑战。在接下来的实验中，我们将车道流量固定为 300 辆/小时，以评估所提方法在混合交通中的适用性。

C. 案例 *2*：不同 *ROP* 下的实验

未来很长一段时间内，网联自动驾驶汽车预计仍将与人工驾驶车辆在交通中共同运行。因此，应全面研究不同情况下网联自动驾驶汽车与人工驾驶车辆之间的交互。基于上述实验，我们采用 NCL 博弈和批处理策略作为控制方法。每种控制方法均在三种 ROP 下进行测试，分别为 100%、60% 和 20%。

图 7 给出了计算并标准化后的平均行驶速度和总延误。随着 ROP 增大，平均行驶速度上升、总延误下降，呈现出明显的趋势，表明系统效率有所提升。此外，在所有 ROP 下，采用 NCL 博弈控制网联自动驾驶汽车时，系统的表现均更好。从另一个角度来看，采用 NCL 博弈控制时的最低平均行驶速度，甚至高于采用批处理策略控制时的最高速度，这证实了高密度连续交通流确实扩大了 NCL 博弈与批处理策略之间的差距。总延误也呈现出相同的结论，证明我们提出的 NCL 博弈优于传统的基于预约的方法，适用于同时包含网联自动驾驶汽车和人工驾驶车辆的混合交通流。

以上分析旨在评估交叉口的效率。由于安全性取决于人们对网联自动驾驶汽车的信任，而这目前是网联自动驾驶汽车普及的一大障碍，因此也应关注安全性。为此，我们引入 PET，从安全性角度评估算法。最初，Allen 等人 [43]将 PET 定义为前车驶离冲突区域或侵入线的时刻与后车到达冲突区域或侵入线的时刻之间的时间。不同 ROP 和控制方法下的 PET 分布如图 8 所示。

根据以上箱线图，随着 ROP 降低，PET 分布变得更加离散，这表明车辆之间的交互更加混乱且效率低下。同时，ROP 下降也导致 PET 最大值和平均值升高，这意味着按照 PET 的定义，车辆通过的间隔更长。换言之，随着人工驾驶车辆占比上升，交互变得更加保守且危险。因此，发生死锁的可能性更高。

另一条证据来自最小值 PET，它表示整个仿真时长内最危险的交互案例。图 8 显示了正相关关-

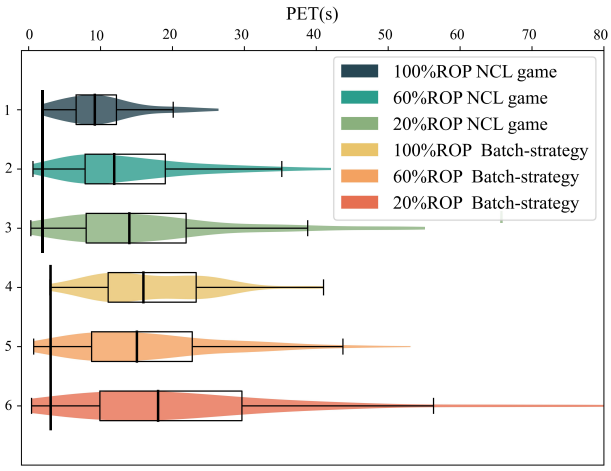

图 8. 不同ROP下PET的分布。

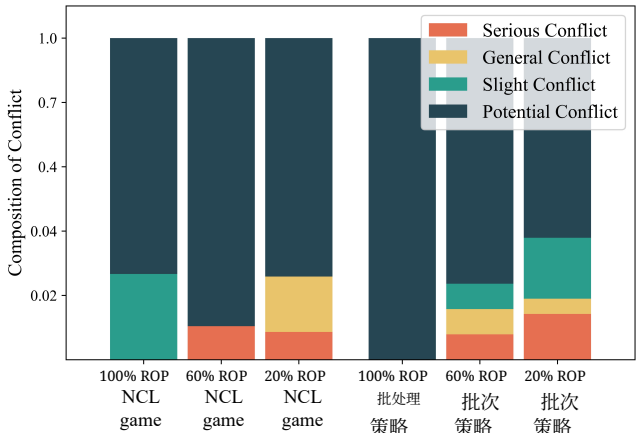

图 9. 不同ROP下的冲突组成。

ROP 与最小 PET 之间的关系，证明增加网联自动驾驶汽车的数量有助于提高系统安全性。

Qi等人 [44] 根据冲突数据确定了适当的PET阈值，并将冲突分为四类。PET<$0.7s$表示严重冲突， $0.7s \leq$ PET<$1.31s$ 表示一般冲突， $1.31s \leq$PET<$2.25s$ 表示轻微冲突，PET$\geq 2.25s$ 表示潜在冲突。此外，我们根据上述划分计算并可视化了图 9 中的冲突组成。

