多波束LEO卫星海洋通信网络联合波位划分的无线资源调配方法

张贤[1,2]，王东东[1,2]，周政[3]，李学华[1,2]
（1.北京信息科技大学信息与通信工程学院，北京 102206；2.北京信息科技大学现代测控技术教育部重点实验室，北京 102206；
3.宁波大学信息科学与工程学院，浙江 宁波 315211）

摘要：针对海洋环境下用户分布稀疏、信道状态动态变化导致覆盖与吞吐量难以兼顾的问题，提出了一种联合波位划分的无线资源调配方法。首先，基于P中心算法，联合优化波位数量、中心位置与覆盖半径；然后，以最大化网络加权平均吞吐量（WAT）为目标，进一步联合优化子信道分配与用户上行发送功率，并设计基于多智能体双延迟深度确定性策略梯度（MATD3）算法的协同资源调配策略进行求解。仿真结果表明，与固定波位半径和K均值算法波位划分方案相比，所提波位划分方法的网络WAT分别提升10.26%和3.85%；与MADQN和MADDPG相比，所提MATD3资源调配方法的网络WAT分别提升20.53%和5.58%。

关键词：低轨卫星通信；波位划分；无线资源调配；多智能体深度强化学习
中图分类号：TN927
文献标志码：A
doi：10.11959/j.issn.1000−436x.TXXB260096

Joint beam position partitioning based on radio resource allocation method in multi-beam LEO satellite maritime communication networks

Zhang Xian[1,2], Wang Dongdong[1,2], Zhou Zheng[3], Li Xuehua[1,2]
1. School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing 102206, China
2. Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing Information Science and Technology University, Beijing 102206, China
3. Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

Abstract: To address the difficulty of balancing coverage and throughput caused by sparse user distribution and dynamic channel conditions in maritime environments, radio resource allocation method with joint beam position partitioning was proposed. First, based on the P-center algorithm, the number of beam positions, center locations, and coverage radii were jointly optimized. Then, with the objective of maximizing the network weighted average throughput (WAT), subchannel allocation and users' uplink transmit power were further jointly optimized, and a cooperative resource allocation strategy based on the multi-agent twin delayed deep deterministic policy gradient (MATD3) algorithm was designed to solve the problem. Simulation results show that, compared with the fixed beam-radius scheme and the K-means beam position partitioning scheme, the proposed beam position partitioning method improves network WAT by 10.26% and 3.85%, respectively. In addition, compared with MADQN and MADDPG, the proposed MATD3-based resource allocation method improves network WAT improvements of 20.53% and 5.58%, respectively.

Key words: LEO satellite communication, beam position partitioning, radio resource allocation, multi-agent deep reinforcement learning

收稿日期：2026−02−14；修回日期：2026−03−27
通信作者：周政，zhouzheng@nbu.edu.cn
基金项目：国家自然科学基金资助项目（No.62401066, No.62101289）
Foundation Items：The National Natural Science Foundation of China (No.62401066, No.62101289)

0 引言

低地球轨道（low Earth orbit, LEO）卫星通信具备大范围连续覆盖和灵活部署能力，已成为补充地面网络覆盖盲区的重要手段[1]。受限于星上功率与频谱等资源，LEO 卫星通信通常采用多波束技术在单星载荷上生成多个高增益窄波束，以实现对目标区域的精细化覆盖，并显著提升频谱复用效率与网络吞吐量[2]。随着 6G 演进与“智慧海洋”建设提速，未来网络对海洋通信从近岸到远洋的覆盖能力提出了更高要求[3-4]。然而，海洋环境下用户分布稀疏，不同业务类型对速率与时延的需求差异显著。受天气、海况和平台运动等因素影响，链路状态呈现较高时变性，全频率复用下的波束间同频干扰与负载波动更加显著，使传统地面无线网络资源调配方法难以直接适用[5-6]。因此，如何合理划分波位，基于 LEO 卫星构建多波束 LEO 卫星海洋通信网络，实现精细化覆盖与吞吐量均衡，对提供稳定的通信服务保障至关重要。

1 相关工作

现有针对多波束 LEO 卫星通信网络无线资源调配的研究，主要围绕功率控制、频率分配和子信道分配等展开。例如，文献[7]通过优化发送功率实现波束负载均衡与链路速率优化，提高了系统吞吐量。文献[8]提出了基于混沌粒子群优化的信道与功率分配方法，通过增强粒子群多样性来缓解标准粒子群易陷入局部最优的问题。文献[9]联合优化带宽分配与发送功率，降低了流量需求与资源供给之间的差异，提升了 LEO 卫星通信系统的频谱利用效率。文献[10]面向高通量卫星通信系统，联合优化发送功率、波束增益和波束位置，在不同地理分布的业务需求场景下实现了更高的流量容纳率。上述研究多在既定覆盖结构下进行资源调配优化，而在海洋用户分布稀疏且非均匀的条件下，固定覆盖结构可能导致覆盖与业务负载不匹配，限制资源利用效率。为此，研究者开始关注波位划分与资源调配的协同设计。例如，文献[11]提出了基于用户业务分布动态调整波位大小的资源管理算法，降低了丢包率，提升了带宽利用率。文献[12]通过灵活调节波位覆盖半径，缓解了多用户组播场景下的资源分配问题，提高了平均数据速率。文献[13]采用 K 均值聚类算法，依据地面用户历史位置进行分组，以提高频谱利用率。文献[14]针对多波束卫星通信系统中的动态波束模式与带宽分配问题，通过优化有限卫星上资源与不均匀和时变业务需求之间的匹配关系，在吞吐量与时延公平性方面均取得提升。尽管上述工作取得了一定进展，但在海洋通信场景下仍面临差异化业务需求与干扰耦合下的高维强耦合决策问题，传统优化方法的在线求解和自适应能力仍然受限。

为解决传统算法复杂度高和自适应能力受限的问题，基于深度强化学习（deep reinforcement learning, DRL）的资源调配方法近年来受到广泛关注。例如，文献[15-16]将资源调配问题建模为马尔可夫决策过程（Markov decision process, MDP），并通过数据驱动的策略迭代实现对时变环境的自适应。文献[17]提出了基于 DRL 的带宽与功率联合优化方法，采用集中训练和分布执行框架，在降低系统功耗的同时提高了带宽利用率。文献[18]提出了基于 DRL 的跳波束与覆盖控制联合优化算法，通过单智能体轮询机制联合优化波束跳变策略和覆盖半径，在提高系统吞吐量的同时降低了丢包率。文献[19]面向多波束 LEO 卫星同频组网场景，提出了基于决策性能评估的功率与带宽联合分配算法，提升了系统吞吐量与公平性。文献[20]提出了基于双重决斗深度 Q 网络（dueling double deep Q-network, D3QN）的 LEO 卫星跳波束资源管理方法，联合优化用户关联与功率分配策略，以最大化系统吞吐量并最小化时延。此外，文献[21]提出了基于多智能体深度确定性策略梯度（multi-agent deep deterministic policy gradient, MADDPG）的资源分配和缓存设计方案，通过联合优化用户关联、功率控制和缓存部署来提升系统能效。文献[22]提出了基于双延迟深度确定性策略梯度（twin delayed deep deterministic policy gradient, TD3）的方法，通过联合优化高轨卫星通信系统中的子信道与功率调配来提升系统性能。文献[23]提出了基于联邦深度强化学习的方法，通过卫星节点分簇与协同训练提高了 LEO 卫星通信网络的路由效率。尽管上述 DRL 方法为动态资源调配提供了自适应思路，但现有研究多面向地面蜂窝网络或星地网络覆盖场景，较少结合海洋通信的信道与业务特性，因此难以直接迁移至多波束 LEO 卫星海洋通信网络。

针对上述研究在海洋通信场景中适配性不足的问题，本文面向多波束 LEO 卫星海洋通信网络，提

出了一种联合波位划分的两阶段无线资源调配方法。第一阶段针对覆盖与吞吐量难以兼顾的问题，提出了基于P中心的波位划分算法，优化波位数量、中心位置和覆盖半径，并结合最小圆覆盖实现几何参数自适应更新，提升覆盖匹配性与资源利用效率。第二阶段在给定波位划分结果的基础上，将子信道分配与发送功率联合建模为MDP，并提出了基于多智能体双延迟深度确定性策略梯度（multi-agent twin delayed deep deterministic policy gradient, MATD3）的多智能体协同资源调配策略，以应对强耦合干扰环境下的复杂决策，提升收敛性能。

