

基于成本函数优化的多波束卫星通信系统 波束调度及资源分配算法

张思雅^{1,2}, 柴 蓉^{1,2}, 梁承超^{1,2}, 陈前斌^{1,2}

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 重庆市移动通信技术重点实验室, 重庆 400065)

摘 要: 多波束卫星通信系统由于其可实现高吞吐量及高效资源利用而受到广泛关注. 本文研究多波束卫星通信系统的波束调度及资源分配问题, 综合考虑用户地理位置及业务特性, 提出基于Optics的用户初始分组算法. 为提升波束覆盖性能, 设计基于最小圆的卫星波位点及覆盖半径优化方案. 基于所确定的用户分组策略, 定义系统成本函数, 并将联合波束调度、子信道选择及功率分配问题建模为系统成本函数最小化问题. 为求解所建模型的优化问题, 首先引入聚合节点描述各用户组特性, 并提出基于参数化深度 Q 网络的联合波束调度及功率分配算法; 然后基于所确定的用户组波束调度及功率分配策略, 提出基于双深度 Q 学习算法及基于近端策略优化算法的联合子信道及功率分配策略. 仿真结果验证了所提算法的有效性.

关键词: 多波束卫星; 用户分组; 波束调度; 子信道及功率分配; PDQN算法

基金项目: 国家自然科学基金(No.62271097)

中图分类号: TN927+.2

文献标识码: A

文章编号: 0372-2112(2025)04-1232-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240116

Cost Function Optimization-Based Beam Scheduling and Resource Allocation Algorithm for Multibeam Satellite Communication Systems

ZHANG Si-ya^{1,2}, CHAI Rong^{1,2}, LIANG Cheng-chao^{1,2}, CHEN Qian-bin^{1,2}

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Chongqing Key Laboratory of Mobile Communications Technology, Chongqing 400065, China)

Abstract: Multibeam satellite communication systems have received widespread attentions due to their high throughput and efficient resource utilization. This paper investigates the beam scheduling and resource allocation problem in multibeam satellite communication system. By jointly considering user position and service characteristics, an optics-based initial user grouping algorithm is proposed. To enhance beam coverage performance, a minimum circle algorithm is proposed to optimally design satellite beam positions and coverage radius. Given the determined user grouping strategy, system cost function is defined and the joint beam scheduling, sub-channel allocation and power allocation problem is formulated as a system cost function minimization problem. To solve the formulated optimization problem, aggregate nodes are introduced to describe the characteristics of user groups, and a parameterized deep Q -network-based joint beam scheduling and power allocation algorithm is proposed. Based on the obtained user group beam scheduling and power allocation strategy, a double deep Q -network algorithm and a proximal policy optimization-based joint subchannel and power allocation strategies are proposed. Simulation results validate the effectiveness of the proposed algorithms.

Key words: multibeam satellite; user grouping; beam scheduling; subchannel and power allocation; parametrized deep Q -network (PDQN) algorithms

Foundation Item(s): National Natural Science Foundation of China (No.62271097)

1 引言

基于高轨道人造卫星提供信息交互服务的高轨卫

星通信系统, 具有覆盖范围广、系统容量大、不受地理条件限制及抗干扰能力强等优点^[1]. 多波束高轨卫星

技术通过在卫星上安装多个天线,形成多个独立的通信波束,可同时为多个用户或地区提供通信服务,从而实现更高的灵活性及系统容量的提升.多波束高轨卫星通信系统的波束调度及无线资源分配技术是实现系统性能增强的重要保障.然而,由于卫星系统的动态特性以及不断增长的用户需求与有限的服务提供能力之间的冲突,多波束卫星系统^[2]的波束调度及资源分配问题面临挑战及困难.

文献[3~15]针对多波束卫星系统的波束调度及资源分配问题开展研究.文献[3,4]研究多波束卫星系统的用户分组策略,提出基于速率最大化的用户分组算法^[3]及基于加权 K 均值聚类方法的用户分组方案^[4].文献[5,6]研究多波束卫星通信系统的功率分配问题,以实现用户传输速率供需比优化^[5]及卫星资源的高效利用^[6].文献[7]设计卫星随机接入网络的用户配对及功率分配方案,实现系统总速率最大化.文献[8]设计多波束卫星通信系统的用户调度及功率分配方案,实现系统吞吐量优化.

文献[9~11]研究了多波束卫星通信系统的联合波束跳变及资源分配问题,提出一种协作的长短期带宽分配策略以提高速率供需比^[9],高效资源利用及动态流量需求匹配方案^[10],以及基于系统传输速率和能效优化的混合波束成形、用户调度及资源分配优化方案^[11].文献[12,13]运用深度强化学习(Deep Reinforcement Learning, DRL)算法设计卫星通信系统资源分配策略,提出了一种基于合作多智能体DRL框架的带宽分配策略^[12],以及双延迟深度确定性策略梯度的联合子信道及功率分配算法^[13].文献[14]基于多智能体深度确定性策略梯度方法,联合优化设计了卫星系统用户关联、功率控制及缓存部署策略.文献[15]提出了一种基于深度强化学习的卫星系统动态波束模式及带宽分配方案,实现吞吐量最大化及小区间延迟公平性最小化.

尽管已有研究考虑了多波束卫星通信系统的资源分配问题,但较少考虑用户分组和动态资源分配策略的联合优化设计,导致系统性能受限^[3~15].此外,现有研究多考虑静态场景的资源分配策略^[3~11],而未充分考虑实际场景中业务随机到达及信道随机性导致的网络动态特性.尽管文献[12~15]研究了动态卫星网络设计资源分配算法,但主要考虑小区/用户组的资源分配,较少考虑面向用户的细粒度资源分配策略.针对现有研究存在的问题,本文研究多波束卫星通信系统的用户分组及动态资源分配策略.首先基于用户间距离及业务特性,设计基于Optics算法的用户初始分组策略;然后基于最小圆算法优化确定卫星波位点及覆盖半径.基于所确定的用户分组策略,本文首先设计面向用户组的波束调度及功率分配策略,将该问题建模为基于成本函数最小化的优化问题.为了求

解所建模的问题,将其转换为一个马尔可夫决策过程(Markov Decision Process, MDP),并提出了一种参数化深度 Q 网络(Parametrized Deep Q -Network, PDQN)算法.给定波束调度策略,进而为各用户组内的用户设计子信道及功率分配策略,分别提出基于双重深度 Q 网络(Double Deep Q -Network, DDQN)算法,基于离散空间及连续空间的近端策略优化(Proximal Policy Optimization, PPO)算法的用户子信道及功率分配策略.

