

空地协同 UDN 中基于 Graph-DRL 的频谱资源分配

孟芸¹, 魏帆², 刘鑫一², 王威², 柴佳辉²

(1. 长安大学电子与控制工程学院, 陕西 西安 710018; 2. 长安大学信息工程学院, 陕西 西安 710018)

摘要: 为了支持各类移动用户实时数据传输和稳定连接, 采用空地协同超密集网络 (UDN) 来提升网络服务能力。在这种新型网络结构中, 空中基站能够更好地支持用户移动性, 而地面 UDN 能够实现更充分的频谱复用。然而, 差异化的信道条件与覆盖范围使得干扰与频谱分配问题面临着新的挑战。因此, 面向新型的空地协同 UDN, 针对该网络模型下移动场景中的干扰协调与频谱效率平衡问题, 提出基于图神经网络 (GNN) 的深度强化学习 (DRL) 算法 Graph-DRL, 动态调度频谱资源并提高算法的实时性。首先, 利用 GNN 构建基站间邻小区的用户移动相关性模型, 通过共享本小区的需求信息来加强合作, 确定带宽复用度; 其次, 设计基于价值的学习网络, 以网络中的平均吞吐量为优化目标, 考虑用户移动性、用户密度和频谱复用度等因素, 利用训练好的 Q 值网络进行资源分配。仿真结果表明, 所提出的算法相较于对比算法显著提高了高移动性用户和小区边缘用户的传输速率。

关键词: 无线通信; 空地协同网络超密集网络; 频谱资源分配; 干扰协调; 基于图神经网络的深度强化学习

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070270

Graph-DRL-Based Spectrum Resource Allocation in Air-Ground Cooperative UDN

MENG Yun¹, WEI Fan², LIU Xinyi², WANG Wei², CHAI Jiahui²

(1. School of Electronics and Control Engineering, Chang'an University, Xi'an 710018, Shaanxi, China;

2. School of Information Engineering, Chang'an University, Xi'an 710018, Shaanxi, China)

【Abstract】 To support real-time data transmission and maintain stable connectivity for diverse mobile users, an air-ground cooperative Ultra-Dense Network (UDN) is adopted to enhance network service capabilities. In this novel network architecture, airborne base stations can better support user mobility, while terrestrial UDNs enable more efficient spectrum reuse. However, differentiated channel conditions and coverage disparities introduce new challenges in interference and spectrum allocation. Therefore, this paper focuses on a novel air-ground cooperative UDN and addresses the challenges of interference coordination and spectrum efficiency balancing in mobile scenarios within this network model. A Graph Neural Network (GNN)-based Deep Reinforcement Learning (DRL) algorithm, Graph-DRL, is proposed to dynamically schedule spectrum resources and improve the real-time performance of the algorithm. First, a GNN is utilized to model the user movement correlations among neighboring base station cells. By sharing the demand information of the local cell, cooperation is enhanced to determine the bandwidth reuse factor. Second, a value-based learning network is designed, with the average throughput of the network as the optimization objective. Considering factors such as user mobility, user density, and spectrum reuse degree, a trained Q value network is employed for resource allocation. Simulation results demonstrate that, compared to benchmark algorithms, the proposed algorithm significantly improves the transmission rates for highly mobile users and cell-edge users.

【Key words】 wireless communication; air-ground cooperative Ultra-Dense Network (UDN); spectrum resource allocation; interference coordination; Deep Reinforcement Learning (DRL) based on Graph Neural Network (GNN)

0 引言

随着第五代 (5G) 无线通信网络进入商用部署阶段, 超第五代 (B5G)/第六代 (6G) 无线通信网络也得到了广泛的研究。在即将到来的大规模随机部署节点的无线网络中, 用户数目呈海量式增加, 并且对网络服务的需求也在急剧增加, 因此推广部署地

面超密集网络 (UDN), 通过缩小网络覆盖范围, 运行有效的频谱资源分配算法来提高频谱利用率和减少相互干扰, 以提升网络服务能力^[1-2]。但其中高移动性用户在地面 UDN 中存在切换频繁甚至中断的问题。此时, 无人机空中基站因其较大的覆盖范围, 尤其是在临时大规模活动网络需求激增的情况下, 能有效减少移动用户在基站之间的频繁切换次数,

基金项目: 国家重点研发计划 (2020YFB1807001); 装备预研教育部联合基金 (8091B032226); 陕西省重点研发计划 (2023-YBGY-212)。

作者简介: 孟芸, 女, 副教授, 主研方向为通信网络资源分配; 魏帆, 硕士研究生; 刘鑫一 (通信作者), 王威, 教授; 柴佳辉, 博士研究生。

收稿日期: 2024-08-19

修回日期: 2024-12-02

E-mail: liuxinyi@chd.edu.cn

提高连接的稳定性和用户体验^[3-4]。然而,由于这种新型网络结构中空中基站与地面基站在信道条件、移动性支持与频谱复用能力上具有差异,因此干扰与频谱分配问题面临着新的挑战。

针对传统静态 UDN 中的带宽资源分配问题,现有研究者已经提出了一系列分配算法。WEI 等^[5]提出了一种两阶段资源管理方案,首先使用贪婪搜索算法在每个集群中执行主资源分配,再利用补充分配算法来分配剩余子信道,确定小小区的实际密集部署方案。研究人员在资源分配过程中动态调整初始分簇结果以适应网络拓扑^[6]变化,通过有效利用用户剩余资源和微调资源块,提高用户业务速率。研究人员针对 D2D 网络提出了一种基于图着色理论的频谱资源^[7]分配算法。该算法提升了网络负载能力和吞吐量并且增强 D2D 用户的公平性。YAO 等^[8]在加权干扰图中对小区进行分簇,子信道优先分配给用户速率较大的簇。文献^[9-12]研究了 UDN 中的联合资源分配。ZHANG 等^[9]针对全双工 UDN 提出了一种低复杂度的启发式方案,采用分时松弛方法解决小区内子信道分配问题,降低了计算复杂度。HABIBINEJAD 等^[10]针对异构网络中飞基站层的子信道分配和功率控制联合问题,提出了分布式算法,最大限度地提高了全双工模式下飞蜂窝的总吞吐量。文献^[11-12]考虑了所有类型的干扰和用户的质量服务(QoS)要求,WANG 等^[12]利用组合拍卖机制来优化频谱分配子问题。文献^[13]利用非合作博弈论进行具体的 femto 层频谱分配以减小干扰。文献^[14]针对用户提出了一种基于图论的低复杂度次优联合干扰对齐和子信道分配干扰管理方案,旨在最大化 UDN 中具有不同自由度需求的满意用户数。ZHANG 等^[15]研究了 UDN 中基于部分簇干扰对齐和簇合并算法的子信道分配。在保证每个用户数据流需求的同时,最大化所有用户的可达数据流。GAO 等^[4]基于流量分布进行子信道分配,以提高系统吞吐量,同时降低小小区之间的干扰。在真实的通信场景下,由于用户的移动性,网络中用户的分布状态会发生变化,在这种情况下,单一的固定接入点可能不适合实时用户分布^[16]。上述已有研究工作均面向地面 UDN 进行,而在空地协同 UDN 中,空地信道与地面信道共存,空基与 UDN 对移动性支持能力存在差异,因此需要在频谱分配中考虑信道以及用户移动性差异。此外,上述研究工作中的传统算法博弈论、小区分簇结合启发式资源分配算法等,由于算法受时间复杂度的限制,因此不适用于考虑用户移动性的动态场

