第45卷第1期 河北科技大学学报 Vol.45,No.1
2024年2月 Journal of Hebei University of Science and Technology Feb. 2024

文章编号:1008-1542(2024)01-0044-08

基于深度强化学习的空中无人机基站资源分配与公平性研究

郭少雄[1],宋志群[1,2],李 勇[1,2]

(1.通信网信息传输与分发技术重点实验室,河北石家庄 050081;
2.中国电子科技集团公司第五十四研究所,河北石家庄 050081)

摘 要:为了提高无人机基站(unmanned aerial vehicle base stations,UAV-BS)为地面多用户服务时的数据速率,提出一种基于决斗深度神经网络(dueling deep Q-network, Dueling-DQN)的深度强化学习(deep reinforcement learning,DRL)算法。采用决斗网络(dueling network,DN)结构以克服动态环境的部分可观测问题,联合优化了UAV-BS的位置和下行链路功率分配,在更符合实际的空地概率信道模型中检验了Dueling-DQN算法的性能。结果表明,相较于对比算法,所提出的Dueling-DQN算法可以提供更高的数据速率和服务公平性,且随着地面用户数量的增大,算法的优势更加明显。Dueling-DQN算法可有效解决复杂非凸性问题,为UAV-BS的资源分配问题提供理论参考。

关键词:无线通信技术;UAV;空中基站;深度强化学习;资源分配;公平性

中图分类号:TN958.98 **文献标识码**:A **DOI**:10.7535/hbkd.2024yx01005

Deep reinforcement learning-based resource allocation and fairness of aerial UAV base stations

GUO Shaoxiong[1], SONG Zhiqun[1,2], LI Yong[1,2]

(1. Science and Technology on Communication Networks Laboratory, Shijiazhuang, Hebei 050081, China;
2. The 54th Research Institute of CETC, Shijiazhuang, Hebei 050081, China)

Abstract: In order to improve the data rate of unmanned aerial vehicle base stations (UAV-BS) when serving multiple users on the ground, a deep reinforcement learning (DRL) algorithm was proposed based on dueling deep Q-network (Dueling-DQN). A dueling network (DN) structure was employed to overcome the partially observable problem of the dynamic environment, and the position of the UAV-BS and the power allocation of the downlink were jointly optimized to satisfy the quality of service (QoS) of the ground users. The performance of the algorithm was examined in a more realistic air-ground probabilistic channel model. The results show that compared with the baseline algorithm, the proposed Dueling-DQN

收稿日期:2023-09-23;修回日期:2023-12-15;责任编辑:冯 民
基金项目:国家自然科学基金(FFX23641X003)
第一作者简介:郭少雄(1994—),男,河北石家庄人,博士研究生,主要从事无线移动自组网方面的研究。
通信作者:宋志群,研究员。E-mail:zhiqun_sy54@126.com

郭少雄,宋志群,李勇.基于深度强化学习的空中无人机基站资源分配与公平性研究[J].河北科技大学学报,2024,45(1):44-51.
GUO Shaoxiong,SONG Zhiqun,LI Yong. Deep reinforcement learning-based resource allocation and fairness of aerial UAV base stations [J]. Journal of Hebei University of Science and Technology,2024,45(1):44-51.

algorithm can provide higher data rate and service fairness, and the advantages are more obvious with the increase in the number of ground users. The Dueling-DQN algorithm is effective to solve the complex non-convexity problem, which provides some theoretical reference for the resource allocation problem of UAV-BS.

Keywords: wireless communication technology; UAV; aerial base stations; deep reinforcement learning; resource allocation; fairness

在地面基础通信设施缺失或损坏的情况下，部署无人机空中基站(unmanned aerial vehicle base stations，UAV-BS)为地面静态或者移动终端提供无线接入服务已经获得了广泛关注[1]。特别是在一些热点区域(如会议中心、露天音乐会和足球场)或临时事件(如救灾和军事行动)中，得益于无人机的灵活移动性与可操控性，通过相应的部署可以与地面用户实现视线(line of sight，LoS)连接[2]，并提高系统的容量和增强区域覆盖[3-5]。然而，无人机的移动性增大了空地信道状态与干扰关系的复杂性，这就需要对无人机的飞行轨迹和通信资源(如功率、带宽等)进行精确管理和协调，因此研究基于 UAV-BS 的资源分配问题具有重要意义[6]。

