控制与决策
Control and Decision

基于博弈论的多车智能驾驶交互决策综述

衣鹏, 潘越, 王文远, 刘政钦, 洪奕光

引用本文:
衣鹏,潘越,王文远,刘政钦,洪奕光. 基于博弈论的多车智能驾驶交互决策综述[J]. 控制与决策, 2023, 38(5): 1159−1175.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.1512

基于博弈论的多车智能驾驶交互决策综述

衣 鹏[1,2†], 潘 越[1], 王文远[1], 刘政钦[1], 洪奕光[1,2]

(1. 同济大学 电子与信息工程学院，上海 200082；2. 上海自主智能无人系统科学中心，上海 201210)

摘 要: 智能驾驶是交通和汽车领域未来发展的重要方向,决策规划作为智能驾驶系统中的关键模块,一直是其重点研究领域之一. 当前的研究热点正在从单车智能驾驶决策向混行交通场景下的多车智能驾驶决策进行拓展,因此,需要在复杂动态场景和多并行任务下生成符合车辆动力学且不与道路边界和其他交通参与者发生碰撞的高质量轨迹. 多车混行驾驶是对道路时空资源的竞争性使用,博弈论可为多车交互决策提供重要的理论与技术手段. 对此,应用博弈论方法进行智能驾驶决策研究的综述,基于滚动时域、微分博弈和马尔科夫博弈这3类常用的博弈模型,对现有相关研究进行归类总结和分析. 首先简要介绍博弈论基础知识;其次,总结常见的智能驾驶场景并分析各场景下交互决策的核心问题;然后,通过3种不同的博弈模型对多车交互决策进行建模,分别介绍它们的求解算法和思路及相关的研究工作;最后,介绍相关的仿真实验和测试方法,同时也对未来的技术发展和挑战给出见解.

关键词: 博弈论；智能驾驶；多车交互；博弈模型；智能决策；综述

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1512

引用格式: 衣鹏,潘越,王文远,等. 基于博弈论的多车智能驾驶交互决策综述 [J]. 控制与决策, 2023, 38(5): 1159-1175.

A review on interactive decision-making of multi-vehicle autonomous driving with a game theoretical perspective

YI Peng[1,2†], *PAN Yue*[1], *WANG Wen-yuan*[1], *LIU Zheng-qin*[1], *HONG Yi-guang*[1,2]

(1. School of Electronic and Information Engineering，Tongji University，Shanghai 200082，China；2. Shanghai Research Institute for Intelligent Autonomous Systems，Shanghai 201210，China)

Abstract: Autonomous driving is an important direction for the future development of the transportation and the automotive field. Planning and decision-making as a key module has always been an important area of autonomous driving research. The research interest of the community is expanding from single-vehicle driving to multi-vehicle driving in hybrid traffic scenarios. The main forthcoming challenge is to generate high-quality trajectories that conform to vehicle dynamics and do not collide with road boundaries and other traffic participants in complicated and dynamical scenarios with multiple tasks. Multi-vehicle driving can be treated as competitive utilization of the spatial-temporal resources of the road, hence game theory can provide an important theoretical and technical tool for multi-vehicle interactive decision-making. This paper provides a review on applying game theory to the decision-making of intelligent driving. We summary and classify the literatures on the interactive decision-making of intelligent driving based on game theory into three commonly used game methods, including the receding horizon games, the differential games, and the Markovian games. At first, the basic knowledge of game theory is introduced, and then the common intelligent driving scenarios are introduced and the key problems of interactive decision-making are summarized. Then, the above three game methods are used to model the multi-vehicle interactive decision-making, and the solution algorithms in literatures are introduced respectively. Finally, the relevant simulation experiments and test methods are summarized. The remaining challenges for the future development of areas are discussed with a personal perspective.

Keywords: game theory；intelligent driving；multi-vehicle interaction；game model；intelligent decision-making；survey

收稿日期: 2022-08-23；**录用日期:** 2023-01-29.

基金项目: 国家自然科学基金项目 (62003239)；科技部重点研发计划项目 (2022YFA1004700).

责任编委: 杨涛.

†通讯作者. E-mail: yipeng@tongji.edu.cn.

0 引 言

随着经济的发展,汽车的保有量逐年增加,车辆行驶安全和道路拥堵已成为亟待解决的两大交通领域难题. 据世界卫生组织统计,全球每年约有130万人因交通事故而死亡,几千万人因此受伤,其中大部分事故是驾驶员的错误判断与操作失误造成的. 百度地图在《2021年度中国城市交通报告》中给出了我国10个省会城市的交通拥堵情况,报告中显示,这些城市的车辆在通勤高峰期的实际速度仅在25～30 km/h之间. 新一代的智能交通系统有望通过数字化和智能化手段解决以上两个问题. 无人驾驶汽车是汽车未来发展的重要方向,将成为道路交通领域具有颠覆性影响的变革性运载工具. 2015年国务院印发了《中国制造2025》,将智能网联汽车列入未来10年国家智能制造发展的重点领域,明确指出了2020年掌握智能辅助驾驶的总体技术及各项关键技术,2025年掌握智能驾驶总体技术及各项关键技术. 2022年3月,我国正式发布了《汽车驾驶自动化分级》标准,将汽车驾驶自动化等级划分为L0～L5级,如表1所示. 该标准根据中国的智能驾驶的发展特点,规定了汽车驾驶自动化分级遵循的原则和技术要求,促进了智能汽车的发展. 随着人工智能、传感器、大数据、云计算等核心技术的突破和完善,智能驾驶汽车将不断提升行驶安全、提高通行效率,逐渐被公众接受,成为未来出行和物流的有效工具.

表1 汽车驾驶自动化分级

等级	名称	控制	支援	运行范围
L0	应急辅助	驾驶员	驾驶员	有限制
L1	部分辅助驾驶	驾驶员/系统	驾驶员	有限制
L2	组合辅助驾驶	系统	驾驶员	有限制
L3	有条件自动驾驶	系统	后援用户	有限制
L4	高度自动驾驶	系统	系统	有限制
L5	完全自动驾驶	系统	系统	无限制

作为典型的自主智能无人系统,智能驾驶汽车是汽车电子信息化和智能化的高科技产物,是集环境感知、决策规划、控制执行等功能于一体的现代运载工具和移动信息处理平台. 一般而言,智能驾驶系统可分为感知层、决策层和控制层3个模块[1]. 感知层通过观测和分析数据,提供有效的环境认知信息;决策层通过综合出行任务和环境认知给出判断和决策并生成规划路径;控制层根据决策层提供的参考轨迹输出底层的控制量. 可以将感知层比作人的眼睛与耳朵,决策层就相当于人的大脑,而控制层相当于人的手和脚,它们之间的关系和类比如图1所示.

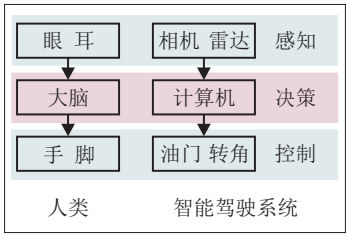

图1 智能驾驶系统功能模块

智能驾驶系统中的3个模块密切耦合协同以实现自主无人的驾驶. 其中,决策层起到连接感知模块和控制模块的关键作用,是从感知智能和认知智能向决策智能的递进. 目前,围绕单车智能驾驶决策已经有了大量的研究[2-4]. 如图2所示,传统的单车决策规划算法可以分为4类:基于图搜索的方法、基于采样的方法、曲线拟合方法、数值优化方法. Dijkstra算法和A*算法是常用的基于图搜索的算法. 荷兰科学家Dijkstra在20世纪60年代提出了用来寻求节点之间的最短路径的经典算法,后被称为Dijkstra算法. 20世纪70年代,斯坦福大学研究者提出了A*算法,该算法是一种启发式的图搜索算法,能实现节点间路径的快速搜索,可以视为Dijkstra算法的拓展. 在将车辆运动空间栅格化后,基于图搜索的算法可以被用于车辆的路径规划. 为解决复杂交通场景下的车辆决策规划问题,研究人员还提出了改进的Dijkstra算法[5-6]和A*算法[7-8]. 快速拓展随机树(RRT)算法是一种基于采样的方法,它由美国科学家LaValle于1998年提出,并得到了广泛的发展与应用[9-11]. 但该算法对整个空间进行均匀采样,无法保证规划的效率. 曲线拟合方法和数值优化方法是局部路径规划算法,其中常用的算法有多项式曲线法[12]、贝塞尔曲线法[13]和直接优化法[14-17].

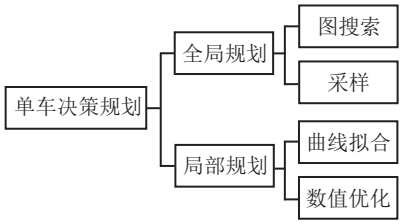

图2 单车决策规划算法分类

随着智能驾驶的发展,有人车、无人车混行交通场景[18]下的多车智能驾驶决策规划得到了关注与研究. 基于单车决策规划算法,在多车场景下可以使用基于规则[19]、优化算法[20]和群智算法[21]等进行决策规划,但是,这些方法求解出的策略无法体现交通参与者之间的动态交互性,难以达到高效决策. 多车决策规划是对道路时空资源的竞争性使用,因此,智能驾驶汽车在道路上的行为会影响其他交通参与者,并

且也会受到其他司机和行人的影响. 例如无论是超车、协商合并还是为了避免发生事故,多车之间存在相互依赖的影响关系,可以视为一种博弈行为. 博弈论是分析多个智能体行为交互的数学理论与方法,为了更好地解决多车交互决策问题,研究人员将其作为重要的智能决策工具[22]. 目前,因其主体利益描述能力和丰富算法[23-27],博弈论方法已经成为一个重要的多车智能决策工具,并得到了日益广泛的关注.

本文对结合博弈模型与方法进行多车交互决策研究的相关工作进行综述与介绍. 首先简介博弈论基础知识;其次讲解智能驾驶多车交互的场景与核心问题;基于3种不同的博弈方法对智能驾驶场景进行建模;阐述不同信息结构下基于博弈的多车决策求解算法;介绍典型的仿真实验与测试方法;最后总结全文,并对未来的研究方向进行了展望.

1　博弈基础介绍

目前,不同类型的博弈方法已被应用于智能驾驶决策,可将其分为3类:滚动时域博弈、迭代微分博弈和马尔可夫博弈. 滚动时域博弈是静态重复博弈与模型预测控制思想的结合,可以通过状态预测刻画动态博弈过程. 微分博弈问题可以视为有多个参与者的最优控制问题,其应用已从零和微分博弈拓展到了一般和微分博弈. 马尔可夫博弈是马尔可夫决策过程向多智能体场景的拓展,是具有竞争或合作目标的多智能体交互决策的博弈论框架. 下面分别介绍3类博弈方法的基础知识.

1.1　滚动时域博弈

1.1.1　静态博弈

在静态博弈中,每个参与者同时选择其行动. 策略式博弈是一种典型的静态博弈,按照参与者的行动是否具有随机性,策略式博弈分为纯策略式博弈和混合策略式博弈.