图 9 显示，在所有网联自动驾驶汽车环境中（ROP为100%），交互均为轻微冲突或潜在冲突。此外，随着ROP降低，严重冲突和一般冲突的比例有所增加。

通过平均行驶速度、总延误和 PET，我们证明了所设计的网联自动驾驶汽车能够推动交叉口系统实现更高的效率和安全性。然而，除了高密度和混合交通之外，人类驾驶员还具有不同的决策偏好，从而形成异质交通。因此，与异质人工驾驶车辆的交互将进一步增强我们网联自动驾驶汽车算法框架的说服力。

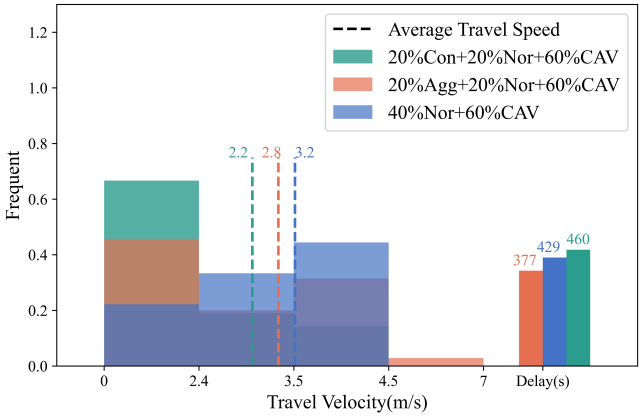

图 10. 不同车辆组成下的平均行驶速度和总延误。

D. 案例*3*：涉及异质人类驾驶员

我们首先在 60% 的 ROP 下进行了三组实验。与案例 2 中让所有人工驾驶车辆在与其他车辆交互时作出普通型决策的情况不同，其中一半人工驾驶车辆被激进型或保守型车辆所替代，这可能导致更难以预测的行为。总延误和平均行驶速度的分布如图 10 所示。

图 10 显示，在仿真环境中，仅有普通型人工驾驶车辆与网联自动驾驶汽车混行时，平均行驶速度显著高于有激进型人工驾驶车辆或保守型人工驾驶车辆参与的情况。此外，当 20% 的普通型人工驾驶车辆被激进型人工驾驶车辆替换时，尽管正如我们在小节 *A* 中所指定的，激进型人工驾驶车辆具有较高的初始速度和目标速度，但低速（速度 $<2.4m/s$）车辆的比例反而有所上升。更不用说，如果 20% 的普通型车辆被 20% 的保守型车辆替换，低速车辆的比例将从 22% 急剧升至 67%。尽管三组实验的总延误差异不大，但上述实验仍凸显了与决策多变的异质人工驾驶车辆交互的难度。

尽管这些发现只能证明我们的网联自动驾驶汽车能够与异质人工驾驶车辆交互（因为没有发生碰撞或死锁），但仍应开展实验，检验我们的网联自动驾驶汽车能否提高系统效率，从而验证其与异质人工驾驶车辆协调的能力。因此，我们根据 XXJH 数据中的人工驾驶车辆组成，构建了一个基于真实世界的环境。最终，实验组由 13% 的激进型人工驾驶车辆、41% 的普通型人工驾驶车辆和 46% 的保守型人工驾驶车辆组成。在对照组中，我们将普通型人工驾驶车辆替换为所设计的网联自动驾驶汽车（请注意，它们具有相同的初始速度、目标速度、奖励构成和奖励权重）。

根据图 11，将普通人工驾驶车辆替换为我们的网联自动驾驶汽车后，系统的各方面性能均有所提升。具体而言，交叉口系统的平均行驶速度提高了 0.3 m/s，高速车辆的比例也有所增加。更显著的变化是总延误减少了 42.5%，这证实了我们的网联自动驾驶汽车算法框架在异质混合交通下的鲁棒性和优越性

