

多智能体协同研究进展综述: 博弈和控制交叉视角

秦家虎¹ 马麒超¹ 李曼¹ 张聪¹ 付维明¹ 刘轻尘¹ 郑卫新²

摘要 多智能体协同应用广泛, 并被列为新一代人工智能 (Artificial intelligence, AI) 基础理论亟待突破的重要内容之一, 对其开展研究具有鲜明的科学价值和工程意义. 随着人工智能技术的进步, 传统的单一控制视角下的多智能体协同已无法满足执行大规模复杂任务的需求, 融合博弈与控制的多智能体协同应运而生. 在这一框架下, 多智能体协同具有更高的灵活性、适应性和扩展性, 为多智能体系统的发展带来更多可能性. 鉴于此, 首先从协同角度入手, 回顾多智能体协同控制与估计领域的进展. 接着, 围绕博弈与控制的融合, 介绍博弈框架的基本概念, 重点讨论在微分博弈下多智能体协同问题的建模与分析, 并简要总结如何应用强化学习算法求解博弈均衡. 选取多机器人导航和电动汽车充电调度这两个典型的多智能体协同场景, 介绍博弈与控制融合的思想如何用于解决相关领域的难点问题. 最后, 对博弈与控制融合框架下的多智能体协同进行总结和展望.

关键词 多智能体系统, 协同控制, 博弈优化, 多移动机器人导航, 电动汽车充电调度

引用格式 秦家虎, 马麒超, 李曼, 张聪, 付维明, 刘轻尘, 郑卫新. 多智能体协同研究进展综述: 博弈和控制交叉视角. 自动化学报, 2025, 51(3): 489–509

DOI 10.16383/j.aas.c240508

CSTR 32138.14.j.aas.c240508

Recent Advances on Multi-agent Collaboration: A Cross-perspective of Game and Control Theory

QIN Jia-Hu¹ MA Qi-Chao¹ LI Man¹ ZHANG Cong¹ FU Wei-Ming¹
LIU Qing-Chen¹ ZHENG Wei-Xing²

Abstract Multi-agent collaboration is widely applied and considered a key component for breakthroughs in next-generation artificial intelligence (AI) foundational theories, possessing significant scientific and engineering value. As AI technology advances, the traditional single-control perspective on multi-agent collaboration is inadequate for large-scale complex tasks. Thus, the integration of game theory and control has emerged, offering greater flexibility, adaptability, and scalability, thereby expanding the potential of multi-agent systems. This paper begins by reviewing progress in the field of multi-agent coordination control and estimation. It then introduces the basic concepts of game theory, focusing on modeling and analysis of multi-agent collaboration within the framework of differential games, and briefly summarizes the use of reinforcement learning algorithms to solve game equilibria. Two typical multi-agent collaboration scenarios, namely multi-robot navigation and electric vehicle charging scheduling, are discussed to demonstrate how the integration of game theory and control addresses key challenges. Finally, the paper summarizes and provides an outlook on multi-agent collaboration within the integrated game-control framework.

Key words Multi-agent system, coordination control, game and optimization, multi-mobile robot navigation, electric vehicle charging scheduling

Citation Qin Jia-Hu, Ma Qi-Chao, Li Man, Zhang Cong, Fu Wei-Ming, Liu Qing-Chen, Zheng Wei-Xing. Recent advances on multi-agent collaboration: A cross-perspective of game and control theory. *Acta Automatica Sinica*, 2025, 51(3): 489–509

收稿日期 2024-07-16 录用日期 2024-11-06

Manuscript received July 16, 2024; accepted November 6, 2024
国家自然科学基金 (U23A20323, 62373341, 62203418, 62303435, 62403444) 资助

Supported by National Natural Science Foundation of China (U23A20323, 62373341, 62203418, 62303435, 62403444)

本文责任编辑 易新蕾

Recommended by Associate Editor YI Xin-Lei

1. 中国科学技术大学自动化系 合肥 230027 中国 2. 西悉尼大学计算机、数据与数学学院 悉尼 NSW 2751 澳大利亚

1. Department of Automation, University of Science and Technology of China, Hefei 230027, China 2. School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney NSW 2751, Australia

多智能体协同是指具备一定计算、通信、感知、学习和执行能力的智能体, 通过协作完成单一智能体无法胜任的大规模复杂任务^[1]. 多智能体协同是人工智能、自动化、通信和机器人等领域交叉融合的产物, 也是多智能体系统领域研究的核心之一. 此外, 多智能体协同还被列为新一代人工智能亟待突破的基础理论瓶颈——自主协同与决策的主要内容之一, 具有重要的科学研究价值. 近年来, 随着人工智能和机器人等技术的迅猛发展, 多智能体系统

协同理论研究也取得长足进步,并广泛应用于仓储物流、电网调度等民用领域,以及战场侦查、空中对抗等军事领域。

早期对于多智能体协同的研究主要从控制视角出发,旨在建立多智能体协同基础理论.沿着“一致性”(定义见第1节)这条主线,研究人员围绕个体动力学以及信息交互两个关键要素,深入探索多智能体系统一致性分析(积分器、高阶线性系统和Lipschitz非线性系统^[2-7]的一致性分析),以及带约束的多智能体系统一致性控制(通信时滞、动态拓扑以及执行器饱和/失效下的一致性控制^[8-15]),揭示了多智能体系统协同的内在机理和特性.正如文献[16]指出,随着智能体变得越来越智能,从而可以完成更复杂任务,考虑更多的主动(Active)而非仅被动(Passive)目标成为必然趋势.在实际场景中,多智能体之间固然需要协同以完成整体任务,每个智能体也会有自身独立的目标^[17],例如每个智能体自然地期望以最小代价实现相互间协同.在这种情形下,可根据每个智能体内在需求构造特定的目标函数,以包含协作信息并反映智能体之间的关联或者制约,引导智能体在对目标函数优化时还能实现整体协同.上述框架难以仅用多智能体协同控制相关概念来描述,但可以结合博弈理论框架进行建模和刻画^[1, 18-20],即每个智能体根据自身目标独立地做出决策,并在决策中考虑其他智能体可能采取的行动,所有智能体执行策略后最终能够实现相互间协同.由此可见,博弈与控制理论的融合具备刻画并实现多智能体复杂协同行为的能力,潜力巨大.此外,通过交叉融合,既可以发挥博弈框架强大的问题建模能力以刻画更丰富的多智能体协同行为,也可以融合协同理论辅助突破博弈分析与求解面临的瓶颈问题,例如利用协同理论分析博弈均衡的可控性^[21],或者求解博弈均衡^[22].近年来,在新一轮人工智能浪潮的推动下,研究人员将更多注意力放到挖掘和发挥多智能体的“智能”属性上,以便满足各种复杂场景下任务执行的需求,因此越来越多的工作将博弈与控制理论方法相融合,旨在突破多智能体优化决策/协同控制难题.

融合博弈与控制的多智能体系统研究已经取得初步进展,例如基于零和博弈的鲁棒控制问题、基于一般和博弈的最优控制问题以及基于团队博弈的协同控制问题研究^[18-19].特别地,目前融合博弈与控制的多智能体协同研究正在蓬勃开展,已经诞生不少研究成果,因而有必要对研究现状进行阶段性梳理.实际上,近年来已有综述开始关注在多智能体系统中博弈、控制与学习的交叉融合.例如文献[19]简要介绍了如何利用博弈框架建模、分析分布式控

制问题,而文献[20]则较为系统地回顾了演化矩阵博弈(Evolutionary matrix game)的分析与控制,但是对多智能体协同关注较少.文献[18]从博弈、控制与学习交叉的视角梳理多智能体系统的相关研究脉络,着眼于探索新的研究范式,虽然涉及基于博弈的多智能体最优协同控制,但是内容较为简略.

本文从交叉融合视角出发,首先回顾多智能体系统一致性分析和控制领域的主要研究框架和核心研究进展,揭示多智能体系统一致性的内在机理、概述一致性控制中不同约束的处理方法,进而探讨基于一致性的多智能体协同估计问题.在一致性分析与控制的基础上,进一步引入微分博弈框架建模多智能体系统协同优化控制问题,并结合最优控制和强化学习,阐述如何对其进行分析和求解.在应用方面,分别选取多移动机器人导航和电动汽车充电调度两个典型场景,介绍博弈与控制融合的思想如何用于解决相关领域的难点问题.本文的总体结构框架见图1.

1 多智能体协同

多智能体通过协同可以完成编队、覆盖、蜂拥等基本控制任务,从而为更复杂任务的完成提供支撑.多智能体一致是协同的核心,编队、覆盖、蜂拥等控制任务均可以在一定程度上转化为多智能体一致性控制问题,例如基于状态一致性的偏置实现编队控制^[23].本节将视野聚焦到多智能体一致性展开叙述.下面首先介绍多智能体系统一致性的数学描述.

多智能体间的交互¹通常使用图论工具来描述.一个动态图 $\mathcal{G}(t)$ 由节点集 \mathcal{N} ,边集 $\mathcal{E}(t)$ 和边的权重 $a_{ij}(t)$ 刻画,其中每个节点对应一个智能体.如果在 t 时刻智能体 i 能够获得智能体 j 的信息,则称智能体 j 是智能体 i 在 t 时刻的邻居;此时存在一条从智能体 j 指向智能体 i 的边 $(j, i) \in \mathcal{E}(t)$,其权重值 $a_{ij}(t) > 0$,否则 $a_{ij}(t) = 0$.

考虑如下由 N 个同构的连续型智能体²组成的系统:

$$\begin{cases} \dot{x}_i(t) = f(x_i(t)) + g(u_i(t)) & (1) \\ u_i(t) = K \sum_{j=1}^N a_{ij}(t)(x_j(t) - x_i(t)) & (2) \end{cases}$$

其中, $x_i \in \mathbf{R}^n$ 表示第 i 个智能体的状态(刻画智能体的位置、电量、速度等信息), $u_i \in \mathbf{R}^m$ 表示输入,并且本节假定 u_i 只能获取局部的信息. $g(\cdot) : \mathbf{R}^m \rightarrow$

¹ 这里的交互指的是信息流动,例如智能体通过通信或者传感装置获取其他个体的信息.

² 除非特别声明,下文均以这里的连续型动力学系统为讨论对象.

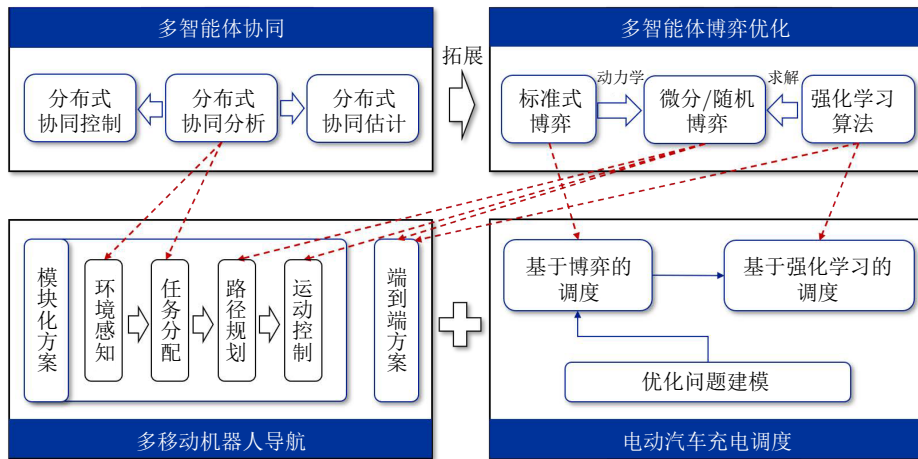


图 1 本文总体结构

Fig.1 General structure of the paper

\mathbf{R}^n 表示具有特定性质的函数, 并视情况做出具体设定. 对于多智能体系统 (1), 如果 $\lim_{t \rightarrow \infty} \|x_i(t) - x_j(t)\| = 0$ 对任意 i, j 和任意初值 $x_i(0)$ 成立, 那么称其实现了全局渐近一致.

对上述多智能体一致性问题开展研究通常需要考虑三个基本要素, 分别是个体自身动力学 (对应 $\dot{x}_i = f(x_i)$)、个体间交互 (对应 $a_{ij}(t)$ 及其相关项) 以及状态耦合 (对应函数 $g(\cdot)$ 和矩阵 K). 在一致性分析中经常假定 $g(\cdot)$ 为输入 u_i 的线性函数, 并围绕上述三个基本要素, 揭示它们对多智能体一致性的本质影响. 而一致性控制则在上述分析的基础上, 从实际场景出发对 $g(\cdot)$ 的结构或特性做出具体约束, 并通过设计参数 (例如反馈矩阵 K) 以实现多智能体渐近一致性.

1.1 动态拓扑下的分布式协同分析

本节围绕动态拓扑带来的挑战, 介绍已有工作如何刻画时变通信拓扑、个体自身动力学、部分状态耦合等多重因素对多智能体一致性的影响, 分别讨论不同阶次的动力学, 并沿着通信拓扑条件如何放松这一主线介绍相关进展.

1.1.1 积分器型智能体: $f(\cdot) = 0$ 且 $g(\cdot) = K = I$

针对积分器型多智能体系统, 研究人员已揭示动态通信拓扑下实现一致性的各种充分必要条件^[2-4, 24-25]. 值得关注的是一类联合 (δ, T) -连通性条件. 这一条件要求给定 $T, \delta > 0$, 对任意时间 t 满足如下条件的边 (i, j) ,

$$(i, j) : \int_t^{t+T} a_{ij}(\tau) d\tau \geq \delta > 0$$

构成的图是连通的. 文献 [24] 揭示了联合 (δ, T) -连通性条件是无向动态通信拓扑下积分器型多智能

体系统以指数速率实现渐近一致的充要条件. 注意, 上述联合 (δ, T) -连通性条件只要求 $a_{ij}(t)$ 是可积函数, 并未对 $a_{ij}(t)$ 施加额外要求, 例如允许 $a_{ij}(t)$ 分段固定但是不要求有正的驻留时间下界, 因此该条件要求宽松、适用范围广. 此外, 以指数速率实现一致性也保证了多智能体系统具有一定抗干扰和不确定性的能力.

1.1.2 高阶多智能体: 具有特定结构的 $g(\cdot)$ 或 $a_{ij}(t)$

近十年来研究人员一直在探索动态通信拓扑下具有高阶个体动力学的多智能体系统一致性问题, 其中一条重要主线致力于探索实现一致性的最弱通信拓扑条件. 现有工作通常对通信拓扑^[7, 12, 26-28] 或系统参数 (包括系统自身动力学 $\dot{x}_i = f(x_i)$ 和耦合函数 $g(\cdot)$)^[29-35] 施加严格的条件. 一方面, 研究人员要求通信拓扑以较高的频率保持足够的连通性 (Frequently connected) 或通信拓扑的时间平均 (Averaged topology) 满足特定连通性条件, 进而有利于开展多智能体系统的一致性分析^[26-27]. 例如, 文献 [7] 假设分段连续通信拓扑的非零权重有一致的下界 (Uniformly bounded), 即通信拓扑的权重可以发生变化, 但通信拓扑的连通性结构需要保持不变. 另一方面, 研究人员还会假设耦合矩阵 (对应线性的 $g(\cdot)$) 是行满秩的^[29] 或多智能体系统自身动力学 (对应 $\dot{x}_i = f(x_i)$) 具有非发散 (Non-expansive) 的性质, 并以此为基础开展一致性分析. 例如, Meng 等^[34] 研究了具有满秩输入矩阵的非发散时变有限维非线性多智能体系统的一致性. 此外, Lu 等^[31]、Meng 等^[32]、Abdessameud^[35] 以及 Liu 等^[33] 考虑了异构多智能体系统的一致性. 解决这类问题的主要思路在于借助内模原理, 通过设计全状态耦

合动态控制器 (Dynamic controller) 或参考信号 (Reference signal), 将带有高阶动力学的多智能体系统一致性问题转化为积分器型多智能体系统的一致性问题^[31-32, 35], 从而简化分析. 文献 [7, 12, 26-35] 中的工作尽管已取得一定的研究进展, 但仍面临一些问题. 例如假定输入矩阵全状态耦合有助于放宽对通信拓扑连通性的要求, 使一致性分析变得更易处理, 但也限制了相关方法的适用范围.

