卫星边缘网络中基于扩散模型的算力分配策略

王兴杰[1]，王侃[1]，费蓉[1]，王怀军[1]，郭银波[2]，兰大鹏[3]，朱晓杰[4]

（1.西安理工大学计算机科学与工程学院，陕西 西安 710048；2.无锡金云智联科技有限公司研发中心，江苏 无锡 214000；
3.奥斯陆大学信息科学系，挪威 奥斯陆 0373；4.阿卜杜拉国王科技大学计算机系，沙特阿拉伯 吉达 23955-6900）

摘　要：通过协同管理地面网络、卫星网络和近地无人机网络等，空天地一体化算力融合网络有望实现全域连接和普适智能，为我国数字经济发展提供有力支撑。而低轨道地球（LEO）卫星具备泛在连接和边缘计算能力，为实现空天地一体化的高效计算体系提供了基础。通过将移动边缘计算（MEC）沉降至 LEO 卫星网络，形成面向业务的“端-边-云”三级计算架构，可将时延敏感型业务从终端卸载到 LEO 卫星侧，以提升该业务的任务完成率。然而，为 LEO 卫星边缘网络制定高效计算卸载和算力分配决策是一个亟待解决的问题。针对高动态性的卫星网络环境和离散-连续的混合动作空间，提出一种基于生成扩散模型的混合近端策略优化（H-PPO）方法。首先，对具有时变特性的无线信道进行建模，并构建不同卸载决策下的服务时延、通信和计算模型。其次，在卸载决策、剩余计算资源和功率控制的多约束条件下，构建最大化平均任务完成率的长期优化问题。然后，建立具有参数化动作的马尔可夫决策过程模型，将生成扩散模型引入为离散动作策略，增强传统深度强化学习（DRL）方法的采样效率和探索能力，并利用所提算法联合优化计算卸载、算力分配和功率控制变量。仿真结果表明，所提算法具有较好收敛性能，并在任务完成率方面优于其他 3 种对比方法。

关键词：空天地一体化网络；低轨道地球卫星；移动边缘计算；深度强化学习；生成扩散模型；计算卸载；资源分配
中图分类号：TP391　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0070019

Computing Power Allocation Strategy Based on Diffusion Model in Satellite Edge Networks

WANG Xingjie[1], WANG Kan[1], FEI Rong[1], WANG Huaijun[1], GUO Yinbo[2], LAN Dapeng[3], ZHU Xiaojie[4]

(1. School of Computer Science and Engineering, Xi′an University of Technology, Xi′an 710048, Shaanxi, China;
2. Research and Development Center, KingLink Science and Technology Ltd., Wuxi 214000, Jiangsu, China;
3. Department of Information Science, University of Oslo, Oslo 0373, Norway;
4. Department of Computer Science, King Abdullah University of Science and Technology, Jeddah 23955-6900, Saudi Arabia)

【Abstract】Via coordinated management of ground networks, satellite networks, and near-earth unmanned aerial vehicle networks, computing-empowered space-air-ground integrated networks can achieve global connectivity and universal intelligence, providing strong support for the development of China′s digital economy. Low Earth Orbit (LEO) satellites have the ability of ubiquitous connectivity and edge computing, which provides the basis for an efficient computing system for space-air-ground integration. By synchronizing the Mobile Edge Computing (MEC) to LEO satellite networks to form a service-oriented end-edge-cloud three-level computing architecture, latency-sensitive tasks can be offloaded from terminals to LEO satellites, which improves the task completion rate. However, methods to make efficient offloading decisions and compute power allocation in LEO satellite edge networks must be developed urgently. Aiming at high dynamics in the satellite network environment and the discrete-continuous hybrid action space, this study proposes a Hybrid Proximal Policy Optimization (H-PPO) method based on the generative diffusion model. First, a wireless channel with time-varying characteristics is modeled and service latency, communication, and computation models under different offloading decisions are constructed. Second, under the multiple constraints of offloading decisions, remaining computing resources, and power control, a long-term optimization problem for maximizing the average task completion rate is constructed. Subsequently, the Markov decision process with parameterized actions is established and the generative diffusion model is introduced as the discrete action policy to improve the sampling efficiency and exploration ability of traditional Deep Reinforcement Learning (DRL) methods. Finally, the proposed method is used to jointly optimize the computing offloading, computing power allocation, and power control. The simulation results show that the

基金项目：国家自然科学基金（61801379，62076200）；陕西省自然科学基金（2023-YBGY-149）；国家自然科学基金国际合作与交流项目（62120106011）；陕西省自然科学基础研究计划（2022SF-353，2022QFY01-03）；西安市科技规划基金项目（2022JH-RYFW-0072）。

作者简介：王兴杰，男，硕士研究生，主研方向为空天地一体化网络、移动边缘计算；王侃，副教授、博士；费蓉（通信作者），教授、博士；王怀军，副教授、博士；郭银波，高级工程师、硕士；兰大鹏，研究员、博士；朱晓杰，助理教授、博士。

收稿日期：2024-06-19　　**修回日期：**2024-08-05　　**E-mail：**annyfei@xaut.edu.cn

计算机工程 www.ecice06.com

proposed method has a better convergence performance and is superior to the three comparison methods in terms of task completion rate.

【Key words】space-air-ground integrated networks; Low Earth Orbit (LEO) satellite; Mobile Edge Computing (MEC); Deep Reinforcement Learning (DRL); generative diffusion model; computing offloading; resource allocation

0 引言

作为国内首创的新型技术,算力网络可将数据转发、处理计算和资源存储等多项计算和网络技术相融合,以期满足未来社会对信息处理的巨大算力需求[1]。通过对闲散计算资源的统一管理和调度,算力网络技术可将分散的计算节点通过网络连接,构建统一资源池,进行计算资源的动态编排管理,并将融合计算资源提供给定制化服务或需求,从而实现资源的高效利用和按需分配。算力网络不仅支持云计算、边缘计算、数据分析和人工智能等应用,还被广泛应用于工业、交通、医疗和教育等多个行业领域[2]。随着 6G 时代到来,通过与新一代通信技术结合,算力网络有望进一步提升网络带宽、降低服务时延并提高连接可靠性,实现全域连接和普适智能,为我国数字经济的发展提供有力支撑[3]。

作为 6G 算力网络的重要远景和组网模式,空天地一体化算力融合网络在提供泛在连接的基础上,旨在进一步提供人工智能应用及服务,以期实现无线信号的泛在连接和人工智能的普惠接入[4]。其中,作为空基网络的重要接入部分,低轨道地球(LEO)卫星具备离地面较近,通常部署在轨道高度为 500～2 000 km 的空间区域,从而具有较低回程时延[5]。传统的 LEO 卫星网络和地面网络相互独立,且受到星上板载能力制约。受益于移动边缘计算(MEC)技术,LEO 卫星网络可在低轨卫星上部署 MEC 服务器[6],并与地面云计算服务器协同,形成面向差异化应用类型的“云-边-端”多级计算架构[7-8]。因此,在 LEO 卫星边缘网络中,LEO 卫星可充当边缘节点,为资源受限的地面终端设备提供计算服务;此外,也可将计算任务卸载到具有丰富计算资源的云服务器以进行处理[9-10]。在 LEO 卫星边缘网络环境中,为应对高动态网络环境、多任务资源竞争和低时延高可靠性要求等挑战,高效计算卸载和资源管理策略成为关键因素。因此,本文拟引入生成扩散模型,将其视作混合近端策略优化(H-PPO)框架的离散动作策略网络,提出一种基于生成扩散模型的 H-PPO 学习框架。本文主要研究内容如下:

1)面向时延敏感型业务,建立 LEO 卫星边缘网络中长期最大化任务完成率的优化问题模型。对具备高动态时变特性的无线信道进行建模,并讨论不同卸载策略下的业务时延模型;在满足卸载决策约束、计算资源约束和发射功率约束的条件下,联合优化计算卸载、算力资源分配和功率控制决策,以最大化接入任务完成率。