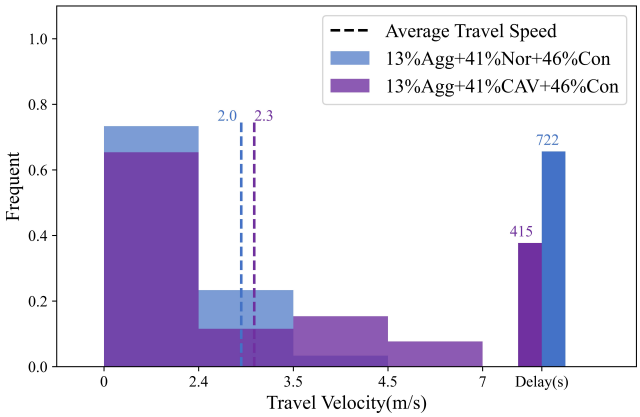

图 11。基于 XXJH 车辆组成的平均行驶速度和总延误。

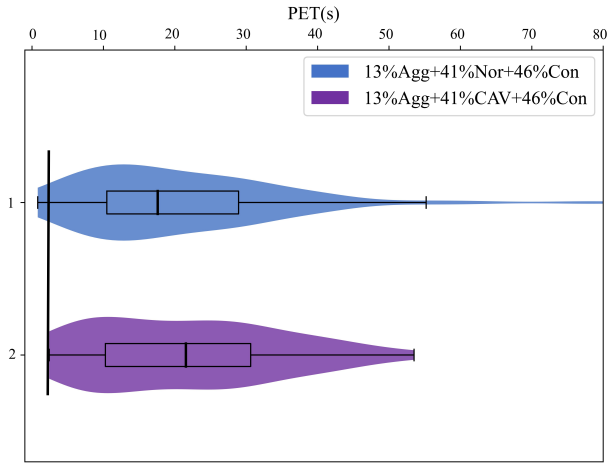

图12。基于 XXJH 车辆组成的 PET 分布。

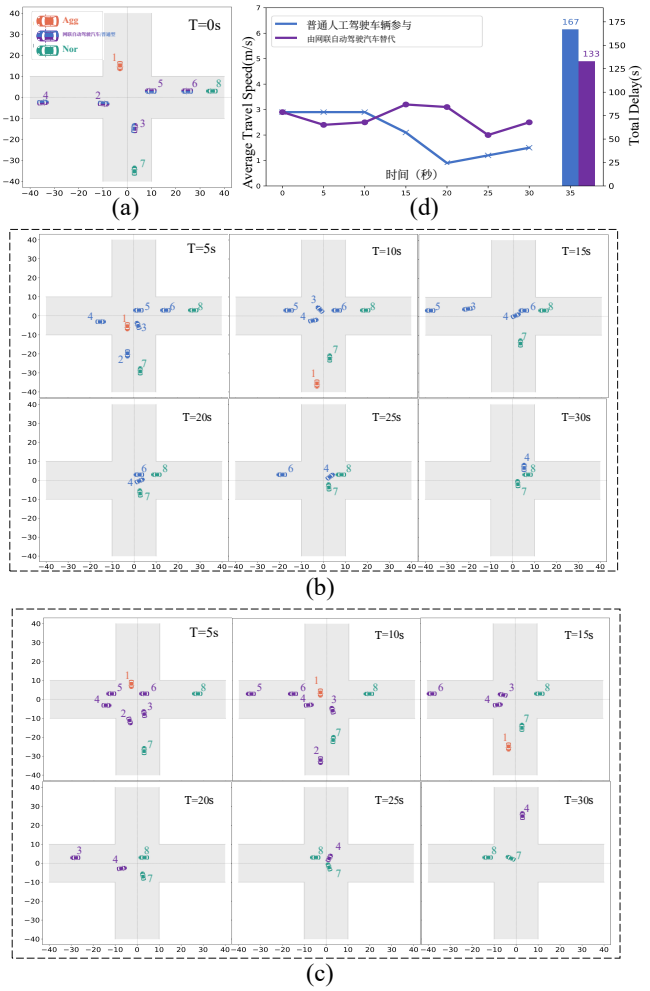

图13. 交互案例：普通人工驾驶车辆与网联自动驾驶汽车参与时的比较。(a) 显示车辆的初始位置，车辆颜色表示其类型；(b) 显示仅有人工驾驶车辆参与交互时后续的六个步骤；(c) 显示将普通人工驾驶车辆替换为网联自动驾驶汽车后的相应步骤；(d) 总结了该交互案例的平均行驶速度和总延误；车辆旁的数字表示它们进入仿真环境的顺序。

