

面向低轨星座边缘计算的博弈强化学习方法综述

谷学强¹, 张万鹏¹, 谭思雨², 罗俊仁¹, 周棣忠¹

(1. 国防科技大学智能科学学院, 湖南 长沙 410073;

2. 湖南先进技术研究院, 湖南 长沙 410205)

摘要: 博弈强化学习作为人工智能领域的新兴范式, 是当前解决低轨星座边缘计算问题的主流方法。融入博弈论的多智能体深度强化学习方法为复杂、动态、不确定性的星座边缘计算问题提供了新思路。通过梳理总结卫星组网、任务卸载以及资源调度3种卫星边缘计算主要研究方向, 详细阐述了博弈强化学习范式基础, 并从博弈模型、深度Q网络、深度确定性策略梯度以及近端策略优化等方面分别阐述了3种研究方向上的典型应用现状, 最后对该领域的前沿挑战进行分析, 期望为博弈强化学习范式与低轨星座边缘计算领域的交叉融合研究提供参考。

关键词: 低轨星座; 边缘计算; 博弈论; 多智能体强化学习

中图分类号: TP181

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202432

Overview on game reinforcement learning methods for edge computing of low-orbit constellation

GU Xueqiang¹, ZHANG Wanpeng¹, TAN Siyu², LUO Junren¹, ZHOU Yanzhong¹

1. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

2. Hunan Institute of Advanced Technology, Changsha 410205, China

Abstract: As a new paradigm in the field of artificial intelligence, game reinforcement learning is an advanced mainstream method to solve the edge computing problem of low-orbit constellation. The multi-agent deep reinforcement learning integrated into the game perspective provides a new idea for dynamic, complex and uncertain constellation edge computing problems. By summarizing the three main research directions of satellite edge computing, namely satellite networking, task unloading and resource scheduling, the basis of game reinforcement learning paradigm is elaborated, and the typical applications in the three research directions are described respectively from the methods of game model, deep Q network, deep deterministic strategy gradient and near-end strategy optimization. In the end, the paper looks forward to the frontier challenges in this field, expected to provide a reference for the cross-fusion research of game reinforcement learning paradigm and low-orbit constellation edge computing.

Key words: low-orbit constellation, edge computing, game theory, multi-agent reinforcement learning

0 引言

低轨星座^[1]是由数百甚至数万颗轨道高度位于

500~2 000 km上的低轨卫星组成的卫星系统, 新型低轨卫星系统能大规模、全覆盖、低时延地支撑全球用户的互联网服务^[2], 具有覆盖范围广、通信带

收稿日期: 2024-05-22; 修回日期: 2024-07-25

通信作者: 张万鹏, wpzhang@nudt.edu.cn

基金项目: 国家自然科学基金项目 (No.92271108, No.62173336)

Foundation Items: The National Natural Science Foundation of China(No.92271108, No.62173336)

宽大、不受地理条件限制以及抗毁性强等优点,在为无法实现地面网络覆盖的区域提供互联网服务方面有着巨大潜力^[3-4]。在传统卫星通信场景^[5]中,卫星的作用仅仅是数据的透明转发,因此用户-卫星的计算过程往往包括任务上传卫星、转发地面云中心、结果回传卫星和卫星将结果发回用户等多个步骤^[6],从而导致用户业务的高时延,难以满足用户的业务需求。随着“太空计算机”概念的出现,一些研究考虑使用卫星作为计算节点为地面用户提供实时计算服务,因此,可考虑将多颗卫星联合起来实现在轨协同计算^[7]。同时,随着星间链路^[8]质量的不断发展和卫星计算能力的不断提升,用户对卫星场景的业务需求也不断增加,业务种类逐渐变得复杂起来,不同的业务对网络的需求也不尽相同,这对新型低轨卫星网络的稳定通信链路^[9]、网络资源^[10]和计算资源的调度能力^[11]提出了挑战^[12]。

近年来,随着新型低轨星座的迅速发展,基于低轨星座边缘计算^[13]类的综述性分析探索也成为热点。Abdelsadek等^[14]对卫星网络体系进行了梳理;Lin等^[15]对边缘计算架构进行了系统阐述;Sadatdiynov等^[16]对边缘计算中的任务卸载优化方法进行了全面的综述;周绪等^[17]对边缘计算中资源调度场景中的架构、建模与方法进行了系统的综述分析;陈全等^[18]针对低轨巨型星座组网中的卫星网络拓扑动态性管理、星地链路切换策略、路由算法设计、地面信关站布局优化设计以及网络仿真与性能评估5个方面系统梳理了各关键技术的主要技术路线和研究进展;王鹏等^[19]详细总结了边缘计算网络中的低轨星座组网方式及架构;Gao等^[20]梳理了星地协同网络中的边缘计算技术研究现状,并对计算卸载技术方案进行汇总和分类对比;吴戌等^[21]对卫星边缘计算架构的物理结构、功能结构和服务流程3个方面进行了综述,并分析了五大边缘计算技术。

强化学习框架^[22]下,多个智能体与环境不断产生交互,逐步学习,产生并接受奖励或惩罚,最终找到目标的最佳策略^[23]。博弈强化学习^[24]作为人工智能领域的新兴范式,是一种模拟智能体与其他智能体及环境交互的学习方法^[25],近年来在众多领域中都展示出卓越性能与适应能力^[26]。博弈视角下的强化学习范式在人工智能领域已硕果累累^[27],在面对众多复杂问题的求解上,博弈强化

学习实现了算法及其变体的不断迭代更新。低轨星座是一个典型的多智能体系统,这类基于博弈理论、建模了交互关系、采用了交互历史的学习方法为低轨星座边缘算法典型任务场景提供了新的解决方案,可以利用多智能体博弈强化学习来开展相关课题研究。

本文的整体架构如图1所示,主要围绕低轨星座边缘计算问题展开,首先归纳分析了卫星组网、任务卸载和资源调度3种卫星边缘计算热点研究方向;其次,详细阐述了博弈强化学习范式基础,提炼总结了博弈模型、深度Q网络(deep Q-network, DQN)、深度确定性策略梯度(deep deterministic policy gradient, DDPG)、近端策略优化(proximal policy optimization, PPO)等方法,并总结归纳了其在低轨星座边缘计算典型场景上的应用情况;最后,展望了该领域的前沿研究。

1 低轨星座边缘计算主要研究方向

本文主要研究的低轨星座边缘计算典型研究方向包括:卫星组网、边缘任务卸载以及资源调度。博弈强化学习是一种模拟智能体与环境交互的学习方法,近年来在众多领域中展示出了其卓越的性能与适应能力。智能体在强化学习框架下,通过不断与环境交互,接受奖励或惩罚,逐步学习,找到目标的最佳策略^[23]。这种自主学习和适应的特性为低轨星座的边缘智能化提供了全新的视角,特别是在处理复杂、动态、不确定问题上,应用效果较好。

1.1 边缘计算概述

边缘计算本质是一种分布式计算。如图2所示,边缘计算场景架构主要包含了四大要素:用户终端、低轨星座网络、云中心以及链路。其中,链路又分为星间链路、地面链路、星间链路与星地链路。在低轨星座网络中主要由低轨卫星上的边缘服务器来承载地面用户的服务需求,且用户终端是边缘计算服务的受众主体,当任务无法在本地进行处理时,可由低轨卫星或云中心来处理运行任务,这个过程就包含了任务卸载。链路的存在使计算可以实现中继传输,任务及资源在星间自由传输的过程就包含了资源调度。卫星组网、任务卸载、资源调度三者合作关系阐述如下:以紧急地球观测任务为例,任务卸载的作用是将图像处理任务从资源紧张的卫星转移到资源较空闲的卫星;资源调度系统需要快速评估各卫星的资源状况,并进行调度决策;

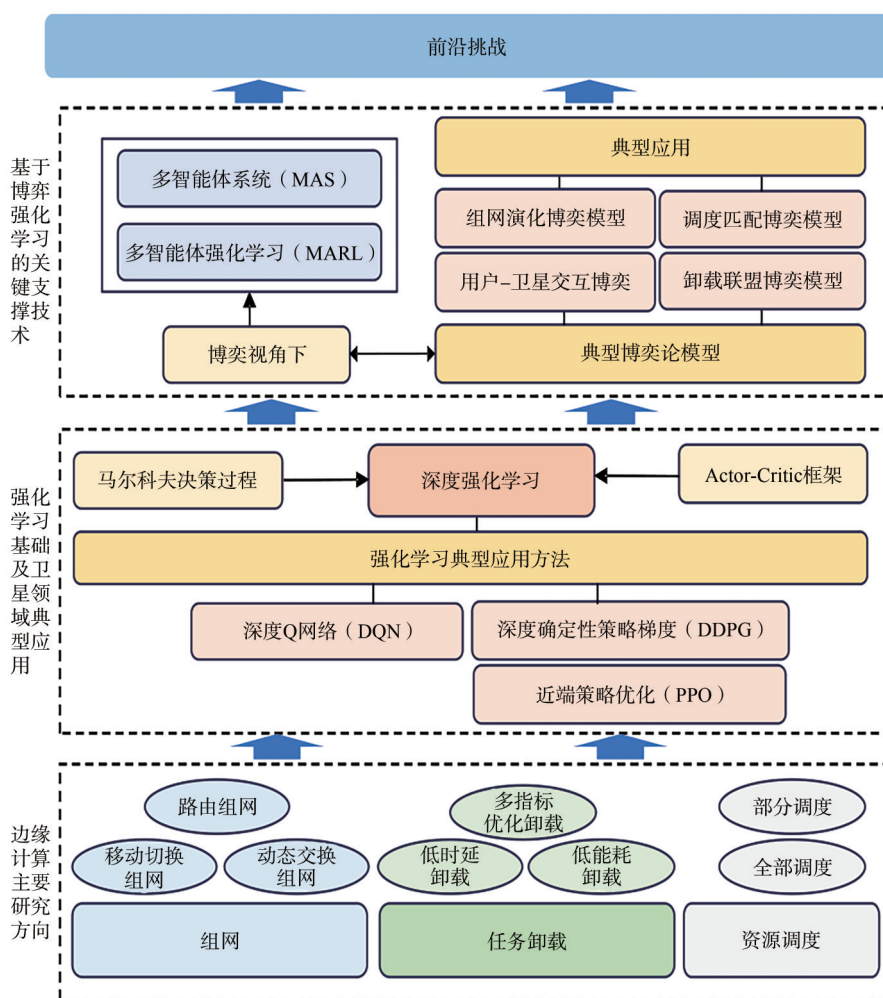


图1 论文整体架构

而卫星组网则通过其高效的通信连接支持这些任务的快速转移和数据的即时交换。

一般来说，这三大研究方向中，使用时延、能耗、负载以及调度效率来作为其方法的评价指标，问题建模的方式一般包括线性规划问题、非线性规划问题以及其他问题。相应的求解方法分为两类，一类是基于启发式算法^[28-30]和博弈论方法的传统经典方法，比如 Dao 等^[31]和滕颖蕾等^[32]采用了基于凸优化的最优解方法来解决边缘任务卸载问题；另一类是基于深度学习^[33]、强化学习^[34]、深度强化学习^[35]的人工智能先进方法。本文重点对博弈强化学习^[36]关键技术进行分析总结与讨论。

1.2 卫星组网

在卫星边缘计算领域中，卫星组网主要通过构建高效的通信和计算网络，支持各种空间任务和地面服务的需求。组网为低轨星座的边缘计算提供网络支撑。由于网络拓扑结构^[37]千差万异，组网的方法