2)设计基于生成扩散模型的 H-PPO 框架。构造具备参数化动作空间性质的马尔科夫决策过程(PAMDP)模型,捕捉卫星边缘网络中卸载决策、资源分配和功率控制等的动态特性;引入生成扩散模型作为离散动作的策略,以增强深度强化学习(DRL)方法的采样效率和探索能力。

3)本文算法具有良好的收敛性能。并在任务完成率方面优于 3 种对比方法,验证了本文算法的有效性。

1 相关研究

目前已有大量工作研究了 LEO 卫星边缘网络中计算卸载和资源分配问题。例如,文献[11]设计了一个多用户混合 LEO 和物联网(IoT)边缘网络场景,其考虑了用户间的任务依赖相关性,联合优化任务的卸载决策和发射功率,并将其建模为一个混合整数规划问题,以最小化能耗和时延的加权和。文献[12]研究了一个多用户多任务的边缘网络场景,考虑了服务缓存、计算卸载和资源分配的联合优化模型,同样将其建模为混合整数规划问题。

为解决卫星边缘网络的时变高动态性等挑战,DRL 技术近年来已在资源分配和调度中被广泛应用,其能够在无需模型先验信息的条件下做出智能决策。文献[13]在星地一体化网络架构中对缓存和计算资源进行统一管理,并利用深度 Q 学习方法对资源进行联合调度。文献[14]在空地一体化边缘网络中,利用 DRL 在线学习最优卸载策略,并使用启发式算法解决资源分配和任务调度的混合整数规划问题。在基于拉格朗日乘子法的基础上,文献[15]首先确定资源分配决策,继而使用 DRL 进行计算卸载。文献[16]同样将计算卸载和资源分配解耦为两个子问题求解;在确定计算卸载决策的基础上,使用拉格朗日乘子法进行网络和计算资源分配。

使用 DRL 方法在求解卫星边缘网络计算卸载

计算机工程 www.ecice06.com

和资源分配联合优化问题的相关研究工作中，大多是将计算卸载和资源分配解耦为两个子问题，分别使用 DRL 和启发式算法进行独立求解。这是因为：已有 DRL 方法通常只针对单一离散或连续动作空间，难以同时对这两种动作变量进行联合策略。然而，卫星边缘网络中的离散卸载决策和连续资源分配组成了连续-离散的混合动作空间；若分别优化，则导致难以找到最优混合策略变量，降低网络接入性能。因此，本文将使用一种可处理混合动作空间的 DRL 方法，即混合近端策略优化（H-PPO）来联合优化计算卸载和资源分配决策。此外，受生成扩散模型强大生成能力的启发，已有工作研究了生成扩散模型与 DRL 之间的有效协作[17]。利用生成扩散模型来增强 DRL 方法的策略表达能力和多模态动作分布的捕捉能力等。因此，本文引入生成扩散模型，将其视作 H-PPO 框架的离散动作策略网络，提出一种基于生成扩散模型的 H-PPO 学习框架。

2 系统模型和问题描述

2.1 卫星边缘网络模型

卫星边缘网络模型由多个地面终端、多颗已部署边缘服务器的 LEO 卫星和云服务器（部署在地面）组成。定义 $\mathcal{L}=\{1,2,\cdots,L\}$ 表示卫星集合，M 表示云服务器；LEO 卫星使用 Ka 频段（27～40 GHz）为地面终端提供服务，且每颗卫星只服务其覆盖范围内的地面终端。此外，定义集合 $\mathcal{K}=\{1,2,\cdots,K\}$，将卫星覆盖范围划分为 K 个区域，且 $\mathcal{I}_k=\{1,2,\cdots,I_k\}$ 表示区域 k 中的终端集合。同时，设每个终端同时只能与一颗卫星通信，而每颗卫星可同时服务多个终端；在同一卫星覆盖范围内的终端距离相近，故这些终端到接入卫星的信道状态几乎相同。云服务器与地面终端设备之间由于距离和地理条件等因素制约，无法建立直连连接，因此地面终端设备需要通过 LEO 卫星转发才可与云服务器建立连接。

为描述由于任务随机到达而动态变化的网络场景，可将系统时隙化，即 $\mathcal{T}=\{1,2,\cdots,t,\cdots,T\}$。定义 $R_{k,i}^{\bar{t}}$ 表示区域 k 中的终端 i 在时隙 $\bar{t}$ 的新到达任务，且 l_k 表示覆盖该区域 k 的卫星，并有 $\bar{t}\in\mathcal{T}$ 和 $l_k\in\mathcal{L}$。此外，假设所有到达任务均是时延敏感型，定义 $t_{\max}$ 表示任务的最大可容忍时延，$c_{k,i}^{\bar{t}}$ 表示任务 $R_{k,i}^{\bar{t}}$ 须处理的数据包大小。然后，定义 $\mathcal{Z}_{k,i}^{\bar{t},t}=\{z_{k,i}^{\bar{t},t,1},z_{k,i}^{\bar{t},t,2},z_{k,i}^{\bar{t},t,3}\}$ 来表示任务 $R_{k,i}^{\bar{t}}$ 在时隙 t 的卸载决策（设 $R_{k,i}^{\bar{t}}$ 从时隙 $\bar{t}$ 开始且到时隙 t 尚未处理完成），且满足 $z_{k,i}^{\bar{t},t,1},z_{k,i}^{\bar{t},t,2},z_{k,i}^{\bar{t},t,3}\in\{0,1\}$ 和 $z_{k,i}^{\bar{t},t,1}+z_{k,i}^{\bar{t},t,2}+z_{k,i}^{\bar{t},t,3}=1$，即每个任务只能选择一种处理方式。具体地，$z_{k,i}^{\bar{t},t,1}=1$ 表示 $R_{k,i}^{\bar{t}}$ 在终端本地处理；$z_{k,i}^{\bar{t},t,2}=1$ 表示 $R_{k,i}^{\bar{t}}$ 卸载至卫星 l_k 处理；$z_{k,i}^{\bar{t},t,3}=1$ 表示 $R_{k,i}^{\bar{t}}$ 卸载至云服务器处理。

2.2 信道模型

鉴于 LEO 卫星的高速移动，卫星和地面终端以及云服务器之间的相对位置变化迅速。星地通信链路所涉及的自由空间路径损耗和大气衰落等诸多参数也随相对位置的变化而改变[18]。故本文考虑卫星与地面终端以及云服务器之间的动态时变信道。

在图 1 所示的通信场景中，将任务由地面终端传输至卫星的通信链路称为上行链路，任务由卫星传输至云服务器的通信链路称为下行链路。当处理结果返回至终端时，由于返回结果数据量极小，可忽略其传输时延。故当 $R_{k,i}^{\bar{t}}$ 卸载至卫星 l_k 处进行处理时，数据传输只经过上行链路；当 $R_{k,i}^{\bar{t}}$ 卸载至云服务器时，数据传输须同时历经上行和下行链路。

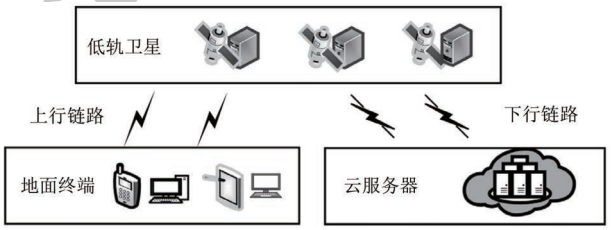


图 1 卫星边缘网络场景

Fig. 1 Satellite edge network scene

随后，将时隙 $\bar{t}$ 时区域 k 中所有终端的上行链路信道状态 $g_k^{\bar{t},L}$ 表示为：

$$g_k^{\bar{t},L}=G_s^E\cdot G_k^{\bar{t},L}\cdot G_r^L \tag{1}$$

式中：G_s^E 表示终端的发射天线增益；$G_k^{\bar{t},L}$ 表示时隙 $\bar{t}$ 时区域 k 中终端与卫星 l_k 之间的信道衰落；G_r^L 表示卫星的接收天线增益。此外，上行链路信道衰落 $G_k^{\bar{t},L}$ 一般包括自由空间路径损耗、大气衰落和服从莱斯分布的小尺度衰落[19]，即：

$$G_k^{\bar{t},L}=\left(\frac{c}{4\pi d_k^{\bar{t},L}f}\right)^2\cdot\phi(\Omega^{\bar{t},L})\cdot\varphi \tag{2}$$