纯策略式博弈由参与者集合、行动集合和效用组成. 对于有N个参与者的博弈,参与者用正整数i表示,参与者集合表示为$I=\{i|i\leqslant N, i\in \boldsymbol{Z}^+\}$. 参与者$i$的行动集合记为$S^i$,包含参与者$i$的全部可选行动. 参与者$i$的决策是在$S^i$中选取行动元素$s^i$. 效用$u^i$是所有参与者的行动组合$s=(s^1,s^2,\dots,s^N)$的函数,$u^i(s)$表示行动给参与者$i$带来的收益或代价.

混合策略式博弈在纯策略式博弈的基础上增加了混合策略集合. 混合策略式博弈中,参与者i决策的内容是行动集合S^i的概率分布σ^i,因此行动具有随机性. 具体而言,设集合S^i中有n个行动元素$S^{i,j}$,其中$j=1,2,\dots,n$,概率分布σ^i给出了选择动作$S^{i,j}$的概率为$\sigma^i(s^{i,j})$. 参与者i的混合策略集合Σ^i是所有满足$\sum\limits_{j=1}^{n}\sigma^i(s^{i,j})=1$的$\sigma^i$的集合. 混合策略式博弈的效用定义为纯策略收益在概率分布下的期望,即

$$E[u^i(\sigma)]=\sum_{s^i\in S^i}\Big(u^i(s^1,s^2,\dots,s^N)\prod_{j=1}^{N}\sigma^j(s^j)\Big). \tag{1}$$

策略式博弈可以用矩阵表示,因此也称矩阵博弈. 策略式博弈中的稳定解概念被称为纳什均衡,纳什均衡的定义表明:每个参与者在纳什均衡中都采取对其他参与者策略的最优响应,任何参与者单方面偏离纳什均衡都会导致其效用减小. 当存在非理性的参与者时,可能出现参与者偏离均衡而导致多个参与者收益下降的情况,故策略式博弈中还存在安全解的概念.

策略式博弈对行动的约束隐含在行动集合中,但这样表示的前提是行动集合S^i不受其他参与者行动的影响,相当于s^i不受s^{-i}的约束. 如果s^i受s^{-i}的约束,则问题就属于广义纳什均衡问题[28]. 此类问题中,参与者i的行动集合$S^i(s^{-i})$是其他参与者行动s^{-i}的函数,每个参与者在给定s^{-i}的条件下求解如下约束优化问题:

$$\begin{aligned}&\min_{s^i} g^i(s^i,s^{-i}),\\&\text{s.t. } s^i\in S^i(s^{-i}),\end{aligned} \tag{2}$$

其中g^i为参与者i的目标函数. 当每个参与者都在给定s^{-i}的条件下最小化目标函数时,行动组合$s^*=(s^{1*},\dots,s^{N*})$为广义纳什均衡. 与矩阵博弈相比,广义纳什均衡问题能处理耦合硬约束,从而保证强制执行安全约束.

1.1.2　重复博弈

静态博弈是一种单次博弈,无法描述动态过程. 如果将静态博弈重复多次进行,则构成了重复博弈. 由于参与者能够根据历史信息进行动态决策挑战,重复博弈中的策略相比于静态博弈更复杂,均衡点的数量也更多.

在重复博弈中,参与者在决策时能观察到过去阶段的全部行动$(s^1,s^2,\dots,s^t)$,定义该向量为历史$h(t)$. $h(t)$定义在阶段博弈重复t次的行动集合S^t上. 参与者在重复博弈中根据历史决策采取行动,因此,参与者i的策略是从历史$h(t)$到当前行动s_i^t(纯策略)或行动概率σ_i^t(混合策略)的映射. 在同一时刻,每个参与者的决策是独立的.

重复博弈根据阶段数分为有限重复博弈和无限

重复博弈. 有限重复博弈适用于建模阶段数确定的短期问题,参与者在每个阶段都无法通过偏离阶段博弈的纳什均衡获得收益,所以阶段博弈的纳什均衡组成的序列是有限重复博弈的纳什均衡. 无限重复博弈适用于建模阶段数不确定的长期问题.

1.1.3 滚动时域博弈

重复博弈在使用相同的静态博弈模型前提下具备了动态决策能力,能够产生决策序列以处理需要连续动态决策的问题. 在智能驾驶决策场景中,博弈模型在各个阶段保持不变是不现实的假设. 更为合理的设想是驾驶行动集合不变,但是,可以根据当前的交通状态在每个阶段更新博弈收益. 将重复博弈与模型预测控制的思想结合,就形成了滚动时域博弈. 此时既保持了重复博弈的形式又引入了状态预测,但决策依然是一个从历史到当前行动的映射. 图3展示了滚动时域博弈的流程,其中目标状态x_f作为输入,在智能驾驶决策中有多种含义,例如车辆的期望速度、终点和车道等.

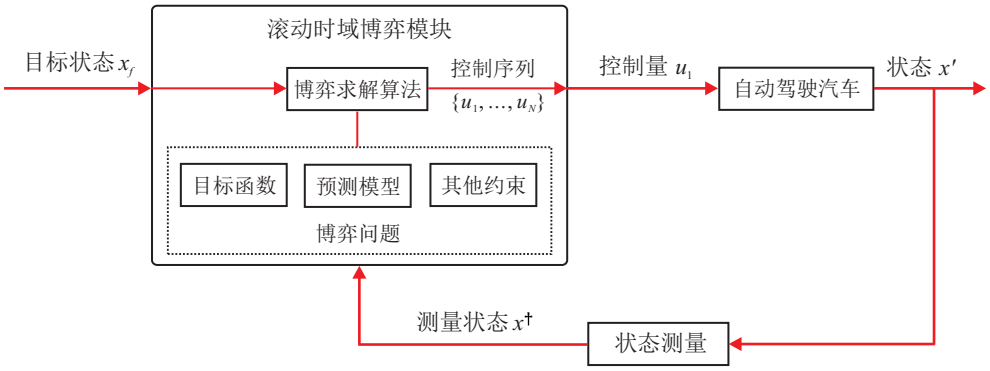

图 3 滚动时域博弈流程

博弈的收益函数一般带有偏离目标状态的惩罚项,从而使车辆尽可能靠近目标状态. 滚动时域博弈的目标是求解未来多个离散时间步的博弈均衡,因此博弈具有预测模型. 预测模型根据t时刻的状态x_t和决策s_t计算$t+1$时刻的车辆状态,记为

$$x_{t+1}=f(x_t,s_t). \tag{3}$$

决策是车辆的控制量,预测模型就是车辆的运动学或动力学模型. 博弈求解算法会求解出所有参与者的预测收益的最优决策序列$(s_1,\ldots,s_N)$,包括从当前时刻s_1到预测终止时刻s_N的决策. 在下一时刻,博弈中所有参与者的状态都被重新测量,然后以测量状态作为新的初始状态重新求解,这种设计给开环的纳什均衡带来了状态反馈. 滚动时域博弈的过程是循环迭代的,预测时域始终是从当前时刻到未来时刻的一段时间,以保证一直对未来有预测效果.

1.2 迭代微分博弈

1.2.1 微分博弈定义

微分博弈问题可以被理解为有多个参与者的最优控制问题,或者把最优控制问题看作是只有一个参与者的微分博弈问题. Isaacs将博弈论引入飞机拦截问题,该问题后来被称为微分博弈中的追逃问题[29-31]. Isaacs认为追逃双方在矛盾中会采取博弈的均衡策略, 1965年,他将研究工作整理后发表于同名著作《微分博弈》中. 早期微分博弈研究的焦点都停留在零和微分博弈领域内,随着管理学、运筹学、工学和经济学的不断发展,研究人员将其推广到了非零和的情况,如Başar等[32]出版了与此相关的书籍,Engwerda[33]的著作专门研究了二次微分博弈.

微分博弈考虑的是一段共同时间内的动态博弈. 通常情况下,智能体共同控制由微分方程描述的动力系统,即

$$\dot{x}=f(t,x,u_t^{1:N}). \tag{4}$$

其中: $x\in\boldsymbol{R}^n$是系统的状态,$u_t^i\in\boldsymbol{R}^{m_i}$, $i\in[N]\equiv\{1,2,\ldots,N\}$是智能体的输入控制变量,$u_t^{1:N}\equiv(u_t^1,u_t^2,\ldots,u_t^N)$.

在博弈的时间范围内,每个智能体都希望优化一个特定的目标函数

$$J_i\triangleq\int_0^T g_i(t,x_t,u_t^{1:N})\mathrm{d}t,\ \forall i\in[N]. \tag{5}$$

智能体利用获取的状态信息和其他智能体的行动信息,通过优化目标函数来确定其行动. 如果只有两个参与者,它们的目标函数求和为零,则称为零和微分博弈;对于其他情况,则为非零和博弈. 智能驾驶决策问题通常被考虑为一种非零和博弈.

1.2.2 微分博弈的信息结构与策略

要解决一个微分博弈问题,不仅需要状态方程和目标函数,还需要系统的状态信息. 常用的信息结构有两种,即开环信息结构和状态反馈信息结构. 设

$v(t)$为时刻t智能体可使用的信息,开环信息结构满足$v(t)=\{x_0,t\}$,即可用信息为当前时刻和初始状态.与开环不同,状态反馈可以获取当前系统的状态,信息结构满足$v(t)=\{x_t,t\}$. 对于一般和的微分博弈,两种不同的信息结构通常会导致不同的微分对策[34].

策略是一种将智能体的动作与可用信息相关联的规则. 对于上述的开环和闭环两种信息结构,可以产生两种不同的策略. 开环策略根据$u_t^j=\gamma^j(x_0,t)$选择控制,状态反馈策略则根据$u_t^j=\gamma^j(x_t,t)$选择策略,文献[35]详细地讨论了两种策略性质的差异. 如果使用开环策略,则意味着在初始时刻智能体已经将之后时刻的轨迹固定,即在每个时刻的控制是预定的. 如果使用状态反馈策略,则决策对系统状态的反应是预先确定的. 开环策略的优点是求解快,但是如果状态受到干扰或者噪声,则状态反馈策略会明显优于开环策略. 其次,开环策略仅能够保证在整个时段为一个纳什均衡. 状态反馈策略计算出的结果不仅在整个时段是一个纳什均衡,而且每一时刻都是一个纳什均衡,也就是子博弈完美纳什均衡,因此是一种更为精炼的均衡.

1.3　马尔可夫博弈

1.3.1　马尔可夫决策过程

近年来,强化学习在解决各类动态决策问题中取得了巨大成功,例如游戏、自然语言处理等领域[36-37]. 强化学习依赖于由Bellman[38-39]开创的马尔可夫决策过程(MDP). 智能体通过与环境交互执行一系列顺序的决策,通过反复试验和学习找到最佳策略,以使其长期的回报最大化. 马尔可夫决策过程由一个5元组M组成,即

$$M=\langle S,A,P,R,\gamma\rangle. \tag{6}$$

其中: S表示环境的状态空间; A表示智能体的动作空间; P表示状态转移概率矩阵,即下一个状态s'关于当前状态s和当前动作a的概率分布; R表示奖励函数,即智能体的收益函数,它是关于当前状态s、当前动作a以及下一步到达状态s'的一个函数; γ是一个折扣因子,它在0～1之间取值. 单智能体与环境的交互如图4所示.