1.1.3 高阶多智能体: 一般形式 $g(\cdot)$ 和 $a_{ij}(t)$

如果高阶多智能体的耦合矩阵 (对应线性 $g(\cdot)$) 非行满秩, 同时仅要求弱连通性条件, 大多数多智能体系统一致性的研究都是基于无向通信拓扑展开的^[36-39]. 在 Su 等^[36]的工作中, 考虑了具有非行满秩输入矩阵的边缘稳定 (Marginally stable)³ 线性系统. 这些线性系统通过分段固定的通信拓扑进行交互, 且具有正的驻留时间 (Dwell time). 文献 [36] 证明, 通过合理的反馈矩阵设计, 只要通信拓扑满足一致连通性 (Uniformly connected) 条件 (联合 (δ, T) -连通性的特殊情况) 并且个体动力学系统可观 (Observable), 就能实现多智能体系统的渐近一致^[36]. 文献 [38-39] 研究了分段固定通信拓扑或连续时变通信拓扑 (即边的权重是时间的连续函数) 下中立型稳定 (Neutrally stable)⁴ 线性多智能体系统的一致性问题, 分别发展了基于子空间分解^[38] 和一致完全可观 (Uniform complete observability) 的多智能体系统一致性分析方法^[39]. 上述方法的优势在于去掉了“正的驻留时间”这一条件, 并给出实现一致性的充分必要条件^[38-39]. 文献 [37] 放宽了对系统矩阵的约束, 使得线性多智能体系统只需满足可控性要求, 允许 $f(\cdot) = A$ 具有不稳定的特征值, 但是仍然假设分段固定的通信拓扑具有正的驻留时间. 文献 [37] 证明, 如果 Lyapunov 指数 (取决于系统动力学参数) 小于同步性指数 (取决于通信拓扑联合连通特性), 就可以实现多智能体系统渐进一致^[37]. 最近, 文献 [40] 在 Wang 等^[37] 研究的基础上, 提出一类图 Laplacian 矩阵函数的预紧性条件, 大大放松了对通信拓扑的要求, 并证明了在上述条件下通信拓扑联合 (δ, T) -连通与个体动力学的可控性是多智能体系统以指数速率实现一致性的充分必要条件. 上述结论在理论上证明了在一定条件下, 联合 (δ, T) -连通性对于高阶多智能体系统以指数速率实现一致性仍然必要, 是对积分器型多智能体系统下结论的推广.

³ 如果矩阵 A 的特征值实部小于等于零, 且实部为零的特征值代数重数等于几何重数, 则称 A 是边缘稳定的.

⁴ 如果矩阵 A 的特征值实部均为零, 且代数重数等于几何重数, 则称 A 是中立型稳定的.

1.2 实际约束下的分布式协同控制

本节主要围绕多智能体系统在实际约束下的一致性控制展开介绍. 所考虑的实际约束包括: 执行器约束、通信约束以及交互约束, 分别对应执行器饱和/失效下的协同控制、基于事件触发的协同控制以及竞合共存下的协同控制. 上述约束均可以反映在函数 $g(\cdot)$ 或通信拓扑权重 $a_{ij}(t)$ 的结构/特性上. 一致性控制需要对系统中的相关参数进行设计 (例如反馈矩阵 K), 并结合一致性分析的方法/结论, 刻画一致性控制的稳定性、鲁棒性等.

1.2.1 执行器饱和和约束下的协同

在实际控制系统中, 所有控制驱动装置都服从饱和约束, 即力、电压、流量等实际物理输入量在控制理论和技术的应用中最终都是有限的^[41].

本部分主要考虑具有有界输入下渐近零可控性质⁵ 的线性系统, 其在执行器饱和下的一致性控制研究主要有两个框架: 1) 半全局一致性控制; 2) 全局一致性控制.

半全局一致性控制指的是给定 $\mathcal{X} \subset \mathbf{R}^n$ 以及 $x_i(0) \in \mathcal{X}$, 在 $u_i(t)$ 的驱动下, 有 $\lim_{t \rightarrow \infty} \|x_i(t) - x_j(t)\| = 0$ 对任意 i, j 成立. 低增益反馈技术广泛用于渐近零可控线性系统在输入有界下的半全局一致性控制^[8, 42], 其背后的原理在于, 通过将反馈增益调节得足够低, 只要初始状态在给定的集合内, 系统将受到输入的线性作用, 从而满足执行器饱和和约束. 利用低增益反馈技术, 已有文献相继深入研究了固定拓扑^[9] 和切换拓扑^[43] 下带执行器饱和和约束的一致性控制问题.

全局一致性控制指的是对任意 $x_i(0) \in \mathbf{R}^n$, 在 $u_i(t)$ 的驱动下, 有 $\lim_{t \rightarrow \infty} \|x_i(t) - x_j(t)\| = 0$ 对任意 i, j 成立. 一方面, 研究表明, 在无执行器饱和和约束下设计的线性控制器可以实现带执行器饱和和约束下的一致性控制^[44-45]; 另一方面, 具有特定形式和性质的非线性控制器也被证明可以实现带执行器饱和和约束下的一致性控制^[46-47]. 例如, 文献 [46] 利用饱和函数 $\tanh(\cdot)$ 设计控制器, 而文献 [47] 则借助 $\delta_{\Delta/2}(\cdot)$ 函数设计基于多跳中继机制的控制器, 其中 $\delta_{\Delta/2}(u) : \mathbf{R} \rightarrow \mathbf{R}$ 定义为 $\delta_{\Delta/2}(a) = \text{sgn}(a) \min\{|a|, \Delta/2\}$, $\text{sgn}(\cdot)$ 表示符号函数.

1.2.2 执行器失效下的协同

由于老化或损坏, 系统执行器、传感器或其他部件不可避免地会出现失效^[48]. 而对于多智能体系统, 执行器、传感器或其他部件的失效问题由于系统规模和复杂度的提高变得更加严峻^[49]. 因此, 在

⁵ 有界输入下渐近零可控的定义见文献 [42].

执行器失效时, 如何有效识别并应对非常关键, 这就要求开展故障识别与隔离 (Fault detection and isolation, FDI) 以及容错控制 (Fault-tolerant control, FTC) 研究。

FDI 的作用是检测并隔离故障, 减小故障对系统的影响或为故障处理作准备。针对 FDI 的研究主要有两种框架: 集中式与分布式。集中式 FDI 需要选取一个中心节点部署失效检测算法, 并搜集系统的所有信息以执行该算法; 与此不同, 分布式 FDI 可以在每一个智能体单独部署, 并仅利用局部信息开展失效检测, 因此通信负担小、可扩展性好^[10]。已有研究围绕分布式 FDI, 针对执行器、传感器、通信等故障开展了较为系统的研究^[10, 49-51], 其中的核心方法之一是利用残差信号对故障进行检测与隔离^[51]。

FTC 的作用是对控制系统进行设计以应对潜在的故障, 从而提高系统可靠性。现有针对多智能体系统容错控制的研究大多聚焦执行器故障^[52-53], 并可粗略地分为两大类: 被动容错控制以及主动容错控制。在被动容错控制中, 控制器参数根据预期的故障类型预先设计, 以提高特定情形下的系统鲁棒性能^[52], 因此这类方法无需系统的实时故障信息; 与被动容错不同, 主动容错控制则自动地检测和识别故障模块, 并据此重新配置系统参数以应对故障^[53]。

1.2.3 基于事件触发的协同

为降低数据传输和控制器更新频次, 节约通信和计算资源, 文献 [11] 将事件触发控制框架^[54]应用到多智能体协同中, 由此引发了大量基于事件触发的多智能体一致性控制研究^[55-56]。

在事件触发机制中, 当测量误差超出设定阈值时相应事件触发, 从而开始数据传输和控制器更新, 因此测量误差和阈值的设计至关重要。一般来说, 测量误差的设计有三种不同形式, 包括: 1) 一般形式^[11]; 2) 组合形式^[57]; 3) 基于模型的形式^[58]。阈值的设计则可划分为: 1) 状态无关形式^[59]; 2) 状态相关形式^[11]; 3) 混合形式^[60]。

1.2.3.1 测量误差设计

假定第 i 个智能体的触发时间序列是 $t_0^i = 0, t_1^i, \dots, t_k^i, \dots$, 那么一般形式的测量误差定义为^[11]: $e_i(t) = x_i(t_k^i) - x_i(t)$ ⁶。采用上述测量误差, 在任意智能体上定义的事件触发时, 所有智能体都要进行信息传输和控制器更新, 这将导致更新频率过高的问题。

为解决上述问题, 文献 [57] 提出了组合形式的

测量 $q_i = \sum_{j=1}^N a_{ij}(x_j - x_i)$, 并由此定义了组合形式的测量误差: $e_i(t) = q_i(t_k^i) - q_i(t)$ 。在这一框架中, 任意智能体一旦检测到事件触发, 只需收集其邻居节点的信息, 并更新自己的控制器, 大大降低了控制器更新频率。这一框架的缺点在于, 一旦通信拓扑是有向的, 上述邻居节点不知道何时发送信息, 或者智能体需要向其邻居节点发送信息传输请求, 从而增加通信代价^[61]。

上述两类测量误差 (一般形式和组合形式) 仅适用于积分器型多智能体系统^[11, 57]。对于一般线性多智能体系统, 通常需要考虑到系统动力学特性的影响。文献 [58] 提出基于模型的测量误差: $e_i(t) = e^{A(t-t_k^i)} x_i(t_k^i) - x_i(t)$ 。式中的矩阵指数 $e^{A(t-t_k^i)}$ 用于估计当前时刻系统的状态, 以避免 $e_i(t)$ 增长过快。

1.2.3.2 触发阈值设计

状态无关形式的阈值通常设计为带有非负下界的递减函数形式^[59]。这样做的好处是可以轻易排除掉 Zeno 现象, 同时在判断触发与否时无需邻居节点的信息。状态无关形式的阈值设计能有效降低通信负担, 但在判断触发时没有利用系统相关信息, 潜在造成了系统性能的改变, 例如系统收敛速率可能取决于阈值本身^[57]。

状态相关形式阈值设计缺乏统一的形式, 按其特性可大致分为两种, 即连续形式和分段固定形式。例如, 连续形式阈值可设计成 $\beta_i \text{sqrt}(\sum_{j=1}^N a_{ij}(x_j - x_i)^2)$ ^[57], 其中, $\beta_i > 0$ 是与个体自身动力学、通信拓扑相关的参数, $\text{sqrt}(\cdot)$ 表示平方根函数。显然, 上述连续形式的阈值需要连续监测邻居状态。为解决这一问题, 还可以采用离散形式的阈值, 将连续的状态量替换为采样状态^[62]。

混合形式的阈值则将连续形式和分段固定形式的阈值相结合^[62], 在有效利用系统信息的同时寻求避免 Zeno 现象, 但是也带来分析和设计上的挑战。

1.2.4 竞合共存下的协同

社会网络可以看作一个典型的多智能体系统。在社会网络中, 个体之间既可以合作, 也可以竞争^[63]。这导致对立子群的出现, 产生了二部一致概念, 即网络分为两个子群, 每个子群内部达到一致, 而子群之间状态则相反。文献 [64] 展示了如何根据个体间合作/竞争、喜好/厌恶等关系构造边权重带符号的图 (简称符号图), 以刻画个体间不同性质的关系。

沿着上述符号图构建的思路出发, 研究人员根据 Laplacian 势能函数的梯度构造了多智能体

⁶ 除非特别说明, 本部分所述时间 $t \in [t_k^i, t_{k+1}^i)$ 。

系统的分布式控制策略以实现二部一致: $\dot{x}_i = -\sum_{j=1}^N |a_{ij}|(x_i - \text{sgn}(a_{ij})x_j)$. 研究发现, 二部一致的必要条件是符号图具有结构平衡的特性 (Structurally balanced)^[63]. 在这一条件下, 可以采用规范变换 (Gauge transformation)^[63] 将二部一致问题转化为一致性问题, 并探索固定^[9] 或时变^[65] 通信拓扑下的二部一致问题. 除了二部一致, 在协作、竞争共存的网络中, 可能存在两个以上的智能体子群, 其中子群内部是合作的关系, 而子群之间合作/竞争并存. 针对这类情形的研究产生了分群一致或聚类一致的概念^[66-70]. 现有工作主要分为两类: 1) 从探索拓扑结构特性入手, 揭示拓扑结构、耦合强度等因素对分群一致的作用^[66-67]; 2) 设计外部输入信号引导子群的演化, 进而刻画聚类一致的收敛性质^[70]. 更详细内容可参考文献 [12].

1.3 基于模型的分布式协同估计

多智能体协同状态估计是指由具备检测、通信以及计算功能的智能体利用检测信息和局部信息交互实现对动态目标状态的估计. 本节主要关注基于模型的协同状态估计问题, 对不同类型的分布式状态估计方法展开介绍, 这些方法的设计均基于式 (2) 中给出的协同控制方法, 利用其实现分布式的计算和信息融合.

在基于模型的协同状态估计问题中, 模型是指目标状态演化模型

$$x_{k+1} = f_t(x_k, u_k, v_k)$$

和智能体检测模型

$$y_{i,k} = h_i(x_k, w_k^i), \quad i = 1, \dots, N$$

其中, k 表示智能体的采样时刻, x 为目标状态, u 为控制输入, y_i 为智能体 i 的检测值, v 和 w^i 是过程和检测噪声. 分布式协同状态估计问题中的一个基本要素是通信拓扑, 它反映了不同智能体之间的信息交互关系, 一般会被建模成图的形式 $\mathcal{G}(k) = (\mathcal{N}, \mathcal{E}(k), \mathcal{A}(k))$, 这里 $\mathcal{G}(k)$ 表示随时间变化的动态拓扑图, 静态拓扑图可以认为是各时刻拓扑均相同的动态拓扑图. 智能体通过通信拓扑向邻居传输本地检测信息或利用检测信息计算出的本地状态估计, 将本地检测或估计信息与获取到的其他个体信息相融合, 获得更加准确的目标状态估计. 因此解决多智能体分布式协同状态估计问题的关键在于设计出有效的检测信息处理方法以及不同智能体间的信息交互和融合方法. 基于此, 研究人员从信息处理及信息融合角度出发设计了一系列多智能体分布式状态估计方法^[71-77], 这些方法从结构出发可以划

分为 Kalman 滤波类方法^[71-74] 和一致-革新类估计方法^[75-77].

1.3.1 Kalman 滤波类的分布式状态估计算法

自 1960 年 Kalman^[78] 在具有里程碑意义的现代滤波理论中提出 Kalman 滤波后, 随着研究的不断发展和深入, 形成了包括扩展 Kalman 滤波^[79]、无迹 Kalman 滤波^[80]、中心差分 Kalman 滤波^[81]、容积 Kalman 滤波^[82] 等在内的一系列估计方法, 统称为 Kalman 滤波类方法, 这类方法均包含基于目标状态模型的预测和基于检测信息的更新两个环节. 基于 Kalman 滤波的分布式状态估计方法就是以 Kalman 滤波类方法为基础, 结合多智能体一致性算法, 使各智能体在每一次估计迭代过程中都能通过实时计算当前状态估计和相应协方差估计的局部平均值修正估计结果. 文献 [71-72] 将经典 Kalman 滤波和扩展 Kalman 滤波的信息形式与一致性算法进行结合, 在每一次完成估计值的预测后, 各智能体会利用通过局部信息交互获得的邻居智能体先验信息对和检测更新信息分别计算二者的局部平均, 并在此基础上计算后验估计. 这两种方法能够在通信拓扑强连通的场景中分别实现对线性和非线性目标状态的分布式估计. 文献 [73] 设计了基于一致的信息分组融合机制, 并将其与扩展 Kalman 滤波的信息形式结合, 在各组智能体的动态通信拓扑分别强连通时能够实现对非线性演化的耦合多目标状态的分布式估计. 文献 [74] 利用一致算法融合检测信息、估计值及相应协方差, 在容积 Kalman 滤波的基础上提出基于混合一致的分布式滤波方法, 实现了高维分布式状态估计.