式中：c 表示光速；$d_k^{\bar{t},L}$ 表示在时隙 $\bar{t}$ 时区域 k 内终端与卫星 l_k 之间的距离；f 表示载波频率；$\phi(\Omega^{\bar{t},L})$ 表示大气衰落；φ 表示服从莱斯分布的小尺度衰落。进一步，大气衰落 $\varphi(\Omega^{\bar{t},L})$ 可表示为：

$$\phi(\Omega^{\bar{t},L})=10^{\frac{3\chi}{10\sin\Omega^{\bar{t},L}}} \tag{3}$$

式中：$\sin\Omega^{\bar{t},L}=H/d_k^{\bar{t},L}$，$H$ 表示 LEO 卫星的轨道高度；χ 是电磁波经过雨和云等的衰减。

类似地，可将下行链路信道状态 $g_k^{\bar{t},M}$ 表示为：

$$g_k^{\bar{t},M}=G_s^L\cdot G_k^{\bar{t},M}\cdot G_r^M \tag{4}$$

式中：G_s^L 表示卫星的发射天线增益；$G_k^{\bar{t},M}$ 表示在时隙 $\bar{t}$ 时卫星 l_k 与云服务器之间的信道衰落；G_r^M 表示云服务器的接收天线增益。与上行链路相同，下行链路的信道衰落 $G_k^{\bar{t},M}$ 同样包括自由空间路径损耗、大气衰落和服从莱斯分布的小尺度衰落，即：

$$G_k^{\bar{t},M}=\left(\frac{c}{4\pi d_k^{\bar{t},M}f}\right)^2\cdot\phi(\Omega^{\bar{t},M})\cdot\varphi \tag{5}$$

式中：$d_k^{\bar{t},M}$ 表示卫星 l_k 与云服务器之间的距离。大气衰落 $\phi(\Omega^{\bar{t},M})$ 可表示为：

$$\phi(\Omega^{\bar{t},M})=10^{\frac{3\chi}{10\sin\Omega^{\bar{t},M}}} \tag{6}$$

2.3 时延模型

当 $R_{k,i}^{\bar{t}}$ 在终端处理时，仅考虑计算时延，即：

$$T_{k,i}^{\bar{t},\mathrm{tol}}=c_{k,i}^{\bar{t}}/C_{k,i}^{\bar{t},E} \tag{7}$$

式中：$c_{k,i}^{\bar{t}}$ 表示任务 $R_{k,i}^{\bar{t}}$ 的数据包大小；$C_{k,i}^{\bar{t},E}$ 表示终端分配给 $R_{k,i}^{\bar{t}}$ 的计算资源。

当 $R_{k,i}^{\bar{t}}$ 卸载至卫星 l_k 处理时，总响应时延由计算、传输和传播时延组成。计算时延可表示为 $c_{k,i}^{\bar{t}}/C_{k,i}^{\bar{t},L}$，其中，$C_{k,i}^{\bar{t},L}$ 为卫星 l_k 分配给 $R_{k,i}^{\bar{t}}$ 的计算资源，传播时延可表示为 $2d_k^{\bar{t},L}/c$。$R_{k,i}^{\bar{t}}$ 的上行传输速率为：

$$s_{k,i}^{\bar{t},t,L}=W_k^L\mathrm{lb}\left(1+\frac{P_{k,i}^{\bar{t},E}g_k^{t,L}}{\sum\limits_{m=1,m\neq k}^{K}\sum\limits_{j=1}^{I}P_{m,j}^{\bar{t},E}g_m^{t,L}+\sigma^2}\right) \tag{8}$$

式中：W_k^L 表示区域 k 内终端与卫星 l_k 之间的链路带宽；$s_{k,i}^{\bar{t},t,L}$ 表示 $R_{k,i}^{\bar{t}}$ 在时隙 t（$R_{k,i}^{\bar{t}}$ 从时隙 $\bar{t}$ 开始，到时隙 t 尚未传输完成）的传输速率；$P_{k,i}^{\bar{t},E}$ 为终端分配给 $R_{k,i}^{\bar{t}}$ 的发射功率；$\sum\limits_{m=1,m\neq k}^{K}\sum\limits_{j=1}^{I}P_{m,j}^{\bar{t},E}g_m^{t,L}$ 表示在当前时隙 t 内其它区域对区域 k 的干扰功率；σ^2 为噪声功率。信道状态逐时隙变化，故任务的传输速率也逐时隙波动。定义 $R_{k,i}^{\bar{t}}$ 在时隙 t 的剩余数据为：

$$y_{k,i}^{\bar{t},t}=\begin{cases}c_{k,i}^{\bar{t}}, & t=\bar{t}\\ y_{k,i}^{\bar{t},t-1}-s_{k,i}^{\bar{t},t-1,L}\cdot\tau, & t>\bar{t}\end{cases} \tag{9}$$

式中：τ 表示时隙所占时间。在每个时隙 t 开始前，须对 $R_{k,i}^{\bar{t}}$ 的状态进行判断，定义 $\eta_{k,i}^{\bar{t},t}$ 来表示 $R_{k,i}^{\bar{t}}$ 在时隙 t 的状态，即：

$$\eta_{k,i}^{\bar{t},t}=\begin{cases}1, & T_{k,i}^{\bar{t},\mathrm{tol}}\leqslant t_{\max}\\ -1, & T_{k,i}^{\bar{t},\mathrm{tol}}>t_{\max},\ T_{k,i}^{\bar{t},\mathrm{genr}}\geqslant t_{\max}\\ 0, & T_{k,i}^{\bar{t},\mathrm{genr}}<t_{\max}\end{cases} \tag{10}$$

式中：$T_{k,i}^{\bar{t},\mathrm{genr}}$ 表示 $R_{k,i}^{\bar{t}}$ 从时隙 $\bar{t}$ 到 t 的已产生时延。$\eta_{k,i}^{\bar{t},t}=1$ 表示 $R_{k,i}^{\bar{t}}$ 处理完成，$\eta_{k,i}^{\bar{t},t}=0$ 表示 $R_{k,i}^{\bar{t}}$ 顺延至下一时隙进行判断，而 $\eta_{k,i}^{\bar{t},t}=-1$ 表示 $R_{k,i}^{\bar{t}}$ 处理失败。

当 $R_{k,i}^{\bar{t}}$ 卸载至云服务器处理时，总响应时延同样包括计算时延、传输时延和传播时延。计算时延可表示为 $c_{k,i}^{\bar{t}}/C_{k,i}^{\bar{t},M}$，其中，$C_{k,i}^{\bar{t},M}$ 是云服务器分配给 $R_{k,i}^{\bar{t}}$ 的计算资源，传播时延为 $2d_k^{\bar{t},L}/c+2d_k^{\bar{t},M}/c$。$R_{k,i}^{\bar{t}}$ 在下行链路中的传输速率可表示为：

$$s_{k,i}^{\bar{t},t,M}=W_k^M\mathrm{lb}\left(1+\frac{P_{k,i}^{\bar{t},L}g_k^{t,M}}{\sum\limits_{m=1,m\neq k}^{K}\sum\limits_{j=1}^{I}P_{m,j}^{\bar{t},L}g_m^{t,M}+\sigma^2}\right) \tag{11}$$