2　系统模型

2.1　网络模型

如图1所示，多波束LEO卫星海洋通信网络由M颗卫星构成，为海面K个用户提供服务。由于海洋用户分布稀疏且不均匀，固定覆盖结构难以同时兼顾覆盖匹配性与资源利用效率。为此，本文首先根据用户空间位置对待覆盖区域进行划分，形成多个波位。波位表示依据用户分布形成的目标覆盖区域，由中心位置和覆盖半径表征。随后，卫星利用相控阵天线生成具有可调指向和波束宽度的物理波束，对波位进行按需覆盖。为提高频谱利用效率，假设网络内各波束采用全频率复用，即所有波束共享整个系统带宽，因此不同波束之间会产生同频干扰。假设系统总带宽为B，将其划分为S个正交子信道，各子信道的带宽为ΔB，满足$B = S \times \Delta B$，定义子信道集合为$\mathcal{S} \triangleq \{ s|s = 1,2,\cdots,S \}$。假设波束内不同用户占用正交子信道进行上行传输，且每个子信道至多分配给一个用户。本文主要考虑海面用户到LEO卫星的上行通信链路，暂不考虑LEO卫星与地面站之间的通信链路。

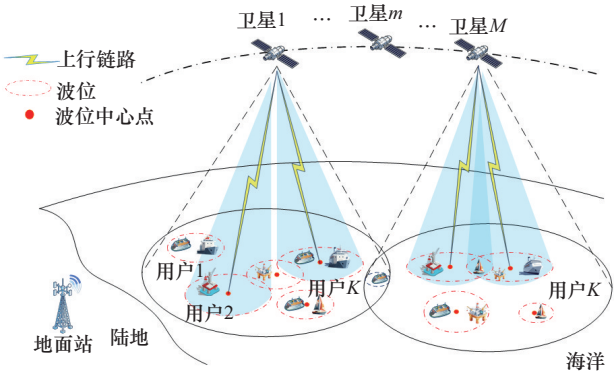


图1　系统模型

2.2　通信模型

基于上述网络模型，下面给出通信模型。对于海洋卫星通信链路，假设$h_{m,n,k,s}^{t}$表示在t时隙内，卫星m覆盖的波位n中，用户k在子信道s上传输数据的信道增益，星海链路的信道模型建模为[24]：

$$h_{m,n,k,s}^{t} = g_{m,n,k}^{t} g_{k}^{t} L_{m,n,k,s}^{t} L_{r}^{t} g_{m,k}^{t} \tag{1}$$

其中，$g_{m,n,k}^{t}$、g_{k}^{t}、$L_{m,n,k,s}^{t}$和L_{r}^{t}分别为卫星接收天线增益、用户发射天线增益、自由空间传播损耗和雨衰损耗。考虑到海洋卫星通信链路以视距分量为主，同时叠加由海面反射和周围环境引起的少量散射分量，因此将小尺度衰落建模为莱斯衰落模型[25]，$g_{m,k}^{t}$可以表示为：

$$g_{m,k}^{t} = \sqrt{\frac{K}{K+1}} \mathrm{e}^{\mathrm{j}\phi(t)} + \sqrt{\frac{1}{K+1}} w_{m,k}^{t} \tag{2}$$

其中，K为莱斯因子，$\phi(t) = 2\pi f_D t \cos\theta_0 + \phi_0$表示直射分量的相位，$w_{m,k}^{t} \sim \mathcal{CN}(0,1)$为零均值的复高斯随机过程，表示非视距多径叠加后的部分。需要说明的是，由于LEO卫星和海上用户的运行轨迹具有一定的确定性，其高移动性引起的多普勒频移可以通过最大似然多普勒估计、卡尔曼滤波和几何分析等方法预先估计和补偿[25-26]，因此本文将在未来工作中进一步考虑多普勒频移的影响。

假设波束为圆形波束，则卫星接收天线增益可以建模为[27]：

$$g_{m,n,k}^{t} = g_{m,n}^{\max} \left[\frac{J_1(u_{m,n,k}^{t})}{2u_{m,n,k}^{t}} + \frac{36 J_3(u_{m,n,k}^{t})}{(u_{m,n,k}^{t})^3} \right]^2 \tag{3}$$

其中，$g_{m,n}^{\max}$表示卫星m覆盖波位n的峰值增益，$g_{m,n}^{\max} = \eta \left(\dfrac{\pi D}{\theta_{m,n}^{t}} \right)^2$，$\eta$表示天线效率，$D$表示卫星相控阵天线的直径，$\theta_{m,n}^{t}$表示卫星$m$覆盖波位$n$的3 dB波束宽度对应的角度。假设卫星$m$覆盖波位$n$的半径为$r_{m,n}$，卫星高度为$h$，则$\theta_{m,n}^{t} = \arctan\left(\dfrac{r_{m,n}}{h} \right)$。$J_1(\cdot)$和$J_3(\cdot)$分别表示第一类和第三类贝塞尔函数，$u_{m,n,k}^{t}$建模为：

$$u_{m,n,k}^{t} = \frac{2.071\,23 \sin(\theta_{m,n,k}^{t})}{\sin(\theta_{m,n}^{t})} \tag{4}$$

其中，$\theta_{m,n,k}^{t}$表示卫星m到其覆盖波位n的中心方向与到用户k方向之间的角度，即离轴角。假设用户

k 的坐标为 $\boldsymbol{u}_k=(x_k,y_k,0)$，卫星 m 覆盖波位 n 的中心点坐标为 $\boldsymbol{\omega}_{m,n}=(a_{m,n},b_{m,n},0)$，则可以得到卫星 m 覆盖波位 n 的中心点与用户 k 之间的距离为 $d_{m,n,k}=\sqrt{(a_{m,n}-x_k)^2+(b_{m,n}-y_k)^2}$，离轴角为 $\theta^t_{m,n,k}=\arctan\left(\dfrac{d_{m,n,k}}{h}\right)$。此外，$L^t_{m,n,k,s}$ 表示时隙 t 内，波位 n 内用户 k 占用子信道 s 传输数据至卫星 m 时的自由空间传播损耗，可以建模为：

$$L^t_{m,n,k,s}=\left(\frac{c}{4\pi d_{m,k}f_s}\right)^2 \tag{5}$$

其中，c 为光速，$d_{m,k}$ 为卫星 m 与用户 k 之间的距离，f_s 为子信道 s 的载波频率。

考虑到用户分布稀疏且业务差异化，首先定义 $\mathcal{A}\triangleq\left\{\alpha^t_{m,n,k}\right\}\in\mathbb{R}^{M\times N\times K}$，$\mathcal{B}\triangleq\left\{\beta^t_{m,n,k,s}\right\}\in\mathbb{R}^{M\times N\times K\times S}$，$\mathcal{P}\triangleq\left\{p^t_{m,n,k,s}\right\}\in\mathbb{R}^{M\times N\times K\times S}$。具体地，定义二元变量 $\alpha^t_{m,n,k}$ 用来描述每个用户所属的波位，如果 $\alpha^t_{m,n,k}=1$，则表示用户 k 属于卫星 m 覆盖的波位 n。由于波束之间可能存在重叠，用户 k 可能被多个波束覆盖，假设每个用户最多只能与一个波位关联。定义二元变量 $\beta^t_{m,n,k,s}$ 表示在 t 时隙内，由卫星 m 覆盖的波位 n 内用户 k 被分配在子信道 s 进行上行传输，如果用户占用子信道，则 $\beta^t_{m,n,k,s}=1$。对应地，定义 $p^t_{m,n,k,s}$ 为 t 时隙内，由卫星 m 覆盖的波位 n 内用户 k 在子信道 s 上的上行发送功率。至此，波位 n 内用户占用子信道 s 与卫星 m 之间的上行接收信干噪比（signal to interference plus noise ratio, SINR）可以表示为：

$$\gamma^t_{m,n,k,s}=\frac{p^t_{m,n,k,s}h^t_{m,n,k,s}}{N_0\Delta B+I^t_{m,n,k,s}} \tag{6}$$

其中，N_0 为噪声功率密度，$I^t_{m,n,k,s}$ 为 t 时隙内，波位 n 内用户 k 占用子信道 s 向卫星 m 发送数据时受到的波束间干扰，可建模为：

$$I^t_{m,n,k,s}=\sum_{m'=1}^{M}\sum_{n'=1,n'\neq n}^{N}\sum_{k'=1,k'\neq k}^{K}\alpha^t_{m',n',k'}\beta^t_{m',n',k',s}p^t_{m',n',k',s}h^t_{m',n',k,s} \tag{7}$$