2 系统模型

2.1 模型设计

本文考虑一个由多波束高轨卫星及多个地面用户组成的卫星通信系统,其中卫星配备了信号收发器,可通过服务链路向用户传输数据,令 $q = (x_s, y_s, h_s)$ 表示卫星的三维坐标.为了提高数据传输性能,卫星可生成多个波束,同时传输数据至地面不同区域,令 K 表示卫星的波束数目.令 U 表示用户总数, q_u 表示用户 u 的位置, G_u 表示第 u 个用户, $1 \leq u \leq U$.令 $\tilde{\eta}_u$ 表示用户 u 的业务需求,本文将用户业务需求建模为均匀分布的随机变量,服从 $\tilde{\eta}_u \in [\tilde{\eta}_{\min}, \tilde{\eta}_{\max}]$.

根据波束的覆盖范围,可将地面用户分为多个用户组,令 N 表示用户组的数目, C_n 表示第 n 个用户组, $q_n = (x_n, y_n)$ 表示 C_n 中心点的坐标, M_n 表示 C_n 的用户数, $1 \leq n \leq N$.令 $U_{n,m}$ 表示 C_n 的第 m 个用户, $1 \leq m \leq M_n$. $q_{n,m} = (x_{n,m}, y_{n,m})$ 表示 $U_{n,m}$ 的坐标.将卫星系统的总带宽划分为 F 个子信道,令 B 表示每个子信道的带宽.为提高频谱利用率,各波束以全频复用的方式占用子信道.本文所考虑的系统模型如图1所示.

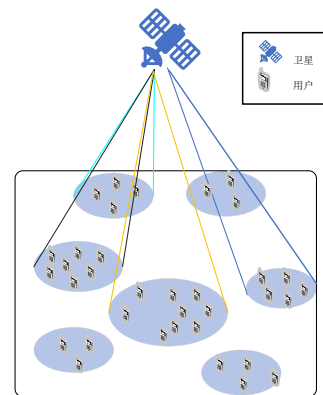


图1 系统模型图

2.2 通信信道模型

令 $h_{n,m,f}$ 表示卫星占用子信道 f 传输数据至 $U_{n,m}$ 对应的信道增益,建模为^[9]

$$h_{n,m,f} = g_{n,m} g_{n,m}^r L_{n,m,f} L \quad (1)$$

其中, $g_{n,m}$ 表示卫星到 $U_{n,m}$ 的发射天线增益, 可建模为

$$[g_{n,m}] = [g_n^{\max}] - 12 \frac{g_n^{\max}}{\eta} \left(\frac{\theta_{n,m}}{70\pi} \right)^2 \quad (2)$$

其中, $[\cdot]$ 表示 dB 值, g_n^{\max} 表示卫星波束照亮 C_n 时对应的最大发射天线增益, η 为天线孔径效率, $\theta_{n,m}$ 表示波束照亮 $U_{n,m}$ 时与波束中心链路之间的夹角, 可建模为

$$\theta_{n,m} = 2 \cos^{-1} \left(\frac{(d_{n,m}^s)^2 + d_n^2 - d_{n,m}^2}{2d_{n,m}^s d_n} \right) \quad (3)$$

其中, $d_{n,m}^s$ 表示卫星到 $U_{n,m}$ 的距离, $d_{n,m}$ 表示 $U_{n,m}$ 到 C_n 中心点的距离, d_n 表示卫星到 C_n 中心点的距离.

式(1)中, $L_{n,m,f}$ 表示卫星占用子信道 f 传输数据到 $U_{n,m}$ 时对应的自由空间链路损耗, 可建模为 $L_{n,m,f} = \left(\frac{c}{4\pi d_{n,m}^s \zeta_f} \right)^2$, 其中, c 为光速, ζ_f 为子信道 f 的载波频率, L 为雨衰系数, 服从对数高斯分布.

3 用户分组及波束半径优化

本节首先提出一种基于 Optics 的用户分组算法, 进而设计基于最小圆算法的波束覆盖半径优化方案.

3.1 基于 Optics 算法的用户分组算法

本节首先提出用户间相似度量, 进而提出基于 Optics 算法的用户分组策略.

3.1.1 相似度量

本节综合考虑用户业务需求及用户之间的距离, 定义用户相似性度量. 令 $\zeta_{u,\hat{u}}$ 表示用户分组的条件, 建模为

$$\zeta_{u,\hat{u}} = \omega_1 \zeta_{u,\hat{u}}^d + \omega_2 \zeta_{u,\hat{u}}^r \quad (4)$$

其中, ω_1 及 ω_2 表示权重系数; $\zeta_{u,\hat{u}}^d$ 表示用户 u 和用户 \hat{u} 之间的归一化欧氏距离, 可建模为 $\zeta_{u,\hat{u}}^d = \frac{\|q_u - q_{\hat{u}}\|}{d_{\max}}$, 其中,

d_{\max} 表示用户之间的最大距离, 也即 $d_{\max} = \max_{1 \leq u \neq \hat{u} \leq U} \|q_u - q_{\hat{u}}\|$; $\zeta_{u,\hat{u}}^r$ 表示用户 u 和用户 \hat{u} 之间的归一化业务需求差, 可建模为 $\zeta_{u,\hat{u}}^r = \frac{|\tilde{\eta}_u - \tilde{\eta}_{\hat{u}}|}{\eta'_{\max}}$, 其中, η'_{\max} 表示

用户之间的最大业务量差, 也即 $\eta'_{\max} = \max_{1 \leq u \neq \hat{u} \leq U} |\tilde{\eta}_u - \tilde{\eta}_{\hat{u}}|$.

3.1.2 用户分组算法步骤

基于所定义的用户相似性度量, 本小节采用 Optics 算法对用户进行分组. 用户分组算法的步骤如下.

(1) 初始化: 令 F_u 表示 G_u 的分组标识, 设置 $F_u = 0$, $1 \leq u \leq U$, 初始化结果列表 $A = \emptyset$, 其中 \emptyset 表示空集, 相似性度量门限值 ζ_{\min} .

(2) 计算用户相似性度量: 基于式(4)计算用户间的相似性度量.

(3) 确定初始用户: 随机选择一个用户, 将其放入结果列表 A 中. 令所选用户为 G_u , 记为当前用户, 更新列表为 $A = [G_u]$, 令 $F_u = 1$.

(4) 寻找最小相似性度量用户: 比较未分组用户与当前用户的相似性度量, 选择具有最小相似性度量的用户, 即, 若 $G_{\hat{u}} = \arg \min_{u' \neq u} \{\zeta_{u,u'}\}$, 将 $G_{\hat{u}}$ 标为当前用户, 并添加至结果列表 A 中, 更新结果列表为 $A = [A G_{\hat{u}}]$, 令 $F_{\hat{u}} = 1$.