上述分布式状态估计方法均针对较理想的估计场景, 智能体间的通信、智能体自身的信息采集等均不受限制. 但是在实际应用场景中总会面对各种局限性和挑战. 这些局限性和挑战针对的对象有所不同, 可以根据其作用对象将协同估计问题分为有限通信下的协同估计^[83-84]、不确定性目标的协同估计^[85-86]、恶意攻击下的协同估计^[87-88] 等类型. 文献 [83] 基于一种涵盖有限带宽、可变延迟和丢包等网络现象的通用通信约束模型, 设计了一种基于 Kalman 滤波递归分布式融合估计器, 解决了一类网络物理系统的分布式协方差交集融合估计问题. 文献 [84] 针对时变线性系统设计了一种基于事件触发的分布式状态估计方法, 每个智能体在事件触发机制下选择性地本地信息传递给相邻节点, 以减少通信带宽需求. 面对目标状态存在不确定性的场景, 文献 [85] 基于带有非线性不确定函数的扩展状态模型设计了基于扩展 Kalman 滤波的分布式预测器; 文献 [86]

遵循一阶广义伪贝叶斯和交互多模型方法将分布式混合信息融合算法与扩展 Kalman 滤波相结合, 解决了目标过程模型对任何智能体都不确定的情况下的分布式状态估计问题. 多智能体系统作为一种典型信息物理系统, 极易受到外部恶意攻击. 面对分布式状态估计过程中最常见的、以改变智能体检测信息为攻击方式的欺骗攻击, 文献 [87] 针对时变线性目标提出一种快速弹性分布式状态估计器, 无论攻击的强度和受损智能体的数量如何, 都会产生有界状态估计误差; 文献 [88] 针对非线性目标, 设计了基于不变扩展 Kalman 滤波的弹性分布式状态估计方法, 在连通的通信拓扑下能够实现欺骗攻击下的目标状态估计.

1.3.2 一致-革新类分布式状态估计算法

基于 Kalman 滤波的分布式状态估计方法均继承了经典 Kalman 滤波方法的特点——只能处理高斯噪声下的状态估计问题. 而一致-革新类分布式状态估计方法能够克服这一局限性, 在一般有界噪声下实现对目标状态的估计. 此类方法中, 分布式估计器的一般形式为

$$\hat{x}_{i, k+1} = f_i(\hat{x}_{i, k}, u_k, 0) + g(y_{i, k}) + F_{i, k} \sum_{j \in \mathcal{U}_{i, k}} (\hat{x}_{j, k} - \hat{x}_{i, k}), \quad i = 1, \dots, N$$

式中, $\mathcal{U}_{i, k}$ 是指 k 时刻智能体 i 所有邻居智能体的集合. 等号右侧的第 2 项为处理检测信息的革新项, 该项的作用是利用检测信息对状态估计进行修正; 第 3 项为基于一致性算法设计的进行分布式信息融合的一致项, 在该项的作用下所有智能体对目标状态的估计将达到一致. 针对不同的场景, 设计合适的函数 $g(\cdot)$ 和系数矩阵 $F_{i, k}$ 可以实现对相应目标状态的分布式估计. 文献 [89] 提出一种用于电力系统广域监测的一致-更新类分布式状态估计算法, 通过与指定的相邻控制区域进行信息交换, 各个控制区域可以在通信拓扑连通的情况下实现对整个电力系统状态的无偏估计. 文献 [75] 通过设计一致-革新类分布式状态估计器, 实现了对移动机器人和地标的定位估计. 文献 [77] 通过引入一组势函数设计了一种新颖的一致-革新类估计方法, 使得可以进行能量收集的传感器网络能在间歇性通信下实现状态估计目标.

在非高斯的有界噪声下同样需要解决第 1.3.1 节指出的实际状态估计过程中需面临的各种局限性和挑战. 针对通信条件的局限性, 文献 [76] 设计的一致-革新类估计方法实现了无线衰落信道下随机哈密顿系统的分布式状态估计; 文献 [90] 提出的基

于事件触发的分布式一致-革新类估计方法通过构造数据调度器在降低通信损耗的同时保证了估计的准确性. 文献 [91] 针对高斯噪声、系统不确定性和非线性影响下的目标状态, 对不确定性和非线性进行统一建模, 设计了用于处理系统动力学和约束的估计器. 面对恶意欺骗攻击, 文献 [92] 针对线性目标状态设计了一种动态目标调节增益估计器, 通过设计一致项的调节增益矩阵和更新项的自适应增益实现欺骗攻击下的分布式估计; 文献 [93] 分析了一致-革新类分布式状态估计算法在联合欺骗攻击下的脆弱性, 给出了部分或全部检测受到攻击时的算法性能被破坏且攻击无法被检测到的充分必要条件.

2 多智能体博弈优化

本节将在多智能体系统中进一步引入效用函数 (Utility function), 以刻画智能体本身的主动目标 (Active objects), 从而构建基于博弈的多智能体优化控制/决策框架, 拓展多智能体系统适用场景和任务. 特别地, 在微分博弈框架中 (见第 2.2 节), 当上述主动目标置为常数时, 这一优化/决策框架可退化为第 1 节的多智能体协同控制问题, 因此第 1 节的研究对象可看成本节的特例.

本节的总体结构如图 2 所示. 第 2.1 节首先给出标准形式的多智能体博弈优化框架, 接着介绍如何融合一致性控制等理论方法开展 Nash 均衡的分布式搜索, 以及不同结构下的多智能体博弈问题; 第 2.2 节在博弈框架下重新表述多智能体协同控制问题, 给出微分博弈框架的定义, 介绍基于最优控制的多智能体微分博弈问题的基本形式和扩展形式, 进一步在上述基础上介绍了多智能体图博弈, 将一致性问题放到微分博弈框架下考虑, 给出基于自适应动态规划 (Adaptive dynamic programming, ADP) 的均衡策略分析与计算方法; 第 2.3 节则针对博弈参与者具有状态概率转移的动态博弈过程, 介绍多智能体随机博弈的基本形式和特殊形式随机博弈均衡解的计算方法, 并探讨不同算法结构下的多智能体强化学习算法.

2.1 标准式博弈与分布式 Nash 均衡搜索

在多智能体博弈问题中, 每个智能体具有不同的优化目标, 主要关注智能体如何依赖自身信息以及与邻居智能体的信息交互优化自身目标函数.

博弈模型 $\Gamma = \{\mathcal{N}, \{\Omega_i\}_{i \in \mathcal{N}}, \{J_i\}_{i \in \mathcal{N}}\}$ 一般包括参与者 (即智能体)、策略集和效用函数三个要素. 一个标准式博弈问题可描述为

$$\min_{x_i \in \Omega_i} J_i(x_i, x_{-i}) \quad (3)$$

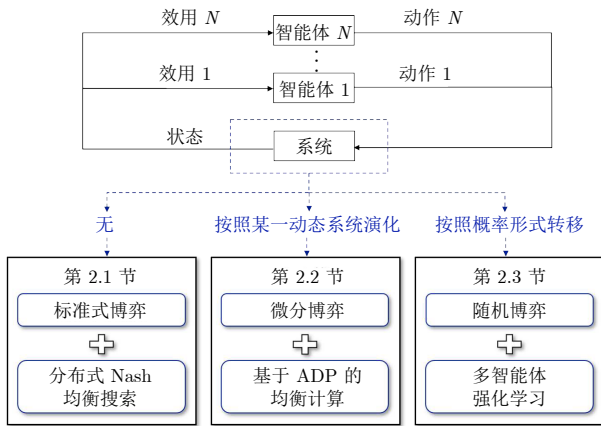


图2 第2节总体结构

Fig.2 General structure of Section 2

其中, x_i 表示智能体 i 的策略, x_{-i} 表示除智能体 i 以外其他所有智能体的策略组合. 为求解该博弈问题, 期望寻找一个 Nash 均衡策略组合 (x_i^*, x_{-i}^*) 满足

$$J_i(x_i^*, x_{-i}^*) \leq J_i(x_i, x_{-i}^*), \quad \forall i \in \mathcal{N}, \quad \forall x_i \in \Omega_i \quad (4)$$

式 (4) 表明, 在 Nash 均衡策略下, 任何智能体单方面改变自己的策略都不会使自身效用函数变小. 当考虑带有耦合约束的多智能体博弈问题^[94], 即智能体 i 的策略集 Ω_i 与其他智能体的策略 x_{-i} 有关, 则称满足式 (4) 的策略组合为广义 Nash 均衡 (Generalized Nash equilibrium) 策略.

2.1.1 分布式 Nash 均衡搜索算法的主要思想

Nash 均衡策略的存在性和唯一性与伪梯度向量的单调性密切相关. 一般地, 当伪梯度向量满足某些单调性条件时, 可保证 Nash 均衡的存在性和唯一性, 相应的多智能体博弈问题也称为单调博弈 (Monotone games). 尽管每个智能体的效用函数受到其他所有智能体策略的影响, 但由于通信能力和通信成本受限, 智能体一般只能与有限数量的其他个体通信. 因此, 相关研究工作用通信拓扑 $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$ 描述智能体之间的局部信息交流关系, 基于一致性控制的思想提出了多种分布式 Nash 均衡搜索算法. 相关算法的主要思想是为每个智能体设计一个关于所有智能体联合行动情况的局部估计, 使用某种一致性协议更新其估计, 并根据所得估计来更新动作. 文献 [95] 基于一致性思想提出基于节点和基于边的完全分布式 Nash 均衡搜索算法, 证明了当通信拓扑为无向连通图时 Nash 均衡具有全局渐近稳定性, 并将基于边的算法扩展到切换无向连通图的情况. 文献 [96–98] 则考虑不同的通信拓扑切换形式, 提出有效的分布式 Nash 均衡搜索算法,

并讨论其收敛至 Nash 均衡或其邻域内的条件. 除一致性控制思想以外, 研究学者还结合控制理论中的分析与综合技术, 如无源理论、小增益理论、输出调节、非光滑动力学等, 提出一系列新颖且有效的分布式 Nash 均衡搜索算法. 如, 文献 [99] 基于对无向图上多智能体协调问题的重述推导了新的动力学, 提出基于增量无源特性的分布式 Nash 均衡搜索算法, 并证明了其在任意连通图下的收敛性. 文献 [100] 提出基于输出调节理论的分布式 Nash 均衡搜索算法, 并通过输入–状态稳定性理论证明了算法收敛性.

2.1.2 不同结构下的多智能体博弈问题

在前述标准式博弈问题的基础上, 相关研究学者结合实际应用背景设置不同形式的效用函数, 探索了多种博弈结构下的多智能体博弈问题. 文献 [101–102] 研究了两个智能体零和博弈 (Zero-sum games) 问题, 其中两个智能体的收入或损失之和总为零或一个常数, 即一个智能体效用函数的增加必然导致另一智能体效用函数的降低. 因此, 零和博弈的效用函数通常表示为 $J(x_1, x_2)$, 其中智能体 1 的策略 x_1 期望最小化效用函数 J , 而智能体 2 的策略 x_2 则期望最大化该效用函数. 当效用函数满足某些凹凸性条件时, 零和博弈的 Nash 均衡, 也称鞍点均衡, 满足 $J(x_1, x_2^*) \geq J(x_1^*, x_2^*) \geq J(x_1^*, x_2)$, $\forall x_1 \in \Omega_1, x_2 \in \Omega_2$. 相关文献基于凹凸函数性质和鞍点性质进行探索, 提出了一系列零和博弈 Nash 均衡搜索算法. 此外, 考虑到在一些实际问题中, 智能体之间通过它们的聚合行为而非局部决策进行交互, 每个智能体 i 的效用函数可以写成 $J_i(x_i, x_{-i}) = \tilde{J}_i(x_i, l(x))$, 其中 $l(x)$ 表示所有智能体策略 x 的聚合行为. 这种具有特殊效用函数结构的博弈称为聚合博弈 (Aggregative games)^[103], 其广泛应用于电力市场、交通网络、意见网络、编队控制等领域中. 针对这种特殊的效用函数结构, 相关文献进一步研究了聚合博弈 Nash 均衡的定义、存在性、唯一性和分布式搜索方法等问题^[104–106]. 当智能体数量非常庞大时, 聚合博弈进一步演变为平均场博弈 (Mean-field games)^[107–108]. 势博弈 (Potential games) 也是一种重要的多智能体非合作博弈, 其要求所有智能体的效用函数满足某种全局特性. 具体地, 当存在势函数 P 使得对于每个智能体 i 和策略集 $x_{-i} \in \Omega_{-i}$ 都满足 $P(x_i, x_{-i}) - P(x'_i, x_{-i}) = J_i(x_i, x_{-i}) - J_i(x'_i, x_{-i})$, $\forall x_i, x'_i \in \Omega_i$, 那么将相应的博弈模型称为势博弈. 基于效用函数特性, 研究学者们提出了有效的势博弈 Nash 均衡计算方法^[109], 并在拥塞控制、分布式优化等多智能体决策问题中广泛应用^[110–111]. 在前述博弈问题中, 每个智能体作为利己个体优化

自身目标函数, 因此智能体之间在某些约束下相互制约, 形成竞争关系. 与此不同, 多集群博弈 (Multi-cluster games/multi-coalition games) 考虑智能体之间同时存在竞争与合作, 其中集群内部的智能体之间为合作关系, 而集群之间为竞争关系. 多集群博弈在一个统一框架下描述了分布式优化和非合作博弈问题. 具体地, 当只有一个集群时, 多集群博弈问题退化为分布式优化, 而当每个集群内只包含一个智能体时, 该问题则退化为标准式博弈. 相关的分布式 Nash 均衡搜索算法见文献 [112–113].

2.2 微分博弈与基于 ADP 的均衡策略计算

美国国家工程院院士、国际动态博弈论学会创始主席 Bašar 等^[114]在其著作中系统地介绍了动态非合作博弈理论 (Dynamic noncooperative game theory), 其本质是将单智能体最优控制问题推广到多智能体系统中, 为刻画智能体间的策略交互提供了一种有效框架. 与第 2.1 节中介绍的多智能体博弈不同, 这里涉及一个在连续时间内演化的动态决策过程, 即参与者的博弈状态按照某一微分方程演化, 因此也称为微分博弈 (Differential game).

2.2.1 微分博弈

一般地, 多智能体微分博弈问题可描述为

$$\begin{aligned} \min_{u_i} J_i(x, u_i, u_{-i}) &= \int_0^{\infty} r_i(x(t), u_i(t), u_{-i}(t)) dt \\ \text{s.t. } \dot{x} &= f(x, u_i, u_{-i}) \end{aligned}$$

其中, $x \in \mathbf{R}^n$ 表示刻画动态决策过程的博弈状态; $u_i \in \mathbf{R}^p$ 表示智能体 i 的控制输入 (控制策略); $u_{-i} = \{u_j, j = 1, \dots, N, j \neq i\}$ 表示除智能体 i 以外其他所有智能体控制输入的集合; f 表示博弈状态 x 的动力学方程; J_i 表示参与者 $i, i = 1, \dots, N$ 的累积性能指标; r_i 表示智能体 i 的立即成本, 其一般设置为关于 x, u_i, u_{-i} 的二次型函数. 上述多智能体微分博弈问题是由多输入系统最优控制演变而来, 其中每个智能体通过各自的控制输入 u_i 影响共同的博弈状态 x .