因此，对于卫星 m 覆盖波位 n 内的用户 k，在 t 时隙内占用子信道 s 时，可达传输速率可表示为：

$$R^t_{m,n,k,s}=\Delta B\mathrm{lb}\left(1+\gamma^t_{m,n,k,s}\right) \tag{8}$$

2.3 两阶段无线资源调配方法概述

为兼顾资源调配复杂度与资源利用效率，本文采用两阶段无线资源调配框架，在不同时间尺度上分别优化覆盖结构与无线资源调配。阶段 1 根据用户空间分布优化波位数量、中心位置和覆盖半径，并结合 P 中心算法与最小圆覆盖形成与用户分布相匹配的覆盖结构。阶段 2 在给定波位划分结果的基础上，对各波位内用户的子信道分配与上行发送功率进行动态联合优化。针对离散子信道决策与连续功率控制并存、波束间干扰强耦合且状态维度高的问题，本文将其建模为 MDP，并采用基于 MATD3 的多智能体协同在线学习方法进行求解。本文所提两阶段方案流程如图 2 所示。需要说明的是，本文波位划分用于确定覆盖结构和几何参数，属于长时间尺度决策，子信道与功率调配属于短时间尺度决策。因此，本文两阶段优化方法能够降低联合决策复杂度，同时兼顾海洋通信场景下覆盖与容量均衡的问题。基于单时隙（准）静态用户快照建模，所得波位划分结果适用于用户分布变化较缓的场景，当用户位置移动性较强时，可通过结合统计特征，对波位划分进行自适应周期性优化与更新。

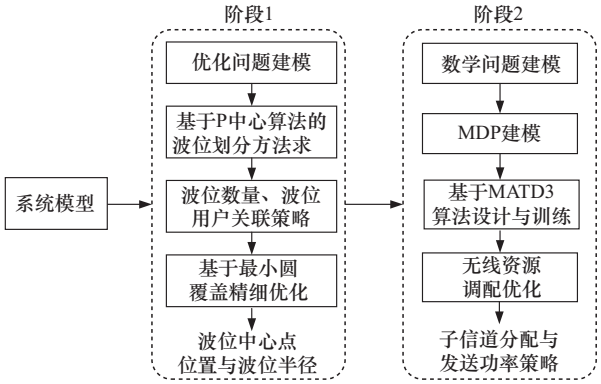


图2 两阶段方案流程

3 基于 P 中心的波位划分优化方法

本节提出了基于 P 中心的波位划分优化算法，联合优化波位数量、中心位置与覆盖半径等几何参数，构建合理的覆盖结构，确定用户与波位的关联关系。

3.1 优化问题建模

基于两阶段无线资源调配框架，在波位划分阶段假设子信道与功率调配策略根据历史经验生成，该阶段优化用户与波位关联关系为二元变量 $\mathcal{A}\triangleq\left\{\alpha^t_{m,n,k}\right\}\in\mathbb{R}^{M\times N\times K}$。考虑到海面用户所在区域通常由多个卫星波束联合覆盖，首先根据用户位置信息对待覆盖区域进行波位划分，随后优化卫星波

束按需覆盖对应波位。为解决覆盖与吞吐量难均衡的问题，在波位划分阶段，优化目标为最大化网络加权平均吞吐量（weighted average throughput, WAT）。因此，阶段1的波位划分优化问题建模为：

$$\begin{aligned}&\text{P1: } \max_{\mathcal{A}} \sum_{k=1}^{K} w_k R_k^t \\ &\text{s.t. C1: } \sum_{m=1}^{M}\sum_{n=1}^{N} \alpha_{m,n,k}^t \leqslant 1, \forall k \in \mathcal{K} \\ &\quad\ \text{C2: } \alpha_{m,n,k}^t \in \{0,1\}, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall k \in \mathcal{K} \end{aligned} \tag{9}$$

其中，w_k表示用户的业务权重，用于刻画不同用户的服务优先级，R_k^t表示t时隙内用户k的上行可达传输速率，可表示为

$$R_k^t = \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{s=1}^{S} \alpha_{m,n,k}^t \beta_{m,n,k,s}^t R_{m,n,k,s}^t \tag{10}$$

在优化问题P1中，约束C1表示在t时隙内，任意用户所在的波位最多关联到一颗卫星，约束C2表示二元变量。

3.2　基于P中心的波位划分算法

需要指出的是，波位数量的选取本质上是覆盖与容量的均衡问题。本文通过波位划分保证覆盖性能，以最大化网络WAT为目标，联合优化覆盖结构与吞吐量性能。当波位数量较小时，为覆盖全部用户需采用较大的波位半径，这会导致卫星3 dB增益角增大、接收增益降低。当波位数量增加时，各波位半径减小，覆盖匹配性和接收增益得到改善，但子信道复用率增加，波束间同频干扰也随之加剧，导致容量增加趋缓甚至下降。因此，最优波位数量通常对应覆盖与容量的平衡。进一步，考虑到最优波位数量与用户数量、空间分布和子信道数量等因素有关，本文采用搜索方法确定波位数量。除波位数量外，波位中心点位置与波位半径决定用户到波位中心的几何距离与离轴角，将影响波束增益与链路质量，从而影响用户最大可达速率。因此，在波位划分阶段优化覆盖结构以最大化网络WAT能够均衡网络覆盖与吞吐量。然而，式(10)中R_k^t与覆盖结构紧密相关且难以量化表达，本文将最大化网络WAT目标下的波位优化近似为P中心优化问题：在二维平面内给定包含K个点的集合$\mathcal{K} \triangleq \{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_K\}$，寻找$P$个中心$\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_P$，使用户到其所属波位中心的距离最小。该转化能够从几何层面抑制过大的波位半径，避免少数远端用户拉大覆盖范围导致整体波束增益下降，从而为提升吞吐量提供合理的覆盖结构。至此，结合对波位数量的搜索，本文采用最小覆盖圆确定波位中心坐标和覆盖半径，设计了基于P中心算法的波位划分方法，具体流程如算法1所示。

算法1　基于P中心的波位划分

输入　给定用户集合坐标$\mathcal{U}_m \triangleq \{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_K\}$

输出　波位数量N，波位划分方案

1) 初始化波位数量$N = 1$，优化目标初始值R_1
2) while $N \leqslant N_{\max}$ do
3)　　$N = N + 1$
4)　　随机选择N个用户坐标作为波位中心，将剩余用户分配到与其距离最近的波位集合$g_{N,m}$
5)　　利用P中心算法更新波位划分方案$\mathcal{G}_m \triangleq \{g_{m,1}, g_{m,2}, \cdots, g_{m,N}\}$，利用最小圆覆盖更新每个波位的中心点坐标$\boldsymbol{\omega}_{m,n}$和波位半径$r_{m,n}$
6)　　计算当前优化目标R_N
7)　　if $R_N < R_{N-1}$ then
8)　　　　break
9)　　end if
10) end while

P中心算法具体步骤如下：首先，初始化波位数量$N = 1$，并计算对应的优化目标初始值R_1；随后，进行迭代搜索，在每一轮迭代中，将当前波位数量更新为$N = N + 1$，并随机选取N个用户位置作为初始波位中心，其余用户依据距离最近准则匹配至相应波位；最后，利用P中心算法迭代调整各波位内的用户匹配问题，以最小化各波位覆盖半径中的最大值，从而更新波位划分方案，并在每次调整后通过最小圆覆盖迭代更新波位半径和波位中心点位置。上述过程不断迭代直至波位划分结果收敛，得到对应波位数量为N时的波位划分方案。随着N的增加，当目标值$R_N = \sum_{k=1}^{K} w_k R_k^t$出现下降时，即停止迭代，确定此时的波位数量、波位中心点坐标和波位半径。