(5) 判断算法是否结束: 判断是否存在未分组用户, 若是, 转至步骤(4); 否则, 转到步骤(6).

(6) 确定分组策略: 输出结果列表 A , 若结果列表 A 中相邻用户之间的相似性度量高于门限值 ζ_{\min} , 则将对应用户划分为一个用户组.

基于上述分组算法, 可得用户分组策略 $C_n = \{U_{n,m}, 1 \leq m \leq M_n, 1 \leq n \leq N\}$.

3.2 基于最小圆覆盖的卫星波位点及覆盖半径优化

基于 3.1 节得到的用户分组策略, 本节确定各用户组所关联卫星波束的波位点及覆盖半径. 由于卫星波束覆盖半径与其发射天线的最大增益成反比, 为了提高用户信道增益, 应在确保波束可覆盖用户组内所有用户的前提下, 实现波束覆盖面积最小化. 本节采用最小圆覆盖法确定每个用户组的波位点及覆盖半径.

最小圆算法的基本思想是: 给定一组用户, 首先确定覆盖部分用户的外接圆; 然后基于迭代算法更新圆的中心及半径, 以得到可覆盖所有用户的最小圆. 不失一般性, 以波束 k 及用户组 $C_n = \{U_{n,m}, 1 \leq m \leq M_n\}$ 为例, 阐述基于最小圆覆盖的卫星波位点及覆盖半径优化算法. 算法步骤如下:

(1) 任意选择用户组 C_n 中的 2 个用户, 令其为 $U_{n,1}$ 和 $U_{n,2}$.

(2) 以连接 2 个用户 $U_{n,1}$ 和 $U_{n,2}$ 的线段作为直径, 线段的中点为圆心, 构造一个圆. 若所得到的圆形区域覆盖了 C_n 中的所有用户, 则算法终止, 对应圆形区域的中心及半径 r_k 即为卫星波束 k 的波位点及覆盖半径; 否则, 选择圆外的一个用户, 令其为 $U_{n,3}$.

(3) 若 $U_{n,1}$ 、 $U_{n,2}$ 和 $U_{n,3}$ 形成的三角形为直角三角形或钝角三角形, 则重新选择圆外一个用户作为 $U_{n,3}$, 并判断所得三角形形状; 若所得三角形为锐角三角形, 则转至步骤(4).

(4) 基于所得三角形构造外接圆, 若该圆包含 C_n 中的所有用户, 则算法结束, 对应圆形区域的中心及半径即为波束 k 的中心点及覆盖半径; 否则, 选择该外接圆外的一点, 将其标记为 $U_{n,4}$, 并选择 $\{U_{n,1}, U_{n,2}, U_{n,3}\}$ 中离 $U_{n,4}$ 最远的点, 重新标记为 $U_{n,5}$. 在外接圆中, 确定过点 $U_{n,5}$ 的圆的直径, 延长该直径后将平面分成 2 个半平

面,找到 $\{U_{n,1}, U_{n,2}, U_{n,3}\}$ 中与用户 $U_{n,4}$ 不在同一半平面中的用户,重新标记为 $U_{n,6}$,令 $\{U_{n,4}, U_{n,5}, U_{n,6}\}$ 为 $\{U_{n,1}, U_{n,2}, U_{n,3}\}$,转至步骤(3)。

4 问题建模

本节首先建模系统成本函数,进而将联合波束调度、子信道及功率分配问题建模为一个满足约束条件的系统成本最小化问题。

4.1 系统成本函数建模

令 \mathcal{G} 表示系统成本函数,将其建模为

$$\mathcal{G} = \sum_{n=1}^N \mathcal{G}_n \quad (5)$$

其中, \mathcal{G}_n 表示 C_n 的成本函数.基于 C_n 的业务需求与可提供容量之间的差异,将 \mathcal{G}_n 定义为 $\mathcal{G}_n = |\eta_n - R_n|$,其中, η_n 表示 C_n 的业务需求,可建模为 $\eta_n = \sum_{G_n \in C_n} \tilde{\eta}_n$; R_n 表示 C_n 的可实现容量,可计算为

$$R_n = \sum_{m=1}^{M_n} \sum_{k=1}^K y_{n,k} R_{n,m} \quad (6)$$

其中, $y_{n,k}$ 表示波束选择变量,若波束 k 照亮 C_n , $y_{n,k} = 1$,否则 $y_{n,k} = 0$; $R_{n,m}$ 表示 $U_{n,m}$ 的可实现容量,可建模为

$$R_{n,m} = \sum_{f=1}^F x_{n,m,f} B \log \left(1 + \frac{P_{n,m,f} h_{n,m,f}}{N_0 B} \right) \quad (7)$$

其中, $x_{n,m,f}$ 表示子信道选择变量,若 $U_{n,m}$ 选择子信道 f 进行数据传输,则 $x_{n,m,f} = 1$,否则 $x_{n,m,f} = 0$; $P_{n,m,f}$ 表示 $U_{n,m}$ 占用子信道 f 的发送功率; N_0 表示加性高斯白噪声的功率谱密度。

4.2 优化问题建模

本文将联合波束调度、子信道选择及用户功率分配问题建模为一个系统成本函数最小化问题,即

$$\begin{aligned} & \min_{y_{n,k}, x_{n,m,f}, P_k, P_{n,m,f}} \mathcal{G} \\ \text{s.t.} & \left\{ \begin{array}{l} \text{C1: } \sum_{k=1}^K y_{n,k} \leq 1 \\ \text{C2: } \sum_{n=1}^N y_{n,k} \leq 1 \\ \text{C3: } x_{n,m,f} \leq y_{n,k} \\ \text{C4: } \sum_{m=1}^{M_n} \sum_{f=1}^F x_{n,m,f} \leq F \\ \text{C5: } \sum_{m=1}^{M_n} x_{n,m,f} \leq 1 \\ \text{C6: } \sum_{f=1}^F x_{n,m,f} \leq 1 \\ \text{C7: } \sum_{k=1}^K P_k \leq P_{\text{tot}} \end{array} \right. \quad (8) \end{aligned}$$

其中,C1和C2为波束调度约束;C3~C6为时频资源分配约束;C7为功率分配约束; P_{tot} 表示卫星的总功率; P_k 表示波束 k 分配的功率值。

5 优化问题求解

式(8)所建模的优化问题为NP难问题,采用传统方法难以求解.本节提出一个两阶段算法:首先为各用户组设计波束调度及功率分配策略;然后针对组内用户设计功率及子信道分配策略。