参与者 i 的 Nash 均衡策略 u_i^* 满足

$$u_i^* = \arg \min_{u_i} H_i(x, u_i, u_{-i}, \nabla V_i^*) \quad (5)$$

其中, $V_i^*(x(t)) = \int_t^{\infty} r_i(x(\tau), u_i^*(\tau), u_{-i}^*(\tau)) d\tau$ 表示智能体 i 的最优值函数, $\nabla V_i^* = \frac{\partial V_i^*}{\partial x}$ 表示 V_i^* 关于 x 的梯度, Hamiltonian 函数 $H_i(x, u_i, u_{-i}, \nabla V_i^*)$ 定义为

$$\begin{aligned} H_i(x, u_i, u_{-i}, \nabla V_i^*) &= \\ & r_i(x, u_i, u_{-i}) + (\nabla V_i^*)^T f(x, u_i, u_{-i}) \end{aligned}$$

由于均衡策略 (5) 依赖于最优值函数 V_i^* , 且 V_i^* 的计算需要使用 u_i^* 和 u_{-i}^* , 故无法直接显式计算均衡策略 u_i^* . 一般可通过求解如下耦合 Hamilton-Jacobian-Bellman (HJB) 方程来同时计算每个智能体的均衡策略:

$$H_i(x, u_i^*, u_{-i}^*, \nabla V_i^*) = 0, \forall i \quad (6)$$

在前述框架下, 文献 [115] 将动态系统的控制输入和外界干扰考虑为两个具有竞争关系的博弈参与者, 基于零和 Nash 博弈研究了动态系统 H_{∞} 控制问题. 文献 [116–117] 针对具有多个控制输入的动态系统, 研究了多智能体非零和 Nash 博弈问题. 考虑到智能体之间的地位和角色可能不同, 其可用于决策的信息并非对称, 文献 [118–120] 研究了多智能体 Stackelberg 博弈问题, 将博弈参与者分为领导者和跟随者, 其中跟随者总是根据观察到的领导者策略做出最优响应, 而领导者能够预测跟随者可能的响应策略并提前采取对自己最有利的策略. 此外, 文献 [121] 考虑博弈参与者数量庞大的情况, 研究了多智能体平均场随机微分博弈 (Mean-field stochastic differential game) 问题, 其中博弈状态动态方程和每个智能体的性能指标函数与所有智能体控制输入的数学期望有关.

2.2.2 图博弈与基于 ADP 的均衡策略计算

与前述工作不同, 文献 [122] 考虑每个智能体的动态性和性能指标只依赖于局部邻居信息的情况, 在多智能体系统一致控制的基础上结合微分博弈理论提出了图博弈的概念. 在图博弈中, 智能体间的交互关系通过局部通信拓扑 \mathcal{G} 来刻画. 基于此, 多智能体图博弈问题可描述为

$$\begin{aligned} \min_{u_i} J_i(\delta_i, u_i, u_{\mathcal{N}_i}) &= \int_0^{\infty} r_i(\delta_i, u_i, u_{\mathcal{N}_i}) dt \\ \text{s.t. } \dot{\delta}_i &= A\delta_i + (d_i + g_i)B_i u_i - \sum_{j \in \mathcal{N}_i} a_{ij} B_j u_j \end{aligned}$$

其中, $\delta_i = \sum_{j \in \mathcal{N}_i} a_{ij}(x_i - x_j) + g_i(x_i - x_0)$ 表示智能体 i 的局部一致性误差, x_i 为智能体 i 的状态且其动力学演化满足 $\dot{x}_i = Ax_i + B_i u_i$, u_i 表示智能体 i 的控制输入, x_0 为领导者状态且其动力学满足 $\dot{x}_0 = Ax_0$, $d_i = \sum_{j \in \mathcal{N}_i} a_{ij}$ 表示智能体 i 的加权入度, $g_i \geq 0$ 为牵引增益, 且当智能体 i 与领导者存在直接耦合时不等号严格成立, $u_{\mathcal{N}_i} = \{u_j, j \in \mathcal{N}_i\}$ 表示智能体 i 的邻居节点的控制输入的集合, $r_i(\delta_i, u_i, u_{\mathcal{N}_i})$ 表示智能体 i 的立即成本.

当通信拓扑 \mathcal{G} 为强连通图, 多智能体图博弈问题存在交互式 Nash 均衡 (Interactive Nash equilibrium) 策略 $\{u_1^*, \dots, u_N^*\}$, 满足

$$J_i(\delta_i, u_i^*, u_{\mathcal{N}_i}^*) \leq J_i(\delta_i, u_i, u_{\mathcal{N}_i}^*), \forall u_i, i \in \mathcal{N} \quad (7)$$

与第 2.1 节中给出的 Nash 均衡的定义不同, 交互式 Nash 均衡保证了每个参与者 k 都存在影响其他参与者 i 的策略, 即 $J_i(\delta_i, u_k^*, u_{\mathcal{N}_k}^*) \neq J_i(\delta_i, u_k, u_{\mathcal{N}_k}^*)$ 在任意 i, k 下成立. 理论结果表明, 均衡策略 u_i^* 能够使局部一致性误差 δ_i 趋于 0, 根据 δ_i 的定义及 Laplace 矩阵的性质可知, 多智能体系统可实现状态一致, 因此, 前述多智能体图博弈问题也称为多智能体系统最优一致控制问题.

在此框架下, 相关研究学者结合零和博弈^[123]、Stackelberg 博弈^[124]、平均场博弈^[125]等思想进行了一系列创新性研究工作. 与前述多智能体微分博弈类似, 这些工作本质上也需要求解一组耦合 HJB 方程来计算博弈均衡策略. 考虑到耦合 HJB 方程组本质上为一组耦合偏微分方程, 其解析解几乎难以计算, 为此, 研究学者们基于自适应动态规划、强化学习提出一系列新颖的算法用于估计耦合 HJB 方程组的解. 文献 [122] 提出一种基于系统动力学模型的在线同步策略迭代算法, 并证明在满足智能体间具有弱耦合的条件时值函数和控制策略可以收敛到相应的最优值. 为了避免算法对系统模型的依赖, 文献 [126] 提出基于执行器-评价器-辨识器结构的强化学习算法, 其中辨识器网络用于估计未知的系统动力学模型. 考虑到系统动力学模型辨识不可避免地存在误差, 而辨识误差的传播可能造成动态系统的演化偏离期望轨迹, 甚至引起系统振荡, 文献 [127-128] 充分利用采样数据, 提出不基于系统动力学模型的 off-policy 策略迭代算法. 策略迭代算法要求初始迭代策略能够使动态系统稳定, 这阻碍了策略迭代算法应用于复杂的多智能体系统协同任务上. 为此, 文献 [124] 提出基于值迭代的积分强化学习算法, 并讨论了通过迭代值函数在线判断系统稳定性的方法.

2.3 随机博弈与多智能体强化学习

20 世纪 50 年代早期, Lloyd Shapley 提出随机博弈 (Stochastic games) 的概念, 它引入了随机性和不确定性, 描述了多个参与者进行的具有状态概率转移的动态博弈过程. 在随机博弈中, 参与者的决策和环境状态的演变都以概率分布的形式呈现, 这种不确定性使得参与者在考虑博弈的同时适应环境的随机性. 受环境非平稳和维度爆炸等问题影响, 随机博弈均衡求解通常比较困难, 而多智能体强化学习 (Multi-agent reinforcement learning) 为随机博弈均衡策略求解提供了一种可行方法^[129]. 另外, 随机博弈在均衡状态解释以及算法收敛性和稳定性分析方面为多智能体强化学习提供了理论框架, 并

为学习框架和算法设计提供灵感.

随机博弈可由 $\langle N, S, \{A^i\}_{i \in N}, P, \{R^i\}_{i \in N}, \gamma \rangle$ 来定义, 其中, $N = \{1, \dots, n\}$ 为博弈参与者的数量, S 表示所有状态的集合, A^i 表示第 i 个参与者的动作空间, 记联合动作空间为 $A = A^1 \times A^2 \times \dots \times A^n$, $P: S \times A \rightarrow [0, 1]$ 表示状态概率转移函数, R^i 表示第 i 个参与者的回报函数, γ 表示折扣系数. 智能体 i 的策略定义为: $\pi_i: S \rightarrow \Omega(A^i)$, 表示从状态空间到智能体动作空间概率分布的映射. $\pi = [\pi_1, \dots, \pi_n]$ 表示所有智能体的联合策略. 初始状态为 s 时, 第 i 个智能体在策略 π 下的累积折扣值函数表示为

$$V_\pi^i(s) = \sum_{t=0}^{\infty} \gamma^t \mathbf{E}_\pi [R_t^i | s_0 = s, \pi] \quad (8)$$

结合 Bellman 方程可将动作值函数 $Q_\pi^i: S \times A^1 \times \dots \times A^n \rightarrow \mathbf{R}$ 表示为

$$Q_\pi^i(s) = R^i(s, a) + \gamma \mathbf{E}_{s' \sim P} [V_\pi^i(s')] \quad (9)$$

2.3.1 特殊形式随机博弈问题

由于动作值函数考虑的是联合状态空间 S 和联合动作空间 A , 均衡策略计算面临环境非平稳和维度爆炸等问题, 使得一般形式随机博弈问题的 Nash 均衡求解非常困难. 因此, 相关学者结合实际背景, 从回报函数设置、状态转移函数设置等角度探究了一些特殊形式随机博弈均衡解的计算问题. 文献 [130] 考虑每个智能体具有相同的回报函数, 即 $R^1 = \dots = R^n$, 研究了智能体之间为完全合作关系的团队博弈 (Team games), 提出了一系列有效的协作多智能体强化学习算法^[131-132]. 文献 [133] 考虑智能体可能具有不同的回报函数, 但其共同利益可由一个相同的势函数来描述的情况, 研究了随机势博弈均衡策略的特性及求解问题. 文献 [134] 考虑简化的线性二次情况, 即状态转移满足线性动力学, 回报函数为状态和动作的二次函数, 给出了丰富的理论结果. 文献 [135] 则基于平均场博弈思想简化值函数表示, 将联合动作空间的维度缩减为 $[a_i, \bar{a}_{-i}]$, 其中 \bar{a}_{-i} 为所有邻居智能体作用效果的平均, 提出平均场多智能体强化学习算法, 并探究了智能体数量接近无穷时所得解偏离 Nash 均衡的程度. 此外, 一些学者在随机博弈的基础上对智能体观察到的信息进行约束, 研究了部分可观随机博弈 (Partially observable stochastic games). 尽管不完美的状态信息进一步加剧了随机博弈均衡策略求解的困难, 一些学者通过引入信念状态或循环神经网络等技术来记忆历史观测并推断当前真实状态, 提出一系列部分可观多智能体强化学习算法. 注意到随机博弈建模了所有智能体同时行动而后获得回报

的情况,然而,在一些实际应用中,参与者往往交替采取行动,为此,一些工作运用树结构描述智能体的交替行动关系,研究了扩展型博弈 (Extensive games),考虑完全信息和不完全信息两种情况,定义了相应的 Nash 均衡并提出针对性的多智能体强化学习算法^[136-137].

2.3.2 不同算法结构下的多智能体强化学习算法

由于状态和动作空间的复杂度随着智能体数量的增加呈指数增长,学习算法结构,即多智能体训练和执行的方式,是多智能体强化学习算法研究中的一个重要问题.一种典型的学习算法结构是集中式训练-分布式执行 (Centralized learning and distributed execution),在这里,智能体通过开放的通信信道获得其他智能体的信息,以便从全局角度进行策略优化和训练,并在训练结束后基于自己的局部观测独立地采取行动,提高了算法在部分可观和具有有限通信能力环境下的适用性.基于这一结构,大量经典的多智能体强化学习算法相继提出,如 MADDPG^[138]、VDN^[139]、QMIX^[140]等.这些算法在智能体规模较小时表现出较好的性能,但是随着智能体数量的增加,集中式训练面临求解力和存储需求高、算法性能差等问题.因此,在大规模多智能体强化学习问题中,完全分布式的学习算法结构更加有效.相关研究工作基于智能体间的局部信息共享提出了一系列完全分布式多智能体强化学习算法.其中,文献^[141]运用有向无环图动态表示智能体间的策略通信关系,实现智能体策略控制相互协调.文献^[142]提出一种新颖的双层递归通信模型,以在有限通信情况下适应不断变化的通信对象.文献^[143]通过共享智能体局部新颖性促进协同策略搜索,在部分可观和稀疏奖励环境下取得较好的协同效果.文献^[144]基于因果推断选择对其他智能体影响最大的反事实通信内容,实现共享信息有限情况下的高效合作.

3 多智能体协同与博弈优化的应用

本节选取多移动机器人导航以及电动汽车充电调度两个典型多智能体协同场景,阐述博弈与控制融合的思想如何用于解决实际领域的难点问题.

3.1 多移动机器人导航

多移动机器人自主导航是一种通过多个机器人在动态和复杂环境中自主进行路径规划和运动控制,从而实现高效协作来完成目标的技术.它依靠传感器数据、控制算法和协调通信,使机器人能够独立或协同决策,灵活应对环境变化并安全避

开障碍物.多移动机器人自主导航技术在仓储物流、工业制造、农业、服务机器人、医疗、军事和救灾等多个领域展现出广泛的应用前景.该技术显著提升了各行业的效率与自动化水平,展示了其巨大的潜力和广泛的适用性.

多移动机器人自主导航有两种主要实现方案:模块化方案和端到端方案.模块化方案通过环境感知、任务分配、路径规划及运动控制等模块的组合或融合以实现导航,其中每个模块相对独立.模块化方案具有高可靠性和可解释性,但其结构较为复杂且依赖精确的环境模型.端到端方案利用深度强化学习技术,直接从原始传感器数据生成导航控制指令,简化了系统架构,具有自适应性强和潜在高性能的优势,但需要大量训练数据和计算资源,训练过程复杂且可解释性差.实际应用中,可根据需求选择合适的方案,或结合两者优势构建混合系统.

3.1.1 模块化方案

多移动机器人协同感知是指移动机器人利用各种传感器获取环境信息,然后借助一致性算法等工具对交互和共享信息协同处理,完成既定感知任务.协同定位和建图是其协同感知中的核心任务,是多移动机器人协同决策和控制的前提.协同定位和建图的过程首先是通过传感器完成原始环境信息的提取并对其进行预处理,其次是多机器人协同地对预处理后的环境信息进行融合,实现协同的位姿估计和全局环境地图构建.考虑到各类传感器的固有缺陷和彼此之间在环境信息提取方面的互补性,人们大多选择同时利用多种类型传感器来进行环境信息及机器人运动状态信息的采集,其中比较常用的组合包括视觉传感器与 IMU (Inertial measurement unit)^[145]、激光雷达与 IMU^[146]、以及视觉传感器和激光雷达及 IMU 三者的组合^[147].由于视觉传感器和激光雷达所提取到的数据规模巨大,通常需要从中提取特征用于定位和建图^[148-149].群组中的各机器人在提取到环境信息并对其进行预处理后,会首先在本地融合各类环境信息估计自身位姿并构建局部环境地图,然后在此基础上通过彼此间的信息交互,利用分布式协同估计等算法实现位姿和地图信息的协同融合.对位姿信息的协同融合多基于优化^[150]或滤波^[151].在协同位姿估计的基础上,通过局部地图的对齐、特征的匹配和融合^[152],就可以实现局部地图的融合,从而完成全局地图的构建.