4　基于波位划分的无线资源调配方法

本节在给定波位划分方案的基础上，设计了基于MATD3的无线资源调配算法。

4.1 数学问题建模

在任意 t 时隙内，给定波位划分方案，联合优化子信道分配与功率控制最大化，网络 WAT 的优化问题可以建模为：

$$\text{P2: } \max_{\mathcal{B},\mathcal{P}} \sum_{k=1}^{K} w_k R_k^t$$

$$\text{s.t. C3: } \beta_{m,n,k,s}^t \in \{0,1\}, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}$$

$$\text{C4: } 0 \leqslant p_{m,n,k,s}^t \leqslant P_{\max}, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall k \in \mathcal{K}, \forall s \in \mathcal{S}$$

$$\text{C5: } \sum_{s=1}^{S} \beta_{m,n,k,s}^t \leqslant 1, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall k \in \mathcal{K}$$

$$\text{C6: } \sum_{k=1}^{K} \alpha_{m,n,k}^t \beta_{m,n,k,s}^t \leqslant 1, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \forall s \in \mathcal{S}$$

$$\text{C7: } R_k^t \geqslant D_{Thr,k}, \forall k \in \mathcal{K} \tag{11}$$

其中，C3 表示子信道分配二元变量约束；C4 表示用户上行发送功率的最大值与最小值约束；C5 表示每个用户在任意 t 时隙内最多只能占用一个子信道；C6 表示波束内用户被分配正交子信道，即波束内每个子信道至多分配给一个用户；C7 表示用户上行最小传输速率约束，其中 $D_{Thr,k}$ 表示满足用户 k 传输需求的最小速率。优化问题 P2 属于混合整数非线性非凸优化问题，难以直接求解，而且依赖先验信息的传统优化方法难以适应复杂且时变的海洋通信环境。因此，本文将设计基于 MATD3 的无线资源调配算法求解该优化问题。

4.2 基于 MATD3 的无线资源调配算法

本节将无线资源调配问题建模为 MDP，提出了一种基于 MATD3 的优化问题求解方法来确定无线资源调配策略。在本文构建的多波束 LEO 卫星海洋通信网络中，每颗 LEO 卫星需根据时变信道状态和用户需求，实时调整子信道分配和用户上行发送功率。该优化问题同时涉及离散与连续动作，并且涉及多颗卫星间的资源竞争与干扰协调，因此具有高维、非凸和强耦合的特性。为实现多卫星协同决策，需要采用支持多智能体并行学习的 DRL 框架。本文将无线资源调配子问题建模为 $(\mathcal{S}_t, \mathcal{A}_t, \mathcal{R}_t)$ 构成的 MDP，其中 $\mathcal{S}_t$ 表示系统状态集，$\mathcal{A}_t$ 表示动作集，$\mathcal{R}_t$ 表示奖励集，具体如下。

（1）状态。令 $\mathcal{S}_t \triangleq \{s_1^t, s_2^t, \cdots, s_M^t\}$ 表示 t 时隙的全局状态，其中 s_m^t 表示在 t 时隙内卫星 m 观测到的局部状态，包括上一个时隙子信道的分配矢量、功率大小以及卫星与用户之间的信道增益，建模为：

$$s_m^t = \left(\boldsymbol{\beta}_{m,n,k}^{t-1}, \boldsymbol{p}_{m,n,s}^{t-1}, \boldsymbol{h}_{m,n,k}^t\right) \tag{12}$$

其中，$\boldsymbol{\beta}_{m,n,k}^{t-1}$ 表示在 $t-1$ 时隙内卫星 m 覆盖的波位 n 内用户 k 的子信道分配矢量，$\boldsymbol{p}_{m,n,s}^{t-1}$ 表示在 $t-1$ 时隙内卫星 m 覆盖的波位 n 内用户 k 的功率矢量，$\boldsymbol{h}_{m,n,k}^t$ 表示在 t 时隙内卫星 m 覆盖的波位 n 内用户 k 的信道增益矢量，分别表示为：

$$\boldsymbol{\beta}_{m,n,k}^{t-1} = \left(\beta_{m,n,k,1}^{t-1}, \beta_{m,n,k,2}^{t-1}, \cdots, \beta_{m,n,k,S}^{t-1}\right) \tag{13}$$

$$\boldsymbol{p}_{m,n,s}^{t-1} = \left(p_{m,n,1,s}^{t-1}, p_{m,n,2,s}^{t-1}, \cdots, p_{m,n,K,s}^{t-1}\right) \tag{14}$$

$$\boldsymbol{h}_{m,n,k}^t = \left(h_{m,n,k,1}^t, h_{m,n,k,2}^t, \cdots, h_{m,n,k,S}^t\right) \tag{15}$$

（2）动作。令 $\mathcal{A}_t \triangleq \{a_1^t, a_2^t, \cdots, a_M^t\}$ 表示在 t 时隙所有卫星智能体对应的全局动作，其中 a_m^t 表示在 t 时隙内卫星 m 的动作，包含其覆盖的波位 n 内用户 k 的子信道分配矢量和上行发送功率，建模为：

$$a_m^t = \left(\boldsymbol{\beta}_{m,n,k}^t, \boldsymbol{p}_{m,n,s}^t\right) \tag{16}$$

其中，$\boldsymbol{\beta}_{m,n,k}^t = \left(\beta_{m,n,k,1}^t, \beta_{m,n,k,2}^t, \cdots, \beta_{m,n,k,S}^t\right)$ 表示在 t 时隙内卫星 m 覆盖的波位 n 内用户 k 的子信道分配矢量，$\boldsymbol{p}_{m,n,s}^t = \left(p_{m,n,1,s}^t, p_{m,n,2,s}^t, \cdots, p_{m,n,K,s}^t\right)$ 表示在 t 时隙卫星 m 覆盖的波位 n 内用户 k 发送功率向量。

（3）奖励。令 $\mathcal{R}_t \triangleq \{r_1^t, r_2^t, \cdots, r_M^t\}$ 表示在 t 时隙内所有智能体对应的奖励集合，对于单个智能体，通过执行动作来获得相应奖励。为了实现协同效果，所有智能体均使用相同的奖励函数，为高效配置功率和频率资源，最大化系统加权平均吞吐量，本文将 t 时隙内卫星 m 的奖励函数建模为：

$$r_m^t = \sum_{k=1}^{K} w_k R_k^t \tag{17}$$

本文构建了基于 DRL 的优化框架，与传统依赖固定规则或先验信道信息的优化方法不同，DRL 能够根据环境变化持续更新策略。针对复杂的海洋通信环境，本文设计了基于 MATD3 的资源优化方法，考虑到多波束 LEO 卫星海洋通信网络中存在的干扰协调和资源竞争等挑战，MATD3 针对 DRL 中的一些关键问题进行了扩展。如图 3 所示，MATD3 采用双 Critic 网络结构，训练两个独立的 Critic 网络，并取其输出的最小值作为目标 Q 值，从而有效解决了 Q 值过估计问题，增强了训练的稳定性。此外，MATD3 引入延迟策略更新机制，确保 Actor 网

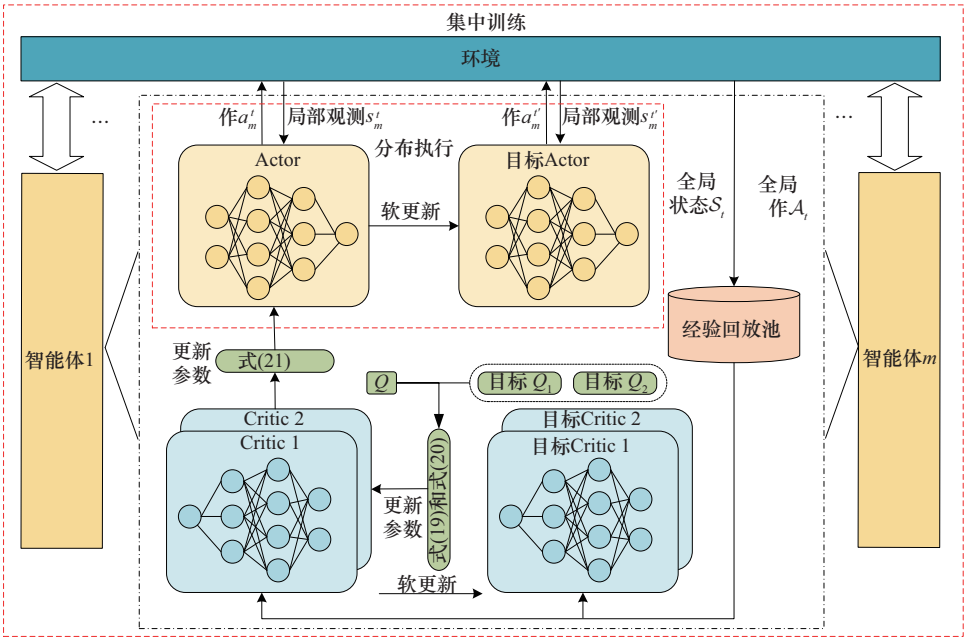


图3　MATD3框架

络的更新频率低于Critic网络，使策略优化更加平滑，减少了梯度振荡。

具体来说，将每颗LEO卫星视为一个智能体，各智能体在共享海洋通信环境中进行资源竞争与协同决策，旨在最大化网络WAT。首先，初始化所有卫星智能体的神经网络、经验回放池D和网络全局状态$\mathcal{S}_t$。对第m个智能体，基于局部观测s_m^t，该智能体的策略网络根据当前策略输出动作，并加入探索噪声，可表示为：