马尔可夫性意味着未来和过去的状态在当前状态下是独立的,未来的结果仅取决于当前状态和执行的行动. 一辆车的动态决策行为是具有马尔可夫性的,车辆从起点行驶到目标点,需要顺序地选择一系列动作. 当把路线的选择看作是一个顺序决策的MDP时,实现智能驾驶的关键就是赋予智能驾驶车辆产生长期驾驶策略的能力.

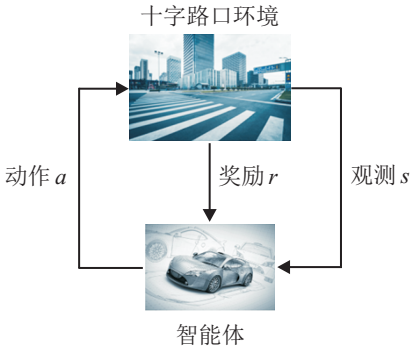

图4　单智能体与环境交互

1.3.2　局部可观测马尔可夫决策过程

在只考虑一辆智能驾驶汽车的场景下,车辆的动态具有马尔可夫性,但是,当下一个状态取决于其他道路使用者(汽车、电动车、行人)的行为时,会违背马尔可夫性[40]. 通常情况下,智能驾驶是一种多车交互的场景,车辆在进行超车、并道和转弯等操作时必须考虑场景中其他车辆的行为并采取相应行动. 采用局部可观测的马尔可夫决策过程(POMDP)[41]建模是一种解决方案. 它是处理交互不确定性的常用选择,此时不确定性被建模为潜在变量,并根据观测轨迹在线估计. 获得了根据隐藏状态分布后,仍然可以假设场景中的智能驾驶车辆具有马尔可夫性. 然而,一个具有复杂状态空间或动作空间的POMDP问题的计算复杂度非常高[42]. 因此,传统的基于POMDP的决策方法通常只考虑两辆车的交互.

1.3.3　马尔可夫博弈

一种解决多车场景的思路是将马尔可夫决策过程与博弈论结合,考虑为随机博弈[43]或马尔可夫博弈[44]. MDP在交互式环境中训练具有丰富的强化学习算法,但它不能直接应用于多智能体决策的场景; 博弈论可以分析多个智能体的行为,但其传统理论为基于模型的分析与设计,无法处理复杂动态的大范围场景. 将两者结合的多智能体博弈中,每个智能体仍然试图通过试错学习来解决顺序决策问题. 不同之处在于,环境状态的转移和每个智能体的奖励函数由所有智能体的联合行动决定. 马尔可夫博弈刻画了多个智能体的交互,是一种将MDP推广到具有竞争或合作目标的多个智能体交互的博弈论框架. 马尔可夫博弈通常由一个6元组M组成,即

$$M=\langle N,S,\{A^i\}_{i\in\{1,\dots,N\}},P,\{R^i\}_{i\in\{1,\dots,N\}},\gamma\rangle. \tag{7}$$

其中: N表示参与博弈的智能体总数,当N为1时,马尔可夫博弈退化为MDP; S表示状态空间; A^i表示第

i个智能体的动作空间;P表示状态转移概率矩阵;R^i表示第i个智能体的奖励函数;γ表示折扣因子. 多个智能体与环境的交互如图5所示.

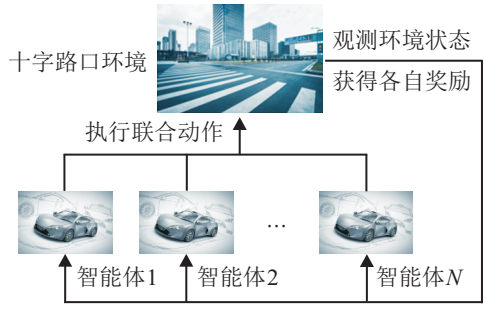

图 5 多智能体与环境交互

马尔可夫博弈通常被定义为一种非合作博弈,其中的智能体都以自身收益最大化为目标. 马尔可夫博弈通常采用纳什均衡解的概念,此时一个智能体的策略是对其他智能体策略的最佳响应. 马尔可夫策略是仅依赖于当前状态的策略. 如果除了该智能体之外的所有参与者都使用马尔可夫策略,则该智能体的最佳响应就是马尔可夫策略. 多智能体强化学习是一种高效且通用的马尔可夫博弈求解方法,可用于求解每个智能体的最优策略,解决多个智能体在共享随机环境中的顺序决策问题. 目前,在多智能体强化学习(MARL)领域[45]已经有多车决策的应用.

2 多车交互场景与核心问题

与单车决策方法研究的交通场景类似,多车智能驾驶决策研究的交通场景也可以分为路口场景、道路场景和赛车场景. 具体的场景分类如图6所示.

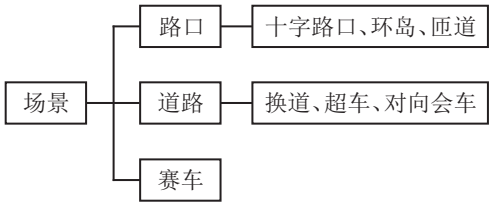

图 6 多车交互场景分类

2.1 路口场景

路口场景可以分为十字路口场景[46-51]、环岛场景[48,51]和匝道合并场景[46,51-53]. 在路口场景中,来自不同道路、不同行驶方向的车辆将会相遇,需要解决车辆相遇时的冲突问题. 十字路口场景下的车辆可以来往于4个方向,车与车之间的交互行为非常丰富. 例如,在不考虑交通信号的情况下,一辆从左往右直行的车辆在通过路口时,不仅要考虑直线行驶的车辆的行为,还要考虑来自上方车道或下方车道的车辆的行为. 对于环岛场景和闸道合并场景,需要考虑从辅道汇入主道时的车辆交互,该场景下主要研究车辆汇入主道时的效率问题,避免出现“冻结机器人”现象[46]. 近年基于博弈方法的路口场景的相关研究如表2所示.

表 2 基于博弈方法的路口场景研究

场景	相关研究
十字路口	文献[46-51,54]
环岛	文献[48,51]
匝道合并	文献[46,51-53]

2.2 道路场景

道路场景可分为换道场景[55-59]、超车场景[48,59]和对向会车场景[49-50]. 换道操作需要考虑主车从当前车道切换到两侧的车道时与前后车的交互. 换道操作在满足安全性的前提下还需要关注车辆换道时的舒适性. 在超车场景中,车辆在换道的基础上还要继续完成后续动作. 因为主车不仅要切换到两侧道路,还要安全地切换回来,所以在超车时车辆之间的安全交互需要着重考虑. 对向会车场景可视为简化的十字路口场景,直行的主车只需考虑前方驶来的车. 近年基于博弈方法的道路场景的相关研究如表3所示.

表 3 基于博弈方法的道路场景研究

场景	相关研究
换道	文献[55-59]
超车	文献[48,59]
对向会车	文献[49-50]

2.3 赛车场景

赛车场景是对常见城市交通场景的一种拓展,已经有研究将博弈论方法应用到赛车、无人机竞速等场景中[60-63]. 赛车场景可以分为车队之间的竞赛[62]与单车之间的竞赛[63]. 文献[62]研究了两个车队之间的竞赛,每一辆车都与队友合作而与对手竞争. 将博弈论方法应用到这一类比赛中,交互的车辆会产生类人的驾驶行为,通过选择适当的“战术”获得胜利. 例如,处于领先地位的车辆会故意干扰后车行驶,使其无法完成超车.

3 基于博弈的多车交互决策建模方法

针对上述场景的多车交互决策问题,可以采用不同的博弈方法进行建模. 在建模过程中,滚动时域博弈和迭代微分博弈的难点是建立目标函数与决策变量. 建立目标函数时可以考虑车辆运动学模型或动力学模型. 对于马尔可夫博弈,建模的难点则是奖励函数的设计,决策变量的建立转化为设计智能体的状态空间和动作空间.

3.1 滚动时域博弈

滚动时域博弈是重复静态博弈与模型预测控制的结合. 在每个离散时刻进行静态博弈,得出未来预测时域内的均衡解,但只采用第1个时刻的行动. 在下一时刻环境状态更新后,再次进行静态博弈. 滚动时域博弈将动态的问题转化为一系列静态的问题求解. 这种开环策略的求解比迭代微分博弈的求解更容易收敛,并且能够保证在预测时间段的策略是一个纳什均衡.

3.1.1 多因素综合的目标函数设计

从乘客的需求出发,智能驾驶决策效果的优劣可以从快速性、安全水平、舒适程度等角度评价. 现有的研究基于以上的3个因素提出了多种不同的刻画方式. 目前有两种基本的目标函数设计方法:一种是特征的线性组合,表示为$J = w^{\mathrm{T}}\theta$. 其中:J是目标函数,w是特征的权重向量,θ是特征向量. 另一种是分段函数,即在不同情况下目标函数的评价因素不同,该方式一般用于突出安全水平的影响.

快速性可以用速度、速度的增量、位置的领先程度等指标来衡量. 在安全水平的评估中,通常在安全约束被违反时激活负收益,或在分段函数中单独列出不安全条件下的目标函数,两种方法都是安全约束的软约束形式. 为保证其有效性,安全水平相关的项应当有超过其他因素的影响效果,表现为高权重、分段函数中固定的低收益或绝对值较大的惩罚项. 舒适程度与诸多因素相关,具有较高的复杂性. 文献[64]研究了乘客对不同情况下车速的偏好,结果表明,道路类型、规定车速、车流密度、天气以及对智能驾驶的信任程度都会对结果产生影响. 目前,基于博弈方法的智能驾驶研究中,舒适程度通常考虑车速的稳定性以及速度随时间的波动、加速度的倒数等指标. 此外,还会考虑与驾驶偏好或机动成功率相关的因素,例如期望的速度、机动的时机等.

文献[56]研究了自主换道问题中换道车辆和目标车道上后方车辆的决策问题,其中换道车辆的收益使用换道前后的速度增量描述,目标车道上后方车辆的收益使用加速度的倒数描述. 双方的目标函数都采用分段形式,在不安全情况下固定为一个负数,远远低于安全条件下的收益. 文献[65]研究了十字路口无保护左转车辆和对向直行车辆的决策问题,设定了避碰和舒适所需的加速度,其中存在碰撞风险时才激活避碰项. 左转车辆的目标函数中引入了与对向车辆的间距和对向车流中下一个间隙长度的比例,以描述车辆对左转时机的选择.