景下实时优化网络中的资源分配。空地协同 UDN 中的频谱分配需要更加高效的算法,以适应动态场景下的资源管理需求。

近些年,深度学习也开始应用于 UDN 中的资源分配。研究人员考虑异构 UDN 中的频谱与功率资源联合控制能效以及用户 QoS 的联合优化问题,提出基于深度强化学习(DRL)^[17]框架的资源分配和功率控制联合优化算法。HU 等^[18]研究了多用户系统中下行链路子信道分配和功率分配的联合问题,与传统的单智能体算法相比,提出了基于多智能体 DRL(MADRL)的算法,采用集中训练和分布式执行模式来解决这个非凸的动态优化问题。文献^[16-17]均结合了分簇思想和 MADRL。WANG 等^[19]规定每个簇内的基站共同为用户提供服务,将问题转化为多智能体 MDP,实现了中心小区的动态分簇、资源分配和双工模式选择。SHARMA 等^[20]采用 Q 值近似解决了超密集环境的主要问题,即大状态动作空间。研究人员基于 MADRL^[21],结合图神经网络(GNN)来加强基站之间的合作,以捕捉在空间上波动的服务需求模式,从而实现网络切片资源的智能管理。YANG 等^[22]利用分布式协调 MADRL 算法,以实现双层异构无线网络的联合设备关联、频谱分配和功率分配策略,在满足设备 QoS 要求的同时最大化长期网络容量。深度学习的引入集中于向用户分配资源,因此考虑将深度学习引入频谱资源分配中同层小区的干扰协调问题中,实现小区间实时协作与资源分配以适应动态的网络场景。

面向移动场景下空地协同的 UDN,考虑减少因用户的移动性导致的切换次数,并且在动态场景下实现快速的资源分配,本文提出一个协调干扰与提升频谱效率的频谱分配策略,主要研究内容如下:

1)针对每个时隙中的网络场景,根据速度阈值将用户划分为空中基站用户与地面基站用户,再利用分簇算法将地面小区分成不同的小区簇,从而按照簇数决定带宽的复用度。

2)按照该时隙小小区的实际用户负载向该小区分配一定比例的带宽数。以减小网络干扰为目标来确定具体的带宽分配策略,本文提出一种基于深度强化学习的资源分配算法,同时结合了图神经网络提高基站之间的空间协作。

3)基于给定的用户移动模型,针对动态的空地协同 UDN 场景来训练价值网络,获得带宽资源的分配策略,在多个随机的动态场景中验证了该分配策略相较于对比算法提高了网络中用户的平均吞吐量。

1 系统模型与问题定义

1.1 系统模型

空地协同 UDN 场景示意图如图 1 所示。本文定义了 1 个无线网络场景,该网络场景由若干个不均匀分布的地面小基站、空中基站以及随机移动的用户构成。每个地面基站覆盖范围是 1 个小区,场景中移动性较高的用户均由空中基站服务,以减少该类用户在地面基站的频繁切换次数。本研究的主要目的是根据每个基站的不同负载需求,基于小区簇实时优化每个基站的带宽分配策略。场景中 $j=\{1,2,\dots,J\}$ 表示微基站的集合, $u=\{U\}$ 表示空中基站。 t 时刻每个基站分配到的子带宽数用 $n^t =$

$\{N_1^t, N_2^t, \dots, N_J^t\}$ 表示,每个子带宽的粒度为 Δ , 用户集合用 $i=\{1,2,\dots,I\}$ 表示。 $\mathbf{E}_{I \times J}^t$ 表示 t 时刻用户与各个地面基站以及空中基站的关联系数矩阵,元素 e_{ij}^t 为 1,表示用户 i 与基站 j 在 t 时刻关联,在空地协同场景下其余未关联到地面基站的用户均由空中基站服务。式(1)表示用户与地面基站之间的关联策略。

$$e_{ij}^t = \begin{cases} 1, & \text{if } v_i \leq v_{\text{threshold}} \text{ and } d_{ij}^t \leq d_{\text{th}} \\ 0, & \text{else} \end{cases} \quad (1)$$

式中: v_i 为用户 i 的移动速度; $v_{\text{threshold}}$ 为划分地面基站用户与空中基站无人机用户的阈值速度; d_{ij}^t 为用户 i 与基站 j 在 t 时刻之间的距离; d_{th} 为基站的最大覆盖范围。