$$a_m^t = \pi_m\left(s_m^t;\theta_m^\mu\right) + N_t \tag{18}$$

其中，θ_m^μ为Actor网络参数，N_t为零均值高斯噪声，用于增强策略探索能力。a_m^t包含关联波束内用户的子信道与功率调配策略变量，所有智能体的动作组合成联合动作$\mathcal{A}_t$。联合动作对各波束内的子信道与功率进行调配，得到网络在t时隙内的全局奖励$\mathcal{R}_t$，系统状态转移至$\mathcal{S}_{t+1}$。随后，将四元组存储到经验回放池D中。当累积样本足够后，随机抽取小批量样本用于更新各智能体的评价和策略网络，从而逐步逼近最优的子信道与功率调配策略。

在MATD3算法中，每个智能体都配备两个独立的Critic网络，用于估计其在全局状态$\mathcal{S}_t$与联合动作$\mathcal{A}_t$下的动作价值，并为每个Critic网络构造对应的目标网络。从经验回放池中采样得到的第k条样本记为$\left(s_t^k, a_t^k, r_t^k, s_{t+1}^k\right)$，为了防止过拟合，首先利用Actor网络为下一时刻生成带平滑噪声的目标联合动作$\tilde{\mathcal{A}}_{t+1}^k$，并定义目标$Q$值为：

$$y_m^k = r_t^k + \gamma \min_{j=1,2} Q_i^{j'}\left(s_{t+1}^k, \tilde{\mathcal{A}}_{t+1}^k\right) \tag{19}$$

其中，$\gamma \in (0,1)$为折扣因子，$\tilde{\mathcal{A}}_{t+1}^k$由各智能体的目标Actor网络输出，并加入裁剪后的高斯噪声，用于实现目标策略的平滑。使用最小均方误差（mean square error, MSE）作为损失函数来更新预测Critic网络参数。对于第i个智能体的第j个Critic网络，其损失函数可表示为：

$$L\left(\theta_m^j\right) = E_{\left(S_t,A_t,R_t,S_{t+1}\right)\sim D}\left[\left(y_m^k - Q_m^j\left(s_t^k, a_t^k\right)\right)^2\right],\quad j = 1,2 \tag{20}$$

通过上述梯度下降步骤更新两个Critic网络参数，获得更准确的价值估计。由于目标Q值取自两个目标Q网络输出的较小者，上述损失函数能够有效缓解单一Critic网络带来的Q值系统性过估计问题，从而提升资源分配策略学习的稳定性。在Critic网络更新的基础上，Actor网络通过最大化Critic网络估计的Q值来更新参数。根据确定性策略梯度定理，第i个智能体的策略梯度可表示为：

$$\nabla_{\theta_m^\mu} J\left(\pi_m\right) = E_{\left(S_t,A_t\right)\sim D}\left[\nabla_{a_m^t} Q_i^1\left(s_t^k, a_1, \cdots, a_N\right)\nabla_{\theta_m^\mu}\pi_m\left(s_m^t\right)\right] \tag{21}$$

在训练过程中，沿着该梯度方向对Actor网络进行梯度上升更新，使Actor网络输出的资源分配动作在当前Critic网络估计下获得更高的预期回报。为进一步提高训练过程的数值稳定性，MATD3算法采用延迟更新策略，即每次采样后更新双Critic网络，而仅在Critic网络更新若干步后才更新一次Actor网络，从而确保策略在相对准确的Q值估计的基础上进行优化。同时，Actor与Critic网络的目标网络参数通过软更新的方式与主网络保持同步，即：

$$\theta_m^{j\prime} \leftarrow \tau\theta_m^j + (1-\tau)\theta_m^{j\prime} \tag{22}$$

$$\theta_m^{\mu\prime} \leftarrow \tau\theta_m^{\mu} + (1-\tau)\theta_m^{\mu\prime} \tag{23}$$

其中，$\tau \in (0,1)$为软更新系数，经过足够多回合的交互与迭代，当损失函数收敛且系统吞吐量趋于稳定时，即可得到相应的子信道与功率联合调配策略。本文提出的基于MATD3的无线资源调配算法如算法2所示。

算法2 基于MATD3的无线资源调配

输入 Actor网络、双Critic网络、经验回放池大小B、训练回合数E、时隙数T、延迟更新步长d、软更新系数等

输出 训练后的Actor网络

1) 初始化所有智能体的Actor网络、Critic网络参数、目标网络、经验回放池
2) for episodes = 1 to E do
3) 重置多波束LEO卫星通信环境，随机生成本回合的信道参数
4) 初始化系统状态，设置动作探索噪声
5) for step = 1 to T do
6) for agent = 1 to M do
7) 观测局部状态
8) 每个智能体根据当前策略选择动作并加入探索噪声 $a_m^t = \pi_m\left(s_m^t;\theta_m^{\mu}\right) + N_t$
9) 组合得到联合动作$\mathcal{A}_t$，并在环境中执行
10) 获得全局奖励$\mathcal{R}_t$以及下一时刻状态$\mathcal{S}_{t+1}$
11) 将四元组$\left(\mathcal{S}_t,\mathcal{A}_t,\mathcal{R}_t,\mathcal{S}_{t+1}\right)$存入经验回放池
12) if $|B| < B_b$ then
13) 跳过参数更新
14) else
15) 从经验回放池中随机采样大小为B_b的经验样本
16) 根据式(19)计算每个智能体目标Q值
17) 根据式(20)更新各智能体双Critic网络参数
18) if step mod d==0 then
19) 根据式(21)计算策略梯度并更新各智能体Actor网络参数
20) 根据式(22)和式(23)更新所有目标Critic网络与目标Actor网络参数
21) end if
22) end if
23) end for
24) end for
25) end for

5 仿真分析

为评估本文方法的性能，开展了仿真实验验证。仿真区域设定为1 200 km×1 200 km，用户在该区域内随机分布，仿真参数如表1所示。

表1 仿真参数

参数名称	参数值
卫星高度/km	500
载波频率/GHz	20
用户终端最大发送功率/W	5
卫星天线效率	0.65
子信道带宽/MHz	20
卫星天线直径/m	2
卫星数量/颗	2
用户数量/个	[10,50]
子信道数量/个	10
噪声功率谱密度/(dBm·Hz^{-1})	−174

为实现本文所提MATD3算法，实验采用Python 3.9和PyTorch 2.5.1，构建多个全连接神经网络，并选用ReLU和Softmax作为激活函数。每个神经网络包含2个隐藏层，每层包含128个神经元；折扣因子设为0.99，经验回放池容量设为100 000，其他参数如表2所示。仿真实验运行环境为云端GPU服务器，具体硬件配置为：CPU Intel Xeon (R) Platinum 8470Q（22核），内存110 GB，GPU

NVIDIA RTX 6000（96 GB显存）；NVIDIA驱动版本为580.95.05，CUDA版本为13.0。

表2　　神经网络模型参数

参数名称	参数值
训练总轮次/次	1 000
回合最大步数/步	100
训练优化器	Adam
优化器学习率	0.001
损失函数	MSE
激活函数	ReLU，Softmax
折扣因子	0.99
经验回放池容量	100 000

在仿真实验中，首先随机生成50个非均匀分布的海面用户。第一阶段基于P中心算法确定用户与波位的关联关系，并采用最小圆覆盖计算波位中心坐标和覆盖半径。第二阶段在此基础上构建多智能体DRL环境，结合用户位置和波位关联信息，计算链路增益与传输速率并输出资源调配动作。通过交互训练，MATD3算法收敛后可得到子信道与用户发送功率的调配策略。