多机器人任务分配是多机器人系统中执行复杂任务的基础,涉及任务的分解、分配和协调.传统的多机器人任务分配方法主要基于市场算法和任务分配图算法,前者将任务分配视为市场交易,机器人

作为买家,任务作为商品,通过模拟市场行为实现任务分配;后者使用图论方法建模任务和机器人之间的关系,通过图的匹配和分割算法实现任务分配^[153].其中,基于市场算法的多机器人任务分配方法可以与一致性这一分布式系统的核心理论工具相结合,形成分布式的任务分配方法.其基本原理是通过让每个机器人对任务进行竞价,采用拍卖机制决定任务分配,并利用一致性算法确保所有机器人对分配结果的理解和执行一致,从而实现高效、公平的任务分配,同时提高系统的鲁棒性和灵活性^[154-156].然而,上述的任务分配方法无法考虑机器人完成任务所需的成本与损耗(资源),更无法处理复杂动态环境中该资源是动态变化的问题.因此,文献^[157]提出一种基于博弈论的多机器人任务分配信息方法,将多机器人任务分配建模为一个博弈问题,每个机器人基于效用函数选择任务,通过 Nash 均衡理论实现任务的动态调整和最优化,并证明了该方法在提高任务分配效率和鲁棒性方面优于传统方法.

传统的单机器人导航通常将路径规划与运动控制模块独立,路径规划主要集中在单个机器人的路径优化和与障碍物的碰撞避免,常见的算法有 A*、Dijkstra 和 RRT 等;在路径生成后,采用传统基于模型的 PID 控制、反馈控制以及模型预测控制等来实现对参考路径的跟踪.然而,多机器人系统导航需要处理多个机器人之间的相互协调和碰撞避免问题,使得问题复杂度显著增加,且实时性要求更高,这也导致在多机器人系统中,越来越多的工作采用融合路径规划和运动控制的思想来设计多机器人运动规划方案^[158-161]以提高效率.而引入博弈来建模多机器人运动规划问题,可以通过动态模型和成本函数的定义,以适应各种复杂的环境和任务要求,并提供灵活且有效的多机器人运动规划方案.文献^[162]提出一种博弈规划器来执行无人机竞速任务,算法在竞速过程中预测对手无人机避免碰撞时可能的让步量,从而生成更具竞争力的竞速轨迹;文献^[163]提出一种可扩展的交互式轨迹优化算法并证明了机器人间的相互作用可以通过博弈论的形式描述,其中交互的结果非常适合用基础动态博弈的均衡状态来进行建模;文献^[164]采用微分博弈的框架来解决多机器人避碰问题,并给出安全性的理论保障.此外,博弈这一工具还被证明可以有效建模复杂的人-集群合作问题,文献^[165]将操作者控制集群机器人避障问题表述为一个 Stackelberg-Nash 博弈,其中假设领导者机器人处于主导地位,直接与人类操作员交互,实现轨迹跟踪并负责引导群体避开障碍物;而跟随机器人始终根据主导机器人的行为做

出最佳响应,目的是实现期望的队形.

3.1.2 端到端方案

近年来,考虑到基于深度强化学习的导航技术可以使机器人具备与复杂环境交互的能力,基于其端到端的导航方法已在移动机器人领域引起广泛的研究兴趣^[166].然而多机器人系统的导航除了与环境的交互,还必须考虑到大规模的机器人之间的交互带来的竞争与合作问题.针对该问题,随机博弈与多智能体强化学习提供了一个强大的框架来处理这种多机器人系统的交互、竞争和合作^[167].以多机器人协同避碰为例,多机器人系统的状态空间包括每个机器人的位置、速度、方向,以及周围环境的感知信息,这对应于随机博弈中的状态空间,其中环境的状态由所有机器人及其感知的信息共同定义.每个机器人可以执行的动作(如前进、后退、左转、右转等)对应于随机博弈中的动作空间,每个智能体选择的动作不仅影响自身状态,还会影响其他智能体的状态和整体环境状态.机器人的状态转移函数基于当前状态和动作来定义.奖励函数包括机器人成功收集物品、避免碰撞等,这些奖励函数引导机器人在复杂环境中高效导航和避障;深度强化学习算法通过神经网络估计 Q 值函数来学习策略,这与随机博弈中智能体学习最优策略的过程一致;采用的集中训练分散执行框架在训练阶段集中利用所有机器人的信息进行训练,但在执行阶段每个机器人独立决策,这种方法实际上是随机博弈的一种实现方式,在训练阶段通过共享全局信息优化各智能体的协作与竞争能力,而在执行阶段每个智能体基于局部信息独立决策,相比于传统的集中式/分布式方法,可以有效提高系统的鲁棒性和可扩展性.综上所述,随机博弈与多智能体强化学习为多智能体导航问题提供了一个结构化的解决方案,通过深度强化学习方法,可以有效地训练多智能体在复杂环境中的导航策略.

基于强化学习的多机器人导航通过自主学习使机器人在复杂和动态环境中实现高效导航,并展现出很强的适应性和协作能力.通过分布式决策,每个机器人能够独立行动,可以显著增强系统的鲁棒性和扩展性.强化学习方法能够处理机器人间的复杂交互,提升整体导航效率.然而,这种方法也面临一些挑战,如缺少安全保障、高计算资源需求、训练过程复杂、探索-利用权衡问题显著、对环境变化的鲁棒性不足以及奖励函数设计难.此外,多智能体强化学习的集中训练分散执行框架也面临非平稳性、通信约束和奖励分配等技术挑战.针对这些问题,单纯使用强化学习方法难以完全克服,需要引入传统方法来设计混合的导航策略,从而全方面提高多

机器人导航算法的安全性、鲁棒性和泛化性. 文献 [168] 首先提出一种将深度强化学习导航策略与传统运动控制策略相结合的混合式导航方法, 针对一些特殊场景设计了针对性的传统控制策略, 如较开放场景的直线加速策略、不安全场景下的避障策略, 并设计了启发式的策略切换规则. 实验结果表明, 这种混合策略能够为大规模非完整机器人系统找到高效且无碰撞的路径, 并且可以安全地推广到未见过的模拟和现实场景中. 文献 [169] 设计了深度强化学习与模型预测控制的混和导航策略, 将深度强化学习策略作为模型预测控制器的全局参考, 由模型预测控制生成控制输入, 从而理论上保证了运动轨迹的安全性. 文献 [170] 进一步引入了速度障碍方法, 使用深度强化学习来优化避障策略, 进一步提高了移动机器人的避障和导航性能. 文献 [171-173] 特别针对多机器人系统引入相互速度障碍法以及其改进方法, 进一步提高了多机器人系统的避障性能. 特别地, 文献 [172] 在神经网络的设计中综合考虑了机器人运动的时空特征, 从而使所训练出的策略具有更高的效率, 有效降低了整个系统的任务完成时间. 文献 [173] 在奖励函数中引入了社会规则的考量, 所训练的导航与避障策略融入了如右侧通行、左侧超车等社会交通规则, 可以更好地应用于人-机-物混合的复杂应用场景.

3.2 电动汽车充电调度

在能源危机和环境问题日益突出的当今时代, 节能减排和可持续发展已经成为全社会关注的焦点. 电动汽车作为一种具备储能功能的移动式电力负载, 具有“源-荷-储”三重属性, 在建设安全经济环保的智能电力系统方面起着重要的作用, 它也是解决交通、能源和环境问题的重要手段^[174]. 电动汽车充电调度是电动汽车研究中的基础问题, 其指通过智能管理系统和优化策略对电动汽车充电行为进行规划和控制, 以实现降低用户成本等目标^[175]. 注意到, 电动汽车充电调度涉及多个参与者, 如大量电动汽车用户、充电站运营商、电网公司等, 多智能体系统博弈优化为这些参与者之间的协调、优化与管理提供了一种有效的方式.

3.2.1 电动汽车充电调度优化目标

将电动汽车充电调度问题建模成优化问题并进行求解是解决该类问题的最主要方式, 其通常通过对充电功率、充电电价等决策变量进行优化, 以实现电动汽车用户充电成本最小化、促进削峰填谷、减少碳排放等目标.

实现电动汽车用户充电成本/能源成本最小化

是电动汽车充电调度的最基本目标. 将电动汽车 i 在时间 t 的充电价格/能源消耗表示为

$$c_{i,t}^{\text{cha}} = p_t P_{i,t} \quad (10)$$

其中, p_t 为电价或为 1, $P_{i,t}$ 为电动汽车 i 的充电功率/能源消耗, 那么其优化目标函数可以记做所有电动汽车在一定时间内充电总价或者消耗的能源之和^[176-178]

$$C_{\text{charging}} = \sum_i \sum_t c_{i,t}^{\text{cha}} \quad (11)$$

对于电动汽车用户个体来说, 还有一些其他因素需要同时考虑, 如, 文献 [176] 在设置优化目标函数时, 除了考虑分布式发电机和变电站的能源消耗, 还增加了额外项来鼓励优先级高的电动汽车在开始的几个时段充电, 并将其建模成混合整数规划问题进行求解; 文献 [177] 同时考虑了电动汽车充电成本、电池退化成本等, 并基于动态规划方法设计了一种多目标优化算法来平衡多种目标; 文献 [178] 在考虑含电动汽车的住宅用户的需求响应问题时, 考虑了电动汽车的充电能源消耗、充电需求未满足导致的不满意度以及电池退化成本三种因素, 并设计了一种无模型的强化学习方法, 解决了未知市场环境下的需求响应管理难题.

除了考虑充电成本外, 还常利用大规模电动汽车协调调度来对电网进行调控, 实现削峰填谷、减少电网波动目标, 以保证电网的稳定性. 将 t 时刻区域电网总负载与平均负载差值的平方和表示为

$$c_t^{\text{grid}} = (P_{\text{load},t} - P_{\text{ave}})^2 \quad (12)$$

其中, $P_{\text{load},t} = \sum_i P_{\text{EV},i,t} + P_{\text{NoEV},t}$ 为所有电动汽车充电负载和其他负载之和, P_{ave} 表示区域电网平均负载情况, 可以是 $P_{\text{load},t}$ 的均值, 那么该优化目标函数可以记做一定时间内区域电网总负载与平均负载差值的平方和^[179-182]:

$$C_{\text{grid}} = \sum_t c_t^{\text{grid}} \quad (13)$$

针对电网负荷优化问题, 文献 [179] 建立了一种统计模型对电动汽车充电负载进行预测, 并在此基础上利用序列二次规划方法给出每个时段电动汽车的充电量; 文献 [180] 考虑了一个非住宅区域的电动汽车调度, 结合历史数据并利用内点法给出了电动汽车的充电功率, 实现了该区域的削峰填谷; 文献 [181] 则将削峰填谷目标和充电成本最小化目标作为模型预测控制的优化目标函数, 设计了电动汽车充电控制算法.

上述优化目标是电动汽车充电调度中主要研究

的两类目标,也有一些其他类型的充电调度目标被研究.例如,文献[183]聚焦于电动汽车生命周期的碳排放,通过数据对电动汽车排放模型拟合,再利用模型对电动汽车群体建立优化模型,最终进行充电规划以减少碳排放;文献[184]额外考虑可再生能源发电,研究包含新能源发电机、储能系统以及电动汽车充电站的配电网的经济管理问题,同时考虑了碳排放因素,并通过优化方法和启发式方法相结合,最终分层求出最优策略;文献[185]考虑了配电网和交通网耦合下的电动汽车充电导航问题,同时考虑了电动汽车的时间成本和充电成本,并提出一种基于分布式有偏最小一致性算法的导航方法.

3.2.2 基于博弈的电动汽车充电调度

电动汽车充电调度方法较多的是采取整体优化的思想,然而对于自私的用户群体来说,难以保证他们都遵循整体最优的策略,因为该策略可能会牺牲个体自身利益.因此,现有一些工作也开始将电动汽车用户及充电站间的交互关系建立为博弈模型.在博弈问题中,各主体*i*仅考虑最小化自身的收益,而并不尝试达成整体的最优:

$$\min C_{\text{user}, i} \quad (14)$$

这里的主体可以是电动汽车用户、充电站运营商以及其他利益相关者.注意这里个体的优化目标和第3.2.1节的优化目标有着一定的联系,例如,电动汽车期望最小化充电成本可以取

$$C_{\text{user}, i} = \sum_t c_{i,t}^{\text{cha}} \quad (15)$$

充电站运营商期望减少自身负载波动可以取

$$C_{\text{user}, i} = \sum_t c_t^{\text{grid}} \quad (16)$$

在博弈模型下,主要研究多利益主体在相关作用下达到的系统稳定状态,即博弈均衡.需要说明的是,在大多数情况下博弈均衡和整体优化的最优解是不同的,当然也有一些工作考虑了特殊的情形,其中博弈均衡与优化最优解一致^[186].

电动汽车用户之间通常可以建模成非合作博弈.文献[187]探究了单个充电站内具有不同充电需求的电动汽车用户间的博弈,这些用户由于受到充电站功率容量上限的耦合约束而构成非合作关系,在设计个体效用函数时,也会引入对应的成本项用于促使电动汽车用户达成均衡.文献[188]则进一步考虑了基于太阳能的充电站,引入马尔科夫链建模太阳能的不确定性,并基于相关决策对博弈模型建模,这样的模型更加贴合实际.

此外,还有一些文献同时探究多个充电站和用户的场景,由于充电站构成竞争关系,往往也会将其建模成非合作博弈.文献[189]基于非合作博弈方法探究了多个充电站的投资者在不同情形下的定价、选址决策,并通过实验分析加入补贴对充电站运营商和电动汽车用户的影响.文献[190]考虑了充电设施有限的情形,并通过多个代理商竞价竞争博弈后签订共享充电桩协议的方式实现各代理商的利润最大化.文献[191]将多充电站的充电规划分为两层:第1层决策总体的功率分配;第2层再通过电动汽车的非合作博弈决策每辆车的充电功率.

注意在用户数量较大的情形下,上述非合作博弈模型常进一步建模为聚合博弈和平均场博弈.这两类博弈中,均会使用一个聚合项表示整个群体的影响.以 x_i 表示个体*i*的决策,并以 $x = (x_1, x_2, \dots, x_N)$ 表示所有智能体的决策,则其聚合项形如

$$G(x) = \sum_{i=1}^N f_i(x_i)$$

平均场博弈是对聚合博弈中大样本的近似,即考虑 $N \rightarrow \infty$ 时的博弈均衡,此时聚合项常形如

$$G(x) = \bar{f}(x_i)$$

对应的博弈问题形式为

$$\min C_{\text{user}, i}(x_i, G(x)) \quad (17)$$

文献[192]使用聚合博弈探究电动汽车充电的日前最优策略,其电价由电动汽车聚合项总充电量决定.文献[193]考虑大样本下电动汽车用户的充电控制问题,因此使用平均场博弈进行建模,聚合项为总用电功率,其效用函数为充电功率的二次函数,并允许反向售电.

除了非合作博弈模型,多个充电站之间也可通过电网构建合作关系,实现多方成本的降低,这便构成了合作博弈.如文献[194–195]探究了多个充电站的合作,允许充电站之间进行电力交流,依此建立了Nash议价模型分配额外利润.

此外,由于电动汽车同时具有电网、路网性质,一些工作将问题背景扩展至车路网系统中.如文献[196]建立博弈模型,研究了对电动汽车用户的充换电站推荐问题.文献[197]探究了路网均衡条件下,电动汽车用户与快速充电站的斯塔克伯格博弈均衡.