上述博弈的目标函数计算了单一阶段的收益,还有一些研究以预测时域内每阶段收益或代价之和作为目标函数. 文献[47]研究了智能驾驶车辆的行为对人类驾驶车辆的影响. 其中人类驾驶员的单一阶段收益由3个安全特征组成,分别是道路边界、车道线和避碰软约束. 智能驾驶车辆的单一阶段收益分为控制和影响两个部分:控制部分包括速度和避碰惩罚;影响部分由实验需求决定,可设置为目标位置、人类驾驶车辆到达的期望位置等. 文献[55]将强制换道问题中换道车辆和目标车道上附近车辆的交互决策建模为斯塔克尔伯格博弈. 目标车道上附近车辆作为跟随者,在已知换道车辆行为的条件下选择最优策略,其目标函数考虑了快速性、安全水平和舒适程度:快速性以归一化的纵向位置超前量表示;安全水平改用绝对值表示,含义由超前量变为车距;舒适程度用负的加速度平方表示. 研究中设置了一个需要估计的攻击性权重β,用于表示车辆在位置领先与安全水平之间的偏好. 文献[52]研究了匝道并线中主道和匝道车辆的决策问题,其中设置了偏离车道的惩罚,并以速度的变化量和换道次数刻画舒适程度. 另外,速度与期望速度、前车速度的误差,以及换道的剩余距离也被考虑在内. 由于建模成合作博弈,目标函数设置为参与者的代价之和.

在赛车场景中,司机关注的因素是位置的领先程度和安全水平. 文献[63]将两车竞速建模成矩阵博弈,并设计了3种博弈模式. 赛车的收益以分段函数表示. 在安全水平方面,赛车离开赛道和碰撞分别得到一个固定的低收益. 领先程度以赛道中线投影对应的已行驶长度计. 另外,在预测时域结束时处于领先的赛车会得到一个附加的收益.

3.1.2 二次型目标函数

二次型目标函数是一种在最优控制中常用的目标函数. 在轨迹优化问题中,有限时域的最小二次型目标函数的基本形式为

$$J = (x_N - x_f)^{\mathrm{T}} Q_N (x_N - x_f) + \sum_{k=0}^{N-1} (x_k - x_f)^{\mathrm{T}} Q (x_k - x_f) + u_k^{\mathrm{T}} R u_k. \tag{8}$$

其中:N为预测时域长度,x_k、u_k分别为k时刻状态和控制,x_f为目标状态,Q_N、Q为实对称半正定矩阵,R为实对称正定矩阵.

二次型目标函数中的状态可以设置为车辆运动学方程的状态,也可以设置为需要跟踪的变量,例如车速. 文献[55]对换道车辆采用以速度为状态的二次型目标函数,并增加了一个安全车距约束的松弛变

量二次项,从而引入安全车距的软约束. 文献[46]提出了一种场景通用的基于广义纳什均衡模型的轨迹规划算法. 目标函数中采用运动学方程的状态,其中包含位置、速度等变量,同样设置了车距的二次惩罚项.

3.1.3 带有敏感性项的目标函数

在赛车场景这样的竞争性场景中,遏制对手与增加自身收益有较大关联性. 为了强化策略对其他智能体收益的影响,除使用耦合安全约束以外,一些研究还在目标函数中设置了对其他智能体策略的敏感性项. 文献[60-61]研究了机器人竞速和赛车场景中的轨迹规划问题,其中目标函数为自主智能体的收益减去其他智能体均衡收益的加权和. 与文献[63]类似,收益通过沿赛道中线的进度计算,证明了添加敏感性项不会改变问题的纳什均衡. 文献[62]考虑了多个智能体分组竞速的情况. 在设计目标函数时,同组智能体的敏感性项给予正权重,不同组则给予负权重,从而使智能体能够表现出与队友合作、与对手竞争的行为.

3.2 微分博弈

微分博弈求解出的策略是一种状态反馈策略. 微分博弈的均衡在每一时刻都是一个纳什均衡,也就是“子博弈完美纳什均衡”. 因此当环境的状态受到干扰时,状态反馈策略会明显优于滚动时域博弈获得的开环策略. 微分博弈可以用于解决驾驶交互决策,文献[49]将智能驾驶交互决策建模为微分博弈.

3.2.1 状态方程

在人车博弈的场景中,最简单的方法是将行人建模为质点或独轮车模型[49],即

$$\dot{p}_{x,i} = v_i \cos(\theta_i), \tag{9}$$

$$\dot{p}_{y,i} = v_i \sin(\theta_i), \tag{10}$$

$$\dot{\theta}_i = w_i, \tag{11}$$

$$\dot{v}_i = a_i. \tag{12}$$

其中:$(p_{x,i}, p_{y,i})$为行人i的坐标位置,θ_i为偏角,v_i为速度. 输入$u_i := (w_i, a_i)$由行人的角速度和加速度组成. 对于车辆,可以将一个四轮阿克曼转向车辆简化为由两个刚性连杆连接的车轮所组成的两轮模型,将车辆建模为自行车模型[48,51,57]. 车辆的状态变量除了增加前轮转角ϕ_i和考虑了车辆的轴间距离之外,其他与独轮车模型相同. 文献[50]还考虑了将车辆建模为增广自行车模型,在自行车模型的基础上对加速度进行求导,将紧急动度作为控制变量.

上述模型都是对单辆车进行建模,而对于一个博弈问题,则需要考虑联合车辆模型,表述为N个智能体组成的非线性系统[32]

$$\dot{x} = f(t, x, u_{1:N}). \tag{13}$$

其中:x为N个智能体的状态变量,$u_{1:N}$为对应的控制变量. 此外,一个离散时间的线性微分模型可以表述为

$$x_{t+1} = A_t x_t + \sum_{j=1}^{N} B_t^j u_t^j. \tag{14}$$

3.2.2 目标函数

微分博弈的目标函数是轨迹规划决策时需要考虑的多项指标的加权,如避免碰撞、限速等. 例如十字路口场景下,智能驾驶汽车决策需要考虑以下几个方面:

1) 位置:与其他车之间的间距,与车道中心线的距离,与车道边界的距离,与终点的距离;

2) 输入:控制输入的上下界;

3) 状态:参考速度,速度上下界.

目标函数需要着重考虑安全性[66-68]. 文献[50]在路口场景的两辆相向会车问题中,给目标函数预先设定一个时间间隔,在此时间间隔内,假设其他车辆暂时分心,以便使获得的均衡轨迹对其他智能体的潜在危险行为具有鲁棒性. 将整个时间范围$[0, T]$划分为两个子区间:在对抗期间,自我车辆假定其他智能体暂时分心并有撞向自己的意愿,进而采取防御行动. 假设自我车辆为$i = 1$,对向车辆的目标函数可以定义为$g_{\text{adv},i}, i \in \{2, \dots, N\}$. 在合作期间,假设对向车辆已经恢复到“正常”或“合作”状态. 在剩余的时间范围内,以非保守的方式选择控制信号,对手车辆的目标函数可以定义为$g_{\text{coop},i}, i \in \{2, \dots, N\}$. 整体的目标函数可以定义为

$$J_i = \int_0^{T_{\text{adv}}} g_{\text{adv},i}(x, u_{1:N})\mathrm{d}t + \int_{T_{\text{adv}}}^{T} g_{\text{coop},i}(x, u_{1:N})\mathrm{d}t,\ i \in \{2, \dots, N\}. \tag{15}$$

随着T_{adv}的增大,自我车辆会假定更为敌对的对向车辆,进而采取的策略也会更保守. 安全驾驶策略在简单两车场景中会有比较好的效果,但是,在多车场景中可能会因为过于保守而出现“冻结机器人”的现象.

为了在不确定性情形下给自车生成具有安全性的行为,对风险敏感汽车的交互进行建模变得至关重要. 基于熵风险的建模将不确定性考虑进目标函数,引入决策风险的概念. 文献[47, 69]表明在确定性情形下,通过博弈建模车辆可以对其行为的影响以及其他车辆的驾驶意图进行推理. 受其启发,在不确定性

的情形下,进一步假设所有车辆在其规划中都考虑了熵决策风险. 此时在交互过程中,车辆或行人表现出风险操作的程度不仅取决于其固有的风险承受能力,还取决于其他车辆对风险的敏感程度.

由于自车对于其他车辆目标函数的不确定性,可以通过非对称责任分配进行目标函数设计. 假设车辆之间的交互仅仅是由于避免碰撞所产生的[57],则不同的避碰责任分配会产生不同的目标函数和决策行为. 这些约束可以是对称的,即所有智能体均参与避碰;也可以进行不对称的责任分配和目标函数设计,也就是说两个智能体中只有一个负任避碰. 从建模的角度来看,非对称的目标函数设计是必要的. 例如,当一辆车在高速公路上从后面接近另一辆车时,后者负责避免碰撞. 一方面,总是将避免碰撞的责任分配给自车可能是最安全的选择,但是这可能导致过度保守的行为,例如无法合并到密集的交通中;另一方面,将避碰责任分配给场景中的其他车辆也可能带来风险,如果错误地假设他车承担避碰责任,则可能会导致自车的危险驾驶行为. 为了使用非对称约束给他车带来的不确定性,可以在两个车辆之间插入约束责任. 考虑动态反馈博弈中的两个车辆,将这两个行动者表示为P_1和P_2,其中P_1为自我车辆. 原始目标函数只依赖于与自身车辆相关的控制和状态变量,即

$$L^1 := \int_0^T g_1(t, x_1, u_1)\mathrm{d}t, \tag{16}$$

$$L^2 := \int_0^T g_2(t, x_2, u_2)\mathrm{d}t. \tag{17}$$

将P_2的目标副本添加到P_1的目标函数中,并通过该假设的赔率进行加权,有

$$J_1 := L^1 + \frac{p}{1-p}L^2, \tag{18}$$

加权项被称为礼貌项. P_2的决策只取决于自身的状态和控制变量. 对于P_1来说,需要同时考虑决策使得P_2的目标得到优化,因此要求P_1考虑P_1与P_2之间的碰撞约束. 如果p很大(大概率发生),礼貌项占优,则P_1对碰撞负主要责任;如果p很小(小概率发生),则P_2对碰撞负主要责任.

3.3 马尔可夫博弈

基于马尔可夫博弈的多车交互决策建模的难点是奖励函数的设计和智能体的状态空间与动作空间设计. 奖励函数往往采用多个性能指标的加权以体现自动驾驶车辆在具体场景中的任务. 基于强化学习的马尔可夫博弈决策方法在处理问题的规模上具有很强的可伸缩性,其不仅可以处理多车交互决策问题,甚至能够解决大规模车辆的交互决策问题. 滚动时域博弈与迭代微分博弈都是基于计算的方法,尽管两者求解出的策略都具有很强的可解释性,但通常只能在少数几辆车的场景下实现高效计算,当车辆规模继续增大至数十辆车时,这种基于计算的方法很难得到理想的效果.

3.3.1 动作空间

车辆动作空间的设置通常与场景中车辆的任务有关. 车辆动作主要分为纵向动作和横向动作. 纵向动作包括加速度、刹车等在纵向上影响车辆动态的行为;横向动作包括左转、右转等在横向上影响车辆动态的行为. 为了简化决策复杂性,动作空间通常可以只考虑纵向的动作[70-73]. 简单任务下的车辆路线是固定的,车辆只需要控制在路线上的行驶速度,即控制加速度或油门[54,72]. 文献[70-71]通过对汽车行驶过程进行分析,增加了车道保持的动作,将汽车的运动分为加速、减速和巡航3种形式. 文献[73]中除了上述3种动作外还增加了停车的动作,以保障车辆的安全.