环境。

与案例 2 类似，我们计算了仿真时长内各交互的 PET。如图12所示，将普通人工驾驶车辆替换为网联自动驾驶汽车后，PET 分布更加集中，且最小 PET 增大，表明我们的网联自动驾驶汽车能够妥善组织其他车辆通行，并提升驾驶安全性。我们进一步选取了一个交互案例，以直观展示普通人工驾驶车辆和网联自动驾驶汽车参与时系统演化过程的差异。

通过对比分析图 6 (d) 和图13 (d) 可以发现，异质人工驾驶车辆的引入导致平均行驶速度降低、总延误增加，这给我们的协同驾驶算法框架带来了更大的挑战。图13 (b)-(c) 详细展示了交互过程。在全人工驾驶车辆环境中，激进型 V_1追求自身利益，并设法尽可能快速地通过。这一现象导致系统在交互初期具有更高的平均行驶速度。然而，由于激进车辆的蛮横操作可能会在较远距离内影响其他车辆，因此在其驶离交叉口后，平均行驶速度大幅下降。

在我们用网联自动驾驶汽车替代普通人工驾驶车辆后，网联自动驾驶汽车能够通过表现出优先通行的意愿，主动促使激进车辆减速（通过比较图13 (b)-(c) 在 $T = 5s$ 和 $T = 10s$ 时的状态）。这归因于我们基于博弈的协同方法。它们的协同行为是自发形成的，而不是由某种预定义协议产生的，这使我们有潜力将其应用于更复杂的场景。尽管根据图13 (d)，整个交叉口的平均行驶速度起初略低，但在后续仿真时长内的速度以及总延误均显著优于没有网联自动驾驶汽车参与的情况。到目前为止，我们的网联自动驾驶汽车协同驾驶框架在各种环境下的有效性和鲁棒性已经得到验证。

总之，我们针对协同驾驶框架的验证与确认开展了三个仿真案例。仿真结果表明，借助所设计的框架，网联自动驾驶汽车能够在高密度、混合交通、无信号控制的

交叉口中与异构车辆进行协调。此外，随着网联自动驾驶汽车占比的增加，交叉口系统的效率和安全性均得到提升。

VI. 结论

通过结合归一化合作博弈与Level-k博弈，本文提出了一种面向网联自动驾驶汽车的协同驾驶框架，用于解决高密度、混合、无信号控制交叉口中的驾驶冲突。不同于基于预约的控制方法，所提出的NCL 博弈论方法能够与其他车辆进行合作，甚至主动协调。也就是说，除了网联自动驾驶汽车之间的合作外，网联自动驾驶汽车还可以与异质人工驾驶车辆协同。本文开展了三个仿真案例进行验证与确认，包括与不同方法的对比分析、不同ROP条件下的案例研究，以及与异质人工驾驶车辆的交互分析。通过平均行驶速度、总延误和PET对交叉口系统的性能进行了分析。实验结果表明，所提出的协同驾驶框架能够应对复杂的混合交通场景，从而提升系统的效率与安全性。