图4展示了3种波位划分方案的可视化对比结果。结果表明，各方案在波位中心位置与覆盖范围上差异显著。其中，不规则布局与可变半径能够提高覆盖 的灵活性，使波位覆盖更贴合用户分布特征。

为确定波位数量，本文分析了不同波位划分方案下网络WAT性能随波位数量变化的趋势，如图5所示。为避免波位划分与子信道、功率分配的耦合优化带来高复杂度，对比实验简化资源调配策略，即各用户上行功率相同，同一波位内用户占用不同子信道。因此，WAT差异主要反映波位划分在覆盖结构、用户关联与干扰抑制方面的影响。由图5可知，当波位数量从2个增至8个时，系统WAT总体上升，表明增加波位数量有助于提升WAT；当波位数量增至9个时，部分方案WAT趋缓甚至下降，说明波位数量过多可能加剧同频干扰并抵消增益。综合覆盖匹配收益与同频干扰代价的权衡，并结合图5所示WAT变化趋势，本文在当前仿真参数设置下取波位数量为8个。

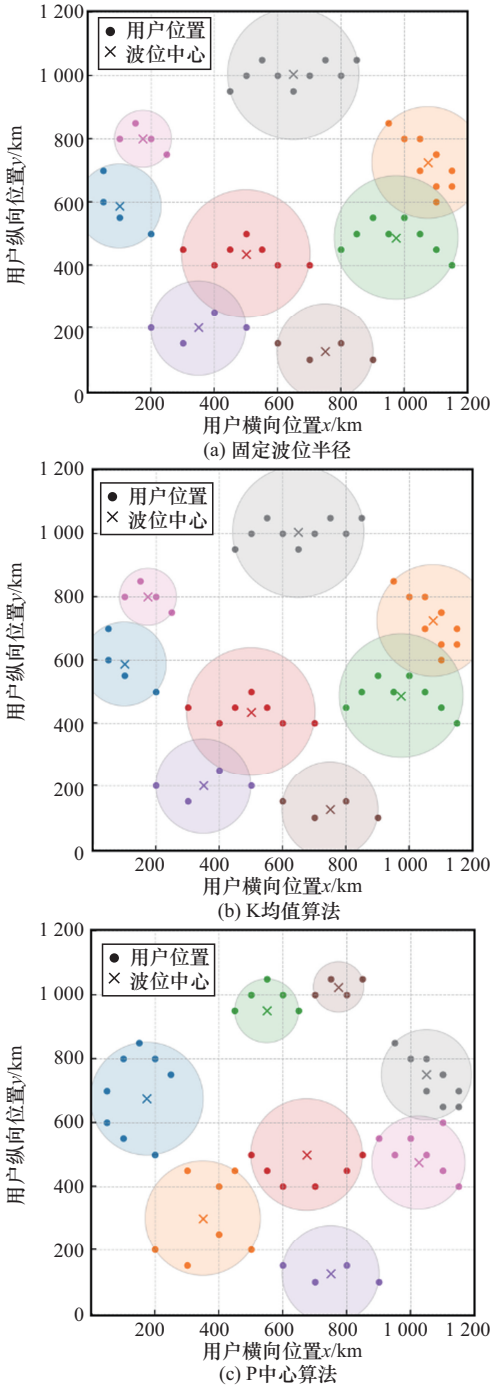


图4　不同波位划分方案

图6展示了在3种波位划分方案的基础上，本文设计的基于MATD3的无线资源调配算法的训练过程与收敛性能。具体地，为了分析波位划分带来的性能增益，分别对比了P中心算法、固定波位半径与K均值算法波位划分3种方案，观察不同波位划分方案对系统吞吐量的影响。从图6可以看出，3种方案的奖励均随训练轮次增加而逐渐提升，并趋于稳定，

说明在各方案对应的波位划分基础上，MATD3 算法均能较好地学习并收敛。3 种方案对应的系统吞吐量由高到低分别为 1 214 Mbit/s、1 169 Mbit/s 和 1 101 Mbit/s。其中，P 中心算法波位划分方案获得最高的系统吞吐量，相较于固定波位半径与 K 均值算法波位划分方案，网络 WAT 分别提升约 10.26% 与 3.85%。原因在于 P 中心算法通过联合优化波位中心位置与覆盖半径，能够形成匹配用户分布的覆盖结构，在一定程度上降低了波束间干扰，进而能够提升网络 WAT；固定波位半径方案的波位位置与覆盖结构预先设定，难以适应用户非均匀分布，因而性能受限；K 均值算法方案依托聚类具有一定自适应能力，其总体表现介于两者之间。综上分析，基于 P 中心算法的波位划分方案能够适配海洋通信场景下的用户分布特性，有助于在覆盖与网络容量之间取得均衡。

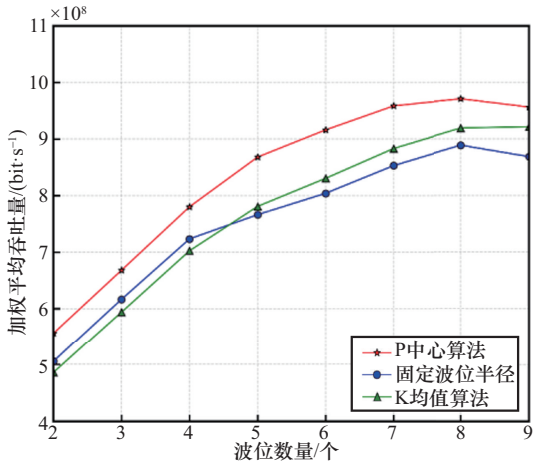


图 5　加权平均吞吐量随波位数量变化

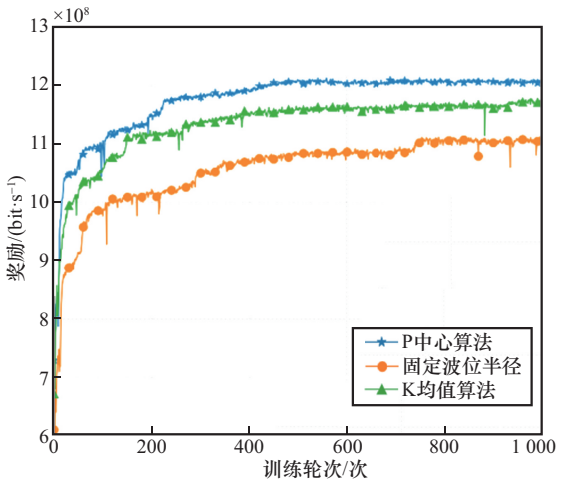


图 6　不同波位划分方案下 MATD3 优化的系统吞吐量收敛曲线

图 7 给出了 MATD3 算法在不同超参数下的收敛性能。图 7(a)展示了 MATD3 算法在 3 种学习率设置下的训练奖励变化趋势。可以观察到，随着学习率增大，系统奖励整体水平逐步提高，且收敛速度加快。当 lr=0.001 时，奖励在训练初期快速上升，在 200 次训练轮次内稳定收敛。图 7(b)给出了不同批量大小设置下 MATD3 算法训练奖励的变化曲线，可以看出，当 bs=256 时算法取得最高奖励，且整体收敛过程更稳定。综合图 7(a)与图 7(b)的对比结果，本文后续实验采用 lr=0.001、bs=256 作为默认超参数配置，以获得更优的奖励性能与训练稳定性。图 7(c)展示了在不同学习步长 d 设置下的训练曲线与收敛性能，其中 d 取值越大代表延迟更新间隔越大，$d = 1$ 时对应取消延迟更新策略，即策略与价值网络几乎同步更新，此时曲线在训练前期上升较快而后收敛。当引入延迟更新并增大学习步长至 $d = 4$ 时，训练前期表现更平滑，且最终收敛奖励值最高，说明适当的延迟更新能够在一定程度上降低价值估计偏差对策略更新的影响，使策略网络在更可靠的价值评估基础上迭代，从而提高收敛稳定性与最终性能。进一步，将学习步长增大至 $d = 8$ 后，曲线最终仍可收敛，但前期波动较大，最终收敛奖励值更低。这表明延迟更新步长过大将降低策略网络的更新频率，使策略改进滞后于环境反馈与价值函数变化，导致学习效率下降和最终性能受限。图 7(d)比较了 MATD3 算法中目标策略平滑的噪声强度与裁剪阈值对训练奖励的影响，其中，σ 表示在计算 TD3 目标 Q 值时，对目标动作添加的零均值高斯噪声标准差；clip 表示将该噪声裁剪到区间 $[-\text{clip}, \text{clip}]$，以限制动作扰动幅度。可以看到，3 组参数下奖励均随训练轮次上升并趋于稳定，说明算法在不同噪声设置下均能收敛。进一步对比可知，σ=0.10 和 clip=0.20 在约 200 轮后优势逐渐明显，且最终稳定奖励最高。该趋势表明，在该场景中适当增大目标平滑噪声及其裁剪范围可增强平滑效果与有效扰动，减弱价值估计误差对策略更新的影响，从而取得更优的收敛结果。