式中：W_k^M 表示卫星 l_k 与云服务器之间的链路带宽；$P_{k,i}^{\bar{t},L}$ 表示卫星 l_k 分配给 $R_{k,i}^{\bar{t}}$ 的功率。假设卫星工作在全双工通信模式，则 $R_{k,i}^{\bar{t}}$ 的传输速率可表示为 $s_{k,i}^{\bar{t},t,M'}=\min\{s_{k,i}^{\bar{t},t,L},s_{k,i}^{\bar{t},t,M}\}$。与上行链路相同，该速率同样逐时隙波动。此外，当 $R_{k,i}^{\bar{t}}$ 卸载至云服务器处理时，对 $R_{k,i}^{\bar{t}}$ 在时隙 t 的剩余数据的计算要考虑全双工速率，即：

$$y_{k,i}^{\bar{t},t}=\begin{cases}c_{k,i}^{\bar{t}}, & t=\bar{t}\\ y_{k,i}^{\bar{t},t-1}-s_{k,i}^{\bar{t},t-1,M'}\cdot\tau, & t>\bar{t}\end{cases} \tag{12}$$

式中：$s_{k,i}^{\bar{t},t-1,M'}$ 表示全双工速率。

2.4 问题描述

鉴于多任务随机到达特性，多个卸载至同一节点（同一卫星或云服务器）的任务将竞争可用资源，须逐时隙联合优化计算卸载和资源分配，以最大化任务完成数量，故将长期优化目标建模为最大化任务完成数的时间平均值，即：

$$\begin{aligned}&\max\ \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i=1}^{I_k}\sum_{\bar{t}=1}^{t}\eta_{k,i}^{\bar{t},t}\\ &\text{s.t. }C_1:\ \sum_{i=1}^{I_k}C_{k,i}^{\bar{t},L}\leqslant\omega_k^{\bar{t},L},\ \forall k,\bar{t}\\ &\qquad C_2:\ 0\leqslant P_{k,i}^{\bar{t},E}\leqslant z_{k,i}^{\bar{t},t,2}\cdot P_{k,i,\max}^{E},\ \forall k,i,\bar{t},t\\ &\qquad C_3:\ \sum_{k=1}^{K}\sum_{i=1}^{I_k}C_{k,i}^{\bar{t},M}\leqslant\omega^{\bar{t},M},\ \forall\bar{t}\\ &\qquad C_4:\ 0\leqslant P_{k,i}^{\bar{t},L}\leqslant z_{k,i}^{\bar{t},t,3}\cdot P_{k,\max}^{L},\ \forall k,i,\bar{t},t\\ &\qquad C_5:\ 0\leqslant C_{k,i}^{\bar{t},L}\leqslant z_{k,i}^{\bar{t},t,2}\cdot\omega_k^{\bar{t},L},\ \forall k,i,\bar{t},t\\ &\qquad C_6:\ 0\leqslant C_{k,i}^{\bar{t},M}\leqslant z_{k,i}^{\bar{t},t,3}\cdot\omega^{\bar{t},M},\forall k,i,\bar{t},t\\ &\qquad C_7:\ \sum_{i=1}^{I_k}z_{k,i}^{\bar{t},t,2}P_{k,i}^{\bar{t},L}\leqslant P_{\max}^{L},\ \forall k,\bar{t},t\\ &\qquad C_8:\ T_{k,i}^{\bar{t},\mathrm{tol}}\leqslant t_{\max},\ \forall k,i,\bar{t}\end{aligned} \tag{13}$$

式中：$\omega_k^{\bar{t},L}$ 和 $\omega^{\bar{t},M}$ 分别表示卫星 l_k 和云服务器 M

计算机工程 www.ecice06.com

在时隙 $\bar{t}$ 的剩余计算资源；$P^{E}_{k,i,\max}$ 和 $P^{L}_{k,\max}$ 分别表示终端和卫星最大传输功率；$P^{L}_{\max}$ 表示卫星的最大转发总功率；C_1 表示当前时隙分配给任务的总计算资源不超过卫星剩余计算资源；C_2 表示分配给任务的发射功率不得超过终端最大发射功率；C_3 表示当前时隙分配给任务的总计算资源不得超过云服务器剩余计算资源；C_4 表示分配给任务的发射功率不超过卫星最大发射功率；C_5 和 C_6 表示当 $z^{\bar{t},t,2}_{k,i}=1$（$z^{\bar{t},t,3}_{k,i}=1$）时，$C^{\bar{t},L}_{k,i}$（$C^{\bar{t},M}_{k,i}$）取正值，否则 $C^{\bar{t},L}_{k,i}=0$（$C^{\bar{t},M}_{k,i}=0$）；C_7 表示卫星上任务的转发功率之和不得超过卫星总转发功率；C_8 表示处理每一个任务 $R^{\bar{t}}_{k,i}$ 所产生的时延不能超过最大容忍时延。

3 基于扩散模型的 H-PPO 方法

现有 DRL 框架大多只考虑单一离散或连续动作空间。为处理离散-连续的混合动作空间，常规方法是将混合动作空间近似化为离散动作空间，或是将离散变量松弛成连续变量[20]。为将混合动作的优化统一在同一个 DRL 框架内，本文使用一种针对离散-连续混合动作空间的框架，即 H-PPO，可直接作用于混合动作空间，而无需近似或松弛。此外，受到生成扩散模型强大生成能力的启发，本文引入生成扩散模型作为 DRL 中的策略网络，以增强该方法捕获环境中复杂关系的能力，并提升其探索能力[21]。

3.1 PAMDP 模型

PAMDP 是标准 MDP 模型的扩展[22]，同样由一个四元组 $(\mathcal{S},\mathcal{A},\mathcal{P},r)$ 表示，其中，$\mathcal{S}$ 表示状态集合，$\mathcal{A}$ 表示动作集合，$\mathcal{P}$ 表示状态转移概率，r 表示奖励。

首先，对每一个状态 $\boldsymbol{s}_t\in\mathcal{S}$，定义 $\boldsymbol{s}_t$ 为：

$$\boldsymbol{s}_t=\{\mathcal{Q}^t,\mathcal{O}^t,\{\omega^{t,L}_k\}_{k\in\mathcal{K}},\omega^{t,M}\} \tag{14}$$

式中：$\mathcal{Q}^t$ 表示时隙 t 新到达任务集合；$\mathcal{O}^t$ 表示时隙 t 正在处理任务集合。

其次，考虑一个参数化的动作空间 $\mathcal{A}$，在整个系统中任务卸载方式共有 3^I 种组合，I 表示终端数量。因此，存在一个有限离散动作集合 $\mathcal{Z}=\{1,2,\cdots,z,\cdots,Z\}$，且 $Z=3^I$。对每一个离散动作 $z\in\mathcal{Z}$，均存在一个连续化的动作参数集合；记该集合为 $\mathcal{W}_z$，并令 $\boldsymbol{w}^t_z=\{C^{t,L}_{k,i},C^{t,M}_{k,i},P^{t,E}_{k,i},P^{t,L}_{k,i},\forall k,i\}\in\mathcal{W}_z$ 表示离散动作 z^t 所关联的连续参数。因此，将时隙 t 的动作定义为：

$$\boldsymbol{a}_t=(z^t,\boldsymbol{w}^t_z)\in\mathcal{A},\forall t \tag{15}$$

$$\mathcal{A}=\bigcup\nolimits_{z\in\mathcal{Z}}\{(z,\boldsymbol{w}_z):\boldsymbol{w}_z\in\mathcal{W}_z\} \tag{16}$$