参考文献

[1] W. Schwarting、A. Pierson、J. Alonso-Mora、S. Karaman 和 D. Rus，《自动驾驶车辆的社会行为》，美国国家科学院院刊，第116卷，第50期，第24 972–24 978页，2019年。[2] C. F. Camerer 和 E. Fehr，《“经济人”何时主导社会行为？》，科学，第311卷，第5757期，第47–52页，2006年。[3] R. Tian、S. Li、N. Li、I. Kolmanovsky、A. Girard 和 Y. Yildiz，《用于环形交叉口自动驾驶车辆控制的自适应博弈论决策》，载于*2018*年*IEEE*决策与控制会议（*CDC*）。IEEE，2018年，第321–326页。[4] C. Xu、W. Zhao、L. Li、Q. Chen、D. Kuang 和 J. Zhou，《一种考虑与交通参与者交互的基于纳什Q学习的运动决策算法》，*IEEE*车辆技术汇刊，第69卷，第11期，第12 621–12 634页，2020年。[5] M. Garz´on 和 A. Spalanzani，《高交通流量场景下自动驾驶车辆汇入机动的博弈论决策》，载于*2019*年*IEEE*智能交通系统会议（*ITSC*）。IEEE，2019年，第3448–3453页。[6] R. Tian、N. Li、I. Kolmanovsky、Y. Yildiz 和 A. R. Girard，《面向自动驾驶车辆控制验证与确认的无信号交叉口网络交通博弈论建模》，*IEEE*智能交通系统汇刊，2020年。[7] H. Wang、Y. Li 和 H. V. Zhao，《基于博弈论和动态Level-k模型的道路交叉口性能分析》，载于*2020*年*IEEE*并行与分布式处理及应用、海量数据与云计算、可持续计算与通信、社会计算与网络国际会议（*ISPA/BDCloud/SocialCom/SustainCom*）。IEEE，2020年，第1112–1119页。[8] H. Yu、H. E. Tseng 和 R. Langari，《用于自动变道的类人博弈论控制器》，《运输研究*C*辑：新兴技术》，第88卷，第140–158页，2018年。[9] N. Li、I. Kolmanovsky、A. Girard 和 Y. Yildiz，《无信号交叉口车辆交互的博弈论建模及其在自动驾驶车辆控制中的应用》，载于*2018*年美国控制会议（*ACC*）。IEEE，2018年，第3215–3220页。[10] S. Karimi 和 A. Vahidi，《利用蒙特卡洛树搜索和Level-k博弈论进行自动变道与汇入的滚动时域运动规划》，载于*2020*年美国控制会议（*ACC*）。IEEE，2020年，第1223–1228页。[11] P. Chu、Y. Yu、J. Yang 和 C. Huang，《基于SOR理论理解年轻驾驶员分心驾驶行为背后的机制》，《交通安全与安保杂志》，第14卷，第10期，第1655–1673页，2022年。[12] Z. el abidine Kherroubi、S. Aknine 和 R. Bacha，《高速公路入口匝道汇入场景下网联自动驾驶车辆的新型决策策略》，*IEEE*智能交通系统汇刊，2021年。

[13] K. Yu、L. Lin、M. Alazab、L. Tan 和 B. Gu，《5G支持的智能交通系统中混合自动驾驶车辆与人工驾驶车辆的基于深度学习的交通安全解决方案》，IEEE智能交通系统汇刊，第22卷，第7期，第4337–4347页，2020年。[14] R. Yao、W. Zeng、Y. Chen 和 Z. He，《混合流交叉口中考虑斜向穿越摩托车冲突的左转车辆行为建模深度学习框架》，《运输研究C辑：新兴技术》，第132卷，第103415页，2021年。[15] L. Bi、G. Gan、J. Shang 和 Y. Liu，《有认知分心任务或无认知分心任务情况下驾驶员横向控制的排队网络建模》，IEEE智能交通系统汇刊，第13卷，第4期，第1810–1820页，2012年。[16] Z. Huang、J. Wu 和 C. Lv，《利用逆强化学习和自然驾驶数据对驾驶行为建模》，IEEE智能交通系统汇刊，2021年。[17] I. V. Chremos、L. E. Beaver 和 A. A. Malikopoulos，《网联自动驾驶车辆社会影响的博弈论分析》，载于2020年IEEE第23届智能交通系统国际会议（ITSC）。IEEE，2020年，第1–6页。[18] S. Mariani、G. Cabri 和 F. Zambonelli，《自动驾驶车辆协调：分类体系与综述》，ACM计算调查（CSUR），第54卷，第1期，第1–33页，2021年。[19] C. Yu、W. Sun、H. X. Liu 和 X. Yang，《孤立交叉口的网联自动驾驶车辆管理：从预约到基于优化的方法》，《运输研究B辑：方法论》，第122卷，第416–435页，2019年。[20] J. Rios-Torres 和 A. A. Malikopoulos，《网联自动驾驶车辆在交叉口的协调及其在高速公路入口匝道汇入的综述》，IEEE智能交通系统汇刊，第18卷，第5期，第1066–1077页，2016年。[21] D. Zhou、Z. Ma、X. Zhao 和 J. Sun，《推理图：混合网联自动驾驶交通环境下无保护转向协作的情境感知框架》，《运输研究C辑：新兴技术》，第143卷，第103815页，2022年。[22] H. Xu、Y. Zhang、L. Li 和 W. Li，《利用树搜索在无信号交叉口协同驾驶》，IEEE智能交通系统汇刊，第21卷，第11期，第4563–4571页，2019年。[23] D. Carlino、S. D. Boyles 和 P. Stone，《基于拍卖的自动驾驶交叉口管理》，载于第16届IEEE智能交通系统国际会议（ITSC 2013）。IEEE，2013年，第529–534页。[24] H. Vu、S. Aknine 和 S. D. Ramchurn，《交叉口交通管理的去中心化方法》，载于IJCAI，2018年，第527–533页。[25] C.-L. Huang 和 G. Nitschke，《协同自动驾驶车辆的进化自动化》，载于2020年IEEE进化计算大会（CEC）。IEEE，2020年，第1–7页。[26] T. Litman，《自动驾驶车辆实施预测》。维多利亚交通政策研究所，加拿大不列颠哥伦比亚省维多利亚市，2017年。[27] M. Wang、S. P. Hoogendoorn、W. Daamen、B. van Arem 和 R. Happee，《预测性换道与跟驰控制的博弈论方法》，《运输研究C辑：新兴技术》，第58卷，第73–92页，2015年。[28] D. Sadigh、N. Landolfi、S. S. Sastry、S. A. Seshia 和 A. D. Dragan，《为与人类协作的汽车进行规划：利用对人类行动的影响进行规划，并主动收集人类内部状态信息》，《自主机器人》，第42卷，第7期，第1405–1426页，2018年。[29] S. Vaskov、S. Kousik、H. Larson、F. Bu、J. Ward、S. Worrall、M. Johnson-Roberson 和 R. Vasudevan，《迈向任意动态环境下可证明无责的自主机器人控制》，arXiv预印本 arXiv:1902.02851，2019年。[30] M. Liu、Y. Wan、F. L. Lewis、S. Nageshrao 和 D. Filev，《自动驾驶车辆的三级博弈论决策框架》，IEEE智能交通系统汇刊，2022年。[31] Y. Ni、M. Wang、J. Sun 和 K. Li，《交叉口行人安全评估：基于人车交互模式的理论框架》，《事故分析与预防》，第96卷，第118–129页，2016年。[32] K.-T. Chen、H.-Y. W. Chen 和 H.-Y. W. Chen，《利用自然驾驶数据进行驾驶风格聚类》，《运输研究记录》，第2673卷，第6期，第176–188页，2019年。[33] Y. Huang、D. J. Sun 和 L.-H. Zhang，《交通拥堵对驾驶员速度选择的影响：基于出租车浮动车数据评估状态激进程度的中介作用》，《事故分析与预防》，第117卷，第318–327页，2018年。[34] Y. Rahmati、M. K. Hosseini 和 A. Talebpour，《辅助自动驾驶车辆完成左转机动：交叉口冲突机动的博弈论决策框架