3.2.3 基于多智能体强化学习的电动汽车充电调度

考虑到电动汽车所处的环境,如电动汽车到达充电站的时间以及相应的电池荷电状态(State of charge, SoC)等具有很强的随机性和不确定性,也

常将电动汽车充电调度问题建模成随机博弈模型, 并利用多智能体强化学习方法进行求解。

以简单的 N 个电动汽车参与充电调度场景为例, 其随机博弈模型六元组 $\langle N, S, A^i, P, R^i, \gamma \rangle$ 中, 状态集合 S 主要包含电动汽车电池的荷电状态, 动作空间 A^i 可以是电动汽车 i 是否充电以及充电功率的集合, 状态转移函数 P 刻画了电动汽车荷电状态的演化动力学, R^i 则是电动汽车 i 的回报函数。回报函数和电动汽车充电调度优化目标是强相关的, 例如, 其可以设置为 $-c_{i,t}^{\text{cha}}$ 。

基于电动汽车充电调度的随机博弈模型, 可以设计多智能体强化学习方法进行求解。文献 [198] 在设计回报函数时考虑了充电成本、用户满意度等因素, 并提出一种多智能体深度强化学习算法, 该算法设计了一种基于极限学习机的前馈神经网络用于电价和电动汽车通勤行为预测, 以及一种 Q 网络用于最优动作值函数逼近; 文献 [199] 则进一步考虑了充电站变压器负载情况以减少充电站电压波动, 在所提出的多智能体深度强化学习算法中使用一种称为 CommNet 的通信神经网络模型来实现全局信息的分布式计算。

鉴于多智能体强化学习中集中式训练-分布式执行框架的优势, 结合该框架下的 MADDPG、VDN、QMIX 等多智能体强化学习算法, 面向不同场景的电动汽车充电调度算法也相继提出。例如, 文献 [200] 针对带有光伏发电板和储能设备的充电站内电动汽车充电调度问题, 基于 COMA 和 MADDPG 算法提出了电动汽车协调充电算法, 并在电动汽车用户满意度和充电成本上得到折中; 文献 [201] 考虑多个充电站的能量采购和分配问题并提出一种多智能体强化学习方法来学习最优能源购买策略, 该方法结合用于学习购买策略的 MADDPG 算法和用于预测电动汽车充电需求的 LSTM (Long short-term memory) 神经网络, 实验表明, 该算法下的经济利润和电动汽车用户满意度方面均得到很好的提升。此外, 文献 [202] 将 VDN 和二进制整数规划相结合, 提出了 VDN-BLP 算法来解决电池换电充电系统中换电站、充电站、电池运载卡车之间的实时联合优化调度问题; 文献 [203] 将竞标机制融入 QMIX 网络中, 设计了一种合作多智能体强化学习算法来将电动汽车充电调度集成到多个充电站实现电网调峰。

4 总结与展望

本文从协同角度入手, 回顾了多智能体协同控制与估计领域的进展。接着围绕博弈与控制的融合,

介绍了博弈框架的基本概念, 重点讨论了在微分博弈下多智能体协同问题的建模与分析, 并简要总结了如何应用强化学习算法求解博弈的均衡。本文还选取了多移动机器人导航和电动汽车充电调度这两个典型的多智能体协同场景, 介绍了博弈与控制融合的思想如何用于解决相关领域的难点问题。

虽然多智能体博弈优化领域已经取得一些研究成果, 但仍面临智能体规模难以扩大、理论结果不完备、博弈均衡搜索算法可解释性差等一系列瓶颈问题。此外, 面对具体复杂任务, 如何设计个体优化目标并有效刻画个体间博弈机制, 从而深入融合目标任务和通信规则设计出符合特定任务要求的高效决策算法, 也是未来多智能体博弈优化的重要发展方向之一。结合机器学习、博弈论、控制论、经济学等领域知识开展跨学科研究, 制定标准化测试基准并加强与实际应用场景的结合, 有助于推进多智能体博弈优化领域发展。

此外, 面向军事需求, 多移动机器人系统需要在开放、未知的战场环境中开展大规模协同对抗, 如何融合博弈与控制理论, 开展复杂的战术和战略规划, 从而提升大规模机器人自主决策能力, 是一个亟待解决的难题。另一方面, 在电动汽车充电调度领域, 随着研究场景中参与者类型和数量的增加, 例如在配电-交通耦合网络等情况下, 如何设计有效的协调调度策略仍是一个亟需解决的挑战。

References

- 1 Yang Tao, Yang Bo, Yin Yun-Qiang, Yu Wen-Wu, Xia Yuan-Qing, Hong Yi-Guang. Guest editorial of special issue on cooperative control and optimization for multi-agent systems. *Control and Decision*, 2023, 38(5): 1153-1158 (杨涛, 杨博, 殷允强, 虞文武, 夏元清, 洪奕光. 多智能体系统协同控制与优化专刊序言. *控制与决策*, 2023, 38(5): 1153-1158)
- 2 Moreau L. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 2005, 50(2): 169-182
- 3 Cao M, Morse A S, Anderson B D O. Reaching a consensus in a dynamically changing environment: Convergence rates, measurement delays, and asynchronous events. *SIAM Journal on Control and Optimization*, 2008, 47(2): 601-623
- 4 Shi G D, Johansson K H. The role of persistent graphs in the agreement seeking of social networks. *IEEE Journal on Selected Areas in Communications*, 2013, 31(9): 595-606
- 5 Qin J H, Gao H J. A sufficient condition for convergence of sampled-data consensus for double-integrator dynamics with nonuniform and time-varying communication delays. *IEEE Transactions on Automatic Control*, 2012, 57(9): 2417-2422
- 6 Qin J H, Zheng W X, Gao H J. Consensus of multiple second-order vehicles with a time-varying reference signal under directed topology. *Automatica*, 2011, 47(9): 1983-1991
- 7 Qin J H, Gao H J, Zheng W X. Exponential synchronization of complex networks of linear systems and nonlinear oscillators: A unified analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(3): 510-521
- 8 Lin Z L. *Low Gain Feedback*. London: Springer, 1999.
- 9 Qin J H, Fu W M, Zheng W X, Gao H J. On the bipartite consensus for generic linear multiagent systems with input saturation.

- tion. *IEEE Transactions on Cybernetics*, 2017, **47**(8): 1948–1958
- 10 Meskin N, Khorasani K. Actuator fault detection and isolation for a network of unmanned vehicles. *IEEE Transactions on Automatic Control*, 2009, **54**(4): 835–840
- 11 Dimarogonas D V, Frazzoli E, Johansson K H. Distributed event-triggered control for multi-agent systems. *IEEE Transactions on Automatic Control*, 2012, **57**(5): 1291–1297
- 12 Qin J H, Ma Q C, Shi Y, Wang L. Recent advances in consensus of multi-agent systems: A brief survey. *IEEE Transactions on Industrial Electronics*, 2017, **64**(6): 4972–4983
- 13 Qin J H, Yu C B, Gao H J. Coordination for linear multiagent systems with dynamic interaction topology in the leader-following framework. *IEEE Transactions on Industrial Electronics*, 2014, **61**(5): 2412–2422
- 14 Zhang Y Y, Li S. Distributed biased min-consensus with applications to shortest path planning. *IEEE Transactions on Automatic Control*, 2017, **62**(10): 5429–5436
- 15 Qin J H, Gao H J, Zheng W X. Second-order consensus for multi-agent systems with switching topology and communication delay. *Systems and Control Letters*, 2011, **60**(6): 390–397
- 16 Zhang J F. Preface to special topic on games in control systems. *National Science Review*, 2020, **7**(7): 1115–1115
- 17 Shamma J S. Game theory, learning, and control systems. *National Science Review*, 2020, **7**(7): 1118–1119
- 18 Wang Long, Huang Feng. An interdisciplinary survey of multi-agent games, learning, and control. *Acta Automatica Sinica*, 2023, **49**(3): 580–613
(王龙, 黄锋. 多智能体博弈、学习与控制. 自动化学报, 2023, **49**(3): 580–613)
- 19 Marden J R, Shamma J S. Game theory and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2018, **1**: 105–134
- 20 Riehl J, Ramazi P, Cao M. A survey on the analysis and control of evolutionary matrix games. *Annual Reviews in Control*, 2018, **45**: 87–106
- 21 Zhang R R, Guo L. Controllability of Nash equilibrium in game-based control systems. *IEEE Transactions on Automatic Control*, 2019, **64**(10): 4180–4187
- 22 Huo W, Tsang K F E, Yan Y, Johansson K H, Shi L. Distributed Nash equilibrium seeking with stochastic event-triggered mechanism. *Automatica*, 2024, **162**: Article No. 111486
- 23 Oh K K, Park M C, Ahn H S. A survey of multi-agent formation control. *Automatica*, 2015, **53**: 424–440
- 24 Anderson B D O, Shi G D, Trunpf J. Convergence and state reconstruction of time-varying multi-agent systems from complete observability theory. *IEEE Transactions on Automatic Control*, 2017, **62**(5): 2519–2523
- 25 Xiao F, Wang L. Asynchronous consensus in continuous-time multi-agent systems with switching topology and time-varying delays. *IEEE Transactions on Automatic Control*, 2008, **53**(8): 1804–1816
- 26 Kim H, Shim H, Back J, Seo J H. Consensus of output-coupled linear multi-agent systems under fast switching network: Averaging approach. *Automatica*, 2013, **49**(1): 267–272
- 27 Back J, Kim J S. Output feedback practical coordinated tracking of uncertain heterogeneous multi-agent systems under switching network topology. *IEEE Transactions on Automatic Control*, 2017, **62**(12): 6399–6406
- 28 Valcher M E, Zorzan I. On the consensus of homogeneous multi-agent systems with arbitrarily switching topology. *Automatica*, 2017, **84**: 79–85
- 29 Qin J H, Gao H J, Yu C B. On discrete-time convergence for general linear multi-agent systems under dynamic topology. *IEEE Transactions on Automatic Control*, 2014, **59**(4): 1054–1059
- 30 Yang T, Meng Z Y, Shi G D, Hong Y G, Johansson K H. Network synchronization with nonlinear dynamics and switching interactions. *IEEE Transactions on Automatic Control*, 2016, **61**(10): 3103–3108
- 31 Lu M B, Liu L. Distributed feedforward approach to cooperative output regulation subject to communication delays and switching networks. *IEEE Transactions on Automatic Control*, 2017, **62**(4): 1999–2005
- 32 Meng H F, Chen Z Y, Middleton R. Consensus of multiagents in switching networks using input-to-state stability of switched systems. *IEEE Transactions on Automatic Control*, 2018, **63**(11): 3964–3971
- 33 Liu T, Huang J. Leader-following attitude consensus of multiple rigid body systems subject to jointly connected switching networks. *Automatica*, 2018, **92**: 63–71
- 34 Meng Z Y, Yang T, Li G Q, Ren W, Wu D. Synchronization of coupled dynamical systems: Tolerance to weak connectivity and arbitrarily bounded time-varying delays. *IEEE Transactions on Automatic Control*, 2018, **63**(6): 1791–1797
- 35 Abdessameud A. Consensus of nonidentical Euler-Lagrange systems under switching directed graphs. *IEEE Transactions on Automatic Control*, 2019, **64**(5): 2108–2114
- 36 Su Y F, Huang J. Stability of a class of linear switching systems with applications to two consensus problems. *IEEE Transactions on Automatic Control*, 2012, **57**(6): 1420–1430
- 37 Wang X P, Zhu J D, Feng J E. A new characteristic of switching topology and synchronization of linear multiagent systems. *IEEE Transactions on Automatic Control*, 2019, **64**(7): 2697–2711
- 38 Ma Q C, Qin J H, Zheng W X, Shi Y, Kang Y. Exponential consensus of linear systems over switching network: A subspace method to establish necessity and sufficiency. *IEEE Transactions on Cybernetics*, 2022, **52**(3): 1565–1574
- 39 Ma Q C, Qin J H, Yu X H, Wang L. On necessary and sufficient conditions for exponential consensus in dynamic networks via uniform complete observability theory. *IEEE Transactions on Automatic Control*, 2021, **66**(10): 4975–4981
- 40 Ma Q C, Qin J H, Anderson B D O, Wang L. Exponential consensus of multiple agents over dynamic network topology: Controllability, connectivity, and compactness. *IEEE Transactions on Automatic Control*, 2023, **68**(12): 7104–7119
- 41 Bernstein D S, Michel A N. A chronological bibliography on saturating actuators. *International Journal of Robust and Nonlinear Control*, 1995, **5**(5): 375–380
- 42 Zhou B, Duan G R, Lin Z L. A parametric Lyapunov equation approach to the design of low gain feedback. *IEEE Transactions on Automatic Control*, 2008, **53**(6): 1548–1554
- 43 Su H S, Chen M Z Q, Lam J, Lin Z L. Semi-global leader-following consensus of linear multi-agent systems with input saturation via low gain feedback. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2013, **60**(7): 1881–1889
- 44 Li Y, Xiang J, Wei W. Consensus problems for linear time-invariant multi-agent systems with saturation constraints. *IET Control Theory and Applications*, 2011, **5**(6): 823–829
- 45 Meng Z Y, Zhao Z Y, Lin Z L. On global leader-following consensus of identical linear dynamic systems subject to actuator saturation. *Systems and Control Letters*, 2013, **62**(2): 132–142
- 46 Ren W, Beard R W. Consensus algorithms for double-integrator dynamics. *Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications*. London: Springer, 2008. 77–104
- 47 Zhao Z Y, Lin Z L. Global leader-following consensus of a group of general linear systems using bounded controls. *Automatica*, 2016, **68**: 294–304
- 48 Zhang Y M, Jiang J. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control*, 2008, **32**(2): 229–252
- 49 Davoodi M R, Khorasani K, Talebi H A, Momeni H R. Distributed fault detection and isolation filter design for a network of heterogeneous multiagent systems. *IEEE Transactions on Control Systems Technology*, 2014, **22**(3): 1061–1069
- 50 Kashyap N, Yang C W, Sierla S, Flikkema P G. Automated