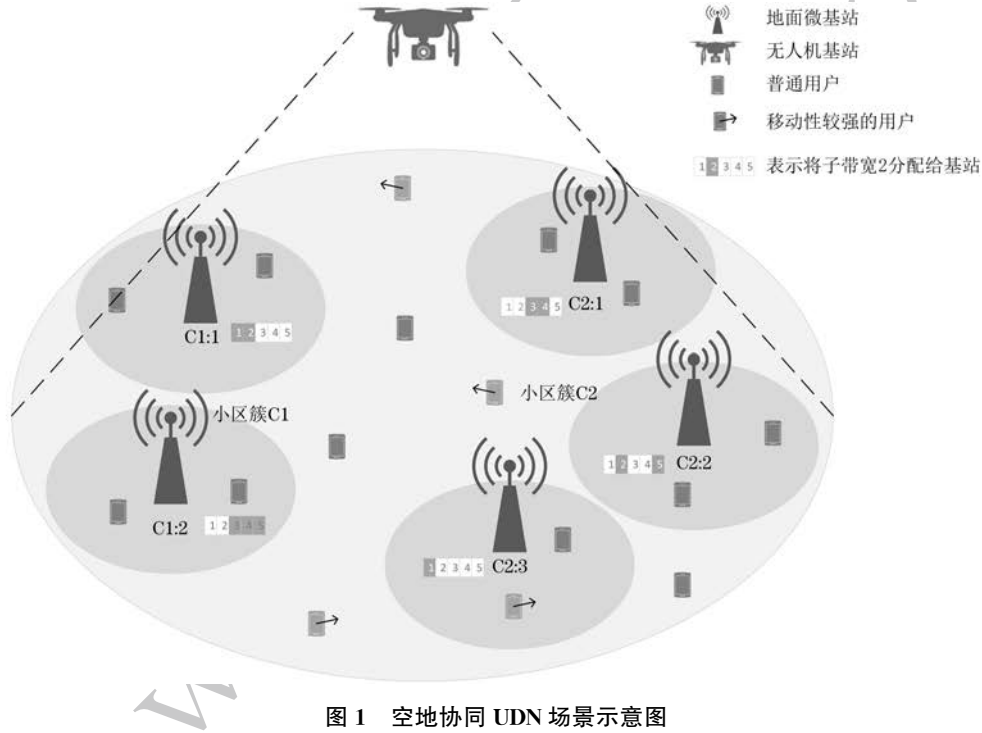


图 1 空地协同 UDN 场景示意图

Fig.1 Schematic diagram of air-ground cooperative UDN scenario

基站之间的带宽需求在时隙模型中根据用户负载更新,每 1 s 根据各基站的负载重新分配带宽。将系统中各小小区边缘用户以及高移动性用户的平均传输速率作为评估算法的指标。因为这两种用户在场景中通信质量较差且易受影响。

t 时刻地面链路用户 i 与基站 j 之间的信道增益模型如下:

$$h_{ij}^t = G - \gamma_{ij}^t - P_s \quad (2)$$

式中: G 表示基站的天线增益; P_s 表示阴影衰落; γ_{ij}^t 表示 t 时刻地面基站 j 与其服务用户 i 之间的路径损失。路径损失模型如下^[22]:

$$\gamma_{ij}^t = 140.7 + 36.7 \lg(d_{ij}^t) \quad (3)$$

空地链路路径损失模型如式(4)所示^[23]。对于

用户和空中基站之间的通信,视距通信(LoS)指的是用户与空中基站之间存在直接的无障碍可视路径。而非视距通信(NLoS)则是指两者之间存在障碍物或无直接可视路径, i 表示与空中基站关联的用户。

$$\gamma_{iu}^t = P_{\text{LoS}}^t \gamma_{iu}^{\text{LoS},t} + P_{\text{NLoS}}^t \gamma_{iu}^{\text{NLoS},t} \quad (4)$$

t 时刻 LoS 链路概率如下:

$$P_{\text{LoS}}^t = \frac{1}{1 + \alpha \exp(-\beta(\theta_i^t - \alpha))} \quad (5)$$

式中: α 和 β 均为环境参数; θ_i^t 为 t 时刻用户 i 与空中基站之间的仰角。

NLoS 链路概率为 $P_{\text{NLoS}}^t = 1 - P_{\text{LoS}}^t$ 。LoS 与 NLoS 链路的路径损失模型分别如下:

$$\begin{aligned}\gamma_{iu}^{\text{LoS},t} &= 20\lg\left(\frac{4\pi f_c}{c}d_{iu}^t\right) + \xi_{\text{LoS}} \\ \gamma_{iu}^{\text{NLoS},t} &= 20\lg\left(\frac{4\pi f_c}{c}d_{iu}^t\right) + \xi_{\text{NLoS}}\end{aligned}\quad (6)$$

式中: d_{iu} 为用户 i 到空中基站之间的距离; ξ_{LoS} 与 ξ_{NLoS} 均为环境参数。

t 时刻由地面基站 j 服务的用户 i 的信干噪比如下:

$$R_{\text{SINR}_{ij}^t} = \frac{\sum_{i \in L_j^t} P_j h_{ij}^t}{\sum_{j' \in J \cup u, j' \neq j, i \in L_j^t} \left(P_{j'} h_{ij'}^t + \frac{P_u}{\lambda \gamma_{iu}^t} \right) + \sigma^2} \quad (7)$$

式中: L_j^t 为 t 时刻由基站 j 服务的用户集合; P_j 为基站 j 的传输功率; P_u 为空中基站的传输功率; $\lambda > 1$ 表示实际调制方案与理论高斯信号之间的信噪比差^[20]; γ_{iu} 为用户 i 与空中基站之间的路径损失; σ^2 为噪声功率。

由空中基站服务用户的信干噪比如下:

$$R_{\text{SINR}_{iu}^t} = \frac{\sum_{i \in L_u^t} P_u \lambda \gamma_{iu}^t}{\sum_{j \in J, i \in L_u^t} P_j h_{ij}^t + \sigma^2} \quad (8)$$

式中: L_u^t 为 t 时刻由空中基站服务的用户集合。

t 时刻平均系统传输速率如式(9)所示。取每个小小区的统计用户平均传输速率,累加每个小小区的平均传输速率获得 $\hat{R}^t$:

$$\begin{aligned}\hat{R}^t &= \sum_{j \in J} \frac{e_{ij}^t}{|L_j^t|} R_{ij}^t + \sum_{i \in L_u^t} \frac{1}{|L_u^t|} R_{iu}^t = \\ &= \sum_{j \in J} \frac{B_j^t}{|L_j^t|} \lg\left(1 + \sum_{i \in L_j^t} \frac{e_{ij}^t}{|L_j^t|} R_{\text{SINR}_{ij}^t}\right) + \\ &= \frac{B_u^t}{|L_u^t|} \lg\left(1 + \sum_{i \in L_u^t} \frac{1}{|L_u^t|} R_{\text{SINR}_{iu}^t}\right)\end{aligned}\quad (9)$$

第 j 个基站对带宽的需求为 N_j^t , 它是 BS 带宽分配的约束条件, 基站 j 的带宽 B_j^t 为 $N_j^t \times \Delta$ 基站服务的所有用户平均分配该基站的带宽, $L^t = \{L_1^t, L_2^t, \dots, L_J^t, L_u^t\}$ 为各个基站的服务用户集合。 T 时间内网络平均吞吐量(T_{TP})为:

$$T_{\text{TP}} = \frac{1}{T} \sum_{0 \leq t \leq T} \hat{R}^t \quad (10)$$

1.2 带宽分配

首先按照阈值速度 $v_{\text{threshold}}$ 将用户划分为由空中基站和地面基站服务的两部分。移动速度大于 $v_{\text{threshold}}$ 的用户由空中基站来服务, 其余用户由地面基站来服务。空中基站使用所有带宽资源,