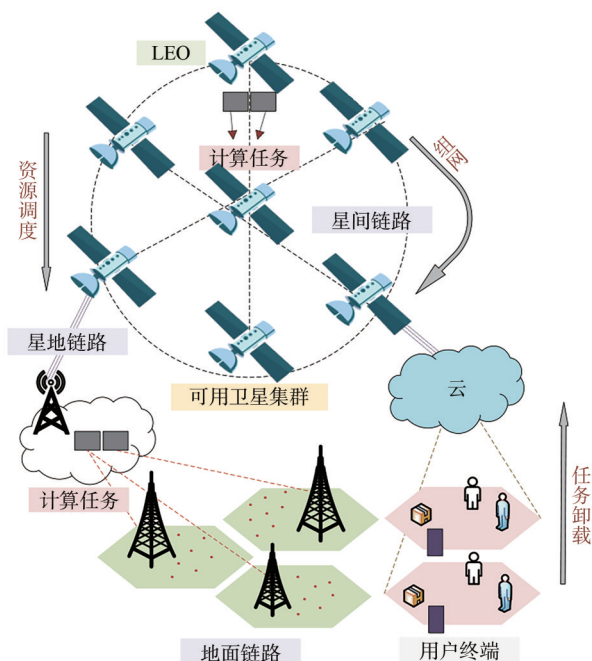


图2 边缘计算场景架构及三大研究方向

式也不尽相同^[38]。卫星组网技术^[39]隶属多学科交叉技术,且组网对象交互方式一般包含卫星与卫星之间(星间)以及卫星与地面之间(星地),星间路由协议^[40]一般作为星间组网技术的重点方向,星地网络拓扑设计和星地链路切换技术一般作为星地组网技术的重点方向。而针对地面,地面网关路由设计技术也十分重要,例如欧阳等^[41]提出了一种新的路由算法来避免组网拥塞。常见的组网方式包括移动切换组网、动态交换组网和路由组网。

1.2.1 移动切换组网

移动切换^[42]主要包括星内切换组网和星地切换组网。其中,星内切换组网是指星地或星间的链路发生动态频繁切换^[43]时,引起的终端用户动态频繁地切换到新链路的过程;星地切换组网是指伴随卫星远离终端用户时而引起的终端用户切换到另一个新的卫星网络的过程^[44]。频繁地移动切换常常会提高整个卫星网络系统的信令开销、切换冲突概率、切换成本^[45]。因此,简洁化切换过程、提高切换可靠性是卫星切换组网的重要需求之一。

1.2.2 动态交换组网

动态交换组网是指针对数据的接收解析而快速完成数据的寻址与分发的过程,其目的是保障星间链路的连通性。动态交换组网一般包括低时延组网技术、组网抗毁重构技术、激光星间链路技术^[46]以及星间组网协议及互操作技术。降低切换时延与信令开销、确保传输信息数据完整是动态交换组网的重要需求之一^[47]。

1.2.3 路由组网

路由是指从原地址起始端到目的地终点端过程中,决定端到端路径的网络范围的进程,而从原地址起始端到目的地终点端的这个过程就包含了星间链路过程与星地链路过程。常见的路由组网方式包括改进地面网络动态路由、基于虚拟拓扑的路由策略、基于虚拟节点的路由策略^[48-49]等。倪少杰等^[50]基于不同的路由决策主体,对集中式、分布式与混合式3种路由组网方式进行了详细的总结描述。增强抗毁鲁棒性、提高繁杂业务类型承载力是路由组网的重要需求之一^[51]。

1.3 任务卸载

在卫星边缘计算领域中,任务卸载^[52]主要涉及将计算任务从地面或较多负载的卫星转移至边缘节点(如其他卫星)执行,以达到优化资源利用和提高响应速度的目的,而卫星组网为任务卸载提供

了所需的基础通信网络。任务卸载指的是通过终端设备把任务按照卸载选择策略^[53],匹配到合适且资源充裕的客户端服务器上,并通过无线网络分配给指定的服务器进行处理,然后服务器架构将计算结果返回的过程^[54]。计算卸载的性能评估指标有很多,时延和能耗的优化是卸载策略及方法的主要评价参数,在针对星间链路间的任务卸载决策的设计中,可以权衡能耗和时延的综合指标^[55]。一般来说,低轨星座的任务卸载分为二进制定卸载和部分卸载。二进制定卸载是指要么在本地设备上执行全部任务,要么将任务完全卸载到边缘服务器;部分卸载是指终端设备可以将任务的一部分传输至边缘服务器,在本地执行余下部分。从任务卸载的要求来看,其又可分为低时延卸载、低能耗卸载以及多指标优化卸载。

1.3.1 低时延卸载

时延的产生包括很多方面,例如设备计算能力、任务的计算总量、卫星网络系统运行状态等^[56]。低时延卸载主要从时延产生的主体出发,比如降低任务上传的时延、降低计算数据回传的延时以及任务在服务器上计算处理的时延。低时延卸载的主要考虑因素包括通信延迟、处理延迟,其中通信延迟时间是指从源设备到目标处理设备的数据传输时间,处理延迟时间指的是目标设备上实际处理任务所需的时间。

1.3.2 低能耗卸载

数据经链路后传输以及边缘服务器计算任务的过程等都会产生能量损耗,即能耗^[57]。数据的体量、传输距离、传输速度等都是影响数据传输过程中能耗高低的因素,而处理器、内存、硬盘、网络等组件的运作主要影响服务器在执行计算任务时所消耗的能量。低能耗卸载主要考虑因素包括设备能耗、服务器能耗,其中设备能耗主要指执行任务时本地设备的能耗与通过网络发送数据到远程服务器的能耗,服务器能耗是指远程处理任务的能耗。

1.3.3 多指标优化卸载

多指标优化卸载^[58]除了包含低时延低能耗指标以外,还包含用户隐私以及用户对卸载质量的满意程度等多个用户QoE(quality of experience)需求指标。例如,任务在卸载过程中消耗过多能耗,那么该卸载是不满足用户的能耗敏感型任务,而任务在卸载过程中产生过多的时延,则该卸载是不满足用户的时延敏感型任务。同时,用户QoE需求

指标常常作为多目标优化卸载中的一个较为重要的指标,其影响因素十分复杂,常见的包含网络吞吐量、网络带宽等。多指标优化卸载已成为任务卸载研究热点,其研究过程通常会被看作博弈过程中的多目标优化问题。

1.4 资源调度

在卫星边缘计算领域,资源调度涉及在卫星网络中高效、合理地分配计算、存储和通信资源。资源调度是指资源进行调度和有效分发管理,从而实现任务的高效执行和资源的优化利用^[59-60]。资源调度对时延、带宽、能耗、负载均衡、安全性以及灵活性等^[61]都提出了苛刻要求,其能够将资源受限的完全或者部分任务调度到资源充足的设备或者环境中进行处理^[62]。例如Yuan等^[63]提出了一种增强的元批评学习算法来有效地调度资源以服务大量连接的用户,从而可以交付/服务更多的数据和用户。针对卫星资源调度的研究方法从最开始的传统经典最优化方法^[64-66]逐渐过渡到了深度强化学习范式方法^[67]。从调度方式来看,资源调度可分为全部调度和部分调度。

1.4.1 全部调度

全部调度是指所有任务的资源都调度到同一个固定不变的服务器上进行处理。在卫星边缘计算网络中,全部调度指的是系统对所有可用资源进行集中管理和调度。在这种模式下,调度系统拥有全局的资源信息,包括每个节点的计算能力、存储空间、能耗状况及其通信能力等。调度决策基于全面的网络状态和资源情况,旨在优化整个系统的性能和提高资源利用率,全部调度适用于那些对资源使用效率要求极高并且可以接受一定延迟的场景。

1.4.2 部分调度

部分调度是指首先将任务划分为多个子任务,然后再将子任务的资源调度到不同服务器上进行处理。通常在部分调度的研究中,任务划分后的子任务之间的依赖关系可以映射为有向无环图,根据依赖关系,进行子任务的重要级别排序^[68],从而确定子任务的优先调度顺序。

2 博弈强化学习基础

本节将详细阐述多智能体系统(multi-agent system, MAS)、多智能体强化学习(multi-agent reinforcement learning, MARL)、马尔可夫决策过程(Markov decision process, MDP)、Actor-Critic

框架等概念,并给出博弈视角下的深度强化学习分析。

2.1 多智能体系统

多智能体系统定义为由多个同类或者异类的智能体与其所在环境互相交互组成的分布式决策系统。从这个定义上看,低轨星座就是一个典型的MAS。MAS通常由智能体、环境、交互体系以及学习方法四大模块组成^[68],其架构如图3所示。其中,智能体模块是MAS中的对象主体,不同智能体与环境的交互通信关系是不同的。智能体模块定义了观测状态和输出的指定动作以及智能体相互之间的交互通信方式,也定义效能函数,该函数用于定义环境状态的偏好以及行动选择策略。状态空间、动作空间、转换函数和奖励函数共同组成了环境模块,状态空间和动作空间的主体是单个智能体。状态空间表示单个智能体处于任意时间域内的一组状态;动作空间表示单个智能体处于任意时间域内的一组动作;转换函数的主体是每个智能体,其定义了固定状态下每个智能体执行动作时需改变的方式;奖励函数依托于状态空间和行动空间,其根据状态-行动的迭代结果来得出奖励反馈信号。交互体系模块的主体是所有智能体,其表示智能体与智能体之间交互的时间长短、交互的对象、交互的频率以及动作的交互顺序。学习方法模块通常包含学习的实体、学习的目的、学习的历史经验数据、学习更新的方式以及学习目标成果。其中,学习实体就是指多智能体;学习目的也就是学习的任务目标,通常映射为评价函数;学习历史经验数据也就是学习实体的学习基础数据,是从哪些信息数据中启发学习的;学习更新就包含了学习过程中学习实体的迭代更新规则。

2.2 多智能体强化学习

多智能体强化学习^[69]的概念约同于多智能体深度强化学习,因为需要对特征信息进行学习,所以其目标函数的构建借助了深度学习神经网络。MARL的工作过程可以简单表述为:主体是多个小集合或大群体的具有自主控制能力和自主决策能力的智能体,依托强化学习方法从而实现序贯决策问题,最终目的是使MAS达到任务需求所要求的最优均衡状态^[70]。

MARL的联合动作定义为所有智能体的动作集合,联合状态 S 和联合奖励 R 经过环境反馈给MARL,最终目的是通过学习得到一组最优的策略

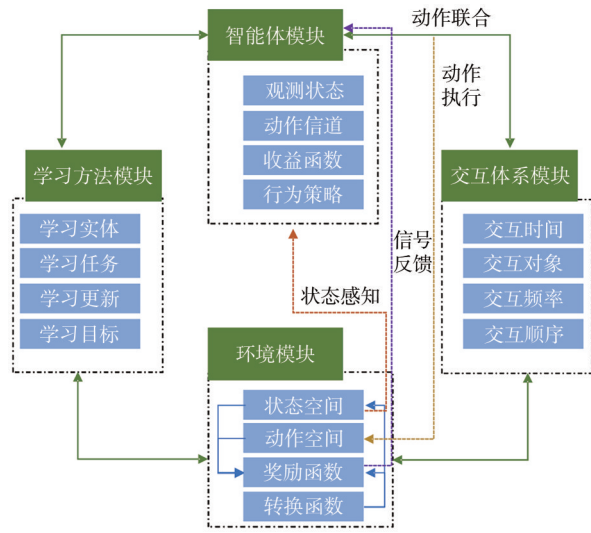


图3 MAS架构

值集合: $\pi^* = \{\pi_1^*, \pi_2^*, \dots, \pi_n^*\}$, MARL 架构如图4所示。MARL联合值函数表示为:

$$Q_{\text{tot}}^{\pi}(s, a) = E \left[\sum_{t=0}^{\infty} \lambda^t \gamma^t | s_0 = s, a_0 = a, \pi \right] \quad (1)$$

联合值函数的动作集合组成了联合策略, 即 $\pi(s) = \arg \max_a Q_{\text{tot}}^{\pi}(s, a)$, 这种策略组合涉及所有参与者的选择, 表达了他们如何在某个具体的博弈场景下一起行动; 在任何应用场景中, MARL 都适配以马尔可夫决策(见2.3小节)为基础的博弈理论。