现阶段基于 UAV-BS 的优化问题主要有：最大化系统能源效率(energy efficiency，EE)[7]，最大化地面用户的数据速率[8]，最大化 UAV-BS 的服务用户数量[9]等。基于这些问题，文献[10]提出了一种基于 Lipschitz 连续性特征的凸规划差分法框架，提高了地面用户的总数据速率。文献[11]提出了一种基于 Dinkelbach 和分块坐标下降的迭代算法来求解整数非凸和非凹的混合问题，提高了系统的能源效率。文献[12]基于下行蜂窝网络，提出了一种二维空间下 UAV-BS 最优位置选择与功率分配方法，对单架 UAV-BS 的部署轨迹与功率分配进行了联合优化，通过控制 UAV-BS 的移动最大化了地面用户的速率。然而这些传统算法在面临复杂动态的信道条件和干扰关系时，在用户的 QoS 保障和 UAV-BS 的能量限制下，无法构造精确地数学模型且存在较高的计算复杂度[7]。深度强化学习(deep reinforcement learning，DRL)近年来在许多领域获得了广泛的应用，它可以根据环境的反馈来调整策略以获取最优性能[13]。文献[14]设计了一种基于多智能体强化学习框架，采用集中训练与分布执行相结合的方法，最大化了对地面用户的服务数量，同时最小化 UAV-BS 的整体能耗；文献[15]基于 DRL 对 UAV-BS 的轨迹和功率控制问题进行了联合优化，在保证地面用户 QoS 的同时最小化了系统的下行传输功率；文献[16]在采用非正交多址(nonorthogonal multiple access，NOMA)接入方式为地面多用户服务时，提出了一种 DRL 方法求解了三维状态下的 UAV-BS 部署和资源分配问题，提高了系统的能量效率。

然而，不管是传统优化算法还是 DRL 算法，为了简化问题，大都忽视了非视距链路(none line of sight，NLoS)的影响与对用户的服务公平性，由于位置和移动性的影响，UAV-BS 与地面用户之间的链路质量具有不确定性。当 UAV-BS 为地面用户提供服务时，存在一定的服务偏向，即 UAV-BS 根据信道状态的好坏来进行服务，因而不能满足一些用户的 QoS 需求。尤其是当 NLoS 链路存在时，这种情况会更加严重，甚至出现“饿死”链路，严重影响了系统的性能。针对上述问题，本文将决斗网络加入到深度神经网络中，并提出了一种 Dueling-DQN(dueling deep Q-network)算法来解决 UAV-BS 的资源管理问题，采用一种更符合实际的概率信道模型，将 UAV-BS 实际的功耗与信道模型和用户的实际约束相结合，可以在保证 UAV-BS 对地面用户的 QoS 保障与服务公平性的情况下，使对地面用户的总数据速率最大化。

1　系统模型和问题表述

本文考虑 UAV-BS 服务下的下行蜂窝网络，通信场景如图 1 所示。所示通信场景中主要包含用户、UAV-BS 2 类对象，$m=\{1,2,\cdots,M\}$架 UAV-BS 为地面的 $n=\{1,2,\cdots,N\}$个用户提供通信服务，其中用户可以是任何地面终端，包括移动电台、智能手机等。当已部署的网络中出现网络覆盖空洞、基站过载或故障等问题时，需要快速恢复无线通信服务，此时，可以通过快速部署 UAV-BS 为用户按需提供无线接入服务。本文假设无人机已经按照相应的部署策略部署在了合适的位置，并对一定范围内的地面用户提供服务。

UAV-BS 的空对地信道主要由三维地形所决定，可能是 LoS 链路或者 NLoS 链路。由于阴影和反射效应，NLoS 链路的路径损耗一般大于 LoS 链路，现有的研究为简化问题大都忽略了 NLoS 链路的存在，缺乏实用性。本文采用一种更符合实际的空对地信道概率模型，将信道损耗建模为 LoS 和 NLoS 分量的概率加