地面基站按照带宽分配策略使用对应的带宽资源^[24], 以下分两步分配地面小区可用的频谱资源。

1) 小小区负载估计。为每个小小区分配与其负载成比例的带宽数目。在 t 时刻, 如果一个用户满足 $\gamma_{ij}^t \geq \gamma_{ij'}^t$, 则用户 i 与基站 j 相关联。基于基站负载与基站分簇来确定每个基站分配到的子带宽数目。小小区分簇使用 mean shift 算法, 以一种平衡干扰和频谱复用的方式定义阈值来划分小小区簇, 较大的阈值可以减小每个单元簇的大小, 从而以更高的干扰和更低的频谱效率为代价实现更高的频谱重用, 每一个小小区簇使用全部频谱。

对于基站 j , 按照分簇算法获取小小区簇 C_j , 满足 $\gamma_{jj'} \geq \gamma_{\text{threshold}}$ 的基站 j' 属于 C_j , 带宽分配如下:

$$\begin{aligned}l_j^t &= \frac{|L_j^t|}{\sum_{j' \in C_j} |L_{j'}^t|} \\ N_j^t &= l_j^t \times |N|\end{aligned}\quad (11)$$

式中: $\sum_{j \in C_j} N_j^t = N$; l_j^t 为基站的相对负载; N 为场景中可供分配的总带宽数目。

2) 在上一步分配好的带宽数目基础上确定具体的带宽分配策略。 $\mathbf{X}^t = [x_{jn}^t]^{J \times N}$ 表示带宽分配矩阵, 其中元素 x_{jn}^t 为基站 j 在第 n 个子带宽上的具体带宽分配策略, 该指数为 1 表示 t 时刻第 j 个基站分配了第 n 个带宽。将基站 j 受到其它所有基站的干扰 I_j^t 作为优化目标。

$$I_j^t = \sum_{n \in (1, N)} x_{jn}^t \left(\sum_{j' \in J, j' \neq j} |G_{jj'}| \cdot x_{j'n}^t + P_N \right) \quad (12)$$

式中: $G_{jj'}$ 为基站 j 与基站 j' 在 1 个子带宽上的干扰强度; P_N 为噪声在 1 个子带宽上的功率。约束条件为:

$$x_{jn}^t \in \{0, 1\}^{|N|}, \sum_{n \in (1, N)} x_{jn}^t = N_j^t \quad (13)$$

$G_{jj'}$ 的表达式如下:

$$G_{jj'} = P_{j'n} g_j g_{j'} \gamma_{jj'} \quad (14)$$

式中: $P_{j'n}$ 为基站 j 在一个子带宽上的发射功率; g_j 和 $g_{j'}$ 分别表示基站天线增益和用户天线增益; $\gamma_{jj'}$ 表示基站 j 和基站 j' 之间的路径损耗。 $P_{j'n}$ 的表达式如下:

$$P_{j'n} = \frac{P_{j'}}{N}$$

式中: $P_{j'}$ 为基站 j' 的总发射功率。

1.3 问题定义

根据上述系统模型, 将场景中所有地面基站

的带宽资源管理问题建模为多智能体马尔科夫决策过程(MG),以便于确定多个基站的资源分配决策过程以及捕捉各基站分配策略之间的相互影响。MG 由元组 $(B, S, O, A, p, R, \gamma)$ 表示,其中 B 表示前面提到的 BS 集合。该元组的其他组件定义如下。

1)系统状态空间 S 和局部观测空间 O 。 S 表示由来自一些智能体处理后的观测数据组成的系统状态空间。为了捕捉服务需求的时间相关性,第 j 个 BS 在时间 t 的局部观测由 $o_j^t = \{N_j^{t-1}, N_j^t\}$ 表示,该观测由其过去和当前的服务需求组成。 s_j^t 表示第 j 个 BS 在时间 t 的系统状态。

2)动作空间 A 。在时间 t ,假定第 j 个 BS 从其候选动作空间 A 中选择 $a_j^t = x_j^t$ 作为具体的资源块分配策略。动作空间的大小由子带宽的粒度 Δ 决定,带宽粒度决定子带宽的个数。

3)转移概率 $p(s_j^{t+1} | s_j^t, a_j^t)$ 表示第 j 个 BS 根据在时间 t 的 j 处动作从状态 s_j^t 转移到下一个状态 s_j^{t+1} 的概率。给定当前状态 s ,智能体执行动作 a ,环境用某个函数 τ 计算出新的状态 $s' = \tau(s, a)$ 。

4)奖励 R 。在每个时间 t 之后,第 j 个 BS 将通过指定的奖励函数从当前环境中获得实时奖励 r_j^t 。考虑到优化目标,奖励函数设计为:

$$r_j^t = -I_j^t \quad (15)$$

5)折扣系数 γ 。 $\gamma \in (0, 1]$ 是奖励计算中的一个超参数,它决定了未来奖励的重要性。设置 $\gamma = 0$ 意味着智能体只考虑当前奖励,而 $\gamma = 1$ 则重视长期的高奖励。根据经验,设定 $\gamma = 0.9$ 。回报函数 U_j^t 如下:

$$U_j^t = \sum_{k=0}^T \gamma^k r_j^{t+k} \quad (16)$$

2 基于图神经网络的多智能体强化学习

在这一部分,本文描述了所提出的基于图神经网络的 MARL 算法,提出了自底向上的网络结构。

1)通过多层感知器(MLP)实现的观测表示,它是由一个包含多个神经元的多层简单网络组成的非线性函数。由于用户的移动性和随之发生的 BS 切换,相邻 BS 对上一步的需求是用于预测当前 BS 在这一步资源管理策略的重要特征。因此,记录过去的需求作为观测的一部分。

2)通过图神经网络进行处理,构造邻小区间的用户移动相关性,共享自己的需求信息来加强空间合作。

3)基于深度 Q 网络实现资源分配。

2.1 观察表示

对于从场景中获取的原始数据,通过 MLP 将从场景中获取的 n 维原始数据映射到 k 维隐空间 ($k > n$) 进行低维输入,通过上述过程可以将低维不可分数据转换为高维可分数据。由于系统的原始数据是低维的,因此通过 MLP 将其映射到更高维来实现观察表示,表示为:

$$o_j^t = \{d_j^{t-1}, d_j^t\}$$

式中: d_j^t 为 t 时刻基站 j 分配到的资源块数目 N_j^t 。

$$h_j^{t-1} = \text{MLP}(d_j^{t-1})$$

$$h_j^t = \text{MLP}(d_j^t)$$

将观测向量分为 d_j^{t-1} 和 d_j^t ,分别在两个 MLP 网络中处理, h_j^{t-1} 需要由 GNN 处理, h_j^t 与 GNN 的输出连接,然后由 DRL 处理。