最后，在每个时隙对集合 $\mathcal{O}^t$ 中的任务进行遍历，并将任务 $R^{\bar{t}}_{k,i}$ 的即时奖励定义为 $r^{\bar{t},t}_{k,i}$。具体地，当 $R^{\bar{t}}_{k,i}$ 在时隙 t 能够处理完成时，将 $r^{\bar{t},t}_{k,i}$ 设为正数；当 $R^{\bar{t}}_{k,i}$ 顺延至下一时隙继续传输时，将 $r^{\bar{t},t}_{k,i}$ 暂设为 0；当 $R^{\bar{t}}_{k,i}$ 处理失败时，将 $r^{\bar{t},t}_{k,i}$ 设为负数。

3.2 生成扩散模型

生成扩散模型的前向过程可建模为一个 N 步的马尔科夫链[23]，前向过程通过对原始数据 $\boldsymbol{x}_0$ 分 N 步来叠加高斯噪声，产生一系列数据 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N$，从 $\boldsymbol{x}_{n-1}$ 到 $\boldsymbol{x}_n$ 的转移概率可建模为：

$$q(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})=\mathcal{N}(\sqrt{1-\beta_n}\boldsymbol{x}_{n-1},\beta_n\boldsymbol{I}_Z),\ \forall n \tag{17}$$

式中：β_n 是高斯分布的方差。此外，定义 $\alpha_n=1-\beta_n$ 和 $\bar{\alpha}_n=\prod_{n'=1}^{n}\alpha_n$，此时 $\boldsymbol{x}_n$ 可进行递归表示，即：

$$\begin{aligned}\boldsymbol{x}_n&=\sqrt{\alpha_n}\boldsymbol{x}_{n-1}+\sqrt{1-\alpha_n}\varepsilon_1=\\&\sqrt{\alpha_n}(\sqrt{\alpha_{n-1}}\boldsymbol{x}_{n-2}+\sqrt{1-\alpha_{n-1}}\varepsilon_2)+\sqrt{1-\alpha_n}\varepsilon_1=\\&\sqrt{\alpha_n\alpha_{n-1}}\boldsymbol{x}_{n-2}+\sqrt{1-\alpha_n\alpha_{n-1}}\varepsilon_2=\cdots=\\&\sqrt{\bar{\alpha}_n}\boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_n}\varepsilon_n,\forall n\end{aligned} \tag{18}$$

式中：$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{n-1}\sim\mathcal{N}(0,\boldsymbol{I}_Z)$ 且 $\varepsilon_2,\varepsilon_3,\cdots,\varepsilon_n\sim\mathcal{N}(0,\boldsymbol{I}_Z)$；随着 n 增加，$\boldsymbol{x}_n$ 逐渐变为纯噪声 $\boldsymbol{x}_N$。同时，在生成扩散模型的反向过程中，通过对纯噪声 $\boldsymbol{x}_N\sim\mathcal{N}(0,\boldsymbol{I}_Z)$ 去噪并推断出最优概率分布 $\boldsymbol{x}_0$。定义 $p(\boldsymbol{x}_{n-1}|\boldsymbol{x}_n)$ 为从 $\boldsymbol{x}_n$ 到 $\boldsymbol{x}_{n-1}$ 的转移概率。虽然分布 $p(\boldsymbol{x}_{n-1}|\boldsymbol{x}_n)$ 是无法直接计算得到的，但其加上 $\boldsymbol{x}_0$ 的后验分布 $p(\boldsymbol{x}_{n-1}|\boldsymbol{x}_n,\boldsymbol{x}_0)$ 可通过计算得到，即：

$$p(\boldsymbol{x}_{n-1}|\boldsymbol{x}_n)=\mathcal{N}(\mu(\boldsymbol{x}_n,n,\boldsymbol{x}_0),\bar{\beta}_n\boldsymbol{I}_Z),\ \forall n \tag{19}$$

式中：可由贝叶斯定理得到 $\bar{\beta}_n=\frac{1-\bar{\alpha}_{n-1}}{1-\bar{\alpha}_n}\beta_n$。当 $\bar{\beta}_n$ 确定后可得在第 n 步的均值为：

$$\mu(\boldsymbol{x}_n,n,\boldsymbol{x}_0)=\frac{\sqrt{\alpha_n}(1-\bar{\alpha}_{n-1})}{1-\bar{\alpha}_n}\boldsymbol{x}_n+\frac{\sqrt{\bar{\alpha}_{n-1}}\beta_n}{1-\bar{\alpha}_n}\boldsymbol{x}_0,\ \forall n \tag{20}$$

根据式(17)和式(19)，均值可进一步改写为：

$$\mu(\boldsymbol{x}_n,n,\boldsymbol{x}_0)=\frac{1}{\sqrt{\alpha_n}}\left(\boldsymbol{x}_n-\frac{1-\alpha_n}{\sqrt{1-\bar{\alpha}_n}}\bar{\varepsilon}_n\right),\ \forall n \tag{21}$$

式中：$\boldsymbol{x}_n$ 确定而 $\bar{\varepsilon}_n$ 未知，须通过神经网络 θ 来估计。令 $\bar{\varepsilon}_\theta(\boldsymbol{x}_n,n,\boldsymbol{x}_0)\approx\bar{\varepsilon}_n$ 且 $\mu_\theta(\boldsymbol{x}_n,n,\boldsymbol{x}_0)\approx\mu(\boldsymbol{x}_n,n,\boldsymbol{x}_0)$，可参数化式(18)为：

$$p_\theta(\boldsymbol{x}_{n-1}|\boldsymbol{x}_n)=\mathcal{N}(\mu_\theta(\boldsymbol{x}_n,n,\boldsymbol{x}_0),\bar{\beta}_n\boldsymbol{I}_Z) \tag{22}$$

使用重参数化技巧得到 $\boldsymbol{x}_{n-1}$ 为：

$$\boldsymbol{x}_{n-1}=\mu_\theta(\boldsymbol{x}_n,n,\boldsymbol{x}_0)+\sqrt{\bar{\beta}_n}\varepsilon,\ \forall n \tag{23}$$

式中：$\varepsilon\sim\mathcal{N}(0,\boldsymbol{I}_Z)$。由 $\boldsymbol{x}_N$ 得到 $\boldsymbol{x}_0$ 的分布，即通过

计算机工程 www.ccie06.com

累乘得到：

$$p_\theta(\boldsymbol{x}_0)=p(\boldsymbol{x}_N)\prod_{n=1}^{N}p_\theta(\boldsymbol{x}_{n-1}\mid\boldsymbol{x}_n) \tag{24}$$

最终，由纯噪声 $\boldsymbol{x}_N$ 生成 $\boldsymbol{x}_0$。进一步将 $\boldsymbol{x}_0$ 输入到 Softmax 函数中，得到离散动作的概率分布，即：

$$\pi_\theta=\left\{\frac{\exp(\boldsymbol{x}_0^z)}{\sum_{k'\in\mathcal{K}}\exp(\boldsymbol{x}_0^{z'})},\ \forall z\right\} \tag{25}$$

3.3 基于生成扩散模型的 H-PPO 方法

与 PPO 相似，H-PPO 同样采用 Actor-Critic 架构并使用状态价值函数作为 Critic[24]。然而，与传统的 Actor-Critic 架构不同，H-PPO 中有两个 Actor 网络，分别作用于离散和连续动作空间，且完整动作由离散动作 $z\in\mathcal{Z}$ 和其连续参数动作 $\boldsymbol{w}_z\in\mathcal{W}$ 组合得到。图 2 所示为基于生成扩散模型的 H-PPO 框架的具体流程。