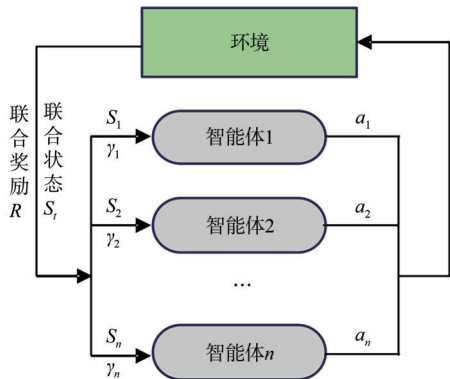


图4 MARL架构

2.3 马尔可夫决策过程

马尔可夫决策过程一般服务于强化学习的建模和决策, 其描述一般映射为五元组 (S, A, P, R, r) 。其中, 决策过程中所有状态的有限集合表示为 S ; 不同智能体的所有动作的有限集合表示为 A ; 转换函数定义为 P ; 回报函数定义为 R ; 折扣系数定义

为 r , 且 $r \in [0, 1]$, 以上函数定义了不同智能体间的瞬时回报和长久回报对总体回报的影响权重因子。

马尔可夫决策过程通过以上定义得到的求解结果是策略 π^* , 也就是总体回报值最大的一组最优策略, 该策略一般表示为:

$$Q^*(s, a) = \max E[R_t | S_t = s, a_t = a, \pi] \quad (2)$$

马尔可夫决策过程不局限于单个智能体, 而是针对多个智能体随着环境的不断动态变化, 智能体随环境产生的回报从而又影响智能体的动态与状态的过程。如图5所示, 马尔可夫决策过程中当前动作的执行会对下一个动作造成影响, 从而形成动作概率间的相互转移过程。

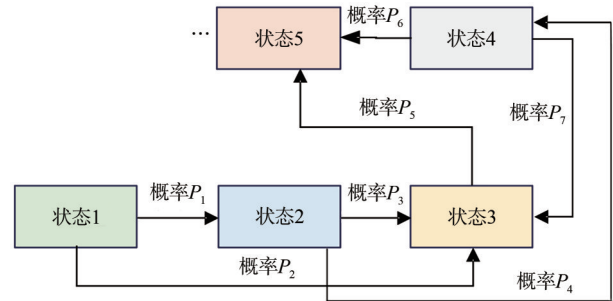


图5 马尔可夫决策过程

在MARL的学习决策过程中, 智能体与环境的关系仅仅是感知和交互的关系, 并不能实现对整体环境的替换和覆盖。因此, 有关MARL的求解建模通常是一个分布式部分可观测的马尔可夫决策过程, 该过程可用一组七元组表示, 即 $(S, A, P, R, \Omega, O, \gamma)$ 。其中, 智能体表示为 $i \in I = \{1, 2, \dots, N\}$, 状态空间表示为 $s \in S$, 动作空间表示为 $a_i \in a, a = A^N$, 转移函数 P 表示为 $P(s'|s, a)$, 奖励值表示为 $R(s, a)$, 观测过程表示为 $a_i \in \Omega$, 观测函数表示为 $a_i \in \Omega \sim O(s, i)$, O 是一组条件观察概率, $\gamma \in [0, 1]$ 是折扣因子。

2.4 Actor-Critic框架

强化学习一般符合Actor-Critic框架, 如图6所示。Actor行动模块表明状态与动作的策略结果, 该模块用来进行动作的选择; Critic评判模块里包含了估计值函数, 该函数用来评估Actor行动模块选择后的动作, 评判后又给予Actor行动模块更新, 从而形成了“状态-动作-奖励-动作”的闭环输入输出框架。

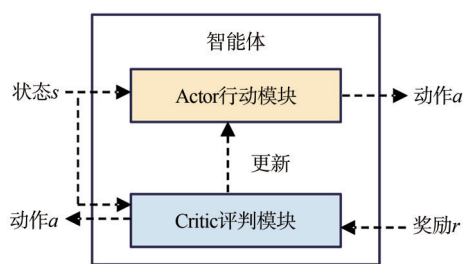


图6 Actor-Critic框架

2.5 博弈视角下的深度强化学习

MAS中智能体与智能体之间的关系通常可以映射为博弈中玩家与玩家之间的关系^[71]，这种关系通常包含竞争、合作以及竞争合作。MAS中智能体间的决策问题通常由学习目标来决定，一般学习目标包含均衡、协同和合作等。博弈论是一门用于处理个体之间交互关系（合作或竞争）的数学理论，其核心观点^[72]在于为玩家间的博弈过程构建一个与之匹配的交互策略模型，均衡解就是让玩家满意的一组交互策略集合，通过这组集合能直观地展示与博弈结果匹配的策略，这种博弈核心观点与MARL是高度契合的。因此可以将博弈论中的均衡概念映射到MARL中智能体之间的交互过程中，并将纳什均衡作为评估指标。纳什均衡描述了一个多参与者决策问题中，每个参与者在给定其他参与者选择的前提下，达到自己策略的最优状态。强调结果需收敛到纳什均衡点，这样使MARL中每个智能体的收益都能相对最大，收敛速度也相对最快。

3 博弈强化学习在低轨星座边缘计算的典型应用

目前基于单智能体任务的强化学习方法已层出不穷。低轨星座包含数量极多的卫星及复杂高动态的环境，是一个典型的分布式多智能体复杂系统，具有十分复杂和不确定的系统特性。博弈强化学习基于传统深度强化学习的基础框架，有效引入博弈论来指导多智能体的动作交互，从而建模并解决深度强化学习下的求解问题。博弈论适配于多个智能体之间任何活动的通用理论及工具，用于协调多智能体间的各异行动从而共同实现同一目标，并能对其协调策略的合理性做出解释。

3.1 博弈模型在低轨星座上的典型应用

近年来，博弈论已经广泛应用于卫星通信网络领域，解决了网络的功率、带宽、信道等资源的分

配问题^[73-75]。卫星网络中的多数节点存在自私性和趋利性，也就是说单个节点不愿消耗自身的资源来帮助其他节点用于数据转发，这种节点间强竞争关系正好可以用博弈论来完美化解。基于博弈论的路由算法或协议能激励这些网络节点相互协作从而形成一个主体网络，最终有效地管理及控制自身网络资源^[76]。博弈论对低轨卫星网络中协作组网以及边缘计算调度等问题已有学者研究^[77]，且博弈论很早就被应用到卫星网络的节点切换研究当中。例如，Peddi等^[78]提出了一种基于博弈论的异构无线网络垂直切换模型；Charilas等^[79]提出了一种将博弈论与网络选择相结合的接入控制方案；Niyato等^[80]建立了无线网络间的合作博弈模型，引入破产博弈解决接入控制问题。基于博弈论的边缘计算方法首先确定其优化模型，而后通过确定不同的博弈类型来采用不同的博弈算法设计边缘计算策略，基于博弈论求解问题的方法步骤如图7（a）所示。

在低轨星座系统中，地面用户和卫星是两大交互主体，用户和卫星需要实时进行需求任务的上传和信息的反馈，其交互过程就是典型的合作博弈过程。如图7（b）所示，首先定义地面用户平面、用户任务平面和卫星计算节点平面3个平面，这3个平面构成了交互博弈的主体过程。用户首先向低轨卫星卸载地面任务，而不同的卫星拥有不同的“金钱库存”，需要给任务“定价”后才能吸引有购买能力的卫星来“购买”此任务，从而获得收益；而低轨卫星则根据自身的“消费能力”来制定合理的购买计划，从而使自己的性价比收益最高。此过程即用户-卫星的交互合作博弈过程^[81-82]，表1所示为博弈论方法在低轨星座边缘计算领域上的研究应用。

以上分析的博弈论方法为解决卫星边缘计算问题提供了建模思路。目前，基于博弈论的方法已广泛应用于地面物联网系统，而使用博弈论在卫星边缘计算相关研究也进入研究热潮。博弈纳什均衡映射为边缘设备与计算资源之间的负载均衡，博弈联盟思想映射为设备、用户任务及计算资源间的均衡匹配等。

3.2 深度强化学习在低轨星座上的典型应用

深度强化学习（deep reinforcement learning, DRL）^[99]是强化学习与深度学习的变异结合^[100]。表2给出了深度强化学习和深度学习与强化学习的

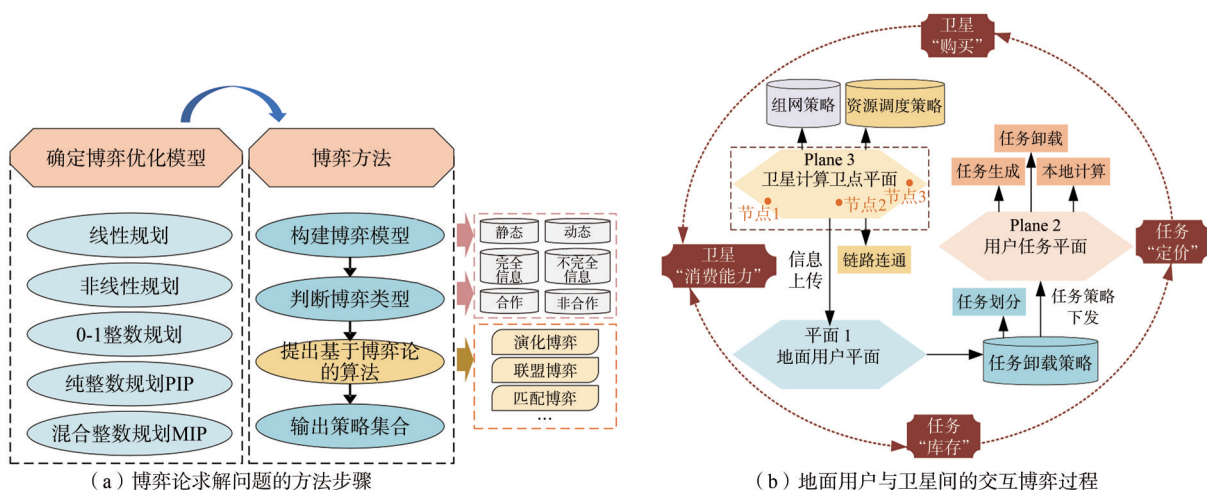


表1 面向低轨星座边缘计算的博弈论方法研究

研究方向	解决问题	方法描述	优势
卫星组网	针对多用户接入,且多颗卫星可提供服务的复杂场景	文献[83]提出一种基于演化博弈的卫星切换方法,利用博弈达到均衡状态,从而实现多用户和多颗卫星间的效用最大化	该方法在多卫星场景下具有明显稳定性和有效性
	针对星座组网过程中星间链路功率、频谱重叠等问题	文献[84-85]分别提出了基于超模博弈论的星间链路分布式功率分配方法以及基于精确潜在博弈的大规模卫星分布式频谱共享方法,建立了基于博弈论的系统模型来描述卫星链路	该方法能节省电源资源,以及在收敛速度和网络整体满意度之间取得了平衡
任务卸载	针对卫星网络边缘计算中存在的资源约束下的计算卸载问题	文献[86-87]研究了卸载过程中的用户开销最小化问题,将卸载问题建模为完全信息静态非合作博弈问题,采用双边-匹配博弈给出了卸载决策的最优解	该方法大幅度节省了卸载决策时间与系统开销
	针对服务器资源紧张的情况	文献[88-89]引入总流经时间最小化的置换流水车间调度思路,联合匹配-联盟博弈来连续迭代得到最优解	该方法有效降低了任务计算代价并获得了更好的增益
资源调度	针对大规模地面用户的超密集卫星网络中的动态数据卸载问题	文献[90-91]采用 Stackelberg 博弈来求解地面用户与卫星之间的交互问题	算法具有明显的时延优势并具有很好的收敛性
	针对联合考虑用户关联、计算资源分配等问题	文献[92-93]将问题建模为具有外部性的多对一匹配博弈,提出了一种结合计算资源分配和功率控制的匹配-联盟方法来寻找最优解	该方法能有效地降低系统总能耗
	针对边缘用户资源分配实现最小总体系统成本问题	文献[94]将问题建模为 NP-hard 的博弈问题,并提出了一个去中心化算法来实现纳什均衡	该方法能有效解决总体系统成本最小的问题
	针对有限的计算资源和移动用户的动态需求的最优化调度问题	文献[95-96]将问题建模为混合整数非线性规划问题,分别提出了基于次梯度的非合作博弈和基于潜在博弈的分散资源调度方法	该方法能最小化请求的响应延迟
	针对多颗对地观测卫星对地无线通信网络中的资源分配问题	文献[97-98]分别提出了基于 Stackelberg 博弈和基于精确势博弈的分布式调度算法	该方法首次从博弈论的角度来分析近地轨道的非地面网络问题