2.2 基于图神经网络的状态预处理

从场景中获取的原始数据用户在各基站之间进行切换,导致各基站在不同调度时段的带宽需求波动较大。资源管理策略需要考虑相邻基站的带宽分配策略, GNN 能够整合不同基站的带宽需求信息,从而获得更加全局和协调的管理策略,在此前提下,需要加强基站之间的合作。GNN 可以将相邻 BS 的带宽需求信息纳入当前状态,有利于资源管理策略,具体实现为将上一步观察值 d_j^{t-1} 经过 MLP 预处理得到的 h_j^{t-1} 以及相邻 BS j' 的 h_j^{t-1} 作为 GNN 的输入,得到 GNN 的第一层输出 h_j' ,按照相同的规则将输出得到的 h_j' 再次输入到第二层 GNN,得到 h_j'' 。 D_j 为基站 j 自定义规则下的相邻基站集合,相邻基站的定义规则为:与基站之间距离 $d_{jj'}$ 小于等于 d_{th} ,即 $d_{jj'} \leq d_{th}, j' \in D_j$ 。

$$h_j' = \text{GNN}^1(h_j^{t-1}, h_m^{t-1}), \forall m \in D_j$$

$$h_j'' = \text{GNN}^2(h_j', h_m'), \forall m \in D_j$$

2.3 基于深度 Q 网络的资源分配

作为基于 GNN 的 DRL 算法中的最后一个模块,本节将标准 DQN 及其变体应用于资源管理策略的优化。多智能体 DRL 系统示意图如图 2 所示。

基于动作价值函数,获得最优策略 $\pi(\cdot | s)$,动作价值函数如式(17)所示。状态 $s_j^t = \{h_j^t, h_j', h_j''\}$,由 MLP、GNN¹、GNN² 的输出级联构成。

$$Q^\pi(s_j^t, a_j^t) = \mathbb{E}_{\pi, P}[U_j^t | S = s_j^t, A = a_j^t] \quad (17)$$

根据 Bellman 方程,可得:

$$Q^\pi(s_j^t, a_j^t) = \mathbb{E}_{\pi, P}[r_j^t + \gamma Q^\pi(s_j^{t+1}, \pi(s_j^{t+1}))] \quad (18)$$

式中: s_j^{t+1} 表示由状态转移概率函数 $p(s_j^{t+1} | s_j^t, a_j^t)$ 确定的系统下一个状态。

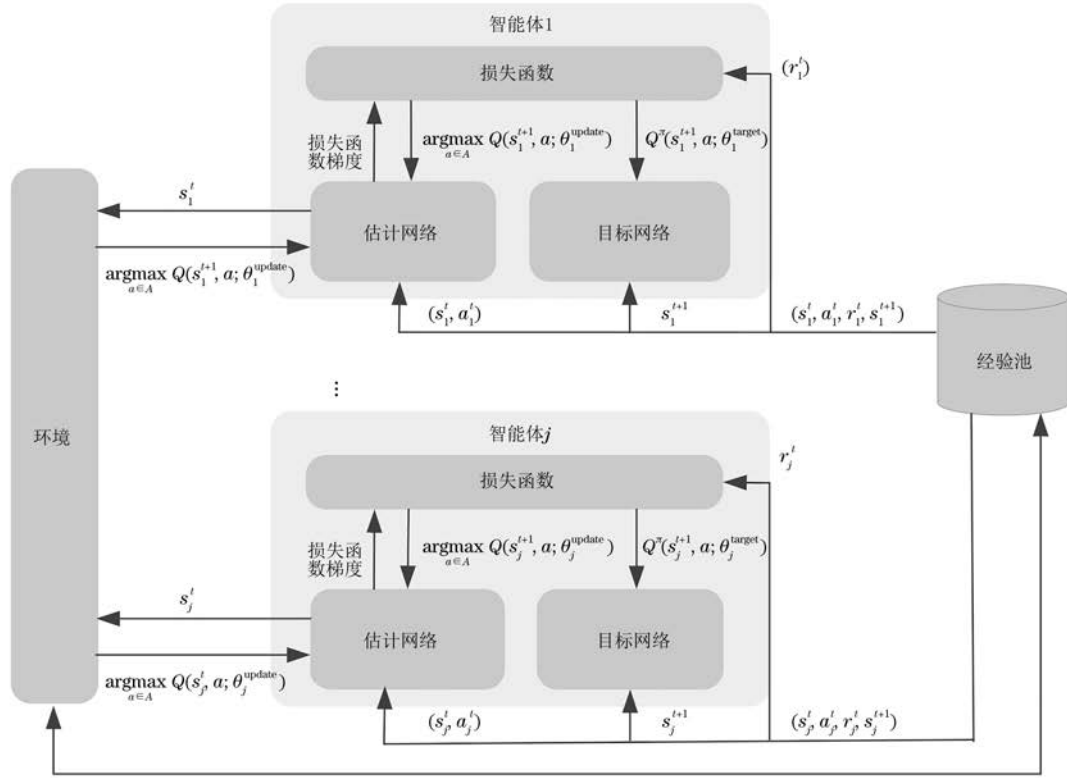


图 2 多智能体 DRL 系统示意图

Fig.2 Schematic diagram of multi-agent DRL system

对于所有 s_j^t 和 a_j^t , 追求最大 $Q^\pi(s_j^t, a_j^t)$ 的最优策略定义为:

$$\pi^* = \underset{\pi}{\operatorname{argmax}} Q^\pi(s_j^t, a_j^t) \quad (19)$$

因此, 相应的动作值函数为:

$$Q^*(s_j^t, a_j^t) = \mathbb{E}_{\pi, P} [r_j^t + \gamma \max_{a \in A} Q(s_j^{t+1}, a; \theta_j^{\text{target}})] \quad (20)$$

式中: θ_j^{target} 为基站 j 目标网络的训练参数。基于 Dueling DQN 的 Q 值函数为:

$$Q(s_j^t, a_j^t; \theta_j^{\text{update}}, \mu_j, v_j) = V(s_j^t; \theta_j^{\text{update}}, \mu_j) + \left[A(s_j^t, a_j^t; \theta_j^{\text{update}}, v_j) - \frac{1}{|A|} \sum_{a \in A} A(s_j^t, a; \theta_j^{\text{update}}, v_j) \right] \quad (21)$$

式中: θ_j^{update} 、 μ_j 、 v_j 分别为基站 j 更新网络的共享卷积编码器、值函数 $V(\cdot)$ 和动作优势函数 $A(\cdot)$ 的可训练参数。Double DQN 的损失函数 $\mathcal{L}(\theta_j^{\text{update}})$ 定义为:

$$Y^{\text{DoubleQ}} = r_j^t + \gamma Q^\pi(s_j^{t+1}, \underset{a \in A}{\operatorname{argmax}} Q(s_j^{t+1}, a; \theta_j^{\text{update}}); \theta_j^{\text{target}}) \quad (22)$$

$$\mathcal{L}(\theta_j^{\text{update}}) = (Y^{\text{DoubleQ}} - Q(s_j^{t+1}, a_j^t; \theta_j^{\text{update}}))^2 \quad (23)$$