- fault location and isolation in distribution grids with distributed control and unreliable communication. *IEEE Transactions on Industrial Electronics*, 2015, **62**(4): 2612–2619
- 51 Teixeira A, Shames I, Sandberg H, Johansson K H. Distributed fault detection and isolation resilient to network model uncertainties. *IEEE Transactions on Cybernetics*, 2014, **44**(11): 2024–2037
- 52 Wang Y J, Song Y D, Lewis F L. Robust adaptive fault-tolerant control of multiagent systems with uncertain nonidentical dynamics and undetectable actuation failures. *IEEE Transactions on Industrial Electronics*, 2015, **62**(6): 3978–3988
- 53 Chen S, Ho D W C, Li L L, Liu M. Fault-tolerant consensus of multi-agent system with distributed adaptive protocol. *IEEE Transactions on Cybernetics*, 2015, **45**(10): 2142–2155
- 54 Tabuada P. Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Transactions on Automatic Control*, 2007, **52**(9): 1680–1685
- 55 Cao M T, Xiao F, Wang L. Event-based second-order consensus control for multi-agent systems via synchronous periodic event detection. *IEEE Transactions on Automatic Control*, 2015, **60**(9): 2452–2457
- 56 Lu W L, Han Y J, Chen T P. Synchronization in networks of linearly coupled dynamical systems via event-triggered diffusions. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(12): 3060–3069
- 57 Fan Y, Feng G, Wang Y, Song C. Distributed event-triggered control of multi-agent systems with combinational measurements. *Automatica*, 2013, **49**(2): 671–675
- 58 Garcia E, Cao Y C, Casbeer D W. Decentralized event-triggered consensus with general linear dynamics. *Automatica*, 2014, **50**(10): 2633–2640
- 59 Seyboth G S, Dimarogonas D V, Johansson K H. Event-based broadcasting for multi-agent average consensus. *Automatica*, 2013, **49**(1): 245–252
- 60 Zhu W, Jiang Z P. Event-based leader-following consensus of multi-agent systems with input time delay. *IEEE Transactions on Automatic Control*, 2015, **60**(5): 1362–1367
- 61 Cheng Y, Ugrinovskii V. Event-triggered leader-following tracking control for multivariable multi-agent systems. *Automatica*, 2016, **70**: 204–210
- 62 Mu N K, Liao X F, Huang T W. Event-based consensus control for a linear directed multiagent system with time delay. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2015, **62**(3): 281–285
- 63 Altafini C. Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, 2013, **58**(4): 935–946
- 64 Cartwright D, Harary F. Structural balance: A generalization of Heider's theory. *Psychological Review*, 1956, **63**(5): 277–293
- 65 Meng Z Y, Shi G D, Johansson K H, Cao M, Hong Y G. Behaviors of networks with antagonistic interactions and switching topologies. *Automatica*, 2016, **73**: 110–116
- 66 Qin J H, Yu C B, Anderson B D O. On leaderless and leader-following consensus for interacting clusters of second-order multi-agent systems. *Automatica*, 2016, **74**: 214–221
- 67 Qin J H, Yu C B. Cluster consensus control of generic linear multi-agent systems under directed topology with acyclic partition. *Automatica*, 2013, **49**(9): 2898–2905
- 68 Ren L, Li M, Sun C Y. Semiglobal cluster consensus for heterogeneous systems with input saturation. *IEEE Transactions on Cybernetics*, 2021, **51**(9): 4685–4694
- 69 Qin J H, Ma Q C, Gao H J, Shi Y, Kang Y. On group synchronization for interacting clusters of heterogeneous systems. *IEEE Transactions on Cybernetics*, 2017, **47**(12): 4122–4133
- 70 Xia W G, Cao M. Clustering in diffusively coupled networks. *Automatica*, 2011, **47**(11): 2395–2405
- 71 Battistelli G, Chisci L, Mugnai G, Farina A, Graziano A. Consensus-based linear and nonlinear filtering. *IEEE Transactions on Automatic Control*, 2015, **60**(5): 1410–1415
- 72 Battistelli G, Chisci L. Stability of consensus extended Kalman filter for distributed state estimation. *Automatica*, 2016, **68**: 169–178
- 73 Zhang C, Qin J H, Li H, Wang Y N, Wang S, Zheng W X. Consensus-based distributed two-target tracking over wireless sensor networks. *Automatica*, 2022, **146**: Article No. 110593
- 74 Chen Q, Yin C, Zhou J, Wang Y, Wang X Y, Chen C Y. Hybrid consensus-based cubature Kalman filtering for distributed state estimation in sensor networks. *IEEE Sensors Journal*, 2018, **18**(11): 4561–4569
- 75 Guo M, Jayawardhana B. Simultaneous distributed localization, formation, and group motion control: A distributed filter approach. *IEEE Transactions on Control of Network Systems*, 2024, **11**(4): 1867–1878
- 76 Sun W W, Lv X Y, Qiu M Y. Distributed estimation for stochastic Hamiltonian systems with fading wireless channels. *IEEE Transactions on Cybernetics*, 2022, **52**(6): 4897–4906
- 77 Chen W, Wang Z D, Ding D R, Yi X J, Han Q L. Distributed state estimation over wireless sensor networks with energy harvesting sensors. *IEEE Transactions on Cybernetics*, 2023, **53**(5): 3311–3324
- 78 Kalman R E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 1960, **82**(1): 35–45
- 79 Ljung L. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 1979, **24**(1): 36–50
- 80 Wan E A, van der Merwe R. The unscented Kalman filter. *Kalman Filtering and Neural Networks*. New York: Wiley, 2001. 221–280
- 81 Julier S J, Uhlmann J K. Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. In: Proceedings of the American Control Conference (IEEE Cat. No. CH37301). Anchorage, USA: IEEE, 2002. 887–892
- 82 Arasaratnam I, Haykin S. Cubature Kalman filters. *IEEE Transactions on Automatic Control*, 2009, **54**(6): 1254–1269
- 83 Chen B, Hu G Q, Ho D W C, Yu L. Distributed covariance intersection fusion estimation for cyber-physical systems with communication constraints. *IEEE Transactions on Automatic Control*, 2016, **61**(12): 4020–4026
- 84 Yu D D, Xia Y Q, Li L, Zhai D H. Event-triggered distributed state estimation over wireless sensor networks. *Automatica*, 2020, **118**: Article No. 109039
- 85 Peng H, Zeng B R, Yang L X, Xu Y, Lu R Q. Distributed extended state estimation for complex networks with nonlinear uncertainty. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(9): 5952–5960
- 86 Wang S C, Ren W, Chen J. Fully distributed dynamic state estimation with uncertain process models. *IEEE Transactions on Control of Network Systems*, 2018, **5**(4): 1841–1851
- 87 Yu F, Dutta R G, Zhang T, Hu Y D, Jin Y E. Fast attack-resilient distributed state estimator for cyber-physical systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2020, **39**(11): 3555–3565
- 88 Zhang C, Qin J H, Yan C Z, Shi Y, Wang Y N, Li M. Towards invariant extended Kalman filter-based resilient distributed state estimation for moving robots over mobile sensor networks under deception attacks. *Automatica*, 2024, **159**: Article No. 111408
- 89 Xie L, Choi D H, Kar S, Poor H V. Fully distributed state estimation for wide-area monitoring systems. *IEEE Transactions on Smart Grid*, 2012, **3**(3): 1154–1169
- 90 Qian J C, Duan P H, Duan Z S, Shi L. Event-triggered distributed state estimation: A conditional expectation method. *IEEE Transactions on Automatic Control*, 2023, **68**(10): 6361–6368
- 91 Duan P H, Wang Q S, Duan Z S, Chen G R. A distributed optimization scheme for state estimation of nonlinear networks with norm-bounded uncertainties. *IEEE Transactions on Automatic Control*, 2022, **67**(5): 2582–2589

- 92 Zhang C, Qin J H, Ma Q C, Shi Y, Li M L. Resilient distributed state estimation for LTI systems under time-varying deception attacks. *IEEE Transactions on Control of Network Systems*, 2023, **10**(1): 381–393
- 93 Wang H J, Liu K, Han D Y, Xia Y Q. Vulnerability analysis of distributed state estimation under joint deception attacks. *Automatica*, 2023, **157**: Article No. 111274
- 94 Facchinei F, Kanzow C. Generalized Nash equilibrium problems. *Annals of Operations Research*, 2010, **175**(1): 177–211
- 95 Ye M J, Hu G Q. Adaptive approaches for fully distributed Nash equilibrium seeking in networked games. *Automatica*, 2021, **129**: Article No. 109661
- 96 Meng Q, Nian X H, Chen Y, Chen Z. Attack-resilient distributed Nash equilibrium seeking of uncertain multiagent systems over unreliable communication networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(5): 6365–6379
- 97 Ye M J, Han Q L, Ding L, Xu S Y, Jia G B. Distributed Nash equilibrium seeking strategies under quantized communication. *IEEE/CAA Journal of Automatica Sinica*, 2024, **11**(1): 103–112
- 98 Zhong Y F, Yuan Y, Yuan H H. Nash equilibrium seeking for multi-agent systems under DoS attacks and disturbances. *IEEE Transactions on Industrial Informatics*, 2024, **20**(4): 5395–5405
- 99 Gadjov D, Pavel L. A passivity-based approach to Nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control*, 2019, **64**(3): 1077–1092
- 100 Romano A R, Pavel L. Dynamic gradient play for NE seeking with disturbance rejection. In: Proceedings of the IEEE Conference on Decision and Control (CDC). Miami, USA: IEEE, 2018. 346–351
- 101 Lou Y C, Hong Y G, Xie L H, Shi G D, Johansson K H. Nash equilibrium computation in subnetwork zero-sum games with switching communications. *IEEE Transactions on Automatic Control*, 2016, **61**(10): 2920–2935
- 102 Lu K H, Jing G S, Wang L. Distributed algorithms for searching generalized Nash equilibrium of noncooperative games. *IEEE Transactions on Cybernetics*, 2019, **49**(6): 2362–2371
- 103 Chen S B, Cheng R S. Operating reserves provision from residential users through load aggregators in smart grid: A game theoretic approach. *IEEE Transactions on Smart Grid*, 2019, **10**(2): 1588–1598
- 104 Zhu Y N, Yu W W, Wen G H, Chen G R. Distributed Nash equilibrium seeking in an aggregative game on a directed graph. *IEEE Transactions on Automatic Control*, 2021, **66**(6): 2746–2753
- 105 Carnevale G, Fabiani F, Fele F, Margellos K, Notarstefano G. Tracking-based distributed equilibrium seeking for aggregative games. *IEEE Transactions on Automatic Control*, 2024, **69**(9): 6026–6041
- 106 Shi Xia-Sheng, Ren Lu, Sun Chang-Yin. Distributed adaptive generalized Nash equilibrium algorithm for aggregative games. *Acta Automatica Sinica*, 2024, **50**(6): 1210–1220 (时侯圣, 任璐, 孙长银. 自适应分布式聚合博弈广义纳什均衡算法. *自动化学报*, 2024, **50**(6): 1210–1220)
- 107 Zhang Y Y, Sun J, Wu C Y. Vehicle-to-grid coordination via mean field game. *IEEE Control Systems Letters*, 2022, **6**: 2084–2089
- 108 Alasseur C, ben Taher I, Matoussi A. An extended mean field game for storage in smart grids. *Journal of Optimization Theory and Applications*, 2020, **184**(2): 644–670
- 109 Martinez-Piazuelo J, Quijano N, Ocampo-Martinez C. Nash equilibrium seeking in full-potential population games under capacity and migration constraints. *Automatica*, 2022, **141**: Article No. 110285
- 110 Zhang J, Lu J Q, Cao J D, Huang W, Guo J H, Wei Y. Traffic congestion pricing via network congestion game approach. *Discrete and Continuous Dynamical Systems—Series S*, 2021, **14**(4): 1553–1567
- 111 Zeng J, Wang Q Q, Liu J F, Chen J L, Chen H Y. A potential game approach to distributed operational optimization for microgrid energy management with renewable energy and demand response. *IEEE Transactions on Industrial Electronics*, 2019, **66**(6): 4479–4489
- 112 Deng Z H, Luo J. Distributed algorithm for nonsmooth multi-coalition games and its application in electricity markets. *Automatica*, 2024, **161**: Article No. 111494
- 113 Meng M, Li X X. On the linear convergence of distributed Nash equilibrium seeking for multi-cluster games under partial-decision information. *Automatica*, 2023, **151**: Article No. 110919
- 114 Bařar T, Olsder G J. *Dynamic Noncooperative Game Theory* (2nd edition). Philadelphia: SIAM, 1999.
- 115 Modares H, Lewis F L, Jiang Z P. H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(10): 2550–2562
- 116 Song R Z, Lewis F L, Wei Q L. Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(3): 704–713
- 117 Odekunle A, Gao W N, Davari M, Jiang Z P. Reinforcement learning and non-zero-sum game output regulation for multiplayer linear uncertain systems. *Automatica*, 2020, **112**: Article No. 108672
- 118 Li M, Qin J H, Freris N M, Ho D W C. Multiplayer Stackelberg-Nash game for nonlinear system via value iteration-based integral reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(4): 1429–1440
- 119 Mukaidani H, Xu H. Stackelberg strategies for stochastic systems with multiple followers. *Automatica*, 2015, **53**: 53–59
- 120 Li Man, Qin Jia-Hu, Wang Long. Seeking equilibrium for linear-quadratic two-player Stackelberg game: A Q-learning approach. *Scientia Sinica Informationis*, 2022, **52**(6): 1083–1097 (李曼, 秦家虎, 王龙. 线性二次二人 Stackelberg 博弈均衡点求解: 一种 Q 学习方法. *中国科学: 信息科学*, 2022, **52**(6): 1083–1097)
- 121 Lin Y N. Necessary/sufficient conditions for Pareto optimality in finite horizon mean-field type stochastic differential game. *Automatica*, 2020, **119**: Article No. 108951
- 122 Vamvoudakis K G, Lewis F L, Hudas G R. Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality. *Automatica*, 2012, **48**(8): 1598–1611
- 123 Jiao Q, Modares H, Xu S Y, Lewis F L, Vamvoudakis K G. Multi-agent zero-sum differential graphical games for disturbance rejection in distributed control. *Automatica*, 2016, **69**: 24–34
- 124 Li M, Qin J H, Ma Q C, Zheng W X, Kang Y. Hierarchical optimal synchronization for linear systems via reinforcement learning: A Stackelberg-Nash game perspective. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(4): 1600–1611
- 125 Li M, Qin J H, Wang Y N, Kang Y. Bio-inspired dynamic collective choice in large-population systems: A robust mean-field game perspective. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(5): 1914–1924
- 126 Kamalapurkar R, Klotz J R, Walters P, Dixon W E. Model-based reinforcement learning in differential graphical games. *IEEE Transactions on Control of Network Systems*, 2018, **5**(1): 423–433
- 127 Li J N, Modares H, Chai T Y, Lewis F L, Xie L H. Off-policy reinforcement learning for synchronization in multiagent graphical games. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(10): 2434–2445
- 128 Qin J H, Li M, Shi Y, Ma Q C, Zheng W X. Optimal synchronization control of multiagent systems with input saturation via off-policy reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, **30**(1): 85–96
- 129 Sun Chang-Yin, Mu Chao-Xu. Important scientific problems of multi-agent deep reinforcement learning. *Acta Automatica Sinica*, 2020, **46**(7): 1301–1312