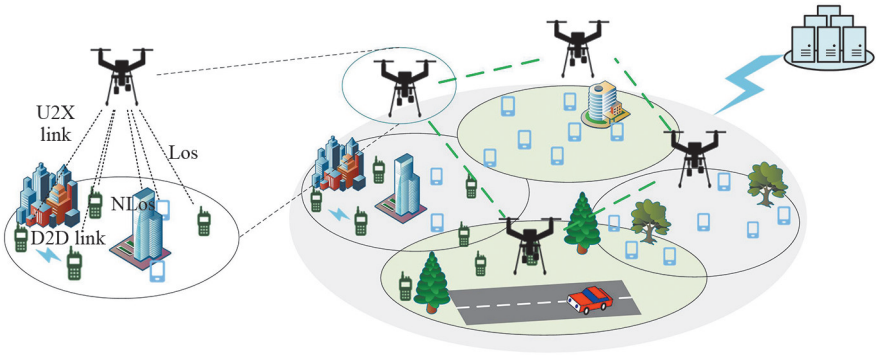


图 1 UAV-BS 辅助通信场景

Fig. 1 Communication scenarios of UAV-BS

权和，当 UAV-BS m 和用户 n 的俯仰角为 $\theta_{m,n}$ 时，两者之间出现 LoS 链路的概率可表示为[17]

$$P(\mathrm{LoS},\theta_{m,n})=\frac{1}{1+\zeta\exp(-\beta(\theta_{m,n}-\zeta))}\ , \tag{1}$$

式中 ζ 和 β 均是由环境（例如农村、城市、密集城市等）决定的常量因子。因此 UAV-BS m 和用户 n 之间的路径损耗可计算为

$$PL_{m,n}=P(\mathrm{LoS},\theta_{m,n})\times L_{\mathrm{LoS}}^{m,n}+(1-P(\mathrm{LoS},\theta_{m,n}))\times L_{\mathrm{NLoS}}^{m,n}, \tag{2}$$

式中 $L_{\mathrm{LoS}}^{m,n}$ 和 $L_{\mathrm{NLoS}}^{m,n}$ 分别表示 LoS 和 NLoS 链路的平均路径损耗[18]，其表达式为

$$\begin{aligned} L_{\mathrm{LoS}}^{m,n}&=20\log\ d_{m,n}+20\log\ f+20\log(4\pi/c)+\eta_{\mathrm{LoS}},\\ L_{\mathrm{NLoS}}^{m,n}&=20\log\ d_{m,n}+20\log\ f+20\log(4\pi/c)+\eta_{\mathrm{NLoS}}, \end{aligned} \tag{3}$$

其中：η_{LoS} 和 η_{NLoS} 分别表示 LoS 和 NLoS 链路的平均额外路径损耗，由环境所决定；f 表示 UAV-BS 与用户之间的载波频率；$d_{m,n}$ 表示 m 和用户 n 之间的距离。

当用户 n 接收到来自 UAV-BS m 的信噪比（signal-to-interference-plus-noise-radio，SINR）$\gamma_{m,n}$ 大于阈值 γ_{th} 时，则称用户 n 被 UAV-BS m 服务，并且其传输速率和 QoS 均得到满足，其计算式为[19]

$$\gamma_{m,n}=\frac{P_{m,n}G_{m,n}}{\sum_{n'\neq n}P_{m,n'}G_{m,n'}+\sigma^2}\geqslant\gamma_{th}\ , \tag{4}$$

式中：$P_{m,n}$ 表示 m 对用户 n 的发送功率；σ^2 表示高斯白噪声功率；$G_{m,n}$ 表示信道增益。根据香农定理，在带宽 B 下，N 个用户的总速率为

$$R_{\mathrm{tot}}=\sum_{i=1}^{n}R_i\ , \tag{5}$$

$$R_i=B\log_2(1+\gamma_{m,n})\ 。 \tag{6}$$

本文目标是在保证对地面用户 QoS 的基础上，最大化地面用户的速率，同时保障 UAV-BS 的服务公平性。据此，优化问题可表示为

$$\begin{cases} \max\ R_{\mathrm{tot}} \\ \text{s. t.}\quad \mathrm{C1}:\forall m,-L\leqslant X_m\leqslant L,-L\leqslant Y_m\leqslant L, \\ \qquad\ \mathrm{C2}:\gamma_{ij}\geqslant\gamma_{th}, \\ \qquad\ \mathrm{C3}:0\leqslant P_{m,n}\leqslant P_{\max},\forall_{i\in m}, \\ \qquad\ \mathrm{C4}:J^F\geqslant J_{\min}, \end{cases} \tag{7}$$