更复杂的任务下,动作空间需要同时考虑纵向动作和横向动作[58-59,74-77]. 文献[76]中车辆的动作有3类:左转、右转和车道保持. 左转分为正常左转、加速左转和减速左转,右转分为正常右转、加速右转和减速右转,所以动作空间中包含7个动作. 文献[77]除了定义横向和纵向上的动作,还定义了主动避免碰撞的动作以增加车辆行驶的安全性. 文献[59]将车辆的动作空间定义为$A = [a_1, a_2]^{\mathrm{T}}$,$a_1$表示车辆的转向角度输入,$a_2$表示加速度. 所以此时车辆在横向上的动作为转向角度的连续输入,通过采用连续输入动作使得车辆的转向行为更加丰富.

3.3.2 状态空间

车辆获取状态空间信息后才能做出相应的动作. 状态空间通常包括车辆的自身状态和环境状态. 自身状态包括车辆的位置、转向角和速度大小等信息. 环境信息通常包括其他车辆状态、障碍物的位置和交通灯信息.

通常情况下,为了简化问题,状态空间只包括车辆的位置信息和速度信息[54,71,74-76],所以状态空间表示为$X = [x, y, v]^{\mathrm{T}}$,$x$和$y$分别表示车辆的横纵坐标位置,$v$表示车辆的速度. 文献[58]中的状态空间包括车辆的位置信息和场景中可以换道的空隙. 文献[59, 72]将状态空间表示为$X = [x, y, \theta, v]^{\mathrm{T}}$,所以状态空间除了包括车辆的位置信息和速度信息,还包括了车辆的转向角大小. 文献[72]中的状态空间还附加了终端状态的信息,它可以有效地防止系统得出错

误的结论. 例如,碰撞后可能达到目标,或者可以多次达到目标.

为了适应场景中特殊的任务,还可以设置更复杂的状态空间. 文献[73]将状态空间分为两类:一类是可以由车载摄像头直接观测到的信息;一类是根据前车的状态定义的,例如前车正在加速或刹车. 文献[70]将状态空间细分为车辆状态和道路状态,车辆状态包括车辆的位置信息(x,y)、速度v、偏转角θ、平均角速度$a_{\rm ave}$、平均加速度$w_{\rm ave}$,道路状态包括交通灯的状态L、持续时间t、前方停车线的位置$(\rm lon, lat)$. 所以,状态空间可表示为

$$X=[x,y,v,\theta,a_{\rm ave},w_{\rm ave},L,t,\ \rm lon,\ lat\ ]^{\rm T}. \tag{19}$$

3.3.3 奖励函数

奖励函数的设置通常包括安全性、高效性和舒适性3方面指标. 安全性是首先考虑的指标,要求车辆之间不发生碰撞并且不与车道边缘或其他障碍物发生碰撞. 高效性要求车辆尽快通过,在最短时间内完成任务. 舒适性要求车辆在行驶过程中的速度变化不能太大,即加速度的绝对值要尽量小.

文献[59, 74]首先考虑避免车辆碰撞、避免驶出车道,还考虑了车辆要尽量行驶在车道线上,即接近车道中心. 文献[59]中的奖励函数被参数化为多个特征的线性组合,可以通过逆强化学习(IRL)的方法得到该参数化的奖励函数,这几个特征包括:

1) $\Phi_1\propto c_1{\rm e}^{-c_2d^2}$:车辆到道路边界的距离. 其中:$d$是车辆与道路边界之间的距离,$c_1$和$c_2$是比例因子.

2) Φ_2:车辆到车道中间的距离,其中规定的功能类似于Φ_1.

3) $\Phi_3=(v-v_{\max})^2$:快速通行. 其中:v是车辆的速度,$v_{\max}$是速度限制.

4) $\Phi_4=\beta_H\boldsymbol{n}$:车辆行驶方向沿道路行驶. 其中:$\beta$是车辆的行驶方向,$\boldsymbol{n}$是沿道路方向的法向量.

5) Φ_5:碰撞约束,避免车辆发生碰撞.

此外,车辆行驶过程中的高效性也得到了考虑[58,70,73,77]. 文献[58]中奖励函数既考虑了纵向的安全,又考虑了横向换道的安全与效率. 文献[73]在保证安全性的前提下,考虑了最小化时间损失,时间损耗最少的行动对应于更高的奖励. 智能驾驶车辆行驶过程中舒适度的研究越来越多[54,70-72,75-77]. 在奖励函数中,舒适度体现为行驶过程中车辆的角度变化平缓或加速度变化平缓,或者行驶过程中角度变化的总和或加速度变化总和尽量小. 在奖励函数中还可以对超速的情况进行惩罚以保证安全舒适[71]. 文献[72, 75-76]中奖励函数的设置除了保障车辆的安全与舒适,还考虑了行驶过程中的经济性,即车辆的驱动输入尽可能小. 上述研究都是在无交通信号下考虑的,在具有交通信号灯的场景中还需要在奖励函数中考虑交通规则带来的激励[54,70,72].

4 不同信息结构下的博弈求解算法

求解滚动时域博弈和迭代微分博弈需要计算纳什均衡,往往采用结合运筹控制理论的计算方法. 滚动时域博弈是对全时段的纳什均衡进行求解,而迭代微分博弈需要求解从每一个时刻开始的纳什均衡. 迭代微分博弈的解更精炼,计算需要的时间也更长. 求解马尔可夫博弈通常采用无模型的强化学习的方法.

4.1 滚动时域博弈

对于具有较小行动集合空间的滚动时域博弈,常用的求解算法是穷举法和搜索算法. 文献[55]提出了一种强制换道场景下的斯塔克尔伯格博弈模型,其中换道车辆作为领导者采用最大最小策略,周边车辆根据领导者行为采取最大收益策略. 博弈的行动集合为一系列给定的换道时间和期望的车头时距系数,优化问题通过二次规划求解. 文献[63]使用基于生存性理论的路径规划模型[78]生成路径树并构成博弈中的行动集合. 为求解纳什均衡或斯塔克尔伯格均衡,基于生存核或识别核的剪枝算法被用于排除不可行轨迹,从而缩小解的空间. 该博弈模型的前提是参与者相互已知路径树,或已知路径规划模型和剪枝算法. 文献[79]提出一种用于无信号灯十字路口的类似斯塔克尔伯格博弈的静态博弈模型,其中领导者对跟随者的唯一优势是对最优策略的认知. 参与者寻求最大最小策略,其中领导者仅考虑跟随者的最优策略,而跟随者考虑领导者的全部策略. 由于博弈的行动集合是一组给定的加速度,使用了树搜索算法求解最大最小策略. 文献[80]也将树搜索作为一种求解K级博弈论中最优决策的基本算法.

对于具有较大行动空间或无限的行动集合的滚动时域博弈,一般需要使用基于优化的算法求解. 迭代最优响应(IBR)是一种被广泛使用的优化框架. IBR算法首先初始化全部参与者的策略s_0,然后在固定其他参与者策略s^{-i}的条件下,依次求解每个参与者i的最优策略s^i,直到所有策略收敛到均衡解. 这里上标表示参与者,下标表示迭代次数. IBR算法在细节上允许不同的实现方式,例如初始策略s_0在此类研究中一般为预测时域内的控制量序列,其生成因预测模型而异. 在对所有参与者的一轮迭代中,参与者的策略可以是串行更新的,即求解s^i_{k+1}时,其他参与

者的策略为

$$s^{-i}=\begin{cases}s_{k+1}^{j},\ j<i;\\ s_{k}^{j},\ j>i.\end{cases}\tag{20}$$

IBR迭代也可以是并行更新的,即固定取上一轮迭代的策略s_k^{-i}. 在求解最优策略时采用的优化方法构成了IBR的核心,可以根据问题的性质设计. 此外还规定了迭代次数上限,以防算法长时间不收敛. 文献[63]提出可使用IBR算法[81]求解合作博弈的情况. 文献[52]提出了一种强制换道的博弈模型,使用庞特里亚金最大化原理给出了纳什均衡的必要条件,并用IBR算法求解. 文献[61]对目标函数中的敏感性项使用一阶泰勒展开,再通过化简拉格朗日函数,从而避免嵌套优化,然后使用IBR算法求解. 使用敏感性项的研究均采用相同的信息结构和避免嵌套优化的方法,仅在IBR框架下选择了不同的优化方式. 文献[62]基于KKT条件,使用高斯-赛德尔迭代法求解优化问题;而文献[60]使用SQP算法进行求解.

针对更大规模行动空间的问题可以采用层次化的决策模型和方法. 文献[82]提出了一种用于轨迹规划的分层博弈模型. 战略层进行闭环斯塔克尔伯格博弈,智能驾驶车辆为领导者,人类驾驶车辆为跟随者. 人类被建模为不完全理性个体并采用玻尔兹曼策略. 战略层使用动态规划计算状态价值函数和最优策略,状态价值函数作为战术层的终端奖励项,可存储为查找表以提高效率. 战术层进行开环轨迹优化,使用OWL-QN算法[83]计算最优策略,并使用IBR算法解决嵌套优化问题. 战术层的求解可以实时运行.

基于IBR的算法的每次迭代嵌套了一个子优化问题,但是也可以采用无需嵌套优化的算法. 作为文献[82]的先前工作,除了使用最大熵原理描述人类的不完全理性因素,文献[47]使用的博弈设置与文献[82]的战略层几乎相同. 其使用的求解算法[83]也是相同的,但未使用IBR框架. 此研究中,智能驾驶车辆具有对人类的奖励函数预测模型,可以通过离线的逆强化学习获得. 文献[46]提出的广义纳什均衡求解算法,使用增广拉格朗日函数法处理约束,其中为不等式约束设置了二次罚函数项. 依据纳什均衡的一阶必要条件,增广拉格朗日函数的梯度在最优点处为零,这与等式约束构成了一个寻根问题,而不等式约束使用二次罚函数项隐式处理. 转化的寻根问题使用拟牛顿法结合线搜索回退法求解. 算法能在10^2次迭代后收敛到纳什均衡解,远远低于文献[84]中算法的$10^3\sim10^4$迭代次数,求解速度是iLQG算法[49]的3倍.

4.2 迭代微分博弈

大多数微分博弈无法获得解析解,而许多数值计算方法会出现“维度灾难”的情况[85]. 对于线性二次结构,文献[32]证明了无论是开环信息结构还是状态反馈信息结构,都可以显式地计算出微分博弈纳什均衡. 类似于最优控制中的开环求解方法,通过构建恰当的哈密顿量,开环信息的线性二次博弈也可以使用Pontryagin极小值原理进行求解. 但是,由于通常需要寻找一个状态反馈的解决方案,Pontryagin极小值原理就不再适用,取而代之的是使用动态规划的方法进行纳什均衡求解.