图 8 给出了 MATD3 算法与两种基准算法 MADDPG 和多智能体深度 Q 网络（multi-agent deep Q-network, MADQN）在训练过程中系统奖励随训练轮次增加的变化曲线。结果表明，3 种算法

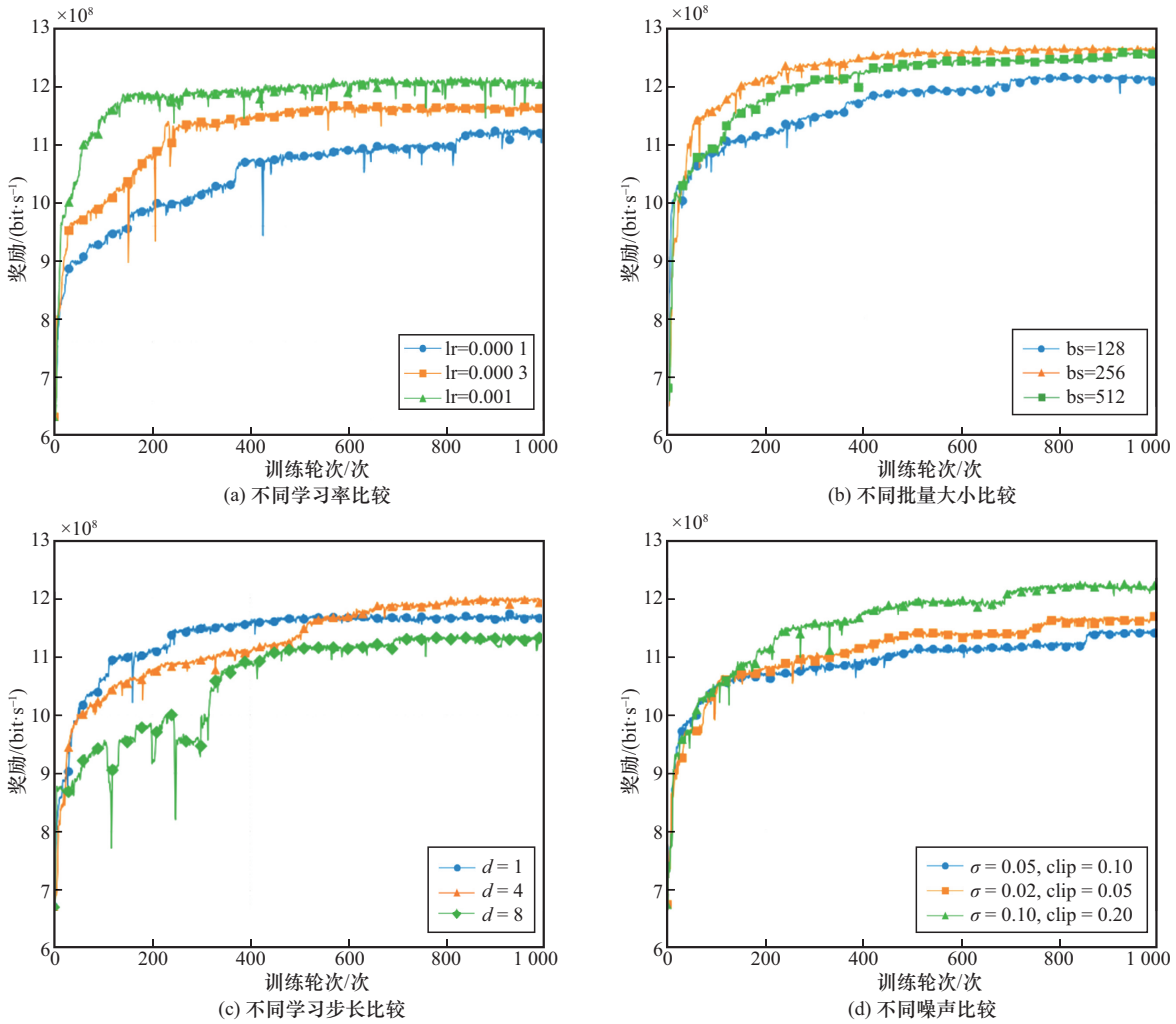


图 7　MATD3 算法在不同超参数下的收敛性能

的奖励均随训练轮次增加而上升并最终稳定，但在收敛速度、稳定性及最终收敛水平方面存在差异，其中，MATD3 算法训练曲线更平滑、收敛更快，且最终吞吐量最高，收敛后达到 1 268 Mbit/s；MADDPG 与 MADQN 算法收敛相对缓慢，最终分别稳定在 1 201 Mbit/s 和 1 052 Mbit/s。由此可得，MATD3 算法相较于 MADDPG 与 MADQN 算法，吞吐量分别提升 5.58% 和 20.53%。其原因在于 MATD3 算法采用双 Critic 网络结构与延迟策略更新，可降低价值过估计并提升训练稳定性，从而获得更优的收敛策略。相比之下，MADDPG 算法缺少上述抑制价值过估计与稳定训练的机制，因而在收敛速度与最终性能上受限。同时 MADQN 算法需要对连续动作进行离散化近似，易造成动作表达受限与价值估计偏差，导致训练波动增大、收敛性能较差。

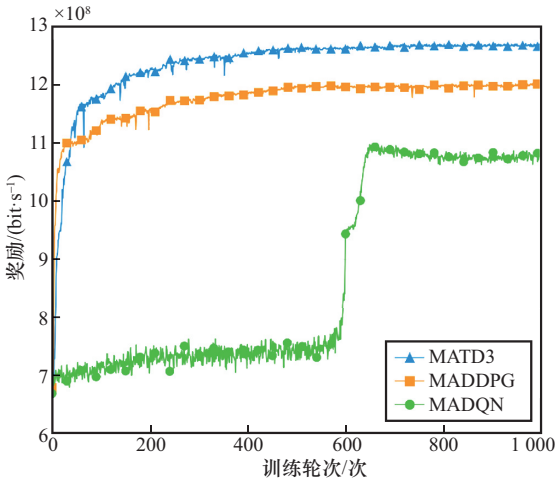


图 8　不同算法的奖励比较

图 9 对比了不同算法在用户数量变化条件下的网络 WAT。结果表明，当用户数量由 10 个增至 30 个时，3 种算法的 WAT 均随用户数量增加而上升，说

明在中等负载范围内用户数量增多能够提升资源利用率并带来更高的 WAT。当用户数量进一步增至 40 个和 50 个时，网络 WAT 出现不同程度回落，表明高负载条件下干扰增强与资源竞争加剧，系统资源利用效率受到限制。在各用户规模下，MATD3 算法始终获得最高的 WAT，且在用户数量继续增加时的下降幅度更小，体现出其在拥塞与强干扰场景中的资源调配鲁棒性与自适应能力。相较之下，MADDPG 与 MADQN 算法的 WAT 值整体偏低，且用户数量增大时下降更为明显，反映其在多用户复杂环境下的策略优化效果与稳定性相对不足。综上，MATD3 算法在复杂的海洋通信环境与不同用户数量下均能够有效地优化无线资源调配策略，提升网络 WAT 和资源利用效率。