,” *IEEE Transactions on*

智能交通系统, 2021. [35] C. Cheng, D. Yao, Y. Zhang, J. Li, and Y. Guo, “基于非合作博弈论的无信号交叉口车辆超车模型”，载于*2019 IEEE*智能交通系统会议（*ITSC*）。IEEE，2019，第2286–2291页。[36] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey 等，“最大熵逆强化学习”，载于*Aaai*，第8卷。美国伊利诺伊州芝加哥，2008，第1433–1438页。[37] X. Wang, S. Zhang, and H. Peng，“环岛场景下高度自动化车辆的综合安全性评估”，*IEEE*智能交通系统汇刊，2022。[38] K. Dresner and P. Stone，“自动驾驶车辆交叉口管理的多智能体方法”，人工智能研究期刊，第31卷，第591–656页，2008。[39] Z. Yang, H. Huang, G. Wang, X. Pei, and D.-y. Yao，“采用合作博弈的无信号交叉口协同驾驶模型”，中南大学学报，第25卷，2181，2018年9月，第2164页。[40] R. Xing, Z. Su, Q. Xu, N. Zhang, and T. H. Luan，“面向网联自动驾驶卡车的安全内容交付：一种联盟形成博弈方法”，*IEEE*智能交通系统汇刊，2022。[41] J. Guo, S. Cheng, and Y. Liu，“常规车辆与自动驾驶车辆汇入和驶离对混合交通的影响”，*IEEE*智能交通系统汇刊，第22卷，1649，2020年3月，第1639页。[42] B. Greenshields, J. Bibbins, W. Channing, and H. Miller，“交通容量研究”，载于公路研究委员会论文集，第1935卷。美国国家研究委员会（美国），公路研究委员会，1935。[43] B. L. Allen, B. T. Shin, and P. J. Cooper，“交通冲突与碰撞分析”，技术报告，1978。[44] W. Qi, W. Wang, B. Shen, and J. Wu，“基于变道特征的城市道路汇入区域修正后侵入时间模型”，*IEEE Access*，第8卷，第72 835–72 846页，2020。