5.1 用户组波束调度及功率分配算法

5.1.1 优化问题建模

给定用户分组策略,本节引入聚合节点描述各用户组中的所有用户业务特性.令 AN_n 表示 C_n 的聚合节点, AN_n 的位置为 C_n 的地理中心,即 $q_n = (x_n, y_n)$, AN_n 的业务定义为 C_n 中所有用户的聚合业务.基于所定义的聚合节点,将用户的波束调度及功率分配问题转换为各聚合节点的波束调度及功率分配问题,可建模为

$$\begin{aligned} & \min_{y_{n,k}, P_k} \mathcal{G} \\ \text{s.t.} & \text{C1, C2, C7} \end{aligned} \quad (9)$$

5.1.2 MDP建模

分析优化问题式(9)可知,上述问题为混合整数非线性规划问题,难以直接进行求解.本节将其转换为MDP,并提出基于PDQN算法确定聚合节点的波束调度及功率分配策略.所构建的MDP模型由环境状态、动作及奖励函数组成。

令 \mathcal{S} 表示环境状态空间.令 $s_t \in \mathcal{S}$ 表示当前环境状态,建模为 $s_t = \{h_{t,1}, h_{t,2}, \dots, h_{t,n}\}$,其中, $h_{t,n}$ 表示当前状态卫星与 AN_n 之间通信链路状态。

令 \mathcal{A} 表示动作空间,定义为各状态下可采取的所有动作的集合.基于当前时刻系统状态 s_t ,输出动作 $a_t \in \mathcal{A}$.为确定聚合节点的波束调度及功率分配策略,将 a_t 定义为 $a_t = \{y_{t,n,k}, P_{t,k}\}$,其中, $y_{t,n,k}$ 表示当前时刻的波束调度决策,若波束 k 照亮 C_n ,令 $y_{t,n,k} = 1$,否则, $y_{t,n,k} = 0$, $P_{t,k}$ 表示当前时刻波束 k 分配的功率值。

给定当前状态及所采取动作,可得到相应奖励函数.为实现系统成本函数优化,建模奖励函数为

$$r_t = \frac{1}{1 + \exp(b_1(\mathcal{G}_t - c_1))} \quad (10)$$

其中, \mathcal{G}_t 表示状态 s_t 下采取动作 a_t 对应的成本函数, b_1 及 c_1 为系统参数。

5.1.3 基于PDQN的联合波束调度及功率分配算法

给定所建模MDP的环境状态,可基于DRL方法,确定动作策略,实现长期奖励函数优化.本节考虑深度Q网络(Deep Q-Network, DQN)及深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法在处理

离散动作空间及连续动作空间方面的优势,提出基于PDQN的波束调度及功率分配算法. 算法主要步骤如下:首先对混合动作空间 $\{y_{l,n,k}, P_{l,k}\}$ 进行处理,将连续动作 $P_{l,k}$ 表示为离散动作 $y_{l,n,k}$ 的函数,即 $P_{l,k}(y_{l,n,k})$;进而基于DDPG方法确定最优连续动作 $P_{l,k}^*(y_{l,n,k})$. 给定当前时刻 t 的系统状态 s_t 及对应的连续动作 $P_{l,k}^*(y_{l,n,k})$,基于DQN方法确定最优离散动作 $y_{l,n,k}^*$. 以下对所提算法进行简要阐述.

给定当前时刻系统状态 s_t 、离散动作 $y_{l,n,k}$ 及连续动作 $P_{l,k}(y_{l,n,k})$,可建模 Q 函数为

$$Q(s_t, y_{l,n,k}, P_{l,k}(y_{l,n,k})) = \mathbb{E} \left[r_t + \gamma \max_{y_{l,n,k}, P_{l,k}(y_{l,n,k})} \sup_{P_{l,k}(y_{l,n,k})} Q(s_{t+1}, y_{l,n,k}, P_{l,k}(y_{l,n,k})) \middle| a_t \right] \quad (11)$$

式中, $a_t = (y_{l,n,k}, P_{l,k}(y_{l,n,k}))$, γ 为折扣因子, $0 < \gamma < 1$. 对于给定的 Q 函数 $Q(s_t, y_{l,n,k}, P_{l,k}(y_{l,n,k}))$ 、状态 s_t 及动作 $y_{l,n,k}$,可将 $\arg \sup_{y_{l,n,k}, P_{l,k}(y_{l,n,k})} Q(s_{t+1}, y_{l,n,k}, P_{l,k}(y_{l,n,k}))$ 视为一个由系统状态到功率分配策略的映射,即 $P_{l,k}^Q(y_{l,n,k}): s \rightarrow P_{l,k}(y_{l,n,k})$. 相应地,可将式(11)重写为

$$Q(s_t, y_{l,n,k}, P_{l,k}(y_{l,n,k})) = \mathbb{E} \left[r_{(t+1)} + \gamma \max_{y_{l,n,k}} \sup_{P_{l,k}(y_{l,n,k})} Q(s_{t+1}, y_{l,n,k}, P_{l,k}^Q(y_{l,n,k}, s_{t+1})) \middle| s_t \right] \quad (12)$$

对于式(12),分别采用神经网络 $Q(w)$ 及 $\pi(\theta)$ 对 $Q(s_t, y_{l,n,k}, P_{l,k}(y_{l,n,k}))$ 及 $P_{l,k}^Q(y_{l,n,k}, s_{t+1})$ 进行估计,其中, $Q(w)$ 为价值网络, w 表示该网络的权值; $\pi(\theta)$ 为基于确定策略梯度的网络, θ 表示该网络的权值. 为确定 w 和 θ 之间的关系,当 w 固定时,对应的 θ 满足:

$$Q\left(\left(s_t, y_{l,n,k}, P_{l,k}(y_{l,n,k}, s_t, \theta)\right); w\right) \approx \sup_{P_{l,k}(y_{l,n,k})} Q\left(s_t, y_{l,n,k}, P_{l,k}(y_{l,n,k}); w\right) \quad (13)$$

基于损失函数最小化可更新 w 及 θ .

5.2 用户子信道及功率分配算法

基于前述面向用户组的波束调度及功率分配策略,本节设计面向用户的子信道及功率分配算法. 假设用户组间的传输干扰较小,可针对每个地面用户组独立设计其子信道及功率分配策略.