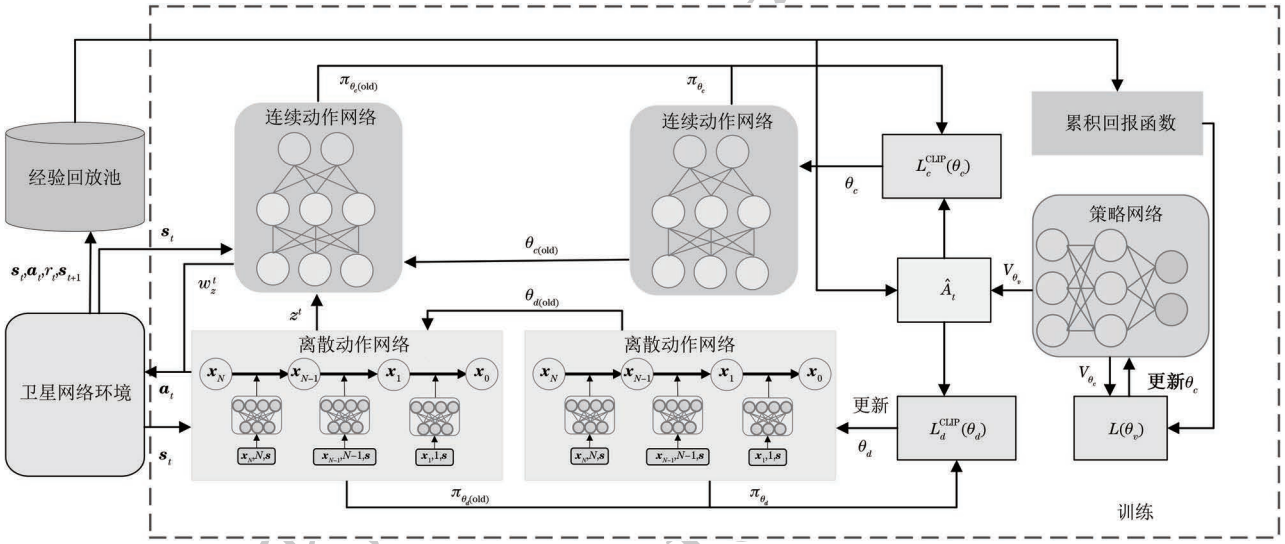


图 2 基于生成扩散模型的 H-PPO 框架

Fig. 2 H-PPO framework based on generative diffusion model

在混合 Actor-Critic 架构中，采用状态价值函数 $V(\boldsymbol{s}_t)$ 作为 Critic，并使用在长度为 $\overline{T}$ 的轨迹上收集到的样本来计算方差缩减的优势函数 $\hat{A}_t$，即：

$$\hat{A}_t=-V(\boldsymbol{s}_t)+r_t+\cdots+\gamma^{\overline{T}-t}V(\boldsymbol{s}_{\overline{T}}),\ \forall t \tag{26}$$

式中：γ 表示折扣因子。

随后，令 $\theta_d:=\theta$ 和 θ_c 分别表示离散动作网络（即式(22)中的去噪网络）和连续动作网络的参数，令 θ_v 表示 Critic 网络的参数。在轨迹收集阶段，H-PPO 与环境交互并收集轨迹 $\mathcal{D}_z=\{\iota\}$，并将 $\mathcal{D}_z$ 保存到经验回放池 $\mathcal{B}$ 中；在训练阶段，$\mathcal{B}$ 中的轨迹被用来训练网络参数 θ_d、θ_c 和 θ_v。

特别地，两个 Actor 网络均通过最大化裁剪替代目标函数来更新 Actor 网络，即：

$$L^{\mathrm{CLIP}}(\theta)=\frac{\mathcal{D}_z}{|\mathcal{D}_z|T}\sum_{\iota\in\mathcal{D}_z}\sum_{t=1}^{T}\min(e_t(\theta)\hat{A}_t,$$

$$\mathrm{clip}(e_t(\theta),1-\xi,1+\xi)\hat{A}_t) \tag{27}$$

式中：$\theta=\theta_d$ 或 $\theta=\theta_c$；ξ 表示裁剪因子；比率 $e_t(\theta_d)$ 和 $e_t(\theta_c)$ 表示当前策略和旧策略的比值。针对离散策略和连续策略，其比率分别为：

$$e_t(\theta_d)=\frac{\pi_{\theta_d}(k_t\mid\boldsymbol{s}_t)}{\pi_{\theta_{d(\mathrm{old})}}(k_t\mid\boldsymbol{s}_t)},\ \forall t \tag{28}$$

$$e_t(\theta_c)=\frac{\pi_{\theta_c}(k_t\mid\boldsymbol{s}_t,k_t)}{\pi_{\theta_{c(\mathrm{old})}}(k_t\mid\boldsymbol{s}_t,k_t)},\ \forall t \tag{29}$$

最后，将价值函数近似为 $V_{\theta_v}(\boldsymbol{s}_t)\approx V(\boldsymbol{s}_t)$，并将其损失函数定义为：

$$L(\theta_v)=\frac{1}{|\mathcal{D}_z|T}\sum_{\iota\in D_z}\sum_{t=1}^{T}[V_{\theta_v}(\boldsymbol{s}_t)-\hat{G}_t]^2 \tag{30}$$

式中：$\hat{G}_t$ 是在时间步 t 处观察到的累积回报函数。此外，本文所提方法的详细过程如算法 1 所示。

算法 1 基于生成扩散模型的 HPPO 方法

1. 初始化网络参数 θ_d、θ_c 和 θ_v，初始化经验回放池 $\mathcal{B}$
2. for each episode do
3. 重置环境和经验回放池
4. for t=1,2,…,T do
5. 观察当前时隙 s_t 状态并初始化 $x_N\sim\mathcal{N}(0,I_Z)$
6. for n=N,…,1 do
7. 根据式(20)得到均值 $\mu(x_n,n,x_0)$
8. 根据均值和式(21)得到 x_{n-1}
9. end for
10. 根据式(24)得到离散策略 π_{θ_d}，使用 π_{θ_d} 随机采样得到离散动作 z^t，随后使用 $\pi_{\theta_c}(\cdot\mid s_t,z^t)$ 得到连续动作 w_z^t
11. 组合 z^t 和 w_z^t 得到动作 $a_t=(z^t,w_z^t)$，执行动作 a_t，获得奖励 r_t 和下一时刻状态 s_{t+1}，将四元组 (s_t,a_t,s_{t+1},r_t) 保存到经验回放池 $\mathcal{B}$ 中

12. end for
13. 更新 $\theta_{d(old)} \leftarrow \theta_d$ 和 $\theta_{c(old)} \leftarrow \theta_c$
14. 从经验回放池 $\mathcal{B}$ 中获取部分轨迹
15. 使用式(25)计算优势函数 $\hat{A}_t$
16. 通过随机梯度下降以最大化裁剪替代目标函数(更新 θ_d 和 θ_c)和最小化损失函数(更新 θ_v)
17. end for

4 仿真实验与结果分析

4.1 仿真参数设置

本文仿真环境包括 Python 3.9.13、PyTorch 1.13.0 和 Satellite Tool Kit(STK)11.6。具体的参数设置如表 1 所示。

表 1 仿真参数设置

Table 1 Simulation parameters setting

参数	参数值
任务大小/bit	$[8\times10^4, 12\times10^4]$
低轨卫星数量/个	3
卫星轨道高度/km	900
卫星最大传输功率/W	100
地面终端最大传输功率/W	20
Ka 波段的载波频率/GHz	30
链路带宽/MHz	25
区域 1 中的终端数/个	14
区域 2 中的终端数/个	12
区域 3 中的终端数/个	8
噪声功率谱密度/(dBm·Hz^{-1})	−174
低轨卫星计算资源/(cycle·s^{-1})	3×10^{10}
云服务器计算资源/(cycle·s^{-1})	1×10^{11}
折扣因子	0.9
裁剪因子	0.2
经验回放池大小	200
Critic 网络的学习率	0.000 1
连续 Actor 网络的学习率	0.000 1

特别地,为获取 LEO 卫星在 200 个时隙中的位置信息,使用 STK 模拟了 3 颗轨道高度为 900 km 的 LEO 卫星,并获得了位置信息报告。此外,CPU 的计算能力被设置为 1 000 cycle/bit,则 LEO 卫星和云服务器上的计算资源(处理能力)可等效地表示为 3×10^7 bit/s 和 1×10^8 bit/s。这里的 bit/s 不是指带宽,而是为了提供一种等效的处理能力表示方式。