优缺点对比。使用 DRL 来解决低轨星座边缘计算领域上的难题已成为国内外研究热点^[101]。本章主要介绍深度强化学习在卫星边缘计算领域上应用较广的几大典型变体^[102]：深度 Q 网络（DQN）^[103]、双重深度 Q 网络（DDQN）、深度确定性策略梯度以及近端策略优化。

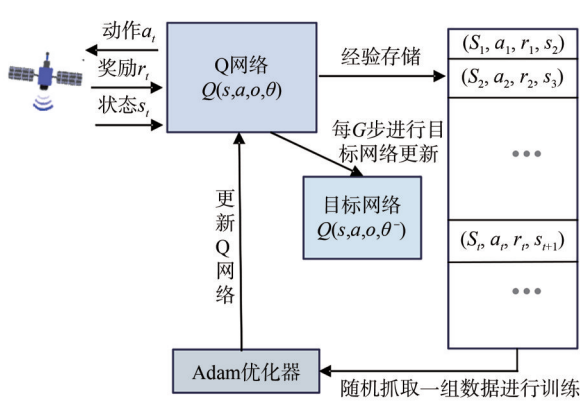
3.2.1 深度Q网络和双重深度Q网络

DQN是DRL领域的开创性鼻祖，由Mnih等^[104]首次提出。DQN由Q-Learning算法变体而来，其在网络训练中能保持很好的稳定性得益于引入了经验回放机制，该机制能增强样本间的独立性，

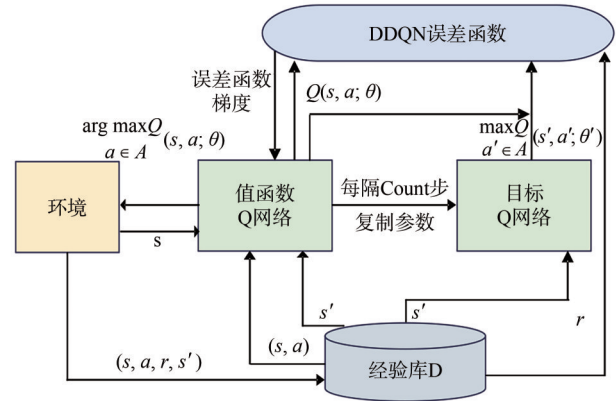
从而提高样本的训练利用率。DQN的算法框架如图8(a)所示,在低轨星座的边缘计算过程中,卫星、资源分配器以及网络数据映射为多个不同的智能体及其环境。当前环境下的初始状态为 s_t ,状态 s_t 输入DQN后产生动作 a_t ,卫星及资源分配器执行动作 a_t 时会与环境产生交互从而更新环境 s_t ,动作 a_t 对应的奖励值为 r_t 。智能体与环境的反复多轮交互产生了DQN的权重参数,每轮交互产生的状态、动作、奖励值都存储在经验池里。目标网络用于更新DQN主网络,从而弱化当前 Q 值与目标 Q 值间的关联性。

表2 三大学习方法优缺点对比

方法	优点	缺点
深度学习	1)非线性关系学习; 2)适配较复杂边缘计算场景; 3)较高准确率; 4)良好的预测能力; 5)高度可扩展性	1)需要较大的数据量支撑; 2)模型参数多且复杂; 3)欠缺较好的可解释性; 4)模型易过拟合
强化学习	1)具备与动态环境自适应学习能力; 2)适配较复杂边缘计算场景; 3)适配动态环境下的智能决策; 4)具备智能化与自主性	1)算法设计要求高; 2)算法训练的硬件设备要求高; 3)需参数调优; 4)可能会不稳定
深度强化学习	1)极强的自适应能力; 2)适配复杂的边缘计算场景; 3)不需要数据的情况下能自主的和动态环境交互学习	1)计算资源支撑; 2)需要实时环境



(a) DQN算法框架



(b) DDQN算法框架

图8 DQN算法框架和DDQN算法框架

DDQN是DQN的一个变体，算法框架如图8(b)所示。DDQN在大致结构上和DQN是相似的，优势在于DDQN中的 Q 值最大动作 a 的计算是在值函数 Q 网络中，即 $a = \arg \max Q(s, a, \theta)$ ，计算出来的动作 a 又输入目标 Q 网络中计算 Q 值，即：

$$y = r + \gamma Q'(s', \arg \max Q(s', a, \theta), \theta') \quad (3)$$

由此可以看出，DDQN这种采用解耦动作的选择和 Q 值的计算方式可以成功地消除估计过度的问题，并在很大程度上克服DQN方法的 Q 值计算估计虚高的问题。

有关DQN及其变体在低轨卫星边缘计算上的应用如表3所示。

以上基于DQN及其变体在边缘计算三大研究方向上的研究侧重点是不尽相同的。在卫星组网方面，大多研究者关注组网过程中路由耗能、流量耗能以及网络中断等问题，在双 Q 网络的计算概念上探讨最佳的路由组网方式；在任务卸载与资源调度方面，更多学者更关注星地协同系统中的任务卸载

与资源调度问题^[124]，然而边缘服务器的计算资源是固定的，而地面用户的计算任务和资源需求是动态变化的，这些信息在实际的低轨星座边缘计算中很难得到。因此，使用马尔可夫决策过程对边缘计算中的动态控制进行建模，并应用强化学习来处理协作边缘计算系统中的问题^[77,125]，是大部分研究的思路。

3.2.2 深度确定性策略梯度

DDPG是一种将DQN和策略梯度有效融合的深度强化学习算法^[126]，算法框架如图9所示。DDPG符合Actor-Critic架构。在目前的研究中，DDPG的输入数据一般来自星间链路参数和用户任务，且星间链路参数和用户任务构成了算法网络的状态空间 s 。训练网络主要包括Actor-Critic网络，而动作空间 a 一般为边缘计算策略与卫星合理范围内的计算能力，通过状态空间 s 和动作空间 a 计算得到奖励值，并筛选得到策略 (s_i, a_i, r_i) ，数据的迭代训练会带来策略的不断改善，从而策略也在不断地更新

表3 面向低轨星座边缘计算的DQN方法研究

研究方向	解决问题	方法描述	优势
卫星组网	针对卫星组网过程中,网络路由的耗能、卫星高速移动带来的流量激增问题	文献[105-107]通过改进DQN,将卫星网络协作缓存过程建模为马尔可夫决策过程,提出了更加智能的路由算法以及协同边缘缓存算法	克服了传统DQN训练慢且难以收敛的问题,提高了边缘缓存的优势
	针对固定的每链路传输功率难以获得最佳的卫星通信性能的问题	文献[108]提出了一种基于参数化深度强化学习的联合优化算法DeepISL,将优化目标建模为POMDP,由于联合优化是一个离散-连续混合动作空间问题,使用DeepISL算法来获得最优决策	大幅提高星座的能源效率,在吞吐量也有很好的性能
	针对卫星组网过程中,网络路由性能问题以及用户连接中断问题	文献[109-110]基于DQN的变体DDQN,提出了具有最短路径特性和负载均衡的路由决策策略 文献[111]提出了一种集中训练和分布式执行的多智能体决斗双深度Q网络(MA3DQN)解决方案	在时延、丢包率和系统吞吐量方面表现更好 最大限度地提高网络总数据速率的同时,最大限度地减少移动引起的切换
任务卸载	针对具有时延带宽和异构设备的常用卸载场景	文献[112-113]把任务卸载建模成马尔可夫决策过程,在DQN网络中引入优先级缓冲机制、专家缓冲机制等	很大程度降低设备能耗、时延以及任务丢包率,提高了样本利用率
	针对动态时变多用户系统下的多层卫星、多用户设备的卸载决策问题	文献[114-117]将卸载问题建模动态混合整数规划问题,在深度强化学习中引入了用户关联和边缘安全决策等	提高了方法的泛化能力,在时延和能量消耗方面都能获得较好的长期回报
资源调度	针对延迟约束以及异构计算任务的动态资源需求	文献[118-119]将问题建模为混合整数非线性规划问题,引入值迭代机制等 文献[120-122]将问题视作多目标优化问题,将资源调度的不同参与主体建模为智能体与环境,以最小化时延和能耗为优化目标,提出将DQN和多目标优化算法耦合的新思路	能显著降低用户设备在不同场景下的能耗 更能满足差异化用户需求,同时满足能耗小、时延低、用户满意度高
	解决卫星网络的动态和异构性质,确保有效的资源管理	文献[123]设计了一个两阶段框架,将深度强化学习(DRL)和局部搜索(LS)方法协同融合,并将链路可靠性纳入算法设计	在CPU容量、链路带宽、节点数据包丢失以及链路可靠性等指标良好

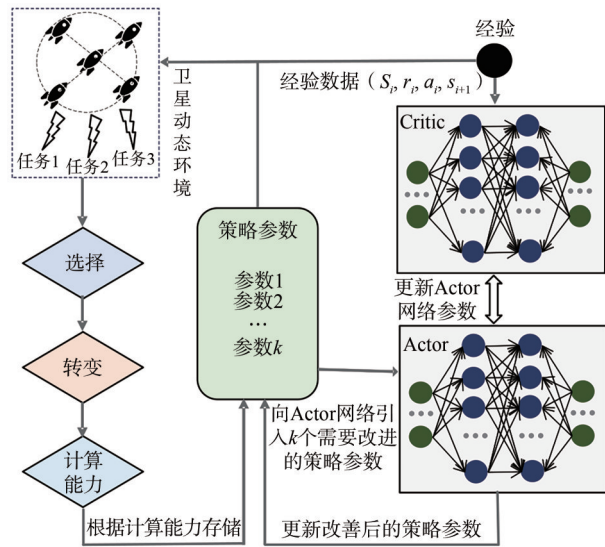


图9 DDPG算法框架

迭代,最终得到最优策略和最优策略下的系统开销。

有关DDPG及其变体在低轨卫星边缘计算上的应用如表4所示。

以上分析可知,大多数研究开始更加专注更复杂的网络场景,例如多用户的多址接入、大量计算密集且时延敏感任务卸载以及突发情况下的用户需求优先等问题。研究者们一方面在DDPG中优化多

约束条件下的奖励函数,期望满足更动态的边缘计算场景,一方面将强化学习与迁移学习、联邦学习相结合,探索更高级别的计算框架。

3.2.3 近端策略优化

PPO网络架构如图10所示。PPO算法以Actor-Critic为基础框架,首先初始化策略参数 θ ,在每次的迭代当中,状态、动作(s_t, a_t)需计算其优势函数 $A_t^{\theta^k}(s_t, a_t)$,且策略参数 θ 在交互过程中不断更新。

其中,重要性采样是PPO算法中的一种核心思想,其思想给每个步骤的数据都加上权重来表达目标策略与行为策略,从而在一定程度上提高数据的利用率。有关PPO及其变体在低轨卫星边缘计算上的应用如表5所示。

PPO算法通常和分布式计算、优势函数变体、多层优化等结合使用,在解决连续控制的应用场景上表现良好,并在处理大规模多智能体协作对抗、多智能体决策控制等方面具有优势^[147]。研究者开始将广义优势估计(GAE)、软Actor-Critic、启发式学习等和PPO算法相结合,旨在解决更加精细的难题。启发式学习等机器学习方法在卫星边缘计算领域已有应用,例如在针对大容量LEO卫星成像数据网络传输调度问题,文献[148]提出的顺序