基于线性二次微分博弈,智能驾驶决策微分博弈的重要算法是iLQG算法,该算法是伯克利大学的Tomlin团队针对非零和动态微分博弈问题而提出的可以求解局部纳什均衡的算法. 该算法借鉴了iLQR算法的基本思想[86-87],其从一个初始轨迹出发,在轨迹附近对状态方程进行一阶泰勒展开,对目标函数进行二阶泰勒展开,将增量问题转变成具有线性二次结构的微分博弈进行求解,并进行轨迹更新. 该算法可以求解局部的纳什均衡,避免了搜索全局纳什均衡时出现的维数灾难. iLQG算法首先利用初始状态x_0和初始的控制策略,通过状态方程获取轨迹. 在当前轨迹附近对状态方程进行线性化. 对于任意状态x_t和控制u_t^j,定义偏差$\delta x_t=x_t-\hat{x}_t$和$\delta u_t^j=u_t^j-\hat{u}_t^j$,进而获得以偏差量为变量的离散时间线性系统近似值以及智能体的目标函数的二次近似值. 使用状态反馈博弈算法对离散线性二次博弈问题进行求解,可得到一组控制策略

$$\tilde{\gamma}_i^k=\hat{u}_t^i-P_t^{i^k}\delta x_t-\alpha_t^{i^k}.\tag{21}$$

其中:$P_t^{i^k}\in\boldsymbol{R}^{m_i\times n},\alpha_t^{i^k}\in\boldsymbol{R}^{m_i}$. 类比于iLQR方法,利用$\tilde{\gamma}_i^k$更新控制策略,有

$$\tilde{\gamma}_i^k=\hat{u}_t^i-P_t^{i^k}\delta x_t-\eta\alpha_t^{i^k}.\tag{22}$$

该算法每次迭代的计算复杂度与iLQR算法相当,其中线性化状态方程都需要计算偏导数,计算复杂度为$\mathcal{O}(n^2)$,对目标函数进行二次化需要偏导数计算,复杂度为$\mathcal{O}(Nn^2)$. 求解线性二次博弈的耦合Riccati方程的复杂度为$\mathcal{O}(N^3n^2)$.

iLQG算法将道路边界约束、速度约束等加入目标函数中,作为软约束处理. 如果参数设计不当,则很容易出现违反约束的问题. 也可将以上约束作为硬约束处理,这样就需要求解微分博弈的广义纳什均衡[28]. 文献[88]将反馈纳什均衡推广到智能体的状

态和输入受约束的情况,提出了一种序列线性二次博弈求解方法.

4.3 马尔可夫博弈

Littman提出将马尔可夫博弈作为不确定环境下多智能体决策的数学框架,并提出了一种Minimax-Q学习算法来求解二人零和博弈的最优策略[44]. 算法中将经典的Q学习算法更新步骤中的“max”算子替换为“minimax”算子,所以Minimax-Q学习算法是对单智能体Q学习算法的一种改进. Littman的这一开创性工作启发了很多后续的多智能体强化学习算法. 例如Littman[89]提出了团队Q学习以处理另一种特殊的马尔可夫博弈,称为团队马尔可夫博弈,其中智能体共享相同的奖励函数,即具有相同的目标. 这项工作指出,Minimax-Q和Team-Q可以被视为一种更为通用的算法的两种特殊情况,称为Nash-Q算法,因为这两种算法都是在合作或竞争场景中用Nash均衡值更新Q函数. 同年,Littman[90]将它们统一为Friend-or-Foe Q(FFQ)学习算法,以处理合作均衡和非合作均衡共存的问题. FFQ是一种能解决多智能体一般和博弈的学习方法,在这种方法中,学习者被告知将其他智能体视为朋友或敌人. 与上述其他学习规则相比,FFQ学习方式的算法可以提供更好的收敛性保证. Hu等[91]提出了一种多智能体Q学习算法,将minimax-Q学习算法从零和博弈扩展到多智能体一般和博弈. 后来他们在文献[92]中又提出了一个同样称为Nash-Q的算法,并证明了该Nash-Q算法的收敛性,即学习过程中的每个阶段的博弈都有一个全局最优解或一个鞍点,并且总是选择均衡来更新Q函数. 以上的学习方法已被用于求解基于马尔科夫博弈建模的多车决策问题.

5 仿真实验与测试方法

智能驾驶测试与评估体系也随着研究的不断深入而逐渐完善[93]. 基于博弈方法的智能驾驶决策算法的仿真测试方法分为3类:全仿真、人机混合仿真、实车测试. 全仿真的测试依赖于智能驾驶仿真平台,是验证算法最基本的方法. 在全仿真的基础上可以进行人机混合的仿真或实车测试. 基于博弈论的多车驾驶决策研究的测试方法如表4所示.

表4 基于博弈论的多车驾驶决策研究的测试方法

场景	相关研究
全仿真	[46-49, 51-52, 55-58, 66, 75, 94]
人机混合	[47, 49, 55, 58, 75]
实车测试	[60, 94]

5.1 全仿真测试

全仿真测试是最为基本的测试要求. 在智能驾驶的背景下,研究人员已经开发了丰富并各具特色的仿真平台[95-98],通过这些平台可以快速地开展决策规划模块的测试研究. CARLA[96]是一个用于智能驾驶研究的开源模拟器,对环境进行了真实渲染并提供了强大的API,允许用户控制与模拟交通场景,包括交通生成、行人行为、天气、传感器等. 文献[71]在CARLA中设置了3种十字路口场景,每个场景下车辆以不同的博弈方法决策. 每一个场景都与传统的方法进行对比后发现,所提出的算法更具有高效性与舒适性. 文献[76]在虚拟交通世界中进行测试,通过在Simulink上运行的VR Sink 3D动画环境进行可视化演示. 所有车辆都行驶在车道中间,对奖励函数做出恰当修改后可以影响车辆换道的行为,例如使智能驾驶车辆总是右转超车或者左转超车,形成独特的偏好.

5.2 混合仿真测试

人机混合仿真也被称为人在环路测试,成为智能驾驶车辆决策规划测试的热点. dSPACE仿真器是一个高保真驾驶模拟器,提供多个模块用于测试和验证智能驾驶算法,文献[55, 58]在dSPACE仿真器中完成了人在环路测试. 文献[55]提出了一种具有攻击性估计的滚动时域博弈决策算法以处理多辆周边车辆存在时的强制换道问题. 人在环路的结果表明,无论是智能驾驶车辆还是人类驾驶车辆,所提出的算法都能够通过正确评估周围车辆的攻击性来安全完成换道. 文献[58]在交通场景中加入了由人类驾驶的车辆,在这种混合交通场景下,智能驾驶车辆可以找到与人类换道行为一致的安全间隙,并且顺利完成换道超车,从而验证了所提出的类人驾驶决策算法. 文献[75]在三维仿真交通流中进行了测试,该环境能够精确模拟传感器数据和车辆物理特性. 测试环境允许研究人员手动驾驶其他车辆,从而保证接近自然的人类驾驶,除主车外的其他车辆由人类通过方向盘手动控制. 实验中智能驾驶车辆表现出了合理的行为. 例如,在适当的距离内跟随速度较慢的车辆,直到迎面而来的车辆通过,以便能够平稳快速地完成后续超车动作. 如果另一辆车突然转向右车道,则会在超车时减速以便能够做出反应.

5.3 实车测试

目前,基于博弈的智能驾驶决策方法进行实车测试的研究还比较缺乏. 文献[60, 94]在真实的交通场景中验证了所提出的算法. 文献[60]提出了一种非

线性的滚动时域博弈算法,用于智能驾驶汽车与其他汽车在竞争性的场景中进行路线规划. 首先在3辆汽车的全仿真环境中验证了该算法,然后在两辆小型的工程车上进行测试,最后在全尺寸的实车上与其他车辆的比赛中进行了验证. 文献[94]首先使用虚拟试驾(VTD)进行虚拟仿真,然后在真实的城市道路场景中测试,验证了所提出的城市环境智能驾驶汽车车道变换的博弈决策方法.

6 总结与展望

近年来,智能驾驶系统技术取得了重大的进展. 车辆的控制技术日渐成熟,车辆感知能力也随着深度学习技术的突破而迅速提升[99]. 但是,由于车与车之间甚至混合交通场景下车与人之间交互的复杂性和不确定性,智能驾驶系统的决策规划依旧具有挑战性. 本文对基于博弈模型与方法的智能驾驶交互决策的研究工作进行了梳理和介绍.

目前,本领域的研究仍有大量的问题需要解决. 人类驾驶车辆和行人运动行为具有随机性和非完全理性[18],多车混行下复杂的信息结构对基于博弈的决策算法带来了多样的挑战. 人车混行下参与者的交通意图极其复杂,在人类有限理性和复杂意图下的博弈决策需要进一步研究. 随着V2X和车联网等技术的发展,需要进一步研究博弈均衡的分布式计算和车路云融合计算算法. 车路云协同可以解决感知范围受限和计算资源不足的问题,提供更加完整的环境信息,有助于智能驾驶汽车路径规划. 这方面也有一些相关工作值得借鉴[100-108]. 马尔可夫博弈通常使用多智能体深度强化学习算法求解[109-110],已经被应用于智能驾驶决策,但算法的可解释性、泛化性能和鲁棒性还需要继续提升. 交互环境下自动驾驶的公共数据集还远远不足,难以支撑大规模交通场景下多车驾驶策略学习.

参考文献(References)

[1] Behere S, Törngren M. A functional reference architecture for autonomous driving[J]. Information and Software Technology, 2016, 73: 136-150.

[2] Feng J Y. Technical change and development trend of automatic driving[C]. The 2nd International Conference on Computing and Data Science. Stanford, 2021: 319-324.

[3] 熊璐, 康宇宸, 张培志, 等. 无人驾驶车辆行为决策系统研究[J]. 汽车技术, 2018(8): 1-9.
(Xiong L, Kang Y C, Zhang P Z, et al. Research on behavioral decision-making system of unmanned vehicles[J]. Automobile Technology, 2018(8): 1-9.)

[4] Huang Y X, Chen D S. Research progress of automatic driving path planning[C]. The 2nd International Conference on Artificial Intelligence and Computer Engineering. Hangzhou, 2022: 95-99.

[5] Parulekar M, Padte V, Shah T, et al. Automatic vehicle navigation using Dijkstra's Algorithm[C]. 2013 International Conference on Advances in Technology and Engineering. Mumbai, 2013: 1-5.

[6] Liu L S, Lin J F, Yao J X, et al. Path planning for smart car based on dijkstra algorithm and dynamic window approach[J]. Wireless Communications and Mobile Computing, 2021, 2021: 1-12.

[7] 王洪斌, 尹鹏衡, 郑维, 等. 基于改进的A*算法与动态窗口法的移动机器人路径规划[J]. 机器人, 2020, 42(3): 346-353.
(Wang H B, Yin P H, Zheng W, et al. Mobile robot path planning based on improved A* algorithm and dynamic window method[J]. Robot, 2020, 42(3): 346-353.)

[8] Tu K B, Yang S S, Zhang H, et al. Hybrid A* based motion planning for autonomous vehicles in unstructured environment[C]. 2019 IEEE International Symposium on Circuits and Systems. Sapporo, 2019: 1-4.

[9] Kuwata Y, Teo J, Fiore G, et al. Real-time motion planning with applications to autonomous urban driving[J]. IEEE Transactions on Control Systems Technology, 2009, 17(5): 1105-1118.

[10] Noreen I, Khan A, Habib Z. Optimal path planning using RRT* based approaches: A survey and future directions[J]. International Journal of Advanced Computer Science and Applications, 2016, 7(11): 97-107.