网络模型结合了 Double DQN 与 Dueling DQN。基于以上损失函数的定义, 考虑训练过程中在损失函数中加入约束式 (13), 调整后的损失函数 $\mathcal{L}'(\theta_j^{\text{update}})$ 为:

$$\mathcal{L}'(\theta_j^{\text{update}}) = \mathcal{L}(\theta_j^{\text{update}}) + w \times \left| \sum_{k \in (1, N)} x_{jk} - N_j \right| \quad (24)$$

式中: w 为常数系数, 用于调整该约束条件对损失函数的影响。

3 仿真及结果数值分析

3.1 仿真场景和参数设置

考虑一般城市场景, $180 \text{ m} \times 180 \text{ m}$ 的场景中有 30 个微小区, 在场景中随机部署。空中基站使用所有带宽资源, 地面基站按照带宽分配策略使用对应的带宽资源。场景参数设置如表 1 所示。

表 1 仿真参数

Table 1 Simulation parameters

场景参数	参数值设置
Δ/MHz	2
载波频率/GHz	2.4
总带宽/MHz	20
子带宽数目/个	10
基站数目/个	30+1(空中基站)
用户速度/($\text{m} \cdot \text{s}^{-1}$)	Uniform[1, 15]
用户数目/个	2 000
基站发射功率/dBm	30
阴影损耗/dB	10
噪声谱密度/($\text{dBm} \cdot \text{Hz}^{-1}$)	-174
基站天线增益/dBi	0, 15

用户移动模型:假设场景中共有 2 000 个用户,每个用户的速度固定,当用户到达场景边缘时,速度大小不变,方向随机变化后继续在场景内移动。假设可供分配的总带宽为 20 MHz,按照 2 MHz 的粒度划分为 10 个子带宽。小区簇划分的路径损失阈值为 $pl_threshold$ 为 -84.8 dB,对应半径 30 m, A-G 链路相关环境参数参考^[25]。价值网络基于以上参数进行训练。

3.2 仿真结果及分析

在仿真验证场景下加入空中基站后,用户在不同地面基站之间的切换比例有所降低,统计与地面基站距离为 30~40 m 以及由空中基站服务的用户,场景中的用户数用 $user_num$ 表示, pl_th 为基站分簇的半径,划分地面基站用户与空中基站用户的阈值速度用 $v_threshold$ 表示。当场景中的用户数为 2 000 个,基站分簇的半径为 30 m 时,不同 $v_threshold$ 下切换连接的用户数如图 3 所示。当空地协同场景 $v_threshold$ 为 13 m/s,基站分簇半径为 30 m 时,在不同用户密度下切换连接的用户数如图 4 所示,场景面积固定,不同用户数表示不同用户密度。

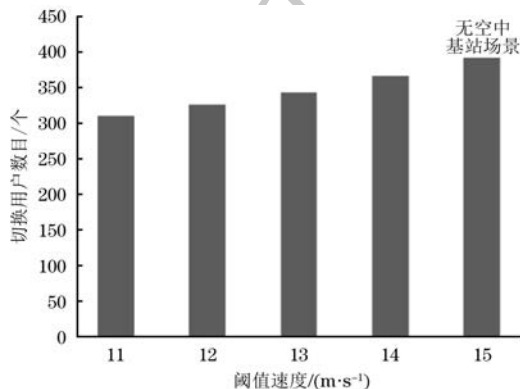


图 3 不同 $v_threshold$ 下切换连接的用户数

Fig.3 Number of users switching connections under different $v_threshold$

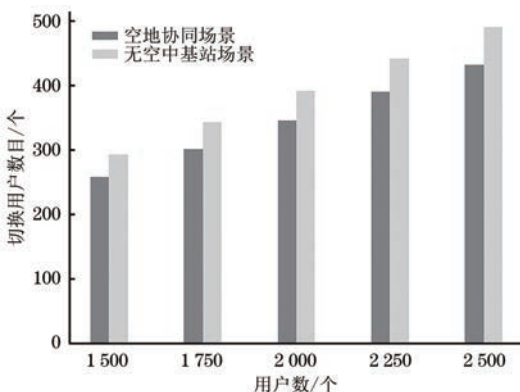


图 4 不同用户密度下切换连接的用户数

Fig.4 Number of users switching connections under different user densities

从图 3 可以看出, $v_threshold$ 越大, 划分给空中基站的高移动性用户就越少, 这部分用户在地面基站之间切换就越频繁。当 $v_threshold$ 为 15 m/s 时, 也就是场景中没有空中基站的情况, 切换的用户数最多。从图 4 可以看出, 在不同的用户密度下, 无空中基站场景中切换用户数均高于空地协同场景。无人机的覆盖范围较大, 将无人机作为基站, 可以服务场景中高移动性的用户, 减少该类用户在地面基站之间的频繁切换次数。

基于训练好的价值网络, 分别仿真分析了该分配策略在不同场景参数下的表现, 考虑了 $v_threshold$ 、用户数、 pl_th 这 3 个因素的影响。使用 Graph-DRL 表示本文所提出的深度强化学习算法, 以文献[13]中的博弈论算法为对比算法, 记为 game theory。

当场景中的用户数为 2 000 个, 基站分簇的半径为 30 m 时, 不同 $v_threshold$ 下算法性能的比较结果如图 5 所示。从图 5 可以看出, 在无空中基站场景下所统计用户的平均吞吐量小于空地协同场景, 各个 $v_threshold$ 下, Graph-DRL 的分配结果均优于博弈论的分配结果。当 $v_threshold$ 为 15 m/s 时, 即由地面基站服务全部用户, 空中基站不参与该场景, 该情况下用户平均吞吐量小于空地协同场景下的用户平均吞吐量。因为 Graph-DRL 的分配决策效率更高, Graph-DRL 相较于对比算法提高了系统的平均吞吐量。

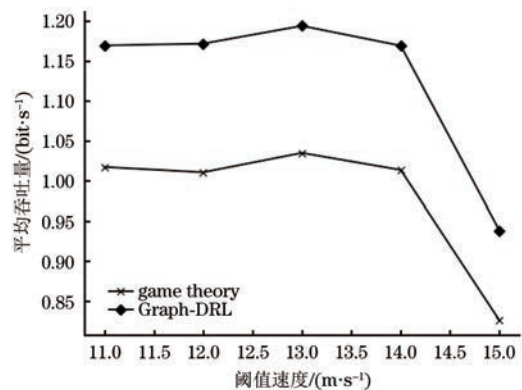


图 5 不同 $v_threshold$ 对统计用户平均吞吐量的影响

Fig.5 Impact on the average throughput of statistics user under different $v_threshold$