表4 面向低轨星座边缘计算的DDPG方法研究

研究方向	解决问题	方法描述	优势
卫星组网	针对卫星路由中最短路径的寻找问题,从而达到节能路由的目的	文献[127-128]在DDPG中融入增强Dijkstra等设计新的路由算法,使卫星节点在没有先验信息的情况下,通过迭代学习找到到达目的节点的最短路径。随后,通过限制算法的跳数和采用动态贪婪系数,在满足每个数据流端到端延迟约束的同时实现路由能耗性能的长期优化	能耗性能显著提高,能满足每个数据流的端到端延迟约束
	现有的路由算法在及时更新路由表、可接受的计算资源消耗和考虑QoS要求等方面存在不足	文献[129]提出了一种动态分布式路由方案(QRLSN)。为了实现低端到端时延和低网络流量负载,QRLSN采用多目标优化方法寻找转发数据包的最优下一跳	实现低端到端时延和低网络流量负载
	卫星用户设备与卫星之间的多址接入问题	文献[130]提出了一种联合迁移学习和深度强化多址学习(FTRMA)相结合的方法,卫星在星地网络中充当联邦服务器,采用联邦学习和迁移学习算法。卫星通过星间链路将全局模型迁移到邻近的卫星上	显著提高信道链路利用率
任务卸载	面向大量计算密集型和延迟敏感型任务卸载问题	文献[131-132]通过优化资源决策参数来改进基于深度强化学习(DRL)和深度确定性策略梯度(DDPG)的动态调度策略,且方法更适配于混合计算卸载需求	适应具有突发条件的动态边缘环境
	面向随机边缘计算能力下每个任务的卸载决策和资源分配问题	文献[133]将问题映射为混合整数非线性规划(MINLP)问题,提出了一个基于Actor-Critic学习结构的深度强化学习框架	显著降低计算复杂度
	针对低轨道卫星边缘计算网络的能量最小化问题	文献[134]提出了一种具有全局奖励的多智能体深度强化学习算法,通过分散的方法优化发射功率、CPU频率、比特分配、卸载决策和带宽分配	大幅度降低地面终端能耗,且具有很好的收敛性
资源调度	针对有突发性紧急用户任务下的资源调度需求	文献[135-136]主要对DDPG中的Actor-Critic训练框架进行优化改进,将截断双Q学习(CDQL)技术融入Critic网络,提出更适配于紧急业务的强化学习资源调度机制	有效提高紧急业务服务率、业务容量和资源调度合理性等指标

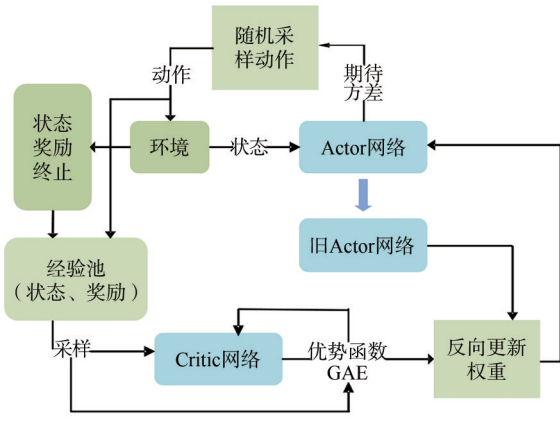


图10 PPO算法框架

两阶段启发式算法在时变卫星间宽带通信网络中也能完成传输链路的高效调度。后续,将机器学习理论方法与强化学习结合起来应用到低轨卫星边缘计算将是大多数研究者集中研究的方向。

4 挑战

人工智能时代,如何将博弈强化学习与低轨卫星边缘计算高效融合仍面临诸多挑战。

4.1 博弈强化学习细粒度求解新范式

博弈强化学习在低轨卫星边缘计算领域已被广泛应用,特别在优化数据卸载策略和满足时延约束的安全资源调度方面展现出高效性。然而,目前的研究往往偏向于理想化的调度场景,未能充分应对

实际操作中的复杂性,如用户需求的动态变化、时延敏感任务的处理,以及星座的多层异构性等。细粒度求解要求对影响任务资源分配的各种因素进行深入分析和详细建模。例如,在动态需求变化的情况下,引入基于预测的调度算法,根据历史数据和实时反馈预测未来需求,从而优化资源分配;对于时延敏感任务,实施优先级调度策略,确保任务能够优先获得所需的计算和通信资源。将这些细粒度元素整合到博弈强化学习框架中,可以发展出能够适应复杂多变条件的多用户、多任务、多维度的边缘任务调度策略。探索基于细粒度的多用户多任务多维度边缘任务调度策略,从而进一步提高低轨星座边缘计算领域的强耦合、高泛化和智能自适应能力^[149]。

4.2 卫星边缘计算智能演进新架构

目前,针对低轨卫星边缘计算的研究主要聚焦在组网、任务卸载和资源调度3个方面。基于博弈论的卫星边缘计算产生了许多模型,包含系统模型、通信模型、任务模型、资源模型等,通过深度强化学习,卫星可以快速基于博弈模型生成智能对抗决策架构。然而,面向数据特征差异、星上资源不均衡、用户隐私等问题,低轨卫星的边缘结构化升级也面临挑战。随着卫星计算能力及存储能力的不断提升,在6G星地网络系统中对卫星的边缘计算能力提出了更高的要求。目前大多数研究仅考虑

表5 面向低轨星座边缘计算的PPO方法总结

研究方向	解决问题	方法描述	特点
卫星组网	针对与卫星网络中的拥塞控制和路由有关的问题	文献[137-138]提出了一种协同多智能体深度强化学习的框架,基于DRL的软行为者批评(SAC)算法的元素,能够灵活地控制LSN的拥塞程度和带宽利用率	模型具有较好的泛化能力 支撑低轨星座网络性能优化和路由决策
	现有LEO多址接入协议针对大规模物联网接入不适配的问题	文献[139]提出了一种基于PPO方法的基于争用的LEO SAT网络大规模物联网随机接入解决方案	适配大规模物联网的接入
	针对卫星组网过程中的轨道提升问题	文献[140]将广义优势估计(GAE)和近端策略优化(PPO)方法相结合,提出了基于强化学习的时间最优小推力变轨策略优化方法	能够保证轨道转移时间的最优性
	优化近地轨道系统的长期加权和速率	文献[141]提出了一种基于PPO算法的深度强化学习框架DRLBPA,寻求在LEO卫星波束功率和GEO用户干扰约束下,使LEO卫星系统的长期加权和率最大化	在不同信道状态和随机到达数据率下具有良好的收敛性和泛化性
任务卸载	针对低轨卫星星间协作处理机制的二次卸载问题	文献[142]将星间协作处理计算任务的场景建模为分布式决策问题,提出了一种基于MAPPO的低轨卫星任务协作处理算法	算法可以收敛到更优状态
	针对地面时延敏感性任务	文献[143]提出了一种基于软Actor-Critic框架的辅助地面网络架构SAC	能耗性能较优
	针对偏远地区设备端能源和计算超时的开销难以承担的问题	文献[144]提出了一种基于异步优势actor-critic的基于卫星辅助边缘计算(SAEC)卸载(ASO)深度强化学习算法,以优化反映延迟和能量的综合奖励,并优化训练决定参与SAEC服务器规模及其计算任务分布的动作集	算法加快收敛速度,提高综合奖励
资源调度	针对卫星网络中功率资源分配问题	文献[121]基于深度强化学习技术的RSMA机制,提出了一种高效的基于PPO的方案	获得最优功率分配策略
	面向卫星波束调度问题	文献[145]将卫星建模为多目标马尔可夫决策过程,提出了一种基于多目标动态深度强化学习的动态波束调度算法(MO-DRL-DBH)	充分满足差异化业务需求,保证时延公平性
	卫星调度合适地面站天线和时间窗的问题	文献[146]将深度强化学习方法与启发式方法相结合,DRL是确定任务分配过程,启发式调度方法可以快速解决单天线调度问题	在不同场景下都表现优异

地面站、卫星以及云服务器等设备的服务,然而空中飞行器、无人机等都可与低轨卫星建立更加强健的通信链路从而辅助卫星边缘计算^[150]。对用户终端的差异化业务需求,以及星地协同网络平台的资源异构性,如何从全局角度,实现多层次、异构、泛在计算资源的融合与统一调度十分重要^[20]。未来,面向大型化和多样化来构建新颖且高效的卫星边缘演进架构是一项重大课题^[151],其将多智能体强化学习与博弈模型合理分布结合,构建满足不同类型节点的高效混合卫星边缘计算演进新架构,从而进一步提高卫星网络的灵活性、服务轻便性和健壮性。

4.3 空天地一体化边缘智能新需求

空天地一体化边缘智能要求整合天上(如卫星)、空中(如飞机、无人机)和地面(如地面服务器)资源,在边缘节点实现数据的即时处理和智能决策。“一体化”的实现关键在于不同层之间的高效互通互联,在通信协议、数据交换格式、接口、卫星互联技术等方面都提出了高要求,需要采用标准化的通信协议(例如5G通信)确保不同设备与平台间的兼容性;需要规范统一的数据格式(例如Json或XML)来确保数据的无缝交换和处理

等。博弈强化学习可以显著助力空天地一体化边缘智能系统的优化和决策过程,特别是在资源分配、任务调度和响应策略的自动化方面。博弈强化学习通过学习各个节点(地面、空中、空间)的使用模式和需求,能够动态优化资源分配,并模拟不同决策者(即系统的不同组成部分)之间的策略和对抗,从而开发出更复杂和有效的决策策略^[152]。通过这些机制,博弈强化学习不仅为空天地一体化边缘智能系统提供了强大的决策支持工具,也极大地提升了其自主性、效率和实时响应能力,使其能够在复杂和动态的环境中保持高效和可靠的运作。

未来,移动边缘计算和博弈强化学习将被进一步引入空天地一体化网络,助力天基、空基、地基一体化边缘网络架构,空天地一体化边缘智能对卫星的能源稳定供应、高中低轨卫星的链路稳定以及并发用户任务的需求都提出了极大的挑战。基于博弈论、分布式等理论,综合利用联邦学习、元博弈、多智能体智能决策、生成式人工智能、学习-优化-反馈演进架构、云原生等技术,设计多层边缘智能优化类方法、强泛化自适应演进框架,生成面向空天地多域协同、高中低轨卫星多类合作、

5G-6G 通信结合等边缘智能需求的自适应、自优化、抗干扰、贴合实际应用的人工智能新方法,为空地一体化网络架构提供全栈支持。

5 结束语

博弈论为多智能体深度强化学习建立了策略学习模型,博弈强化学习为低轨星座的边缘计算赋能“AI 智慧”。本文围绕低轨星座边缘计算3种主要研究方向的需求展开,详细阐述了博弈强化学习范式基础,并总结归纳了博弈模型、DQN、DDPG、PPO 等方法及其变体在低轨卫星边缘计算典型场景下的应用情况,最后基于该领域的前沿研究进行展望,为博弈强化学习在空天领域的边缘计算问题求解提供参考。

参考文献:

- [1] ZHI F K, SHI S H, YU L L, et al. Research on the development status and key technologies of large LEO remote sensing constellations[J]. *Spacecraft Recovery & Remote Sensing*, 2023, 44(1): 93-101.
- [2] LEE J Y, LIM B, KO Y C. Performance analysis of multi-hop low earth orbit satellite network over mixed RF/FSO links[J]. *ICT Express*, 2024.
- [3] 于帆, 垣信. 卫星融资 67 亿元布局低轨卫星互联网[N]. *中国信息化周报*, 2024-02-26(011).
YU F, YUAN X. Satellite raised 6.7 billion yuan to lay out the low-orbit satellite Internet[N]. *China Information Weekly*, 2024-02-26(011).
- [4] LIN X Q, CIONI S, CHARBIT G, et al. On the path to 6G: embracing the next wave of low earth orbit satellite access[J]. *IEEE Communications Magazine*, 2021, 59(12): 36-42.
- [5] JIANG W W. Software defined satellite networks: a survey[J]. *Digital Communications and Networks*, 2023, 9(6): 1243-1264.
- [6] WANG B, FENG T, HUANG D Y. A joint computation offloading and resource allocation strategy for LEO satellite edge computing system[C]// *Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT)*. Piscataway: IEEE Press, 2020: 649-655.
- [7] CUI J J, LIU Y W, NALLANATHAN A. Multi-agent reinforcement learning-based resource allocation for UAV networks[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(2): 729-743.
- [8] SALAS M, FERNÁNDEZ J, VAN DEN IJSSEL J. Precise relative navigation in medium earth orbits with global navigation satellite systems and intersatellite links for black hole imaging[J]. *Advances in Space Research*, 2024, 74(6): 2599-2614.
- [9] 朱振才, 张晟宇, 胡海鹰. 分布式空间系统组网技术综述[J]. *航天器工程*, 2022, 27(6): 77-83.
ZHU Z C, ZHANG S Y, HU H Y. Overview of networking technology of distributed space system[J]. *Spacecraft engineering*, 2022, 27(6): 77-83.
- [10] JIAN W, MIN J, LIANG Z. Research on 6G-oriented LEO satellite edge computing architecture[J]. *Application of Electronic Technique*, 2022, 48(6): 5-9, 13.
- [11] ZHUANG Z L, LI Y, SUN Y N, et al. Network-based dynamic dispatching rule generation mechanism for real-time production scheduling problems with dynamic job arrivals[J]. *Robotics and Computer-Integrated Manufacturing*, 2022, 73: 102261.
- [12] PROL F S, FERRE R M, SALEEM Z, et al. Position, navigation, and timing (PNT) through low earth orbit (LEO) satellites: a survey on current status, challenges, and opportunities[J]. *IEEE Access*, 2022, 10: 83971-84002.
- [13] CASSARÁ P, GOTTA A, MARCHESE M, et al. Orbital edge offloading on mega-LEO satellite constellations for equal access to computing[J]. *IEEE Communications Magazine*, 2022, 60(4): 32-36.
- [14] ABDELSEDEK M Y, CHAUDHRY A U, DARWISH T, et al. Future space networks: toward the next giant leap for humankind[J]. *IEEE Transactions on Communications*, 2023, 71(2): 949-1007.
- [15] LIN L, LIAO X F, JIN H, et al. Computation offloading toward edge computing[J]. *Proceedings of the IEEE*, 2019, 107(8): 1584-1607.
- [16] SADATDIYNOV K, CUI L Z, ZHANG L, et al. A review of optimization methods for computation offloading in edge computing networks[J]. *Digital Communications and Networks*, 2023, 9(2): 450-461.
- [17] 周绪, 苗辉, 杨静, 等. 边缘计算资源调度: 历史、架构、建模与方法分析[J]. *计算机集成制造系统*, 已录用, doi: 10.13196/j.cims.2023.0132.
ZHOU X, MIAO H, YANG J, et al. Edge computing resource scheduling: History, architecture, modeling and Method analysis [J]. *Department of Computer Integrated Manufacturing*, 2024: accepted, doi: 10.13196/j.cims.2023.0132.
- [18] 陈全, 杨磊, 郭剑鸣, 等. 低轨巨型星座网络: 组网技术与研究现状[J]. *通信学报*, 2022, 43(5): 177-189.
CHEN Q, YANG L, GUO J M, et al. Low orbit giant constellation network: Networking technology and research status [J]. *Journal of Communication*, 2022, 43(5): 177-189.
- [19] 王鹏, 张佳鑫, 张兴, 等. 低轨卫星智能多接入边缘计算网络: 需求、架构、机遇与挑战[J]. *移动通信*, 2021, 45(5): 35-46.
WANG P, ZHANG J X, ZHANG X, et al. Low earth orbit satellite intelligent multi-access edge computing networks: requirements, architecture, opportunities and challenges[J]. *Mobile Communications*, 2021, 45(5): 35-46.
- [20] 高媛, 方海, 赵扬, 等. 星地协同网络中的边缘计算技术综述[J]. *空间电子技术*, 2023, 20(2): 15-21.
GAO Y, FANG H, ZHAO Y, et al. A survey on edge computing in satellite-terrestrial networks[J]. *Space Electronic Technology*, 2023, 20(2): 15-21.
- [21] 吴戎, 秦大国, 李阳, 等. 遥感卫星边缘计算架构设计与关键技术分析[J]. *信息工程大学学报*, 2023, 24(3): 336-342.
WU X, QIN D G, LI Y, et al. Remote sensing satellite edge computing architecture design and key technology analysis [J]. *Journal of Information Engineering University*, 2023, 24(3): 336-342.
- [22] KALOEV M, KRASTEV G. Experiments focused on exploration in deep reinforcement learning[C]// *Proceedings of the 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. Piscataway: IEEE Press, 2021: 351-355.
- [23] KALOEV M, KRASTEV G. Comprehensive review of benefits from the use of sparse updates techniques in reinforcement learning: experimental simulations in complex action space environments[C]// *Proceedings of the 2023 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)*. Piscataway: IEEE Press, 2023: 1-7.
- [24] LIU W, ZHANG S C, ZHANG X P, et al. Distributed game coordina-

- tion control method based on reinforcement learning for multi-microgrid[J]. *Energy Reports*, 2023, 9: 912-921.
- [25] KAZEMI N, AZGHANI M. Secure spectrum sharing and power allocation by multi agent reinforcement learning[J]. *Digital Signal Processing*, 2024, 146: 104369.
- [26] ZHU Y Y, YIN X, CHEN C L. Extracting decision tree from trained deep reinforcement learning in traffic signal control[J]. *IEEE Transactions on Computational Social Systems*, 2023, 10(4): 1997-2007.
- [27] ANTHONY T, ECCLES T, TACCHETTI A, et al. Learning to play No-press diplomacy with best response policy iteration[EB]. *arXiv preprint*, 2020, arXiv: 2006.04635.
- [28] JIANG C L, CAO T F, GUAN J F. Intelligent task offloading and collaborative computation over D2D communication[J]. *China Communications*, 2021, 18(3): 251-263.
- [29] HONG Z C, CHEN W H, HUANG H W, et al. Multi-hop cooperative computation offloading for industrial IoT-edge-cloud computing environments[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2019, 30(12): 2759-2774.
- [30] TRAN T X, POMPILI D. Joint task offloading and resource allocation for multi-server mobile-edge computing networks[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(1): 856-868.
- [31] DAO M, LIU Z, GUO S, et al. A computation offloading and resource allocation mechanism based on minimizing devices energy consumption and system delay[J]. *Journal of Electronics & Information Technology*, 2019, 41(11): 2684-2690.
- [32] 滕颖蕾, 刘薇, 欧阳卫平, 等. 无线MEC系统中队列状态感知的卸载和传输联合优化[J]. *北京邮电大学学报*, 2019, 42(3): 14-20.
- TENG Y L, LIU W, OUYANG W P, et al. Joint offloading and transmission optimization of queue state awareness in wireless MEC systems[J]. *Journal of Beijing University of Posts and Telecommunications*, 2019, 42(3): 14-20.
- [33] 邓志龙, 张琦玮, 曹皓, 等. 一种基于深度强化学习的调度优化方法[J]. *西北工业大学学报*, 2017, 35(6): 1047-1053.
- DENG Z L, ZHANG Q W, CAO H, et al. A scheduling optimization method based on depth intensive study[J]. *Journal of Northwestern Polytechnical University*, 2017, 35(6): 1047-1053.
- [34] CUI Y P, LIANG Y J, WANG R Y. Resource allocation algorithm with multi-platform intelligent offloading in D2D-enabled vehicular networks[J]. *IEEE Access*, 2019, 7: 21246-21253.
- [35] MENG H, CHAO D C, GUO Q Y, et al. Delay-sensitive task scheduling with deep reinforcement learning in mobile-edge computing systems[J]. *Journal of Physics: Conference Series*, 2019, 1229(1): 012059.
- [36] HU Z B, WANG S Z, HU L S, et al. Optimization of task unloading strategy based on Game Theory in cloud edge collaborative system[C]// *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*. New York: ACM, 2022: 228-234.
- [37] DARWISH T, KURT G K, YANIKOMEROGLU H, et al. LEO satellites in 5G and beyond networks: a review from a standardization perspective[J]. *IEEE Access*, 2022, 10: 35040-35060.
- [38] XIAO Z Y, YANG J Y, MAO T Q, et al. LEO satellite access network (LEO-SAN) toward 6G: challenges and approaches[J]. *IEEE Wireless Communications*, 2024, 31(2): 89-96.
- [39] JI B, CAO X Y, YANG K. Research and simulation of satellite networking technology based on link detection mechanism[J]. *Electronic Information Warfare Technology*, 2022, 37(1): 90-96.
- [40] RAJAGOPAL A, RAMACHANDRAN A, SHANKAR K, et al. Optimal routing strategy based on extreme learning machine with beetle antennae search algorithm for Low Earth Orbit satellite communication networks[J]. *International Journal of Satellite Communications and Networking*, 2021, 39(3): 305-317.
- [41] OUYANG M, ZHANG R, WANG B Q, et al. Network coding-based multipath transmission for LEO satellite networks with domain cluster[J]. *IEEE Internet of Things Journal*, 2024, 11(12): 21659-21673.
- [42] FENG L, LIU Y F, WU L, et al. A satellite handover strategy based on MIMO technology in LEO satellite networks[J]. *IEEE Communications Letters*, 2020, 24(7): 1505-1509.
- [43] LI Y T, ZHOU W Y, ZHOU S L. Forecast based handover in an extensible multi-layer LEO mobile satellite system[J]. *IEEE Access*, 2020, 8: 42768-42783.
- [44] DU P P, LI J D, BAI W G, et al. Dual location area based distributed location management for hybrid LEO/MEO mega satellite networks[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(2): 2307-2321.
- [45] ZHU H T, GUO Q. User group based multi-satellite handover strategy for LEO satellite networks[J]. *Telecommunications Science*, 2022, 38(4): 39-48.
- [46] ROTH M, BRANDT H, BISCHL H. Implementation of a geographical routing scheme for low Earth orbiting satellite constellations using intersatellite links[J]. *International Journal of Satellite Communications and Networking*, 2021, 39(1): 92-107.
- [47] CHEN Q, CHEN X Q, YANG L, et al. A distributed congestion avoidance routing algorithm in mega-constellation network with multi-gateway[J]. *Acta Astronautica*, 2019, 162: 376-387.
- [48] WU S G, WANG H Y, WANG Y, et al. Research and analysis on routing technology of LEO satellite network[J]. *Satellite & Network*, 2021(9): 66-74.
- [49] ZHOU J, GONG X T, SUN L J, et al. Adaptive routing strategy based on improved double Q-learning for satellite Internet of Things[J]. *Security and Communication Networks*, 2021, 5530023.
- [50] 倪少杰, 岳洋, 左勇, 等. 卫星网络路由技术现状及展望[J]. *电子与信息学报*, 2023, 45(2): 383-395.
- NI S J, YUE Y, ZUO Y, et al. The status quo and prospect of satellite network routing technology[J]. *Journal of Electronics & Information Technology*, 2023, 45(2): 383-395.
- [51] LIU Y, WANG L N, et al. LEO satellite intelligent-sensing routing algorithm based on a dendrite network[J]. *Chinese Journal of Engineering*, 2023, 45(3): 465-474.
- [52] ZHU D L, LIU H T, LI T, et al. Deep reinforcement learning-based task offloading in satellite-terrestrial edge computing networks[C]// *Proceedings of the 2021 IEEE Wireless Communications and Networking Conference (WCNC)*. Piscataway: IEEE Press, 2021: 1-7.
- [53] LI C C, ZHANG Y S, XIE R C, et al. Integrating edge computing into low earth orbit satellite networks: architecture and prototype[J]. *IEEE Access*, 2021, 9: 39126-39137.
- [54] XIE R C, TANG Q Q, WANG Q N, et al. Satellite-terrestrial integrated edge computing networks: architecture, challenges, and open issues[J]. *IEEE Network*, 2020, 34(3): 224-231.
- [55] WANG J X, ZHANG F, WANG X M, et al. AN edge-to-edge collaborative approach for AI tasks in LEO networks[C]// *Proceedings of the 2023 8th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*. Piscataway: IEEE Press, 2023: 300-304.
- [56] ABDERRAHIM W, AMIN O, ALOUINI M S, et al. Latency-aware offloading in integrated satellite terrestrial networks[J]. *IEEE Open Journal of the Communications Society*, 2020, 1: 490-500.
- [57] BI J, YUAN H T, DUANMU S F, et al. Energy-optimized partial com-