5.2.1 基于DDQN的联合子信道及功率分配策略

为不失一般性,以用户组 C_n 为例,建模其子信道及功率分配问题如下:

$$\begin{aligned} & \min_{x_{n,m,f}, P_{n,m,f}} \mathcal{G}_n \\ & \text{s.t. } C3, C4, C5, C6, C8 \end{aligned} \quad (14)$$

该问题混合非整型规划问题,难以直接求解,本节将该优化问题建模为MDP,并提出了基于DDQN的联合子信道及功率分配策略.

令 \mathcal{S} 表示环境状态空间, $s_t \in \mathcal{S}$ 表示当前时刻 t 的环境状态,建模为 $s_t = \{h_{l,n,1,1}, h_{l,n,1,2}, \dots, h_{l,n,M_n,F}\}$,其中, $h_{l,n,m,f}$ 表示当前时刻 t 卫星与 $U_{n,m}$ 之间通信链路状态集合. 令 \mathcal{A} 表示动作空间, $\hat{a}_t \in \mathcal{A}$ 表示当前时刻的动作,定义为 $\hat{a}_t = \{X_{l,m}, p_{l,m}\}$,其中, $X_{l,m} \in \{0, 1, \dots, F\}$ 表示 $U_{n,m}$ 的子信道选择策略,若为 $U_{n,m}$ 分配子信道 f ,则设置 $X_{l,m} = f$,否则, $X_{l,m} = 0$; $p_{l,m}$ 为 $U_{n,m}$ 的功率分配策略. 由于功率分配变量 $p_{l,m}$ 为连续变量,本节利用离散化机制将其转换为离散功率阶数. 假设波束 k 照亮用户组 C_n ,将波束的发送功率 P_k 平均划分为 J 阶. 令 \hat{P}_j 表示第 j 级的发送功率,可得 $\hat{P}_j = \frac{P_k j}{J}, 1 \leq j \leq J$. 令 $\hat{p}_{l,m} \in \{\hat{P}_1, \hat{P}_2, \dots, \hat{P}_J\}$ 表示卫星波束的离散功率分配策略,则时刻 t 的动作可重写为 $a_t = \{X_{l,m}, \hat{p}_{l,m}\}$.

给定当前状态 s_t 及所采取的动作 a_t ,可得到对应奖励函数,建模为

$$r_t = \frac{1}{1 + \exp(b_2(\vartheta_t - c_2))} \quad (15)$$

其中, b_2 及 c_2 为系统参数.

基于所建模的MDP,本节提出一种基于DDQN的联合子信道及功率分配算法. DDQN由两个 Q 网络组成,即评价网络及目标网络. 在训练过程中,目标网络与评价网络合作,完成网络参数的优化. 令 $Q(s_t, a_t; \pi_t), \tilde{Q}(s_t, a_t; \tilde{\pi}_t)$ 分别表示评价网络及目标网络的 Q 值,其中, π_t 及 $\tilde{\pi}_t$ 分别表示评价网络及目标网络的参数. 令 $\bar{Q}(s_t, a_t)$ 表示用于更新评价网络的网络参数的标签值,可建模为

$$\bar{Q}(s_t, a_t) = r_t + \gamma \tilde{Q}\left(s_{t+1}, \arg \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \pi_t); \tilde{\pi}_t\right) \quad (16)$$

以最小均方误差作为损失函数,对评估网络参数进行优化,并使用梯度下降算法更新 π_t . 经过一定次数的迭代,得到更新后的 π_t ,替换目标网络的 $\tilde{\pi}_t$. 基于所建模的MDP,训练DDQN模型,直至模型收敛,可得到对应的子信道及功率分配策略.

5.2.2 基于PPO的联合子信道及功率分配算法

(1) PPO算法概述

PPO是一种基于策略梯度的强化学习算法,属于策略-价值网络结构. 在一定的状态下,策略网络输出一个特定的动作,而价值网络提供值函数,以评估策略网络的性能. 定义策略网络的损失函数为

$$L^c(\theta_t) = \mathbb{E} \left[\min \left(r_t(\theta_t) \hat{A}_t, c(r_t(\theta_t), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right] \quad (17)$$

其中, c 为剪切率; $r_t(\theta_t)$ 表示策略更新幅度, 定义为状态 s_t 当前策略与之前策略采取动作 a_t 的概率之比, 可建模为 $r_t(\theta_t) = \frac{\pi_\theta(a_t|s_t)}{\hat{\pi}_\theta(a_t|s_t)}$, 其中, $\pi_\theta(a_t|s_t)$ 表示当前策略在状态 s_t 时采取动作 a_t 的概率, $\hat{\pi}_\theta(a_t|s_t)$ 表示之前策略在状态 s_t 下采取动作 a_t 的概率; \hat{A}_t 为优势函数; $c(x, b, c)$ 为剪裁函数, 表示将 x 限制在区间 $[b, c]$ 内; θ_t 为策略网络参数, 可采用随机梯度上升算法对其进行更新。

价值网络的损失函数定义为

$$L^v(\phi_t) = \frac{1}{2} \mathbb{E} \left[(V_\phi(s_t) - V_t)^2 \right] \quad (18)$$

其中, V_t 为当前时刻 t 的目标价值函数, 表示当前状态的期望回报; $V_\phi(s_t)$ 表示目标状态 s_t 的目标价值函数的值, 也即状态 s_t 的未来累积奖励; ϕ 为价值网络参数, 可采用随机梯度下降法对 ϕ 进行更新。

(2) 基于 PPO 的子信道及功率分配策略

DDQN 算法可处理具有离散动作空间的 MDP 问题, 但存在 Q 值估计不够稳定, 可能陷入局部最优解等问题. 本节提出基于 PPO 算法的子信道及功率分配策略. 基于所建模的 MDP 模型, 采用离散动作 PPO 算法确定用户子信道及功率分配策略. 给定状态 s_t , 对于离散动作 a_t , 采用 Softmax 函数确定动作执行概率函数, 即

$$\pi_\theta(a_t|s_t) = \frac{e^{\chi(s_t, a_t)^T \theta}}{\sum_{a'_t \in \mathcal{A}} e^{\chi(s_t, a'_t)^T \theta}} \quad (19)$$

其中, $\chi(s_t, a_t)$ 为状态-动作相对的特征向量. 给定卫星波束与用户之间的信道状态特性 s_t , 根据 $\pi_\theta(a_t|s_t)$ 可得采取不同离散动作 a_t 对应的概率, 进而可确定用户子信道及功率分配策略。