式中：L 表示 UAV-BS 的服务覆盖半径；(X_m,Y_m) 表示 UAV-BS 的坐标；γ_{th} 表示 UAV-BS 与用户之间的信噪比阈值；$P_{\max}$ 表示 UAV-BS 的最大发射功率；J^F 表示简式公平性度量，$J^F\in[0,1]$，J^F 值越大，系统的公平性越高，其表达式为

$$J^F=\frac{\left(\sum_{i=1}^{n}R_i\right)^2}{n\sum_{i=1}^{n}R_i^2}\ 。 \tag{8}$$

在这个优化框架中，C1 表示每个 UAV-BS 在各自的通信覆盖范围内；C2 表示 UAV-UE 对之间信噪比约束，当 UAV-UE 之间满足最小信噪比要求时，认为用户 UE 的 QoS 得到满足；C3 表示 UAV 基站的发射功率约束；C4 是基于简式公平性的公平性约束。从式(8)可以看出，该优化问题是一个非凸混合规划问题，采用传统方法进行求解较为困难，不容易得到最优解。因此，本文提出了一种基于 Dueling-DQN 的深度强化学习算法来进行 UAV-BS 的资源管理。

2　基于 Dueling-DQN 的资源分配算法

本文中采用 Dueling-DQN 深度强化学习算法来联合优化 UAV-BS 系统中的无人机位置和对地面用户的功率分配，在相应的约束条件下，实现地面多用户总速率最大化，具体的 Dueling-DQN 算法包括经验回放、深度神经网络和强化学习 3 个部分，具体结构如图 2 所示。

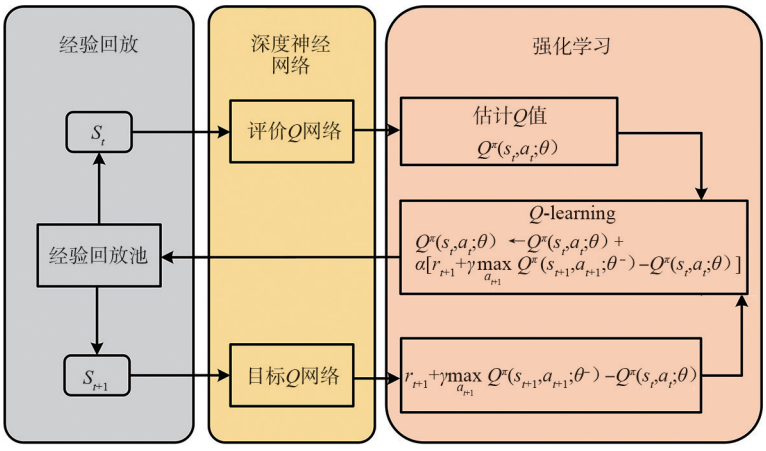


图 2　Dueling-DQN 结构

Fig. 2　Structure of Dueling-DQN

1)经验回放

在 UAV-BS 与环境进行交互过程中，经验回放机制将 UAV-BS 当前状态、动作、奖励以及下一个状态存储为一个元组$(s_t,a_t,r_{t+1},s_{t+1})$作为经验并存放在一个经验回放池(experience replay buffer)中，经验回放缓冲区有着合适的大小和最小的批处理数据长度，在训练 UAV-BS 获取最优策略过程中，从缓冲区中随机抽取一批经验样本来进行训练，这种离线学习模式可以有效地减小样本间的相关性和数据的相关性，提高样本的有效利用率，并且能够更好地平衡探索和利用的策略。

2)深度神经网络

深度神经网络由于包含多个隐藏层，每个隐藏层含有一定数量的神经元并通过相应的权值联系，采用合适的激活函数可以实现较强的非线性映射和拟合能力。然而，对于传统的深度神经网络，每个隐藏层的输出仅取决于当前的输入。为了适应局部观测，本文通过增加决斗网络结构来改进所提出算法中的深度神经网络，结构见图 3。