4.2 结果分析

为验证本文提出的基于生成扩散模型的 H-PPO 方法的有效性,须首先验证本文方法的收敛性能。其次,将本文方法与以下 3 种方法进行对比:

1)未与生成扩散模型结合的 H-PPO 方法(即原始 H-PPO 方法)。

2)文献[24]中基于平均资源分配方法和深度 Q 网络(DQN)的计算卸载方法相结合的方法(即 DQN 卸载-平均资源分配方法)。

3)文献[25]中随机卸载方法和基于深度确定性策略梯度(DDPG)资源分配方法相结合的方法(即随机卸载-DDPG 资源分配方法)。

图 3 所示为本文方法在不同学习率下的收敛性能。具体地,学习率分别设置为 0.01、0.001、0.000 1 和 0.000 01。一方面,较高学习率(如 0.01 和 0.001)将会以较大的步长更新模型参数,从而导致模型参数跳过全局最优点,陷入局部最优。另一方面,较小学习率(如 0.000 01)将导致收敛速率较慢,性能较差。当学习率为 0.000 1 时,在保证收敛速度的同时,又能得到较大奖励值。值得注意的是,在模型初期的随机状态下,因为初始的仿真参数(计算资源和最大容忍时延等)设置较大,所以没有出现成功率很低、奖励值为负的情况。

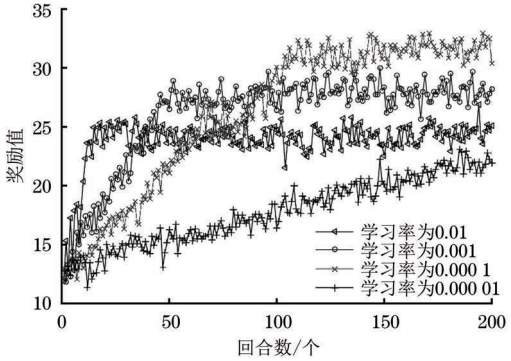


图 3 不同学习率下的收敛性能

Fig.3 Convergence performance under different learning rates

图 4 所示为本文方法在不同扩散步设置下的收敛性能,扩散步的取值分别为 10、20、30 和 40。由图 4 可知,随着扩散步的持续增加,奖励先增大后减小。首先,扩散步从 10 增加到 20 时奖励值增加,这反映了扩散模型初始的不确定性,且随着扩散步的增大和学习过程的进行,所提方法逐渐可以预测出最优离散动作的概率分布。然而,当扩散步从 25 增加到 30 和 40 的过程中,奖励值持续减小。原因在于:随着扩散步的进一步增加,方法的探索能力下降,导致陷入了次优解。

图 5 所示为在卫星不同计算资源下的任务完成率。显然,在卫星计算资源充足和匮乏时,本文方法

计算机工程 www.ecice06.com

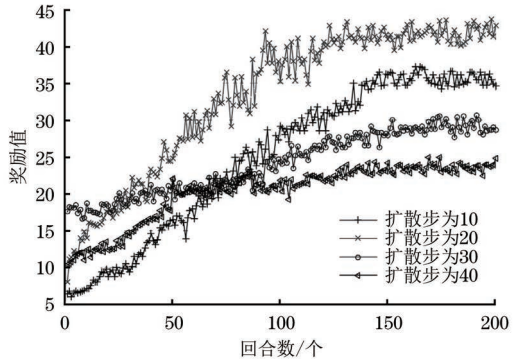


图 4 不同扩散步下的收敛性能

Fig. 4 Convergence performance under different diffusion steps

的性能始终优于对比方法。原因在于：首先，与 DQN 卸载-平均资源分配方法和随机卸载-DDPG 资源分配方法相比，本文方法针对于混合动作空间，同时优化卸载决策和资源分配，而这两种对比方法均是针对单一的离散或连续动作空间；其次，本文方法与原始 H-PPO 方法相比，引入了生成扩散模型作为离散策略，增强了算法捕获环境中复杂关系的能力，并提升了 DRL 的探索能力。此外，当卫星计算资源充足时，DQN 卸载-平均资源分配方法的任务完成率可提升至 74%，而随机卸载-DDPG 资源分配方法的任务完成率却不足 60%。这是因为：后者采用随机计算卸载方式，即使卫星上计算资源充足，但其卸载到卫星上的任务数量有限，导致任务完成率难以提升。

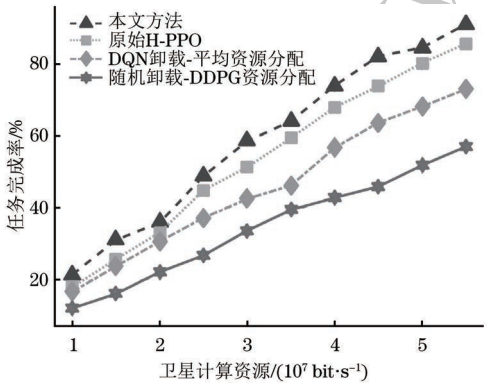


图 5 卫星不同计算资源下的任务完成率

Fig. 5 Task completion rate under different computing resources of satellites

图 6 所示为在云服务器不同计算资源下的任务完成率。与图 5 的情况相似，随着云服务器计算资源的增加，4 种方法的任务完成率都持续提升。但与图 5 不同的是，当云服务器计算资源较大时，DQN 卸载-平均资源分配方法和随机卸载-DDPG 资源分配方法下的任务完成率十分接近。这是因为：DQN 卸载-平均资源分配方法即使有较多任务卸载至云服务器，但在平均资源分配方法下仍有许多任务无法得到充足计算资源；随机卸载-DDPG 资源分配方法下任务完成率低的原因与对图 5 的分析类似。

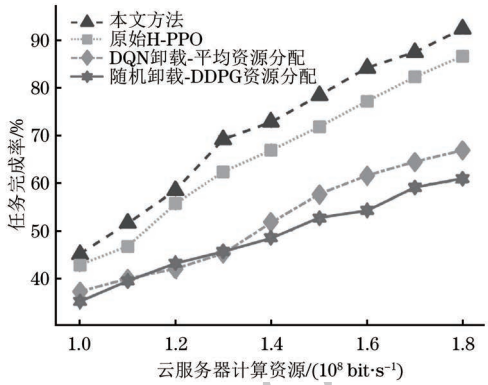


图 6 云服务器不同计算资源下的任务完成率

Fig. 6 Task completion rate under different computing resources of cloud server

图 7 所示为不同最大容忍时延下的任务完成率。随着最大可容忍时延的增加，4 种方法的任务完成率均呈现线性增长的趋势。与图 5 和图 6 不同，DQN 卸载-平均资源分配方法和随机卸载-DDPG 资源分配方法均能够达到较高数值。这是因为：当最大容忍时延较长时，即使不存在较好卸载策略或者资源分配方式，绝大多数任务也可在最大容忍时延内完成。

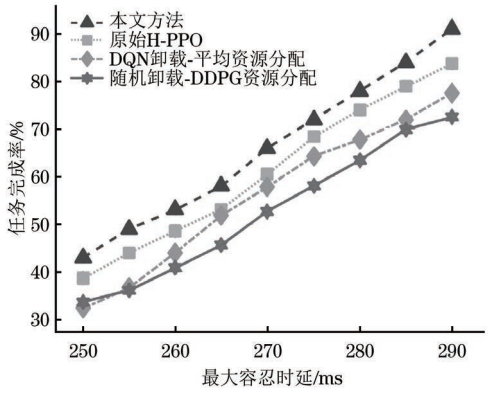


图 7 不同最大容忍时延下的任务完成率

Fig. 7 Task completion rate under different maximum tolerance latencies

图 8 所示为不同地面终端总数量设置下的任务完成率。随着终端数量的增加，4 种方法的任务完成率均呈下降趋势。当终端数量较多时，随机卸载-DDPG 资源分配方法性能超过了 DQN 卸载-平均资源分配方法，这表明在终端数量较多的设置下，任务数量随之增加；无论采取哪种卸载方式，每个节点须处理的任务数量均会增多，此时资源分配方法对任务完成率的影响将高于卸载决策。