- (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. *自动化学报*, 2020, **46**(7): 1301–1312)
- 130 Arslan G, Yüksel S. Decentralized Q-learning for stochastic teams and games. *IEEE Transactions on Automatic Control*, 2017, **62**(4): 1545–1558
- 131 Shao J Z, Lou Z Q, Zhang H C, Jiang Y H, He S C, Ji X Y. Self-organized group for cooperative multi-agent reinforcement learning. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2022. Article No. 413
- 132 Wang L, Zhang Y P, Hu Y J, Wang W X, Zhang C J, Gao Y, et al. Individual reward assisted multi-agent reinforcement learning. In: Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA: PMLR, 2022. 23417–23432
- 133 Leonardos S, Overman W, Panageas I, Piliouras G. Global convergence of multi-agent policy gradient in Markov potential games. In: Proceedings of the 10th International Conference on Learning Representations. Virtual Event: ICLR, 2022.
- 134 Zhang K Q, Hu B, Bašar T. On the stability and convergence of robust adversarial reinforcement learning: A case study on linear quadratic systems. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2020. Article No. 1850
- 135 Yang Y D, Luo R, Li M N, Zhou M, Zhang W N, Wang J. Mean field multi-agent reinforcement learning. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018. 5571–5580
- 136 Ben-Porath E. Rationality, Nash equilibrium and backwards induction in perfect-information games. *The Review of Economic Studies*, 1997, **64**(1): 23–46
- 137 Brown N, Sandholm T. Reduced space and faster convergence in imperfect-information games via pruning. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: PMLR, 2017. 596–604
- 138 Lowe R, Wu Y, Tamar A, Harb J, Abbeel P, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6382–6393
- 139 Sunehag P, Lever G, Gruslys A, Czarniecki W M, Zambaldi V, Jaderberg M, et al. Value-decomposition networks for cooperative multi-agent learning based on team reward. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. Stockholm, Sweden: International Foundation for Autonomous Agents and Multiagent Systems, 2018. 2085–2087
- 140 Rashid T, Samvelyan M, de Witt C S, Farquhar G, Foerster J, Whiteson S. Monotonic value function factorisation for deep multi-agent reinforcement learning. *The Journal of Machine Learning Research*, 2020, **21**(1): Article No. 178
- 141 Ruan J Q, Du Y L, Xiong X T, Xing D P, Li X Y, Meng L H, et al. GCS: Graph-based coordination strategy for multi-agent reinforcement learning. In: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2022. 1128–1136
- 142 Li X S, Li J C, Shi H B, Hwang K S. A decentralized communication framework based on dual-level recurrence for multiagent reinforcement learning. *IEEE Transactions on Cognitive and Developmental Systems*, 2024, **16**(2): 640–649
- 143 Jiang H B, Ding Z L, Lu Z Q. Settling decentralized multiagent coordinated exploration by novelty sharing. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2024. 17444–17452
- 144 Wang H, Yu Y, Jiang Y. Fully decentralized multiagent communication via causal inference. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(12): 10193–10202
- 145 van Goor P, Mahony R. EqVIO: An equivariant filter for visual-inertial odometry. *IEEE Transactions on Robotics*, 2023, **39**(5): 3567–3585
- 146 Shan T X, Englot B, Meyers D, Wang W, Ratti C, Rus D. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, USA: IEEE, 2020. 5135–5142
- 147 Shan T X, Englot B, Ratti C, Rus D. LVI-SAM: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China: IEEE, 2021. 5692–5698
- 148 Zhang Z Y, Wang L, Zhou L P, Koniusz P. Learning spatial-context-aware global visual feature representation for instance image retrieval. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023. 11216–11225
- 149 Harris C, Stephens M. A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference. Manchester, UK: Alvey Vision Club, 1988. 1–6
- 150 Fang S, Li H. Multi-vehicle cooperative simultaneous LiDAR SLAM and object tracking in dynamic environments. *IEEE Transactions on Intelligent Transportation Systems*, 2024, **25**(9): 11411–11421
- 151 Zhu F C, Ren Y F, Kong F Z, Wu H J, Liang S Q, Chen N, et al. Swarm-LIO: Decentralized swarm LiDAR-inertial odometry. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). London, UK: IEEE, 2023. 3254–3260
- 152 Zhang Z J, Wang S, Hong Y C, Zhou L K, Hao Q. Distributed dynamic map fusion via federated learning for intelligent networked vehicles. In: Proceedings of the IEEE International conference on Robotics and Automation (ICRA). Xi'an, China: IEEE, 2021. 953–959
- 153 Khamis A, Hussein A, Elmogy A. Multi-robot task allocation: A review of the state-of-the-art. *Cooperative Robots and Sensor Networks 2015*. Cham: Springer, 2015. 31–51
- 154 Choi H L, Brunet L, How J P. Consensus-based decentralized auctions for robust task allocation. *IEEE Transactions on Robotics*, 2009, **25**(4): 912–926
- 155 Bai X S, Fielbaum A, Kronmüller M, Knoedler L, Alonso-Mora J. Group-based distributed auction algorithms for multi-robot task assignment. *IEEE Transactions on Automation Science and Engineering*, 2023, **20**(2): 1292–1303
- 156 Shorinwa O, Haksar R N, Washington P, Schwager M. Distributed multirobot task assignment via consensus ADMM. *IEEE Transactions on Robotics*, 2023, **39**(3): 1781–1800
- 157 Park S, Zhong Y D, Leonard N E. Multi-robot task allocation games in dynamically changing environments. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China: IEEE, 2021. 8678–8684
- 158 Soria E, Schiano F, Floreano D. Predictive control of aerial swarms in cluttered environments. *Nature Machine Intelligence*, 2021, **3**(6): 545–554
- 159 Saravanos A D, Aoyama Y, Zhu H C, Theodorou E A. Distributed differential dynamic programming architectures for large-scale multiagent control. *IEEE Transactions on Robotics*, 2023, **39**(6): 4387–4407
- 160 Yao W J, de Marina H G, Sun Z Y, Cao M. Guiding vector fields for the distributed motion coordination of mobile robots. *IEEE Transactions on Robotics*, 2023, **39**(2): 1119–1135
- 161 Chen Y D, Guo M, Li Z K. Deadlock resolution and recursive feasibility in MPC-based multirobot trajectory generation. *IEEE Transactions on Automatic Control*, 2024, **69**(9): 6058–6073
- 162 Spica R, Cristofalo E, Wang Z J, Montijano E, Schwager M. A real-time game theoretic planner for autonomous two-player drone racing. *IEEE Transactions on Robotics*, 2020, **36**(5): 1389–1403
- 163 Williams Z, Chen J S, Mehr N. Distributed potential iLQR: Scalable game-theoretic trajectory planning for multi-agent interactions. In: Proceedings of the IEEE International Confer-

- ence on Robotics and Automation (ICRA). London, UK: IEEE, 2023. 1–7
- 164 Chen M, Shih J C, Tomlin C J. Multi-vehicle collision avoidance via Hamilton-Jacobi reachability and mixed integer programming. In: Proceedings of the IEEE 55th Conference on Decision and Control (CDC). Las Vegas, USA: IEEE, 2016. 1695–1700
- 165 Li M, Qin J H, Li J C, Liu Q C, Shi Y, Kang Y. Game-based approximate optimal motion planning for safe human-swarm interaction. *IEEE Transactions on Cybernetics*, 2024, **54**(10): 5649–5660
- 166 Zhu K, Zhang T. Deep reinforcement learning based mobile robot navigation: A review. *Tsinghua Science and Technology*, 2021, **26**(5): 674–691
- 167 He Z C, Dong L, Song C W, Sun C Y. Multiagent soft actor-critic based hybrid motion planner for mobile robots. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(12): 10980–10992
- 168 Fan T X, Long P X, Liu W X, Pan J. Distributed multi-robot collision avoidance via deep reinforcement learning for navigation in complex scenarios. *The International Journal of Robotics Research*, 2020, **39**(7): 856–892
- 169 Brito B, Everett M, How J P, Alonso-Mora J. Where to go next: Learning a subgoal recommendation policy for navigation in dynamic environments. *IEEE Robotics and Automation Letters*, 2021, **6**(3): 4616–4623
- 170 Xie Z T, Dames P. DRL-VO: Learning to navigate through crowded dynamic scenes using velocity obstacles. *IEEE Transactions on Robotics*, 2023, **39**(4): 2700–2719
- 171 Han R H, Chen S D, Wang S J, Zhang Z Q, Gao R, Hao Q, et al. Reinforcement learned distributed multi-robot navigation with reciprocal velocity obstacle shaped rewards. *IEEE Robotics and Automation Letters*, 2022, **7**(3): 5896–5903
- 172 Chen L, Wang Y N, Miao Z Q, Feng M T, Zhou Z, Wang H S, et al. Reciprocal velocity obstacle spatial-temporal network for distributed multirobot navigation. *IEEE Transactions on Industrial Electronics*, 2024, **71**(11): 14470–14480
- 173 Qin J M, Qin J H, Qiu J X, Liu Q C, Li M, Ma Q C. SRL-ORCA: A socially aware multi-agent mapless navigation algorithm in complex dynamic scenes. *IEEE Robotics and Automation Letters*, 2024, **9**(1): 143–150
- 174 Li Y, Davis C, Lukszo Z, Weijnen M. Electric vehicle charging in China's power system: Energy, economic and environmental trade-offs and policy implications. *Applied Energy*, 2016, **173**: 535–554
- 175 Chandra I, Singh N K, Samuel P. A comprehensive review on coordinated charging of electric vehicles in distribution networks. *Journal of Energy Storage*, 2024, **89**: Article No. 111659
- 176 Franco J F, Rider M J, Romero R. A mixed-integer linear programming model for the electric vehicle charging coordination problem in unbalanced electrical distribution systems. *IEEE Transactions on Smart Grid*, 2015, **6**(5): 2200–2210
- 177 Das R, Wang Y, Busawon K, Putrus G, Neaimh M. Real-time multi-objective optimisation for electric vehicle charging management. *Journal of Cleaner Production*, 2021, **292**: Article No. 126066
- 178 Wan Y N, Qin J H, Yu X H, Yang T, Kang Y. Price-based residential demand response management in smart grids: A reinforcement learning-based approach. *IEEE/CAA Journal of Automatica Sinica*, 2022, **9**(1): 123–134
- 179 Zhang P, Qian K J, Zhou C K, Stewart B G, Hepburn D M. A methodology for optimization of power systems demand due to electric vehicle charging load. *IEEE Transactions on Power Systems*, 2012, **27**(3): 1628–1636
- 180 Ioakimidis C S, Thomas D, Rycerski P, Genikomsakis K N. Peak shaving and valley filling of power consumption profile in non-residential buildings using an electric vehicle parking lot. *Energy*, 2018, **148**: 148–158
- 181 van Kriekinge G, de Cauwer C, Sapountzoglou N, Coosemans T, Messagie M. Peak shaving and cost minimization using model predictive control for uni- and bi-directional charging of electric vehicles. *Energy Reports*, 2021, **7**: 8760–8771
- 182 Gong J B, Fu W M, Kang Y, Qin J H, Xiao F. Multi-agent deep reinforcement learning based multi-objective charging control for electric vehicle charging station. In: Proceedings of the 7th Chinese Conference on Swarm Intelligence and Cooperative Control. Nanjing, China: Springer, 2023. 266–277
- 183 Tu R, Gai Y J, Farooq B, Posen D, Hatzopoulou M. Electric vehicle charging optimization to minimize marginal greenhouse gas emissions from power generation. *Applied Energy*, 2020, **277**: Article No. 115517
- 184 Adetunji K E, Hofsajer I W, Abu-Mahfouz A M, Cheng L. An optimization planning framework for allocating multiple distributed energy resources and electric vehicle charging stations in distribution networks. *Applied Energy*, 2022, **322**: Article No. 119513
- 185 Ran L L, Qin J H, Wan Y N, Fu W M, Yu W W, Xiao F. Fast charging navigation strategy of EVs in power-transportation networks: A coupled network weighted pricing perspective. *IEEE Transactions on Smart Grid*, 2024, **15**(4): 3864–3875
- 186 Wan Y N, Qin J H, Li F Y, Yu X H, Kang Y. Game theoretic-based distributed charging strategy for PEVs in a smart charging station. *IEEE Transactions on Smart Grid*, 2021, **12**(1): 538–547
- 187 Zhang L, Li Y Y. A game-theoretic approach to optimal scheduling of parking-lot electric vehicle charging. *IEEE Transactions on Vehicular Technology*, 2016, **65**(6): 4068–4078
- 188 Kabir M E, Assi C, Tushar M H K, Yan J. Optimal scheduling of EV charging at a solar power-based charging station. *IEEE Systems Journal*, 2020, **14**(3): 4221–4231
- 189 Zavvos E, Gerding E H, Brede M. A comprehensive game-theoretic model for electric vehicle charging station competition. *IEEE Transactions on Intelligent Transportation Systems*, 2022, **23**(8): 12239–12250
- 190 Chen J, Huang X Q, Cao Y J, Li L Y, Yan K, Wu L, et al. Electric vehicle charging schedule considering shared charging pile based on generalized Nash game. *International Journal of Electrical Power and Energy Systems*, 2022, **136**: Article No. 107579
- 191 Yan D X, Yin H, Li T, Ma C B. A two-stage scheme for both power allocation and EV charging coordination in a grid-tied PV-battery charging station. *IEEE Transactions on Industrial Informatics*, 2021, **17**(10): 6994–7004
- 192 Liu Z X, Wu Q W, Huang S J, Wang L F, Shahidehpour M, Xue Y S. Optimal day-ahead charging scheduling of electric vehicles through an aggregative game model. *IEEE Transactions on Smart Grid*, 2018, **9**(5): 5173–5184
- 193 Lin R Z, Chu H Q, Gao J W, Chen H. Charging management and pricing strategy of electric vehicle charging station based on mean field game theory. *Asian Journal of Control*, 2024, **26**(2): 803–813
- 194 Wang Y F, Wang X L, Shao C C, Gong N W. Distributed energy trading for an integrated energy system and electric vehicle charging stations: A Nash bargaining game approach. *Renewable Energy*, 2020, **155**: 513–530
- 195 Pahlavanhoseini A, Sepasian M S. Optimal planning of PEV fast charging stations using Nash bargaining theory. *Journal of Energy Storage*, 2019, **25**: Article No. 100831
- 196 Ran L L, Wan Y N, Qin J H, Fu W M, Zhang D F, Kang Y. A game-based battery swapping station recommendation approach for electric vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2023, **24**(9): 9849–9860
- 197 Zeng H T, Sheng Y J, Sun H B, Zhou Y Z, Xue Y X, Guo Q L. A conic relaxation approach for solving Stackelberg pricing game of electric vehicle charging station considering traffic equilibrium. *IEEE Transactions on Smart Grid*, 2024, **15**(3): 3080–3097
- 198 Wan Y N, Qin J H, Ma Q C, Fu W M, Wang S. Multi-agent DRL-based data-driven approach for PEVs charging/discharging scheduling in smart grid. *Journal of the Franklin Institute*, 2022, **359**(4): 1747–1767

- 199 Zhang Z L, Wan Y N, Qin J H, Fu W M, Kang Y. A deep RL-based algorithm for coordinated charging of electric vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2022, **23**(10): 18774–18784
- 200 Park K, Moon I. Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid. *Applied Energy*, 2022, **328**: Article No. 120111
- 201 Zhang Y, Yang Q Y, An D, Li D H, Wu Z Z. Multistep multi-agent reinforcement learning for optimal energy schedule strategy of charging stations in smart grid. *IEEE Transactions on Cybernetics*, 2023, **53**(7): 4292–4305
- 202 Liang Y C, Ding Z H, Zhao T Y, Lee W J. Real-time operation management for battery swapping-charging system via multi-agent deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2023, **14**(1): 559–571
- 203 Wang L, Liu S X, Wang P F, Xu L M, Hou L Y, Fei A G. QMIX-based multi-agent reinforcement learning for electric vehicle-facilitated peak shaving. In: Proceedings of the IEEE Global Communications Conference. Kuala Lumpur, Malaysia: IEEE, 2023. 1693–1698

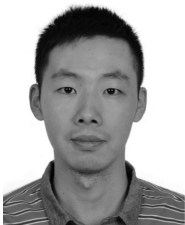


秦家虎 中国科学技术大学自动化系教授。主要研究方向为网络化控制系统, 自主智能系统, 人-机交互。本文通信作者。

E-mail: [jqin@ustc.edu.cn](mailto:jhqin@ustc.edu.cn)

(**QIN Jia-Hu** Professor in the Department of Automation, Uni-

versity of Science and Technology of China. His research interest covers networked control systems, autonomous intelligent systems, and human-robot interaction. Corresponding author of this paper.)



马麒超 中国科学技术大学自动化系副研究员。主要研究方向为多智能体系统协同决策与控制, 及其在机器人系统中的应用。

E-mail: qema@ustc.edu.cn

(**MA Qi-Chao** Research associate professor in the Department of

Automation, University of Science and Technology of China. His research interest covers collaborative decision and control of multi-agent systems, with applications to robotics.)



李曼 中国科学技术大学自动化系副研究员。主要研究方向为多智能体博弈, 强化学习, 人-机交互。

E-mail: man.li@ustc.edu.cn

(**LI Man** Research associate professor in the Department of Automation, University of Science and

Technology of China. Her research interest covers

multi-agent games, reinforcement learning, and human-robot interaction.)



张聪 中国科学技术大学自动化系博士后。主要研究方向为多智能体协同, 分布式状态估计, 移动机器人同步定位与建图。

E-mail: cong_zhang@ustc.edu.cn

(**ZHANG Cong** Postdoctor in the

Department of Automation, University of Science and Technology of China. Her research interest covers multi-agent cooperation, distributed state estimation, and simultaneous localization and mapping (SLAM) for mobile robots.)



付维明 中国科学技术大学自动化系副研究员。主要研究方向为多智能体系统协同与智能电网能量管理。

E-mail: fwm1993@ustc.edu.cn

(**FU Wei-Ming** Research associate professor in the Department of Automation, University of Science

and Technology of China. His research interest covers collaborative in multi-agent systems and energy management in smart grids.)

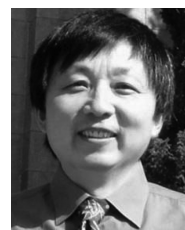


刘轻尘 中国科学技术大学自动化系教授。主要研究方向为网络化系统, 多机器人系统, 基于学习的控制。

E-mail: qingchen_liu@ustc.edu.cn

(**LIU Qing-Chen** Professor in the Department of Automation, Uni-

versity of Science and Technology of China. His research interest covers networked systems, multi-robotics system, and learning based control.)



郑卫新 澳大利亚西悉尼大学杰出教授。主要研究方向为系统辨识, 网络化控制系统, 多智能体系统, 神经网络, 信号处理。

E-mail: w.zheng@westernsydney.edu.au

(**ZHENG Wei-Xing** Distinguished professor at Western Sydney Uni-

versity, Australia. His research interest covers system identification, networked control systems, multi-agent systems, neural networks, and signal processing.)