[11] 阮晓钢, 郭威, 黄静, 等. 机器人信息增益RRT环境探索算法[J]. 控制与决策, 2021, 36(11): 2683-2689.
(Ruan X G, Guo W, Huang J, et al. Robot RRT based on information gain for environment exploration[J]. Control and Decision, 2021, 36(11): 2683-2689.)

[12] 闫尧, 李春书, 唐风敏. 基于五次多项式模型的自主车辆换道轨迹规划[J]. 机械设计, 2019, 36(8): 42-47.
(Yan Y, Li C S, Tang F M. Lane-changing trajectory planning of the autonomous vehicle based on the quintic polynomial model[J]. Journal of Machine Design, 2019, 36(8): 42-47.)

[13] 陈成, 何玉庆, 卜春光, 等. 基于四阶贝塞尔曲线的无人车可行轨迹规划[J]. 自动化学报, 2015, 41(3): 486-496.
(Chen C, He Y Q, Bu C G, et al. Feasible trajectory generation for autonomous vehicles based on quartic Bẽzier curve[J]. Acta Automatica Sinica, 2015, 41(3): 486-496.)

[14] 吴伟, 刘洋, 刘威, 等. 自动驾驶环境下交叉口车辆路径规划与最优控制模型[J]. 自动化学报, 2020, 46(9): 1971-1985.
(Wu W, Liu Y, Liu W, et al. A novel autonomous vehicle trajectory planning and control model for connected-and-autonomous intersections[J]. Acta Automatica

Sinica, 2020, 46(9): 1971-1985.)

[15] 任秉韬, 王淅淅, 邓伟文, 等. 基于混合A*和可变半径RS曲线的自动泊车路径优化方法[J]. 中国公路学报, 2022, 35(7): 317-327.
(Ren B T, Wang X X, Deng W W, et al. Path optimization algorithm for automatic parking based on hybrid A* and reeds-shepp curve with variable radius[J]. China Journal of Highway and Transport, 2022, 35(7): 317-327.)

[16] Wang Z P, Li G B, Jiang H J, et al. Collision-free navigation of autonomous vehicles using convex quadratic programming-based model predictive control[J]. IEEE/ASME Transactions on Mechatronics, 2018, 23(3): 1103-1113.

[17] 魏民祥, 滕德成, 吴树凡. 基于Frenet坐标系的自动驾驶轨迹规划与优化算法[J]. 控制与决策, 2021, 36(4): 815-824.
(Wei M X, Teng D C, Wu S F. Trajectory planning and optimization algorithm for automated driving based on Frenet coordinate system[J]. Control and Decision, 2021, 36(4): 815-824.)

[18] 胡宏宇, 刁小桔, 高菲, 等. 自动驾驶汽车-行人交互研究综述[J]. 汽车技术, 2021(9): 1-9.
(Hu H Y, Diao X J, Gao F, et al. Review of interaction between autonomous vehicles and pedestrians[J]. Automobile Technology, 2021(9): 1-9.)

[19] Wang N N, Wang X, Palacharla P, et al. Cooperative autonomous driving for traffic congestion avoidance through vehicle-to-vehicle communications[C]. 2017 IEEE Vehicular Networking Conference. Turin, 2017: 327-330.

[20] Kessler T, Knoll A. Cooperative multi-vehicle behavior coordination for autonomous driving[C]. 2019 IEEE Intelligent Vehicles Symposium. Paris, 2019: 1953-1960.

[21] Zhang D H, You X M, Liu S, et al. Dynamic multi-role adaptive collaborative ant colony optimization for robot path planning[J]. IEEE Access, 2020, 8: 129958-129974.

[22] Belgioioso G, Grammatico S. Semi-decentralized Nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions[J]. IEEE Control Systems Letters, 2017, 1(2): 400-405.

[23] Lei J L, Shanbhag U V. Stochastic Nash equilibrium problems: Models, analysis, and algorithms[J]. IEEE Control Systems Magazine, 2022, 42(4): 103-124.

[24] Liang S, Yi P, Hong Y. Distributed Nash equilibrium seeking for aggregative games with coupled constraints[J]. Automatica, 2017, 85: 179-185.

[25] Yi P, Pavel L. An operator splitting approach for distributed generalized Nash equilibria computation[J]. Automatica, 2019, 102: 111-121.

[26] Lei J L, Shanbhag U V, Pang J S, et al. On synchronous, asynchronous, and randomized best-response schemes for stochastic Nash games[J]. Mathematics of Operations Research, 2020, 45(1): 157-190.

[27] Tang Y T, Yi P, Zhang Y Q, et al. Nash equilibrium seeking over directed graphs[J]. Autonomous Intelligent Systems, 2022, 2(1): 7.

[28] Facchinei F, Kanzow C. Generalized Nash equilibrium problems[J]. 4OR, 2007, 5(3): 173-210.

[29] Isaacs R. Games of pursuit[M]. Santa Monica: RAND Corporation, 1951: 1-14.

[30] Isaacs R. Differential games I: Introduction[M]. Santa Monica: RAND Corporation, 1954: 18-20.

[31] 刘坤, 郑晓帅, 林业茗, 等. 基于微分博弈的追逃问题最优策略设计[J]. 自动化学报, 2021, 47(8): 1840-1854.
(Liu K, Zheng X S, Lin Y M, et al. Design of optimal strategies for the pursuit-evasion problem based on differential game[J]. Acta Automatica Sinica, 2021, 47(8): 1840-1854.)

[32] Baar T, Olsder G J. Dynamic noncooperative game theory[M]. Philadelphia: SIAM, 1998: 17-76.

[33] Engwerda J. LQ dynamic optimization and differential games[M]. Hoboken: John Wiley & Sons, 2005: 359-421.

[34] Baar T, Haurie A, Zaccour G. Nonzero-sum differential games[C]. Handbook of Dynamic Game Theory. Cham: Springer International Publishing, 2018: 61-110.

[35] Starr A W, Ho Y C. Further properties of nonzero-sum differential games[J]. Journal of Optimization Theory and Applications, 1969, 3(4): 207-219.

[36] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1): 2094-2100.

[37] Sharma A R, Kaushik P. Literature survey of statistical, deep and reinforcement learning in natural language processing[C]. 2017 International Conference on Computing, Communication and Automation. Greater Noida, 2017: 350-354.

[38] Bellman R. Dynamic programming and Lagrange multipliers[J]. Proceedings of the National Academy of Sciences of the United States of America, 1956, 42(10): 767-769.

[39] Bellman R. Introduction to the mathematical theory of control processes: Linear equations and quadratic criteria[M]. Netherlands: Elsevier, 2016: 101-135.

[40] Shou Z, Chen X, Fu Y, et al. Multi-agent reinforcement learning for Markov routing games: A new modeling paradigm for dynamic traffic assignment[J]. Transportation Research — Part C: Emerging Technologies, 2022, 137: 103560.

[41] Cassandra A R. A survey of POMDP applications[C]. Working Notes of AAAI 1998 Fall Symposium on Planning with Partially Observable Markov Decision Processes. Orlando: AAAI, 1998: 1724.

[42] Monahan G E. State of the art — A survey of partially

observable Markov decision processes: Theory, models, and algorithms[J]. Management Science, 1982, 28(1): 1-16.

[43] Shapley L S. Stochastic games[J]. Proceedings of the National Academy of Sciences of the United States of America, 1953, 39(10): 1095-1100.

[44] Littman M L. Markov games as a framework for multi-agent reinforcement learning[C]. Machine Learning Proceedings 1994. Amsterdam: Elsevier, 1994: 157-163.

[45] 孙长银, 穆朝絮. 多智能体深度强化学习的若干关键科学问题[J]. 自动化学报, 2020, 46(7): 1301-1312.
(Sun C Y, Mu C X. Important scientific problems of multi-agent deep reinforcement learning[J]. Acta Automatica Sinica, 2020, 46(7): 1301-1312.)

[46] Le Cleac'h S, Schwager M, Manchester Z. ALGAMES: A fast augmented Lagrangian solver for constrained dynamic games[J]. Autonomous Robots, 2022, 46(1): 201-215.

[47] Sadigh D, Sastry S, Seshia S A, et al. Planning for autonomous cars that leverage effects on human actions[C]. Robotics: Science and Systems. 2016, 2: 1-9.

[48] Fridovich-Keil D, Rubies-Royo V, Tomlin C J. An iterative quadratic method for general-sum differential games with feedback linearizable dynamics[C]. 2020 IEEE International Conference on Robotics and Automation. Paris, 2020: 2216-2222.

[49] Fridovich-Keil D, Ratner E, Peters L, et al. Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games[C]. 2020 IEEE International Conference on Robotics and Automation. Paris, 2020: 1475-1481.

[50] Chiu C Y, Fridovich-Keil D, Tomlin C J. Encoding defensive driving as a dynamic Nash game[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). New York: ACM, 2021: 10749-10756.

[51] Wang M Y, Mehr N, Gaidon A, et al. Game-theoretic planning for risk-aware interactive agents[C]. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, 2021: 6998-7005.

[52] Ladino A, Wang M. A dynamic game formulation for cooperative lane change strategies at highway merges [J]. IFAC-PapersOnLine, 2020, 53(2): 15059-15064.

[53] Toghi B, Valiente R, Sadigh D, et al. Cooperative autonomous vehicles that sympathize with human drivers[C]. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague, 2021: 4517-4524.

[54] Hubmann C, Schulz J, Becker M, et al. Automated driving in uncertain environments: Planning with interaction and uncertain maneuver prediction[J]. IEEE Transactions on Intelligent Vehicles, 2018, 3(1): 5-17.

[55] Zhang Q Y, Langari R, Tseng H E, et al. A game theoretic model predictive controller with aggressiveness estimation for mandatory lane change[J]. IEEE Transactions on Intelligent Vehicles, 2020, 5(1): 75-89.

[56] Meng F L, Su J Y, Liu C J, et al. Dynamic decision making in lane change: Game theory with receding horizon[C]. 2016 UKACC 11th International Conference on Control. Belfast, 2016: 1-6.

[57] Laine F, Fridovich-Keil D, Chiu C Y, et al. Multi-hypothesis interactions in game-theoretic motion planning[C]. 2021 IEEE International Conference on Robotics and Automation. Xi'an, 2021: 8016-8023.

[58] Coskun S, Zhang Q Y, Langari R. Receding horizon Markov game autonomous driving strategy[C]. 2019 American Control Conference. Philadelphia, 2019: 1367-1374.

[59] Sadigh D, Landolfi N, Sastry S S, et al. Planning for cars that coordinate with people: Leveraging effects on human actions for planning and active information gathering over human internal state[J]. Autonomous Robots, 2018, 42(7): 1405-1426.

[60] Wang M Y, Wang Z J, Talbot J, et al. Game-theoretic planning for self-driving cars in multivehicle competitive scenarios[J]. IEEE Transactions on Robotics, 2021, 37(4): 1313-1325.

[61] Wang Z J, Spica R, Schwager M. Game theoretic motion planning for multi-robot racing[C]. Distributed Autonomous Robotic Systems. Cham: Springer International Publishing, 2019: 225-238.