当 $v_threshold$ 为 13 m/s, pl_th 为 30 m 时, 图 6 展示了不同用户密度下算法的性能比较结果。从图 6 可以看出, 随着场景中用户密度的增加, 网络服务质量会相应降低, 因为通信网络的总带宽是有限的, 用户数量的增加会导致用户分配到的带宽减少, 从而限制了数据传输速度和效率。仿真结果表明, 基于深度 Q 网络资源分配的算法

可以适应不同用户密度下的频谱动态分配,图 6 验证了用户分布密度在 $0.046 \sim 0.077$ 个/ m^2 范围内动态变化时,Graph-DRL 相较于博弈论具有更好的分配结果,并且在空地协同场景下网络性能更好。

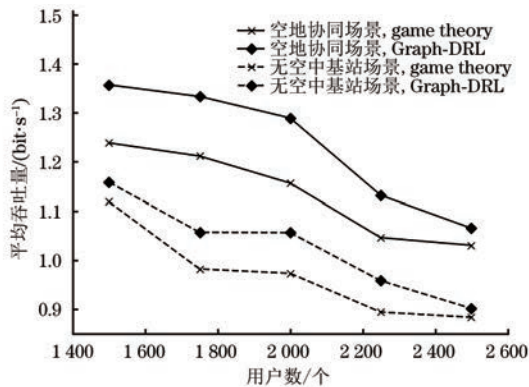


图 6 不同用户密度对统计用户平均吞吐量的影响

Fig.6 The impact of different user densities on average throughput of statistics user

图 7 为当 pl_th 为 30 m 时基站分簇的示意图。当空地协同场景的 $v_threshold$ 为 13 m/s,用户数为 2 000 个时,图 8 展示了不同分簇半径对统计用户平均吞吐量的影响,其横坐标代表了基站分簇半径,坐标从左到右对应的小区簇数(频谱复用度)降低,依次为(12,10,10,10,9)。其中,分簇半径 24 m 对应频谱复用度最高,但干扰也会随之增强,用户平均吞吐量降低,当分簇半径为 27 m 和 30 m 时,频谱复用度降低,但是相应的干扰也减少,用户平均吞吐量增加。当分簇半径为 30~36 m 时,相较于干扰频谱复用度降低对网络的影响更大,因此用户平均吞吐量降低。

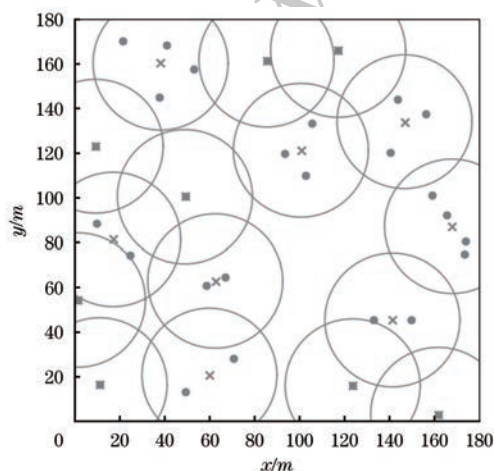


图 7 基站分簇示意图

Fig.7 Schematic diagram of base station clustering

此外,不同用户数和不同分簇半径下算法的运行时间如表 2 和表 3 所示。从表 2 和表 3 可以看

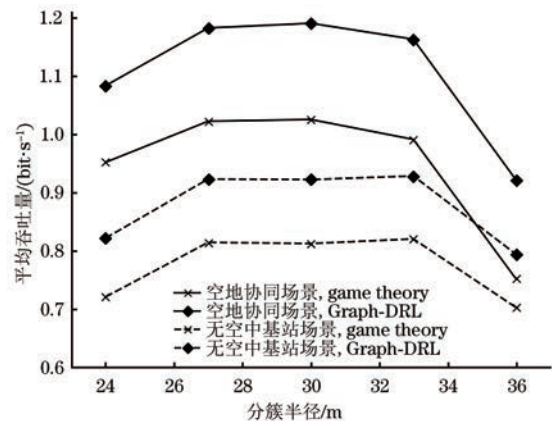


图 8 不同 pl_th 对统计用户平均吞吐量的影响

Fig.8 The impact of different pl_th on the statistical average user throughput

出,利用 Q 值网络做出分配决策的速度远高于利用博弈论算法做出分配决策的速度。Graph-DRL 的推理时间通常取决于神经网络的规模。对于较深的网络或较复杂的任务,推理复杂度为 $O(M)$,其中 M 是神经网络的参数量。在所提方案中,使用神经网络包含 DNN 观察表示层、图神经网络预处理层 $\times 2$ 、Q 网络资源分配层 $\times 2$,各层的神经元个数分别为 32、64、256、512、1022,因此推理复杂度为 $30 \times O(1887)$,其中 30 为地面基站总数目。对于对比算法,计算纳什均衡的复杂度通常为 $O(S \wedge N)$,参数 N 表示玩家数,参数 S 表示每个玩家 S 种策略,对于 N 个基站,每个基站有 S 种分配策略,利用博弈论分配的复杂度为 $O(1022 \wedge 30)$ 。算法运行时间具体的仿真过程为:按照小小区负载确定向相应基站分配的带宽数目之后,将该信息作为 Q 网络的输入以及博弈论的约束条件,以该时刻作为计时的开始,分别利用两种算法进行进一步的分配,在整个网络分配结束之后,再次记录时间,得到算法分配带宽的运行时间。

从表 2 和表 3 可以看出,博弈论单次分配时间需要 40 s,而利用 Graph-DRL 训练好的 Q 值网络可以在 1 s 内做出决策。

表 2 不同用户数下算法的运行时间比较

Table 2 Comparison of algorithms running time under different numbers of users

用户数/个	运行时间/s	
	game theory	Graph-DRL
1 500	49.32	0.11
1 750	49.50	0.11
2 000	49.16	0.11
2 250	46.33	0.11
2 500	47.10	0.11

表 3 不同分簇半径下算法的运行时间比较
Table 3 Comparison of algorithms running time
under different clustering radii

分簇半径/m	运行时间/s	
	game theory	Graph-DRL
24	44.78	0.10
27	49.50	0.10
30	49.56	0.09
33	49.83	0.09
36	48.06	0.09