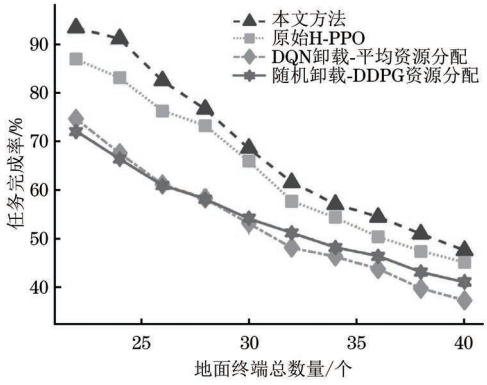


图 8 不同地面终端总数量下的任务完成率

Fig.8 Task completion rate under different total numbers of ground terminals

5 结束语

本文设计一种基于生成扩散模型的 H-PPO 方法，并将其应用于 LEO 卫星边缘网络中计算卸载和资源分配的联合优化问题。与基于离散动作空间的方法（如 DQN）和基于连续动作空间的方法（如 DDPG）不同，本文方法运行在离散-连续的混合动作空间，无须将混合空间近似为离散空间或将其松弛为连续空间。本文考虑了时变信道特性，并联合优化了功率控制、计算卸载和资源分配，以最大化长期任务完成率。为解决该问题，采用 PAMDP 模型来捕捉 LEO 卫星边缘网络的动态特性，引入生成扩散模型作为离散动作的策略，最终通过仿真实验验证本文算法的有效性。实验结果表明，该方法不仅具有良好收敛性能，而且在任务完成率方面优于已有 DRL 方法。下一步将引入不同卫星轨道数据并使用迁移学习方法，以提升模型的鲁棒性以及应用广泛性。

参考文献

[1] 何涛，曹畅，唐雄燕，等. 面向 6G 需求的算力网络技术[J]. 移动通信，2020，44(6)：131-135.
HE T，CAO C，TANG X Y，et al. Research on computing power network technology for 6G requirements[J]. Mobile Communications，2020，44(6)：131-135.（in Chinese）

[2] 段晓东，姚惠娟，付月霞，等. 面向算网一体化演进的算力网络技术[J]. 电信科学，2021，37(10)：76-85.
DUAN X D，YAO H J，FU Y X，et al. Computing force network technologies for computing and network integration evolution[J]. Telecommunications Science，2021，37(10)：76-85.（in Chinese）

[3] 贾庆民，丁瑞，刘辉，等. 算力网研究进展综述[J]. 网络与信息安全学报，2021，7(5)：1-12.
JIA Q M，DING R，LIU H，et al. Survey on research progress for compute first networking[J]. Chinese Journal of Network and Information Security，2021，7(5)：1-12.（in Chinese）

[4] 邓平科，张同须，施南翔，等. 星算网络-空天地一体化算力融合网络新发展[J]. 电信科学，2022，38(6)：71-81.
DENG P K，ZHANG T X，SHI N X，et al. Computing satellite networks-the novel development of computing-empowered space-air-ground integrated networks [J]. Telecommunications Science，2022，38(6)：71-81.（in Chinese）

[5] 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.（in Chinese）

[6] ZHU X M，JIANG C X. Integrated satellite-terrestrial networks toward 6G：architectures，applications，and challenges[J]. IEEE Internet of Things Journal，2021，9(1)：437-461.

[7] BRIK B，FRANGOUDIS P A，KSENTINI A. Service-oriented MEC applications placement in a federated edge cloud architecture[C]//Proceedings of the IEEE International Conference on Communications. Washington D. C.，USA：IEEE Press，2020：1-6.

[8] MEHRABI M，YOU D，LATZKO V，et al. Device-enhanced MEC：multi-access edge computing aided by end device computation and caching：a survey[J]. IEEE Access，2019，7：166079-166108.

[9] HAO Y X，CHEN M，HU L，et al. Energy efficient task caching and offloading for mobile edge computing[J]. IEEE Access，2018，6：11365-11373.

[10] 王奎宇，宋晓勤，缪娟娟，等. 基于 SDN 的高性能 QoS 保障低轨道卫星星间路由算法[J]. 计算机工程，2022，48(5)：185-190，199.
WANG K Y，SONG X Q，MIAO J J，et al. SDN-based high-performance and QoS guaranteed inter-satellite routing algorithm for low-earth orbit satellites [J]. Computer Engineering，2022，48(5)：185-190，199.（in Chinese）

[11] LYU Y F，LIU Z，FAN R F，et al. Optimal computation offloading in collaborative LEO-IoT enabled MEC：a multiagent deep reinforcement learning approach[J]. IEEE Transactions on Green Communications and Networking，2023，7(2)：996-1011.

[12] WANG C M，YU F R，LIANG C C，et al. Joint computation offloading and interference management in wireless cellular networks with mobile edge computing[J]. IEEE Transactions on Vehicular Technology，2017，66(8)：7432-7445.

[13] QIU C，YAO H P，YU F R，et al. Deep Q-learning aided networking，caching，and computing resources allocation in software-defined satellite-terrestrial networks [J]. IEEE Transactions on Vehicular Technology，2019，68(6)：5871-5883.

[14] CHENG X W，LYU F，QUAN W，et al. Space/aerial-assisted computing offloading for IoT applications：a learning-based approach [J]. IEEE Journal on Selected Areas in Communications，2019，37(5)：1117-1129.

[15] 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.

[16] WANG B，XIE J C，HUANG D Y，et al. A computation offloading strategy for LEO satellite mobile edge computing system[C]//Proceedings of the 14th International Conference on Communication Software and Networks. Washington D. C.，USA：IEEE Press，2022：75-80.

[17] DU H Y，WANG J C，NIYATO D，et al. AI-generated incentive mechanism and full-duplex semantic communications for information sharing[J]. IEEE Journal on

计算机工程 www.ccicc.com

Selected Areas in Communications, 2023, 41(9): 2981-2997.
[18] ZHOU D, SHENG M, WANG Y X, et al. Machine learning-based resource allocation in satellite networks supporting Internet of remote things[J]. IEEE Transactions on Wireless Communications, 2021, 20(10): 6606-6621.
[19] 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.
[20] XIONG J, WANG Q, YANG Z, et al. Parametrized deep Q-networks learning: reinforcement learning with discrete-continuous hybrid action space[EB/OL]. [2024-05-10]. https://arxiv.org/abs/1810.06394.
[21] DU H, ZHANG R, LIU Y, et al. Beyond deep reinforcement learning: a tutorial on generative diffusion models in network optimization[EB/OL]. [2024-05-10]. https://arxiv.org/pdf/2308.05384v1.
[22] MASSON W, RANCHOD P, KONIDARIS G. Reinforcement learning with parameterized actions[J]. Artificial Intelligence, 2016, 30(1): 1934-1940.
[23] DU H, LI Z, NIYATO D, et al. Generative AI-aided optimization for AI-generated content services in edge networks[EB/OL]. [2024-05-10]. https://arxiv.org/abs/2303.13052v1.
[24] FAN Z, SU R, ZHANG W N, et al. Hybrid actor-critic reinforcement learning in parameterized action space[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Washington D. C., USA: IEEE Press, 2019: 2279-2285.
[25] JIANG W, FENG D Q, SUN Y, et al. Joint computation offloading and resource allocation for D2D-assisted mobile edge computing[J]. IEEE Transactions on Services Computing, 2022: 16(3): 1949-1963.

文字编辑 索书志
栏目编辑 宋 圆

计算机工程 www.ecice06.com