- putation offloading in mobile-edge computing with genetic simulated-annealing-based particle swarm optimization[J]. *IEEE Internet of Things Journal*, 2021, 8(5): 3774-3785.
- [58] TANG Q Q, FEI Z S, LI B, et al. Computation offloading in LEO satellite networks with hybrid cloud and edge computing[J]. *IEEE Internet of Things Journal*, 2021, 8(11): 9164-9176.
- [59] SHENG S. A resource collaborative scheduling strategy based on cloud edge framework[C]//*Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. Piscataway: IEEE Press, 2021: 1740-1744.
- [60] GONG Y K, YAO H P, WANG J J, et al. Multi-agent driven resource allocation and interference management for deep edge networks[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(2): 2018-2030.
- [61] LI J L, SHANG Y, QIN M, et al. Multiobjective oriented task scheduling in heterogeneous mobile edge computing networks[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(8): 8955-8966.
- [62] AMINI MOTLAGH A, MOVAGHAR A, RAHMANI A M. Task scheduling mechanisms in cloud computing: a systematic review[J]. *International Journal of Communication Systems*, 2020, 33(6): e4302.
- [63] YUAN Y X, LEI L, VU T X, et al. Adapting to dynamic LEO-B5G systems: meta-critic learning based efficient resource scheduling[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(11): 9582-9595.
- [64] YANG F, YANG S W, CHEN Y K, et al. Efficient pencil beam synthesis in 4-D antenna arrays using an iterative convex optimization algorithm[J]. *IEEE Transactions on Antennas and Propagation*, 2019, 67(11): 6847-6858.
- [65] LIU Z J, PANG S, PU C Y, et al. Research of reverse iterative control algorithm[C]//*Proceedings of the 2018 Chinese Automation Congress (CAC)*. Piscataway: IEEE Press, 2018: 3428-3433.
- [66] LIU H, WANG K P, LIU Z K. Application and research of heuristic search algorithm in crawler field[C]//*Proceedings of the 2020 2nd International Conference on Information Technology and Computer Application (ITCA)*. Piscataway: IEEE Press, 2020: 41-44.
- [67] NIU L W, CHEN X F, ZHANG N, et al. Multiagent meta-reinforcement learning for optimized task scheduling in heterogeneous edge computing systems[J]. *IEEE Internet of Things Journal*, 2023, 10(12): 10519-10531.
- [68] GENG L W, ZHAO H B, WANG J Y, et al. Deep-reinforcement-learning-based distributed computation offloading in vehicular edge computing networks[J]. *IEEE Internet of Things Journal*, 2023, 10(14): 12416-12433.
- [69] 王尧, 罗俊仁, 周桢忠, 等. 面向策略探索的强化学习与进化计算方法综述[J]. *计算机科学*, 2024, 51(3): 183-197.
- WANG Y, LUO J R, ZHOU Y Z, et al. Review of reinforcement learning and evolutionary computation methods for strategy exploration[J]. *Computer Science*, 2024, 51(3): 183-197.
- [70] DANESHFAR F, BEVRANI H. Load-frequency control: a GA-based multi-agent reinforcement learning[J]. *IET Generation, Transmission & Distribution*, 2010, 4(1): 13.
- [71] ZOU Q J, JIANG Y J, GAO B, et al. An overview of cooperative multi-agent deep reinforcement learning[J]. *Aero Weaponry*, 2022, 29(6): 78-88.
- [72] LUO J R, ZHANG W P, XIANG F T, et al. Survey on intelligent wargaming: tactical & campaign wargame and strategic game from game-theoretic perspective[J]. *Journal of System Simulation*, 2023, 35(9): 1871-1894.
- [73] FAN W H, HUA M Y, ZHANG Y Y, et al. Game-based task offloading and resource allocation for vehicular edge computing with edge-edge cooperation[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(6): 7857-7870.
- [74] SAKIP A, YERSAINOV R, ATASHIKOVA M, et al. Lightweight energy-efficient offloading framework for mobile edge/cloud computing[C]//*Proceedings of the 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. Piscataway: IEEE Press, 2023: 1-8.
- [75] LIANG Z Z, LIUT Y, HUANG K B, et al. I/O interference aware multiuser computation offloading for virtualized edge computing[C]//*Proceedings of the 2019 IEEE International Conference on Communications (ICC)*. Piscataway: IEEE Press, 2019: 1-6.
- [76] ZHAO L R, DANG Z H, ZHANG Y L. Orbital game: concepts, principles and methods[J]. *Journal of Command and Control*, 2021, 7(3): 215-224.
- [77] XU X S, DANG Z H, SONG B, et al. Method for cluster satellite orbit pursuit-evasion game based on multi-agent deep deterministic policy gradient algorithm[J]. *Aerospace Shanghai (Chinese & English)*, 2022, 39(2): 24-31.
- [78] PEDDI S B, PATIL S R. Game theory based vertical handoff decision model for media independent handover in heterogeneous wireless networks[C]//*Proceedings of the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. Piscataway: IEEE Press, 2016: 719-724.
- [79] CHARILAS D E, MARKAKI O I, VLACHEAS P T. Admission control as a non-cooperative multi-stage game between wireless networks[C]//*Proceedings of the 2009 16th International Conference on Systems, Signals and Image Processing*. Piscataway: IEEE Press, 2009: 1-5.
- [80] NIYATO D, HOSSAIN E. Dynamics of network selection in heterogeneous wireless networks: an evolutionary game approach[J]. *IEEE Transactions on Vehicular Technology*, 2009, 58(4): 2008-2017.
- [81] WANG H L, ZUO Y, CHANG Q. Target protection using multiple unmanned aerial vehicles based on Stackelberg security game[C]//*Proceedings of the 2023 4th International Conference on Intelligent Design(ICID)*. Piscataway: IEEE Press, 2023: 169-174.
- [82] CHEN J X, WU Q H, XU Y H, et al. A multi-leader multi-follower stackelberg game for coalition-based UAV MEC networks[J]. *IEEE Wireless Communications Letters*, 2021, 10(11): 2350-2354.
- [83] 底晓强, 于力伟, 刘旭, 等. 一种基于演化博弈的低轨卫星切换算法研究[J]. *南京大学学报*, 2018, 54(4): 855-862.
- DI X Q, YU L W, LIU X, et al. Research on a Low orbit satellite Switching Algorithm based on Evolutionary game[J]. *Journal of Nanjing University*, 2018, 54(4): 855-862.
- [84] ZHANG H M, LI Q, ZHANG Y J, et al. Game theory based power allocation method for inter-satellite links in LEO/MEO two-layered satellite networks[C]//*Proceedings of the 2021 IEEE/CIC International Conference on Communications in China (ICCC)*. Piscataway: IEEE Press, 2021: 398-403.
- [85] LI W, JIA L L, CHEN Q, et al. A game theory-based distributed downlink spectrum sharing method in large-scale hybrid satellite constellations[J]. *IEEE Transactions on Communications*, 2024, 72(8): 4620-4632.
- [86] 冯彤. 面向低轨卫星的移动边缘计算卸载策略研究[D]. 桂林: 桂林电子科技大学, 2021.
- FENG T. Research on unloading strategy of mobile edge computing for low-orbit satellites[D]. Guilin: Guilin University of Electronic Sci-

- ence and Technology, 2021.
- [87] FANG H, JIA Y Y, WANG Y L, et al. Matching game based task offloading and resource allocation algorithm for satellite edge computing networks[C]//Proceedings of the 2022 International Symposium on Networks, Computers and Communications (ISNCC). Piscataway: IEEE Press, 2022: 1-5.
- [88] 谢杰成. 面向低轨卫星系统的多星边缘计算卸载策略研究[D]. 桂林: 桂林电子科技大学, 2023.
- XIE J C. Research on multi-star edge computing unloading strategy for low-orbit satellite system [D]. Guilin: Guilin University of Electronic Science and Technology, 2023.
- [89] TONG M L, WANG X X, LI S, et al. Joint offloading decision and resource allocation in mobile edge computing-enabled satellite-terrestrial network[J]. Symmetry, 2022, 14(3): 564.
- [90] XIN K J, FEI X, LIN P N. Research on task offloading strategy based on stackelberg game in LEO satellite network[J]. Journal of Xi'an Technological University, 2023, 43(4): 393-404.
- [91] WANG D Z, WANG W, KANG Y H, et al. Distributed data offloading in ultra-dense LEO satellite networks: a stackelberg mean-field game approach[J]. IEEE Journal of Selected Topics in Signal Processing, 2023, 17(1): 112-127.
- [92] XU C, ZHENG G Y, TANG L R. Energy-aware user association for NOMA-based mobile edge computing using matching-coalition game[J]. IEEE Access, 2020, 8: 61943-61955.
- [93] 夏玮玮, 胡静, 宋铁成. 低地球轨道卫星边缘计算场景中任务卸载与资源分配联合优化算法[J]. 通信学报, 2024, 45(07): 48-60.
- XIA W W, HU J, SONG T C. Joint optimization algorithm for task offloading and resource allocation in low earth orbit satellites edge computing scenario[J]. Journal of Communication, 2024, 45(07): 48-60.
- [94] HE Q, CUI G M, ZHANG X Y, et al. A game-theoretical approach for user allocation in edge computing environment[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 31(3): 515-529.
- [95] HU S H, LI G H. Dynamic request scheduling optimization in mobile edge computing for IoT applications[J]. IEEE Internet of Things Journal, 2020, 7(2): 1426-1437.
- [96] GAO X Q, LIU R K, KAUSHIK A. Virtual network function placement in satellite edge computing with a potential game approach[J]. IEEE Transactions on Network and Service Management, 2022, 19(2): 1243-1259.
- [97] IM G, JUNG D H, RYU J G. Enhancing the Connectivity of Satellite IoT Devices in Terrestrial-Non terrestrial Integrated Networks based on the Stackelberg Game Approach[C]//Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC). Piscataway: IEEE Press, 2020: 1783-1785.
- [98] FENG R, LIN Z Y, WU P, et al. Potential game based distributed task scheduling for multi-earth observation satellites[C]//Proceedings of the 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS). Piscataway: IEEE Press, 2020: 411-416.
- [99] 金堃, 邓向阳, 于柯远. 深度强化学习之近端策略优化研究[J]. 物联网技术, 2023, 13(7): 69-75.
- JIN K, DENG X Y, YU K Y. Research on near-end strategy optimization of deep reinforcement learning[J]. Internet of Things Technologies, 2023, 13(7): 69-75.
- [100] MALEKI H, BAŞARAN M, DURAK-ATA L. Handover-enabled dynamic computation offloading for vehicular edge computing networks[J]. IEEE Transactions on Vehicular Technology, 2023, 72(7): 9394-9405.
- [101] CAO Z L, ZHOU P, LI R X, et al. Multiagent deep reinforcement learning for joint multichannel access and task offloading of mobile-edge computing in industry 4.0[J]. IEEE Internet of Things Journal, 2020, 7(7): 6201-6213.
- [102] MUSTAFA E, SHUJA J, BILAL K, et al. Reinforcement learning for intelligent online computation offloading in wireless powered edge networks[J]. Cluster Computing, 2023, 26(2): 1053-1062.
- [103] YANG Y J, CHEN X, CHEN Y, et al. Green-oriented offloading and resource allocation by reinforcement learning in MEC[C]//Proceedings of the 2019 IEEE International Conference on Smart Internet of Things (SmartIoT). Piscataway: IEEE Press, 2019: 378-382.
- [104] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[EB]. arXiv preprint, 2013, arXiv: 1312.5602.
- [105] LIU J H, ZHAO B K, XIN Q, et al. DRL-ER: an intelligent energy-aware routing protocol with guaranteed delay bounds in satellite mega-constellations[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(4): 2872-2884.
- [106] CHEN S W, YAO Z, JIANG X F, et al. Multi-agent deep reinforcement learning-based cooperative edge caching for ultra-dense next-generation networks[J]. IEEE Transactions on Communications, 2021, 69(4): 2441-2456.
- [107] WANG S Y, HUANG Y, CLERCKX B. Dynamic air-ground collaboration for multi-access edge computing[C]//Proceedings of the ICC 2022 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2022: 5365-5371.
- [108] LI Y, LUO J T, RAN Y Y, et al. DeepISL: joint optimization of LEO inter-satellite link planning and power allocation via parameterized deep reinforcement learning[C]//Proceedings of the 2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 3977-3982.
- [109] WANG C, WANG H W, WANG W D. A two-hops state-aware routing strategy based on deep reinforcement learning for LEO satellite networks[J]. Electronics, 2019, 8(9): 920.
- [110] 刘治国, 张姣姣, 潘成胜. 基于 Dueling-DDQN 的星上带宽资源预留算法研究[J]. 兵器装备工程学报, 2023, 44(12): 272-277, 284.
- LIU Z G, ZHANG J J, PAN C S. Research on on-board bandwidth resource reservation Algorithm based on Dueling DDQN[J]. Journal of Ordnance Engineering, 2023, 44(12): 272-277, 284.
- [111] BIRABWA D J, RAMOTSOELA D, VENTURA N. Multi-agent deep reinforcement learning for user association and resource allocation in integrated terrestrial and non-terrestrial networks[J]. Computer Networks, 2023, 231: 109827.
- [112] HSU C K. A dueling DQN-based computational offloading method in MEC-enabled IIoT network[J]. The Computer Journal, 2023, 66(12): 2887-2896.
- [113] CHEN N, ZHANG S, QIAN Z Z, et al. When learning joins edge: real-time proportional computation offloading via deep reinforcement learning[C]//Proceedings of the 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS). Piscataway: IEEE Press, 2019: 414-421.
- [114] LI J, GAO H, LV T J, et al. Deep reinforcement learning based computation offloading and resource allocation for MEC[C]//Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2018: 1-6.
- [115] 左珮玲, 侯少龙, 郭超, 等. 基于强化学习的多层卫星网络边缘安全决策方法[J]. 通信学报, 2022, 43(6): 189-199.
- ZUO P L, HOU S L, GUO C, et al. Security decision method for the edge of multi-layer satellite network based on reinforcement learning[J].