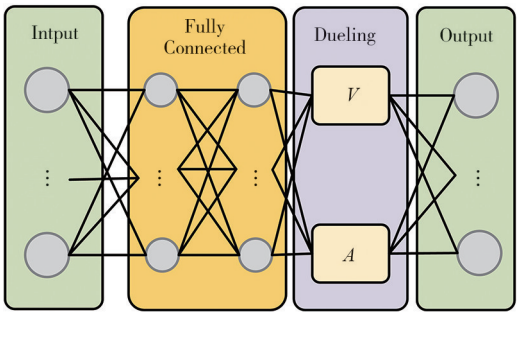


图 3　深度神经网络(DNN)结构

Fig. 3　Deep neural network (DNN) architecture

神经网络中存在动态环境的可观测性问题。如当 UAV-BS 为多个用户服务时，无论采取何种操作，状态都可能是好的或坏的。因此，需要通过每个动作的状态值 $V(s_t;\theta)$ 和优势值 $A(s_t,a_t;\theta)$ 对状态-行为值函数 $Q(s_t,a_t;\theta)$ 进行估计，表示为

$$Q(s_t,a_t;\theta)=V(s_t;\theta)+A(s_t,a_t;\theta)\ 。\tag{9}$$

由图 3 可知，全连接层(fully connected，FC)的输出分为 V 和 A，其中 V 代表值函数，A 代表优势函数，

通过对值函数和优势函数的组合与运算，得到最终的 Q 值。如此，可以更精确地评估状态-动作值函数，以此来指导 UAV-BS 采取更加合适的动作，如通过 UAV-BS 的位置和功率分配系数的调整来获得最大奖励，减少动作采样中的冗余操作，通过最小化损失函数指导发现最优策略 $\pi^* = \arg\max_a Q^*(s_\tau, a_\tau)$。损失函数设置为

$$L(\theta) = \mathbb{E}\left[(r_\tau + \gamma \max Q'(s_{\tau+1}, a_{\tau+1}) - Q(s_\tau, a_\tau))^2\right]。\tag{10}$$

3)强化学习

本文提出的深度强化学习框架是基于 Q-learning 的，通过马尔科夫决策过程(Markov decision process，MDP)将每个 UAV-BS 建模为一个智能体进行分析。MDP 四元组可表示为$\{S, A, P, R\}$，它们分别表示状态空间、动作空间、状态转移概率空间和奖励空间，其中 $s^t \in S$，$a^t \in A$，$r^t \in R$ 则表示某一状态、动作、奖励值。奖励值 r^t 表示 UAV-BS 状态转移后获得的奖励，UAV-BS 从时间步长 τ 开始，依据贝尔曼方程来更新 $Q(s_\tau, a_\tau, \theta)$，并最终确定最优策略 $\pi^* = \arg\max_a Q^*(s_\tau, a_\tau)$。贝尔曼方程可表示为

$$Q^*(s_\tau, a_\tau) = \mathbb{E}\left[r_\tau + \gamma \max_{a_{\tau+1}} Q^*(s_{\tau+1}, a_{\tau+1})\right]。\tag{11}$$

UAV-BS 的状态空间包括相对于地面用户的欧式距离、服务用户的功率系数、与服务用户之间的信道增益 3 部分，如式(12)所示。

$$\boldsymbol{s}_\tau = \left[\boldsymbol{s}_{1,\tau}^{\mathrm{T}}, \boldsymbol{s}_{2,\tau}^{\mathrm{T}}, \cdots, \boldsymbol{s}_{N_{\mathrm{UE}},\tau}^{\mathrm{T}}\right]^{\mathrm{T}},\tag{12}$$

其中

$$\boldsymbol{s}_{i,\tau} = \left[\Delta x_{\mathrm{UAV}-i,\tau}, \Delta y_{\mathrm{UAV}-i,\tau}, \alpha_{i,\tau}, g_{i,\tau}(d_{i,\tau})\right]^{\mathrm{T}}, \quad \forall i \in \{1,2,\cdots,N_{\mathrm{UE}}\}。\tag{13}$$