尽管可采用基于离散动作的 PPO 算法确定用户的子信道及功率分配策略, 但因功率分配策略的离散化会导致策略精度较低及算法性能受限. 为了进一步改进算法的性能, 可采用基于连续动作的 PPO 算法. 该算法使用参数化的策略网络, 输出连续动作的均值及方差, 从而更好地适应高维、连续动作空间. 为应用基于连续动作的 PPO 算法确定用户的子信道及功率分配策略, 需对所建模 MDP 的混合动作空间 $\hat{\mathcal{A}}_t = \{X_{t,m}, p_{t,m}\}$ 进行连续化处理. 因子信道选择动作 $X_{t,m}$ 为离散策略, 需将其转换为连续动作, 令 $X'_{t,m}$ 表示已进行连续化处理的 $U_{n,m}$ 的子信道选择变量, $X'_{t,m}$ 为 $0 \sim F$ 之间取值的连续变量, 若 $f \leq X'_{t,m} < f+1$, 则为 $U_{n,m}$ 分配子信道 f . 令 $a'_t = \{X'_{t,m}, p_{t,m}\}$ 表示时刻 t 的连续动作。

给定状态 s_t , 对于连续动作 a'_t , 采用高斯策略确定动作执行概率函数, 定义如下:

$$\tilde{\pi}_\theta(a'_t|s_t) = \frac{1}{\sqrt{2\pi} \sigma_\theta} e^{-\frac{a'_t - \mu_\theta}{2\sigma_\theta^2}} \quad (20)$$

其中, σ_θ 和 μ_θ 分别表示连续动作空间 a'_t 的标准差和均值. 给定卫星波束与用户之间的信道状态特性 s_t , 根据 $\tilde{\pi}_\theta(a'_t|s_t)$ 可得采取不同离散动作 a'_t 对应的概率, 进而可确定用户子信道及功率分配策略。

6 仿真结果分析

本文采用 Python 仿真软件对所提算法性能进行验证和评估, 并与相关文献所提算法进行比较. 仿真场景由一个多波束高轨卫星及 K 个通信用户组成, 其中, 仿真区域的大小为 $1\ 280\ \text{km} \times 1\ 280\ \text{km}$, 高轨卫星在地面的投影为 $(750\ \text{km}, 750\ \text{km})$ 处, 通信用户随机分布在仿真区域内, 仿真使用的相关参数如表 1 所示。

表 1 仿真参数设置

仿真参数	数值
卫星高度(H)	35 786 km
接收天线增益($g_{n,m}^r$)	52 dBi
功率谱密度(N_0)	-174 dBm/Hz
卫星最大功率(P_{tot})	500 W
系统带宽(B)	25 MHz
雨衰因子(L)	(0.5, 0.3)
载波频率(ξ_f)	19.5~20.5 GHz
用户数量(M)	50
子信道数量(F)	10
卫星波束(K)	4

图 2 展示了通信用户数为 50 时, 使用本文提出的用户分组算法得到的结果, 其中, 图 2(a) 展示了使用 Optics 算法进行用户分组的结果, 图 2(b) 展示了使用最小圆算法对用户组进行调整后的结果, 图 2(c) 展示了用户分组后续调整结果. 如图 2 所示, 使用 Optics 算法综合考虑距离因素和业务量因素, 将相似度较高的用户分配到同一个用户组. 应用最小圆算法可减少用户组之间的重叠区域, 进一步优化用户组分配结果. 由于部分用户组的覆盖区域内可能存在不属于该用户组的用户, 通过调整用户分组策略, 可实现用户组对覆盖区域用户的全覆盖。

图 3 为采用本文提出的基于 PDQN 的波束选择及功率分配算法得到的系统累积奖励与算法训练轮数关系图, 其中系统累积奖励定义为各次奖励的总和. 从图 3 中可看出, 随着训练步数的增加, 算法达到收敛, 证明了算法的有效性。

图 4 为采用本文所提出的用户子信道及功率分配算法得到的系统累积奖励与训练轮数关系图. 对比起见, 图 4 中绘制了本文所提的基于 DDQN、离散动作空间的 PPO 及连续动作空间的 PPO 算法对应的系统累积奖励. 由图 4 可见, 随着训练步数的增加, 奖励值逐渐趋于稳定, 说明三种算法均可收敛, 证明了所提算法的

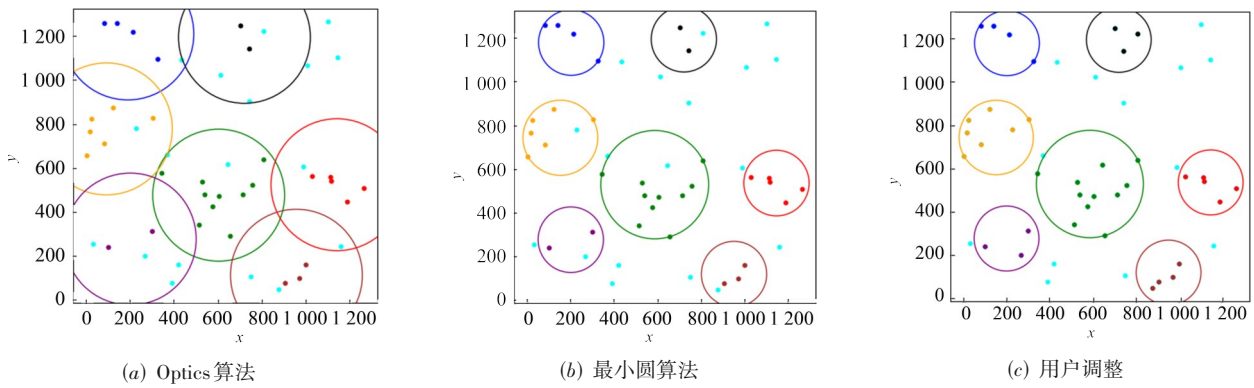


图2 用户分组示例

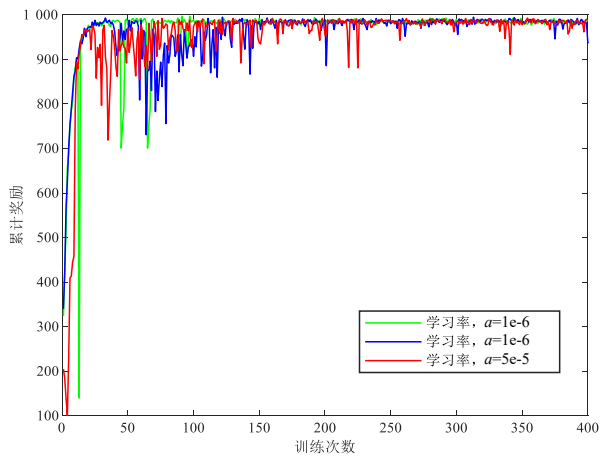


图3 系统累积奖励与算法训练轮数关系图

有效性. 对比三种算法对应的性能可以看出, 本文所提的基于连续动作空间 PPO 的子信道及功率分配算法对应的奖励值高于其他两种算法, 原因是 DDQN 及离散动作空间的 PPO 算法无法有效处理连续动作空间, 对连续动作进行离散化导致动作选择准确度受限, 系统奖励值降低.