4 结束语

针对空地协同 UDN 场景,以减小干扰和提高移动用户的传输速率为目的,本文提出一种基于深度强化学习的密集网络频谱资源分配算法,同时结合图神经网络提高了基站之间对于不同移动性用户的空间协作服务。该算法包括前向传播和反向训练 2 个过程。在前向传输过程中,主要构件 DNN 与 GNN,以最小化系统干扰;在反向训练过程中,通过在估计网络与目标网络之间构建时间差分误差,再结合分配策略的约束条件来获得损失函数,进而反向训练网络的权重参数。仿真结果表明,本文提出的算法在提高系统的平均吞吐量方面优于对比算法,且空中基站的引入相较于传统纯地面基站场景也更有利于提高网络性能。下一步将加入流量预测,基于预测结果进行频谱分配,可以解决蜂窝通信中流量非平稳性的问题。

参考文献

- [1] XU Y J, GUI G, GACANIN H, et al. A survey on resource allocation for 5G heterogeneous networks: current research, future trends, and challenges [J]. IEEE Communications Surveys & Tutorials, 2021, 23(2): 668-695.
- [2] ALZUBAIDI O T H, HINDIA M N, DIMYATI K, et al. Interference challenges and management in B5G network design: a comprehensive review [J]. Electronics, 2022, 11(18): 2842.
- [3] DAI M H, HUANG N, WU Y, et al. Unmanned-aerial-vehicle-assisted wireless networks: advancements, challenges, and solutions [J]. IEEE Internet of Things Journal, 2023, 10(5): 4117-4147.
- [4] GAO X L, HOU Y T, SUN K. Channel allocation strategy based on traffic distribution in ultra-dense networks [C] // Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). Washington D. C., USA: IEEE Press, 2022: 213-217.
- [5] WEI R, WANG Y, ZHANG Y. A two-stage cluster-based resource management scheme in ultra-dense networks [C] // Proceedings of the IEEE/CIC International Conference on Communications in China (ICCC). Washington D. C., USA: IEEE Press, 2015: 738-742.
- [6] 黄舒琳, 章志明, 杨伟. 基于环结构的无线传感器网络基站位置隐私保护路由协议 [J]. 计算机工程, 2025, 51(4): 198-207.
- [7] 刘新梦, 谢健骊, 李翠然, 等. 基于 A3C 的认知物联网通信干扰消除算法 [J]. 计算机工程, 2024, 50(10): 281-290.
- [8] LIU X M, XIE J L, LI C R, et al. A3C-based interference cancellation algorithm for cognitive internet of things communication [J]. Computer Engineering, 2024, 50(10): 281-290. (in Chinese)
- [9] YAO W, LI J R, TAN B W, et al. Interference management scheme of ultra dense network based on clustering [C] // Proceedings of the IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Washington D. C., USA: IEEE Press, 2018: 374-377.
- [10] ZHANG G B, ZHANG H J, HAN Z, et al. Spectrum allocation and power control in full-duplex ultra-dense heterogeneous networks [J]. IEEE Transactions on Communications, 2019, 67(6): 4365-4380.
- [11] HABIBINEJAD Z, RASTI M. Joint subchannel allocation and power control in OFDMA femtocell networks [C] // Proceedings of the 9th International Conference on Computer and Knowledge Engineering (ICCKE). Washington D. C., USA: IEEE Press, 2020: 219-224.
- [12] YANG L H, ZHAO J H, GAO F F, et al. Cluster-based joint resource allocation with successive interference cancellation for ultra-dense networks [J]. Mobile Networks and Applications, 2021, 26(3): 1233-1242.
- [13] WANG Q, HUANG Y H, MA Q X. Low complexity joint spectrum resource and power allocation for ultra dense networks [J]. China Communications, 2023, 20(5): 104-118.
- [14] KHAWAM K, LAHOUD S, EL HELOU M, et al. Coordinated framework for spectrum allocation and user association in 5G HetNets with mmWave [J]. IEEE Transactions on Mobile Computing, 2022, 21(4): 1226-1243.
- [15] LIU W, LIU K, TIAN L G, et al. Joint interference alignment and subchannel allocation in ultra-dense networks [J]. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7287-7296.
- [16] ZHANG H, YANG K D, ZHANG S, et al. Efficient subchannel allocation based on partial clustered interference alignment and clusters merging in dense small cell networks [J]. IEEE Transactions on Vehicular Technology, 2020, 69(12): 16191-16196.
- [17] GUO X Z, LI B, LIU K B, et al. Joint placement optimization and RNC in UAV-based wireless multicast networks [C] // Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops). Washington D. C., USA: IEEE Press, 2020: 1-6.
- [18] 叶宝林, 王欣, 李灵犀, 等. 基于深度强化学习 PPO 的车辆智能控制方法 [J]. 计算机工程, 2025, 51(7): 385-396.
- [19] YE B L, WANG X, LI L X, et al. Vehicle intelligent control method based on deep reinforcement learning PPO [J]. Computer Engineering, 2025, 51(7): 385-396. (in Chinese)
- [20] HU J, WANG X M, LI D P, et al. Multi-agent DRL-based resource allocation in downlink multi-cell OFDMA system [C] // Proceedings of the International Conference on Wireless Communications and Signal Processing. Washington D. C., USA: IEEE Press, 2020: 257-262.
- [21] WANG D, HUANG C, ZHANG H, et al. Dynamic clustering and resource allocation using deep reinforcement learning for smart-duplex networks [J]. IEEE Internet of

- Things Journal, 2023, 10(1): 42-56.
- [20] SHARMA N, KUMAR K. Energy efficient clustering and resource allocation strategy for ultra-dense networks: a machine learning framework [J]. IEEE Transactions on Network and Service Management, 2023, 20(2): 1884-1897.
- [21] SHAO Y, LI R P, HU B, et al. Graph attention network-based multi-agent reinforcement learning for slicing resource management in dense cellular network [J]. IEEE Transactions on Vehicular Technology, 2021, 70(10): 10792-10803.
- [22] YANG H L, ZHAO J, LAM K Y, et al. Distributed deep reinforcement learning-based spectrum and power allocation for heterogeneous networks [J]. IEEE Transactions on Wireless Communications, 2022, 21(9): 6935-6948.
- [23] LYU Y, WANG W, CHEN P. Fixed-wing UAV-based air-to-ground channel measurement and modeling at 2.7 GHz in a rural environment [J]. IEEE Transactions on Antennas and Propagation, 2025, 73(4): 2038-2052.
- [24] HUANG Y W, MEI W D, XU J, et al. Cognitive UAV communication via joint maneuver and power control [J]. IEEE Transactions on Communications, 2019, 67(11): 7872-7888.
- [25] LI J X, ZHAO H T, WANG H J, et al. Joint optimization on trajectory, altitude, velocity, and link scheduling for minimum mission time in UAV-aided data collection [J]. IEEE Internet of Things Journal, 2020, 7(2): 1464-1475.

文字编辑 薛晋栋
栏目编辑 宋 圆