基于状态空间，根据 ε-greedy 策略采取相应的动作，动作空间设置为

$$\boldsymbol{\delta}_a = \left[d_1\delta_1, d_3\delta_3, \cdots, d_{N_{\mathrm{UE}}-1}\delta_{N_{\mathrm{UE}}-1}\right]^{\mathrm{T}}, \quad d_i \in \{1, -1\}。\tag{14}$$

在深度强化学习中，奖励函数的设置是十分重要的，它直接影响到智能体的学习过程和性能。奖励函数应该根据具体问题的特点来设计，以引导智能体向期望的行为进行学习，在本文中，奖励函数设置为

$$\begin{aligned} r_\tau = {} & w_r \times \frac{R_N}{W} \times \prod_{i=1}^{N} I_{R_i \geqslant R_{\min}} + w_f \times J^F \times I_{R_i < R_{\min}} + w_g \times G_{ij}^{\mathrm{tot}} + \\ & w_s \times \sum_{i=1}^{N} I_{R_i \geqslant R_{\min}} + w_u \times \sum_{i=1}^{N} \frac{R_i}{W} I_{R_i < R_{\min}}, \end{aligned}\tag{15}$$

式中：G_{ij}^{tot} 为所有用户的信道增益；w_r，w_f，w_g，w_s，w_u 分别为总速率、公平性、总信道增益、满足最小速率、不满足最小速率所对相应的权值。通过设计奖励函数，使得在所有用户满足最小速率约束后继续增加速率，公平性奖励在用户没有达到最低速率要求的情况下才有意义，以此实现速率补偿并避免“饿死”链路的产生，UAV-BS 在 ε-greedy 策略的指引下采取相应的动作，并获得相应的奖励和新的状态。

4)训练过程

在深度神经网络的训练过程中，主要分为 4 个步骤：第一，初始化 UAV-BS 的位置、信道增益和功率分配系数，具体为在本文的概率信道模型下，基于 MATLAB 初始化 UAV-BS 的位置坐标并计算出相应的信道增益，功率分配系数的设置满足用户的 QoS 需求，在本文中表示为 $\gamma_{m,n} \geqslant \gamma_{th}$；第二，将 UAV-BS 当前状态 s_t 输入到估计网络中，并输出动作，包括新的 UAV-BS 的位置以及更新后的功率分配；第三，UAV-BS 在 ε-greedy 策略的指引下选择一个动作 a_t，采取动作后得到相应的观测值 o_t 和奖励 r_{t+1} 并达到下一个状态 s_{t+1}，此时将得到的经验组$(s_t, a_t, r_{t+1}, s_{t+1})$存放到经验回放池中；第四，计算目标 Q 值并训练估计网络，在特定时间步长内更新目标网络的超参数。重复第二、三、四步直到收敛或者达到最大迭代次数。此外，采用速率补偿机制来防止在 NLoS 链路存在的情况下，UAV-BS 对某些用户的服务偏向，这通过奖励函数的设置来实现，具体见式(15)。即当用户没有达到 QoS 要求时，检测并增加其速率，直到其至少满足最低的通信速率要求，在最大化系统奖励的同时，实现了多用户的 QoS 与公平性保障。