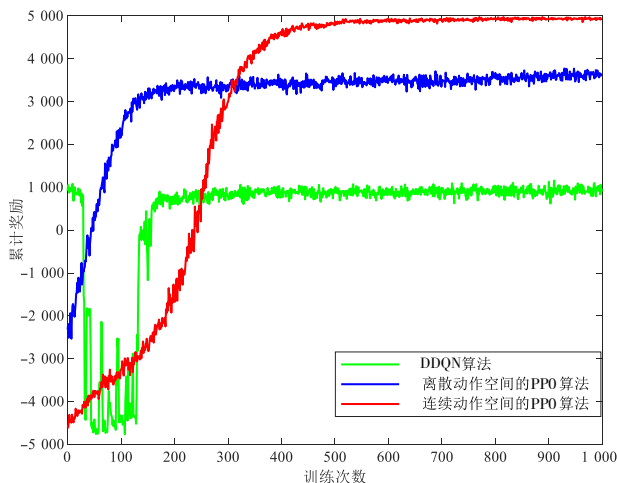


图4 子信道及功率分配收敛图

图5给出了子信道带宽对系统成本的影响的评估结果, 并将本文提出的方案与文献[15]中的方案进行了比较. 从图5中可以观察到, 子信道带宽取值较低时, 随着子信道带宽的增加, 系统成本下降; 当子信道带宽增加到一定值后, 随着子信道带宽的增加, 系统成本逐渐增加. 这是因为在带宽较小的情况下, 增加带宽可以提高服务提供能力, 使得用户需求与服务提供能力之间的差异较小. 然而, 当带宽达到一定值后, 进一步增加带宽会导致系统服务提供能力过高, 使得用户需求与服务提供能力之间的差异增大, 从而增加了成本. 从图5中还可观察到, 更低的噪声功率谱密度会产生更好的系统性能. 与文献[15]中提出的算法相比, 本文提出的算法具有较好的性能.

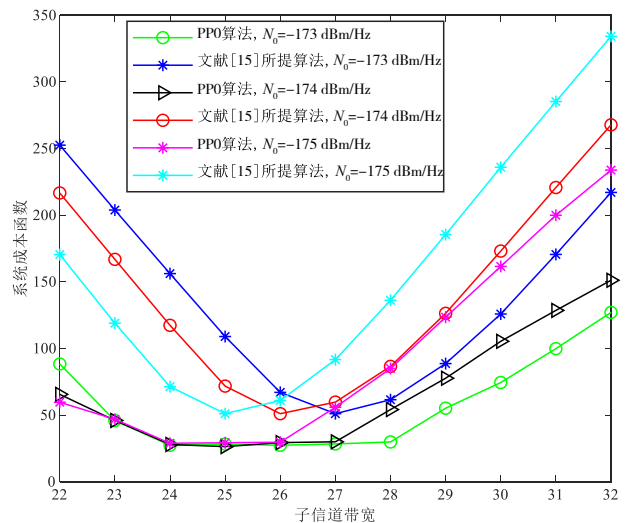


图5 子信道带宽及系统成本函数关系

图6为最大发送功率及系统成本函数关系图, 并将本文提出的方案与文献[15]中的方案进行了比较. 从图6中可观察到, 随着波束最大发送功率的增加, 系统成本相应降低. 这是因为当最大发送功率较小时, 服务提供能力可能无法满足用户需求, 导致用户需求与服务

务提供能力之间的差异较大. 随着最大发送功率的增加,可以实现发送功率的优化分配,从而减小用户需求与服务提供能力之间的差异,降低系统成本. 从图6中还可以观察到,随着子信道带宽的增加,系统成本函数降低. 这是因为较高的子信道带宽能够提供更大的传输容量,满足用户需求. 与文献[15]中提出的算法相比,本文提出的算法具有较好的性能.

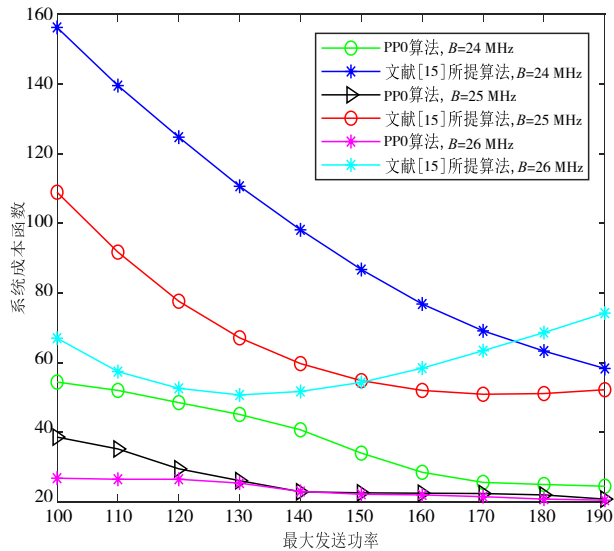


图6 波束最大功率及系统成本函数关系图

图7为不同用户业务特性所对应的系统成本函数与子信道带宽关系图. 对比较见,图7中给出了三类用户业务特性对应的仿真结果,并将本文提出的方案与文献[15]中所提的方案进行对比. 图7中的三类用户业务特性如表2所示. 从图7中可以观察到,三类业务特性对应的系统成本函数随着子信道带宽的增加均呈现先下降后增加的趋势. 这是因为若带宽值较低,系统难以满足用户需求. 此时,随着带宽的增加,系统可提供容量相应增加,导致成本函数降低. 当带宽达到一定值后,系统可提供容量与用户需求基本匹配,进一步增加带宽将导致系统服务提供能力过高,致使用户需求与服务提供能力之间的差异增大,从而导致成本函数增加. 对比不同业务需求可以观察到:子信道带宽较低时,业务特性1对应的系统成本函数较低,而业务特性3对应的系统成本函数较高. 相比于文献[15]中提出的算法,本文所提出的算法表现出更好的性能.

图8为最大发送功率与系统可提供容量之间的关系图. 由图8可以观察到,随着最大发射功率增加,系统可提供容量相应增加. 这是因为随着最大发送功率的增加,可以实现发送功率的优化分配,从而提高系统的可提供容量. 此外,从图8中还可以观察到,随着子信道带

宽的增加,系统可提供容量也随之增加. 与文献[15]中提出的算法相比,本文提出的算法具有更优性能.