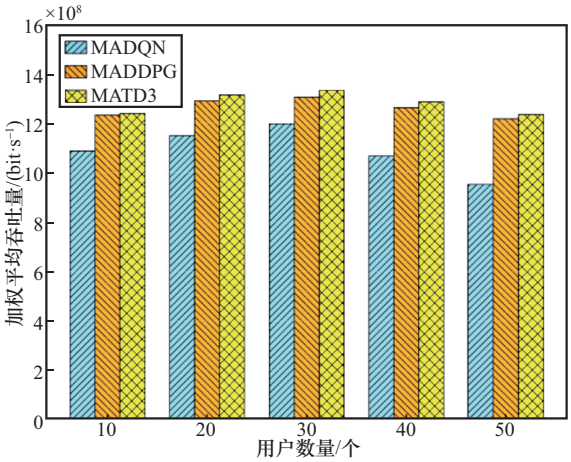


图 9 不同用户数量对加权平均吞吐量的影响

6 结束语

本文针对多波束 LEO 卫星海洋通信网络覆盖与吞吐量难均衡、多波束间干扰强耦合导致无线资源调配策略复杂度高和资源利用率低的挑战，提出了一种联合波位划分的两阶段无线资源调配方法。该方法第一阶段基于 P 中心算法与最小圆覆盖对波位结构进行优化以保证覆盖性能，第二阶段基于 MATD3 算法对多波束耦合下的用户子信道与功率调配进行联合优化，以提升网络容量。仿真结果表明，本文方法在网络加权平均吞吐量和收敛稳定性方面均优于基准方案，验证了其有效性。本文主要聚焦于海面用户到 LEO 卫星节点的上行传输资源调配问题，未来将进一步研究空天地海一体化场景，考虑接入链路与回程传输耦合等问题，重点研究多维资源协同调配与智能优化方法。

参考文献：

[1] 朱剑锋, 孙耀华, 冯昕澳, 等. 星地融合网络中的干扰管理与频谱共享: 研究进展、关键技术及未来挑战[J]. 电信科学, 2025, 41(3): 108-127.
Zhu J F, Sun Y H, Feng X A, et al. Interference management and spectrum sharing in integrated satellite-terrestrial networks: progress, key technologies and future directions[J]. Telecommunications Science, 2025, 41(3): 108-127.

[2] 彭木根, 张世杰, 许宏涛, 等. 低轨卫星通信遥感融合: 架构、技术与试验[J]. 电信科学, 2022, 38(1): 13-24.
Peng M G, Zhang S J, Xu H T, et al. Communication and remote sensing integrated LEO satellites: architecture, technologies and experiment[J]. Telecommunications Science, 2022, 38(1): 13-24.

[3] Rafiq G, Bose P, Orten P. 5G cellular communication for maritime applications[J]. IEEE Access, 2025, 13: 109451-109472.

[4] Shang Z Q, Zhang X, Li X H. Maritime communication networks: a survey on architecture, key technologies, and challenges[J]. Computer Communications, 2025, 241: 108255.

[5] 吴佳丽, 刘向南, 孙春蕾, 等. 海洋移动通信网络无线资源管控[J]. 移动通信, 2024, 48(11): 77-85.
Wu J L, Liu X N, Sun C L, et al. Wireless resource management for marine mobile communication networks[J]. Mobile Communications, 2024, 48(11): 77-85.

[6] Yao X, Gao R F, Hu Y D, et al. Online learning-based beam allocation for maritime communications with fairness constraints[C]//Proceedings of the 2023 IEEE 23rd International Conference on Communication Technology (ICCT). Piscataway: IEEE Press, 2023: 1030-1035.

[7] Sun Z B, Wang C, Zhang C, et al. User position-driven resource allocation for beam hopping in low earth orbit satellite[C]//Proceedings of the 2024 5th International Conference on Electronic Engineering and Signal Processing (EESP). Piscataway: IEEE Press, 2024: 15-20.

[8] Zhang R, Sun M C, Pan C S. Micro-nano satellite resource allocation algorithm based on chaos-filled PSO[C]//Proceedings of the 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT). Piscataway: IEEE Press, 2021: 204-208.

[9] Tang J Y, Bian D M, Li G X, et al. Resource allocation for LEO beam-hopping satellites in a spectrum sharing scenario[J]. IEEE Access, 2021, 9: 56468-56478.

[10] Takahashi M, Kawamoto Y, Kato N, et al. DBF-based fusion control of transmit power and beam directivity for flexible resource allocation in HTS communication system toward B5G[J]. IEEE Transactions on Wireless Communications, 2022, 21(1): 95-105.

[11] Zhang T, Zhang L X, Shi D Y. Resource allocation in beam hopping communication system[C]//Proceedings of the 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). Piscataway: IEEE Press, 2018: 1-5.

[12] Liu B K, Jiang C X, Kuang L L, et al. Joint user grouping and beamwidth optimization for satellite multicast with phased array antennas[C]//Proceedings of the GLOBECOM 2020 - 2020 IEEE Global Communi-

cations Conference. Piscataway: IEEE Press, 2020: 1-6.

[13] Shen B W, Goh S Q, Lam K Y, et al. Efficient and secure federated dynamic spectrum access for LEO satellite Internet-of-things[C]//Proceedings of the 2024 IEEE 24th International Conference on Communication Technology (ICCT). Piscataway: IEEE Press, 2024: 366-371.

[14] 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.

[15] Hu X, Wang Y, Liu Z J, et al. Dynamic power allocation in high throughput satellite communications: a two-stage advanced heuristic learning approach[J]. IEEE Transactions on Vehicular Technology, 2023, 72(3): 3502-3516.

[16] Hu X, Liao X L, Liu Z J, et al. Multi-agent deep reinforcement learning-based flexible satellite payload for mobile terminals[J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 9849-9865.

[17] Li Y J, Yang D W, Yang H W, et al. Joint frequency assignment and power allocation based on multi-agent deep reinforcement learning for multi-beam satellite systems[C]//Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring). Piscataway: IEEE Press, 2023: 1-6.

[18] 许国良, 谭峰, 冉泳屹, 等. 面向多波束卫星系统的波束跳变与覆盖控制联合优化算法[J]. 通信学报, 2023, 44(4): 78-86.
Xu G L, Tan F, Ran Y Y, et al. Joint beam hopping and coverage control optimization algorithm for multibeam satellite system[J]. Journal on Communications, 2023, 44(4): 78-86.

[19] 王朝炜, 庞明亮, 王粟, 等. 基于决策性能评估的多波束低地球轨道卫星网络资源分配算法[J]. 通信学报, 2024, 45(7): 37-47.
Wang C W, Pang M L, Wang S, et al. Resource allocation algorithm for multi-beam LEO satellite based on decision performance evaluation[J]. Journal on Communications, 2024, 45(7): 37-47.

[20] 刘文骏, 周家恩, 王丁, 等. 基于D3QN的低轨卫星跳波束资源管理方法[J]. 无线电工程, 2026, 56(1): 110-120.
Liu W J, Zhou J E, Wang D, et al. Beam hopping resource management for LEO satellite based on D3QN[J]. Radio Engineering, 2026, 56(1): 110-120.

[21] 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.

[22] 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.

[23] 李学华, 廖海龙, 张贤, 等. 面向低轨卫星通信网络的联邦深度强化学习智能路由方法[J]. 电子与信息学报, 2025, 47(8): 2652-2664.
Li X H, Liao H L, Zhang X, et al. Federated deep reinforcement learning-based intelligent routing design for LEO satellite networks[J]. Journal of Electronics & Information Technology, 2025, 47(8): 2652-2664.

[24] 柴蓉, 刘磊, 梁承超, 等. 面向用户多样化业务需求的多波束卫星系统动态资源分配算法[J]. 电子学报, 2024, 52(7): 2438-2448.
Chai R, Liu L, Liang C C, et al. Diverse user service requirement-oriented dynamic resource allocation algorithm for multi-beam satellite systems[J]. Acta Electronica Sinica, 2024, 52(7): 2438-2448.

[25] Zhou Z H, Chen X M, Ying M, et al. Unified design of space-air-ground-sea integrated maritime communications[J]. IEEE Transactions on Communications, 2025, 73(12): 13441-13455.

[26] Xiong K W, Chen X M, Ying M. On the performance of integrated satellite-terrestrial maritime communications[J]. IEEE Internet of Things Journal, 2025, 12(17): 36183-36196.

[27] Tang J Y, Bian D M, Li G X, et al. Optimization method of dynamic beam position for LEO beam-hopping satellite communication systems[J]. IEEE Access, 2021, 9: 57578-57588.

[作者简介]



张贤（1992– ），男，博士，北京信息科技大学讲师、硕士生导师，主要研究方向为空天通信、星地融合网络、雾化协同通信理论与方法等。



王东东（1998– ），男，北京信息科技大学硕士生，主要研究方向为卫星通信、强化学习、资源管理等。



周政（1990– ），男，博士，宁波大学讲师、硕士生导师，主要研究方向为低轨卫星通信、资源调配、无人机通信等。



李学华（1977– ），女，博士，北京信息科技大学教授、博士生导师，主要研究方向为智能计算与无线通信等。