[62] Yuan Q, Li S, Wang C, et al. Cooperative-competitive game based approach to the local path planning problem of distributed multi-agent systems[C]. 2020 European Control Conference. St. Petersburg, 2020: 680-685.

[63] Liniger A, Lygeros J. A noncooperative game approach to autonomous racing[J]. IEEE Transactions on Control Systems Technology, 2020, 28(3): 884-897.

[64] Delmas M, Camps V, Lemercier C. Effects of environmental, vehicle and human factors on comfort in partially automated driving: A scenario-based study[J]. Transportation Research — Part F: Traffic Psychology and Behaviour, 2022, 86: 392-401.

[65] Rahmati Y, Talebpour A. Towards a collaborative connected, automated driving environment: A game theory based decision framework for unprotected left turn maneuvers[C]. 2017 IEEE Intelligent Vehicles Symposium. Los Angeles, 2017: 1316-1321.

[66] Jamgochian A, Menda K, Kochenderfer M J. Multi-vehicle control in round abouts using decentralized game-theoretic planning[J/OL]. 2022, arXiv: 2201.02718.

[67] Margellos K, Lygeros J. Hamilton-jacobi formulation for reach-avoid differential games[J]. IEEE Transactions on Automatic Control, 2011, 56(8): 1849-1861.

[68] Pilipovsky J, Tsiotras P. Chance-constrained optimal covariance steering with iterative risk allocation[C].

2021 American Control Conference. New Orleans, 2021: 2011-2016.

[69] Spica R, Cristofalo E, Wang Z J, et al. A real-time game theoretic planner for autonomous two-player drone racing[J]. IEEE Transactions on Robotics, 2020, 36(5): 1389-1403.

[70] Cheng W, Wang G, Wu S B, et al. A human-like longitudinal decision model of intelligent vehicle at signalized intersections[C]. 2017 IEEE International Conference on Real-time Computing and Robotics. Okinawa, 2018: 415-420.

[71] Li G F, Li S L, Li S, et al. Continuous decision-making for autonomous driving at intersections using deep deterministic policy gradient[J]. IET Intelligent Transport Systems, 2022, 16(12): 1669-1681.

[72] Brechtel S, Gindele T, Dillmann R. Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs[C]. The 17th International IEEE Conference on Intelligent Transportation Systems. Qingdao, 2014: 392-399.

[73] El Hamdani S, Loudari S, Novotny S, et al. A Markov decision process model for a reinforcement learning-based autonomous pedestrian crossing protocol[C]. The 3rd IEEE Middle East and North Africa Communications Conference. Agadir, 2022: 147-151.

[74] Liu K W, Li N, Tseng H E, et al. Cooperation-aware decision making for autonomous vehicles in merge scenarios[C]. The 60th IEEE Conference on Decision and Control. Austin, 2022: 5006-5012.

[75] Brechtel S, Gindele T, Dillmann R. Probabilistic MDP-behavior planning for cars[C]. The 14th International IEEE Conference on Intelligent Transportation Systems. Washington, 2011: 1537-1542.

[76] Coskun S, Langari R. Predictive fuzzy Markov decision strategy for autonomous driving in highways[C]. 2018 IEEE Conference on Control Technology and Applications. Copenhagen, 2018: 1032-1039.

[77] Guo C Z, Kidono K, Terashima R, et al. Toward human-like behavior generation in urban environment based on Markov decision process with hybrid potential maps[C]. 2018 IEEE Intelligent Vehicles Symposium. Changshu, 2018: 2209-2215.

[78] Liniger A, Lygeros J. Real-time control for autonomous racing based on viability theory[J]. IEEE Transactions on Control Systems Technology, 2019, 27(2): 464-478.

[79] Li N, Yao Y, Kolmanovsky I, et al. Game-theoretic modeling of multi-vehicle interactions at uncontrolled intersections[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(2): 1428-1442.

[80] Tian R, Li N, Kolmanovsky I, et al. Game-theoretic modeling of traffic in unsignalized intersection network for autonomous vehicle control verification and validation[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(3): 2211-2226.

[81] Liniger A. Path planning and control for autonomous racing[M]. Zurich: ETH Zurich, 2018: 80-109.

[82] Fisac J F, Bronstein E, Stefansson E, et al. Hierarchical game-theoretic planning for autonomous vehicles[C]. 2019 International Conference on Robotics and Automation. Montreal, 2019: 9590-9596.

[83] Andrew G, Gao J F. Scalable training of L^1-regularized log-linear models[C]. Proceedings of the 24th International Conference on Machine Learning. Corvalis, 2007: 33-40.

[84] Di B L, Lamperski A. Local first-order algorithms for constrained nonlinear dynamic games[C]. 2020 American Control Conference. Denver, 2020: 5358-5363.

[85] Bellman R. Dynamic programming[J]. Science, 1966, 153(3731): 34-37.

[86] Li W, Todorov E. Iterative linear quadratic regulator design for nonlinear biological movement systems[C]. International Conference on Informatics in Control, Automation and Robotics (ICINCO). Setúbal: Springer, 2004: 222-229.

[87] Todorov E, Li W W. A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems[C]. Proceedings of the 2005 American Control Conference. Portland, 2005: 300-306.

[88] Laine F, Fridovich-Keil D, Chiu C Y, et al. The computation of approximate generalized feedback Nash equilibria[J/OL]. 2021, arXiv: 2101.02900.

[89] Littman M L. Value-function reinforcement learning in Markov games[J]. Cognitive Systems Research, 2001, 2(1): 55-66.

[90] Littman M L. Friend-or-foe Q-learning in general-sum games[C]. International Conference on Machine Learning. New York: ACM, 2001: 322-328.

[91] Hu J, Wellman M P. Multiagent reinforcement learning: Theoretical framework and an algorithm[C]. International Conference on Machine Learning. New York: ACM, 1998, 98: 242-250.

[92] Hu J, Wellman M P. Nash Q-learning for general-sum stochastic games[J]. Journal of Machine Learning Research, 2003, 4(11): 1039-1069.

[93] 朱冰, 张培兴, 赵健, 等. 基于场景的自动驾驶汽车虚拟测试研究进展[J]. 中国公路学报, 2019, 32(6): 1-19. (Zhu B, Zhang P X, Zhao J, et al. Review of scenario-based virtual validation methods for automated vehicles[J]. China Journal of Highway and Transport, 2019, 32(6): 1-19.)

[94] Ulbrich S, Maurer M. Probabilistic online POMDP decision making for lane changes in fully automated driving[C]. The 16th International IEEE Conference on Intelligent Transportation Systems. Hague, 2014: 2063-2067.

[95] Shah S, Dey D, Lovett C, et al. AirSim: High-fidelity visual and physical simulation for autonomous vehicles[J/OL]. 2017, arXiv: 1705.05065.

[96] Dosovitskiy A, Ros G, Codevilla F, et al. CARLA: An open urban driving simulator[J/OL]. 2017, arXiv: 1711.03938.

[97] Lopez P A, Behrisch M, Bieker-Walz L, et al. Microscopic traffic simulation using SUMO[C]. The 21st International Conference on Intelligent Transportation Systems. Maui, 2018: 2575-2582.

[98] Vinitsky E, Kreidieh A, Le Flem L, et al. Benchmarks for reinforcement learning in mixed-autonomy traffic[C]. Conference on Robot Learning. New York: PMLR, 2018: 399-409.

[99] 段续庭, 周宇康, 田大新, 等. 深度学习在自动驾驶领域应用综述[J]. 无人系统技术, 2021, 4(6): 1-27.
(Duan X T, Zhou Y K, Tian D X, et al. A review of deep learning applications for autonomous driving[J]. Unmanned Systems Technology, 2021, 4(6): 1-27.)

[100] Liang S, Yi P, Hong Y G, et al. Exponentially convergent distributed Nash equilibrium seeking for constrained aggregative games[J]. Autonomous Intelligent Systems, 2022, 2(1): 6.

[101] Liu T F, Qin Z Y, Hong Y G, et al. Distributed optimization of nonlinear multiagent systems: A small-gain approach[J]. IEEE Transactions on Automatic Control, 2022, 67(2): 676-691.

[102] Fu W M, Ma Q C, Qin J H, et al. Resilient consensus-based distributed optimization under deception attacks[J]. International Journal of Robust and Nonlinear Control, 2021, 31(6): 1803-1816.

[103] Yang T, Yi X, Wu J, et al. A survey of distributed optimization[J]. Annual Reviews in Control, 2019, 47: 278-305.

[104] Liu P, Li R. Distributed optimization for a class of uncertain nonlinear multiagent systems with arbitrary relative degree subject to exogenous disturbances[J]. International Journal of Robust and Nonlinear Control, 2022, 32(8): 4631-4647.

[105] You K Y, Tempo R, Xie P. Distributed algorithms for robust convex optimization via the scenario approach[J]. IEEE Transactions on Automatic Control, 2019, 64(3): 880-895.

[106] 杨涛, 徐磊, 易新蕾, 等. 基于事件触发的分布式优化算法[J]. 自动化学报, 2022, 48(1): 133-143.
(Yang T, Xu L, Yi X L, et al. Event-triggered distributed optimization algorithms[J]. Acta Automatica Sinica, 2022, 48(1): 133-143.)

[107] Yi P, Pavel L. Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms[J]. IEEE Transactions on Control of Network Systems, 2019, 6(1): 299-311.

[108] Yuan D, Hong Y, Ho D W C, et al. Optimal distributed stochastic mirror descent for strongly convex optimization[J]. Automatica, 2018, 90: 196-203.

[109] Aradi S. Survey of deep reinforcement learning for motion planning of autonomous vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(2): 740-759.

[110] Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 4909-4926.

[111] 郭戈, 许阳光, 徐涛, 等. 网联共享车路协同智能交通系统综述[J]. 控制与决策, 2019, 34(11): 2375-2389.
(Guo G, Xu Y G, Xu T, et al. A survey of connected shared vehicle-road cooperative intelligent transportation systems[J]. Control and Decision, 2019, 34(11): 2375-2389.)

[112] 丁飞, 张楠, 李升波, 等. 智能网联车路云协同系统架构与关键技术研究综述[J]. 自动化学报, 2022, 48(12): 2863-2885.
(Ding F, Zhang N, Li S B, et al. A survey of architecture and key technologies of intelligent connected vehicle-road-cloud cooperation system[J]. Acta Automatica Sinica, 2022, 48(12): 2863-2885.)

作者简介

衣鹏 (1988−), 男, 教授, 博士生导师, 从事多智能体系统、分布式优化及博弈论等研究, E-mail: yipeng@tongji.edu.cn;

潘越 (1999−), 男, 博士生, 从事自动驾驶决策与博弈论的研究, E-mail: 2211302@tongji.edu.cn;

王文远 (1998−), 男, 硕士生, 从事自动驾驶决策与博弈论的研究, E-mail: 2230816@tongji.edu.cn;

刘政钦 (2000−), 男, 硕士生, 从事自动驾驶决策与博弈论的研究, E-mail: 2230709@tongji.edu.cn;

洪奕光 (1966−), 男, 教授, 博士生导师, 从事控制理论、优化及人工智能等研究, E-mail: yghong@tongji.edu.cn.

(责任编辑: 李君玲)