表2 业务特性设置

单位: $\text{Mbit} \cdot \text{s}^{-1}$

业务特性类型	$\tilde{\eta}_{\min}$	$\tilde{\eta}_{\max}$
业务特性1	240	280
业务特性2	270	310
业务特性3	300	340

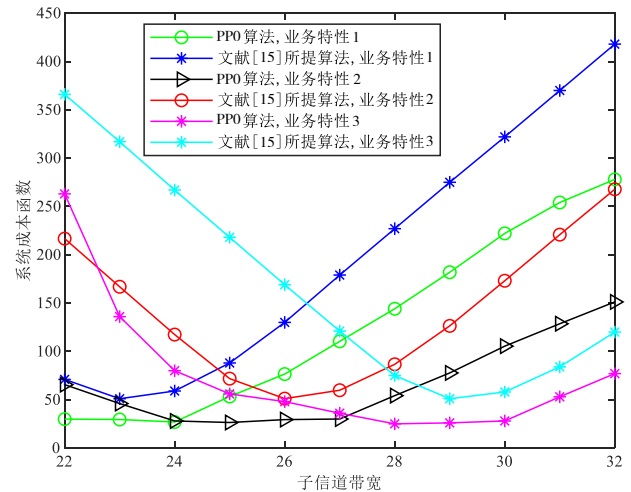


图7 子信道带宽及系统成本函数关系图

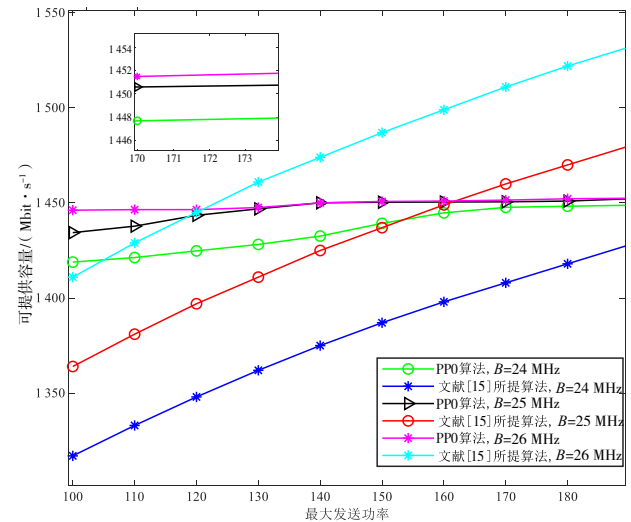


图8 最大发送功率与可提供容量之间的关系图

7 结论

本文研究了多波束卫星通信系统中的用户分组、波束调度及资源分配问题,首先提出了基于Optics和最小圆的用户分组算法. 给定用户分组策略,进而设计面向用户组的基于PDQN的波束调度和功率分配算法以及面向用户的基于DDQN和PPO算法的子信道及功率分配策略. 数值结果表明,用户分组算法能够综合考虑

距离因素和业务量因素,将相似度较高的用户分配到同一个用户组;本文所提波束调度、子信道及功率分配算法与现有算法相比,具有更优性能,可实现系统所提供的容量与用户通信需求之间更好的匹配。

参考文献

- [1] CENTENARO M, COSTA C E, GRANELLI F, et al. A survey on technologies, standards and open challenges in satellite IoT[J]. *IEEE Communications Surveys & Tutorials*, 2021, 23(3): 1693-1720.
- [2] XIAO A L, WANG X C, WU S, et al. Mobility-aware resource management for integrated satellite-maritime mobile networks[J]. *IEEE Network*, 2022, 36(1): 121-127.
- [3] CHEN H J, QI C H. User grouping for sum-rate maximization in multiuser multibeam satellite communications[C]// *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. Piscataway: IEEE, 2019: 1-6.
- [4] HONNAIAH P J, MATURO N, CHATZINOTAS S, et al. Demand-based adaptive multi-beam pattern and footprint planning for high throughput GEO satellite systems[J]. *IEEE Open Journal of the Communications Society*, 2021, 2(1): 1526-1540.
- [5] EFREM C N, PANAGOPOULOS A D. Dynamic energy-efficient power allocation in multibeam satellite systems[J]. *IEEE Wireless Communications Letters*, 2020, 9(2): 228-231.
- [6] ABDU T S, KISSELEFF S, LAGUNAS E, et al. Flexible resource optimization for GEO multibeam satellite communication system[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(12): 7888-7902.
- [7] ZHAO B, DONG X D, REN G L, et al. Optimal user pairing and power allocation in 5G satellite random access networks[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(6): 4085-4097.
- [8] VAN CHIEN T, LAGUNAS E, TA T H, et al. User scheduling and power allocation for precoded multi-beam high throughput satellite systems with individual quality of service constraints[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(1): 907-923.
- [9] WANG A Y, LEI L, LAGUNAS E, et al. Joint optimization of beam-hopping design and NOMA-assisted transmission for flexible satellite systems[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(10): 8846-8858.
- [10] ZHAO D, QIN H, XIN N, et al. Flexible resource management in high-throughput satellite communication systems: A two-stage machine learning framework[J]. *IEEE Transactions on Communications*, 2023, 71(5): 2724-2739.
- [11] PENG D Y, BANDI A, LI Y, et al. Hybrid beamforming, user scheduling, and resource allocation for integrated terrestrial-satellite communication[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(9): 8868-8882.
- [12] LIAO X L, HU X, LIU Z J, et al. Distributed intelligence: A verification for multi-agent DRL-based multibeam satellite resource allocation[J]. *IEEE Communications Letters*, 2020, 24(12): 2785-2789.
- [13] DENG D H, WANG C W, PANG M L, et al. Dynamic resource allocation with deep reinforcement learning in multibeam satellite communication[J]. *IEEE Wireless Communications Letters*, 2023, 12(1): 75-79.
- [14] LI X N, ZHANG H J, ZHOU H, et al. Multi-agent DRL for resource allocation and cache design in terrestrial-satellite networks[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(8): 5031-5042.
- [15] 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.

作者简介



张思雅 女,2000年出生.现为重庆邮电大学硕士研究生.主要研究方向为空天地一体化通信、卫星通信、无线资源分配等.
E-mail: s220132218@stu.cqupt.edu.cn



柴蓉 女,1974年出生.现为重庆邮电大学教授、博士生导师.主要研究方向为空天一体化网络架构及关键技术、无线资源管理及移动性管理技术等.
E-mail: chairong@cqupt.edu.cn