- Journal on Communications, 2022, 43(6): 189-199.
- [116] CUI G F, LI X Y, XU L X, et al. Latency and energy optimization for MEC enhanced SAT-IoT networks[J]. IEEE Access, 2020, 8: 55915-55926.
- [117] PANG S C, ZHENG J Y, WANG M, et al. Minimize average tasks processing time in satellite mobile edge computing systems via a deep reinforcement learning method[J]. Journal of Cloud Computing, 2023, 12(1): 159.
- [118] 李天宇. 基于强化学习的云计算资源调度策略研究[J]. 上海电力学院学报, 2019, 35(4): 399-403.
- LI T Y. Research on cloud computing resource scheduling strategy based on reinforcement learning[J]. Journal of Shanghai University of Electric Power, 2019, 35(4): 399-403.
- [119] JIANG K, ZHOU H, LI D W, et al. A Q-learning based method for energy-efficient computation offloading in mobile edge computing[C]// Proceedings of the 2020 29th International Conference on Computer Communications and Networks (ICCCN). Piscataway: IEEE Press, 2020: 1-7.
- [120] 彭明阳. 面向低轨星座的跳波束机制及资源分配研究[D]. 南京: 南京邮电大学, 2023.
- PENG M Y. Research on beam hopping mechanism and resource allocation for low orbit constellation[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.
- [121] 张沛, 刘帅军, 马治国, 等. 基于深度增强学习和多目标优化改进的卫星资源分配算法[J]. 通信学报, 2020, 41(6): 51-60.
- ZHANG P, LIU S J, MA Z G, et al. Improved satellite resource allocation algorithm based on DRL and MOP[J]. Journal on Communications, 2020, 41(6): 51-60.
- [122] ZHANG Y C, HU X, CHEN R, et al. Dynamic beam hopping for DVB-S2X satellite: a multi-objective deep reinforcement learning approach[C]//Proceedings of the 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence(DSCI) and Smart Computing, Networking and Services(SmartCNS). Piscataway: IEEE Press, 2019: 164-169.
- [123] ZHANG P Y, CUI Z X, KUMAR N, et al. Local search resource allocation algorithm for space-based backbone network in Deep Reinforcement Learning method[J]. Ad Hoc Networks, 2024, 163: 103575.
- [124] YAO Y, DIAO X B, WANG X D, et al. Joint offloading strategy and trajectory optimization for aerial-ground cooperative mobile edge computing systems[C]//Proceedings of the IET 8th International Conference on Wireless, Mobile & Multimedia Networks. London: IET, 2019: 31-37.
- [125] NIE Y W, ZHAO J H, GAO F F, et al. Semi-distributed resource management in UAV-aided MEC systems: a multi-agent federated reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology, 2021, 70(12): 13162-13173.
- [126] LAI J Y, LIU H S, SUN Y S, et al. Multi-agent deep reinforcement learning aided computing offloading in LEO satellite networks[C]// Proceedings of the ICC 2023 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2023: 3438-3443.
- [127] WANG X T, DAI Z Q, XU Z. LEO satellite network routing algorithm based on reinforcement learning[C]//Proceedings of the 2021 IEEE 4th International Conference on Electronics Technology (ICET). Piscataway: IEEE Press, 2021: 1105-1109.
- [128] HSU Y H, LEE J N, XU F M. A deep reinforcement learning based routing scheme for LEO satellite networks in 6G[C]//Proceedings of the 2023 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2023: 1-6.
- [129] HUANG Y X, WU S F, KANG Z Y, et al. Reinforcement learning based dynamic distributed routing scheme for mega LEO satellite networks[J]. Chinese Journal of Aeronautics, 2023, 36(2): 284-291.
- [130] MA T, XIN X R, MA X, et al. An intelligent multiple access protocol: federated transfer reinforcement learning for satellite-ground network[C]//Proceedings of the 2023 IEEE 11th International Conference on Information, Communication and Networks (ICIN). Piscataway: IEEE Press, 2023: 218-223.
- [131] NATH S, WU J X. Dynamic computation offloading and resource allocation for multi-user mobile edge computing[C]//Proceedings of the 2020 IEEE Global Communications Conference. Piscataway: IEEE Press, 2020: 1-6.
- [132] CAO S H, CHEN S, CHEN H, et al. HCOME: research on hybrid computation offloading strategy for MEC based on DDPG[J]. Electronics, 2023, 12(3): 562.
- [133] YAN J, BI S Z, ZHANG Y J A. Offloading and resource allocation with general task graph in mobile edge computing: a deep reinforcement learning approach[J]. IEEE Transactions on Wireless Communications, 2020, 19(8): 5404-5419.
- [134] LI H, YU J Y, CAO L L, et al. Multi-agent reinforcement learning based computation offloading and resource allocation for LEO Satellite edge computing networks[J]. Computer Communications, 2024, 222: 268-276.
- [135] 谭惠东. 低轨星座宽带网络边缘计算任务调度策略研究[D]. 成都: 电子科技大学, 2023.
- TAN H D. Research on edge computing task scheduling strategy for low-orbit constellation broadband networks[D]. Chengdu: University of Electronic Science and Technology of China, 2023.
- [136] 周碧莹. 基于强化学习的中低轨卫星网络系统资源调度技术研究[D]. 合肥: 安徽大学, 2020.
- ZHOU B Y. Research on Resource scheduling technology of low-medium orbit satellite network system based on reinforcement learning [D]. Anhui University, 2020, Anhui, China.
- [137] LIN Z Y, NI Z Y, KUANG L L, et al. Dynamic beam pattern and bandwidth allocation based on multi-agent deep reinforcement learning for beam hopping satellite systems[J]. IEEE Transactions on Vehicular Technology, 2022, 71(4): 3917-3930.
- [138] ZHU J C, SUN X B, CHEN D J. Deep reinforcement learning for congestion control and routing optimization in LEO satellite networks[C]// Proceedings of the 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA). Piscataway: IEEE Press, 2023: 739-742.
- [139] LEEL J H, SELVAM D P, MOLISCH A F, et al. Reinforcement learning empowered massive IoT access in LEO-based non-terrestrial networks[C]//Proceedings of the 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). Piscataway: IEEE Press, 2022: 1347-1350.
- [140] 韩明仁, 王玉峰. 基于强化学习的全电推进卫星变轨优化方法[J]. 系统工程与电子技术, 2022, 44(5): 1652-1661.
- HAN M R, WANG Y F. Optimization method for orbit transfer of all-electric propulsion satellite based on reinforcement learning[J]. Systems Engineering and Electronics, 2022, 44(5): 1652-1661.
- [141] XU J, FAN S M, ZHAO Z T, et al. Deep reinforcement learning-based dynamic multi-beam power allocation for GEO-LEO co-existing satellites[J]. Acta Astronautica, 2024, 223: 197-209.

- [142] 秦泽宇. 低轨卫星物联网中任务卸载策略优化方法研究[D]. 北京: 北京邮电大学, 2022.
QIN Z Y. Research on Optimization Method of mission unloading Strategy in Low-orbit Satellite Internet of Things[D]. Beijing: Beijing University of Posts and Telecommunications, 2022.
- [143] WANG Q T, CHEN S Y, ZONG J Y, et al. Resource Allocation in LEO Satellites assisted Terrestrial Network for Latency-Sensitive Task[C]// Proceedings of the 2024 International Wireless Communications and Mobile Computing(IWCMC). Piscataway: IEEE Press, 2024: 836-841.
- [144] ZHU Y, LU D W. Satellite-assisted edge computing management based on deep reinforcement learning in industrial Internet of Things[J]. Computer Networks, 2023, 237: 110050.
- [145] HUANG J F, YANG Y, YIN L F, et al. Deep reinforcement learning-based power allocation for rate-splitting multiple access in 6G LEO satellite communication system[J]. IEEE Wireless Communications Letters, 2022, 11(10): 2185-2189.
- [146] OU J W, XING L N, YAO F, et al. Deep reinforcement learning method for satellite range scheduling problem[J]. Swarm and Evolutionary Computation, 2023, 77: 101233.
- [147] DURYEA E, GANGER M, HU W. Exploring deep reinforcement learning with multi Q-learning[J]. Intelligent Control and Automation, 2016, 7(4): 129-144.
- [148] LIN X, CHEN Y N, XUE J H, et al. Large-volume LEO satellite imaging data networked transmission scheduling problem: model and algorithm[J]. Expert Systems with Applications, 2024, 249: 123649.
- [149] LI X H, CHEN B, KANG J W, et al. Coalition game-assisted joint resource optimization for digital twin-assisted edge intelligence[J]. Computer Science, 2023, 50(2): 42-49.
- [150] 吴树范, 王伟, 温济帆, 等. 低轨互联网星座发展研究综述[J]. 北京航空航天大学学报, 2024, 50(1): 1-11.
WU S F, WANG W, WEN J F, et al. Review on the development of low-orbit Internet constellation[J]. Journal of Beijing University, 2024, 50(1): 1-11.
- [151] 王雨晗. 基于卫星边缘计算的任务卸载与服务放置策略研究[D]. 北京: 北京邮电大学, 2023.
WANG Y H. Research on mission unloading and service placement Strategy based on satellite edge computing[D]. Beijing: Beijing University of Posts and Telecommunications, 2023.
- [152] WEI P, FENG W, WANG Y M, et al. Joint mobility control and MEC offloading for hybrid satellite-terrestrial-network-enabled robots[J]. IEEE Transactions on Wireless Communications, 2023, 22(11): 8483-8497.
- [153] ZHU X M, JIANG C X. Delay optimization for cooperative multi-tier computing in integrated satellite-terrestrial networks[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(2): 366-380.

[作者简介]



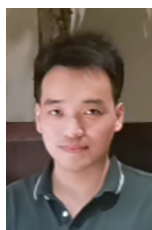
谷学强 (1983-), 男, 博士, 国防科技大学智能科学学院副研究员, 主要研究方向为智能规划与决策、边缘智能等。



张万鹏 (1981-), 男, 博士, 国防科技大学智能科学学院研究员, 博士生导师, 主要研究方向为大数据智能、智能演进等。



谭思雨 (1996-), 女, 湖南先进技术研究院算法工程师, 主要研究方向为深度学习与影像智能处理。



罗俊仁 (1989-), 男, 国防科技大学智能科学学院博士生, 主要研究方向为不完备信息博弈、多智能体学习等。



周核忠 (2000-), 男, 国防科技大学智能科学学院硕士研究生, 主要研究方向为大模型规划、智能演进。