3 仿真设置与结果分析

Dueling-DQN 算法采用了 2 个全连接层，每层神经元数量为 128 个，使用的激活函数为 ReLU 函数[20]，

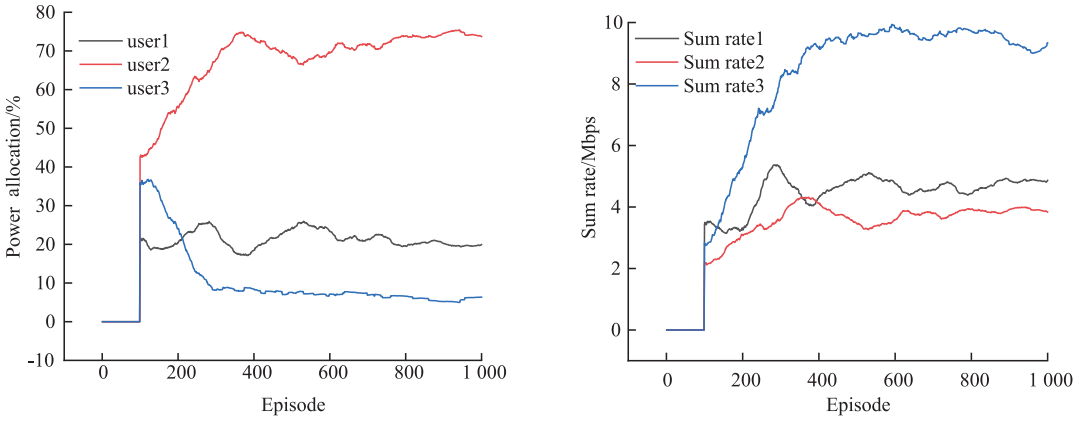
并使用 Adam 算法[21]进行随机梯度下降来更新 θ，智能体的学习率 α 设为 0.001，将经验回放池大小设置为 1 000，并以先输入先输出的方式更新经验回放池，最小经验结构大小设置为 64。表 1 列出了更详细的仿真环境和算法参数设置。

设置地面用户数量为 3 并随机分布，对提出的基于 Dueling-DQN 的功率分配情况进行仿真分析，图 4 展示了基于 Dueling-DQN 算法的功率分配与用户获得的总速率图。从图 4 可以看出，针对不同位置信道状况不同的各地面用户，采用不同的资源分配率，在满足各自 QoS 的基础上，产生了差异化的速率，并均达到了最低速率要求，证明了所提算法的有效性。

表 1　仿真设置

Tab. 1　Simulation setup

参数	值
R_{min}/Mbps	1.48
P_{max}/dBm	23
h/m	100
L/m	500
γ_{th}/dB	−9
ζ	9.611 7
β	0.158 1
v_{max}/(m/s)	10
σ^2/dBm	−140
B/MHz	2
γ	0.9
α	0.001



a）功率分配　　b）用户获得的总速率

图 4　3 用户功率分配与可达速率

Fig. 4　Power allocation and reachable rate of 3 users

为进一步检验 Dueling-DQN 的性能，采用了 4 种对比算法，分别是深度神经网络 DQN 算法、文献[12]中的 LOPA 算法、最大功率算法（Max）、随机功率算法（Random）。图 5 展示了 5 种算法在不同地面用户数量下的平均总数据速率情况。从图中可以看出，最大功率算法与随机功率算法提供的数据速率最低；相比于传统算法 LOPA，随着地面用户数量的增多，DQN 算法和 Dueling-DQN 算法可以提高更大的数据速率，尤其是当地面用户的数量大于 10 个时这种优势更加明显。所提出的 Dueling-DQN 算法较于 DQN 算法，将动作价值分解为价值和优势 2 部分，并分别对不同动作的优势进行估计，从而更有效地处理多个动作的情况，并根据实际的信道状况，取得了更优的资源分配策略，因而相较于 DQN 算法平均总数据速率可提高约 11%。

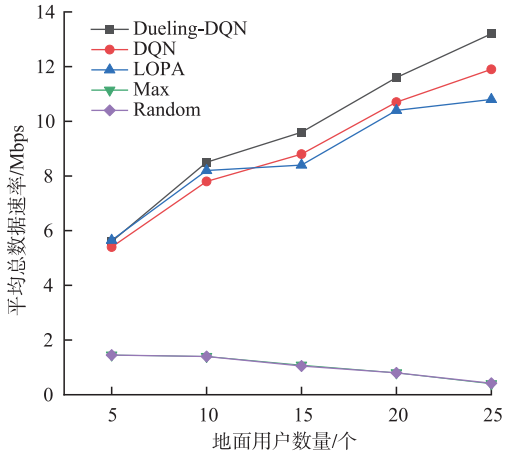


图 5　不同频谱效率下的平均总速率

Fig. 5　Average total rate at different spectral efficiencies

图 6 展示了不同位置用户的可达速率情况，UAV-BS 被部署在了不同的位置，用来检测不同信道状态下的服务性能。从图中可看出，最大功率与随机功率算法的用户可达速率最低；基于数学模型的 LOPA 算法服务总速率与 DQN 算法相比存在差距；DQN 算法在计算动作价值时采用最大化操作容易引起价值估计过高的问题；决斗网络提供了更准确的动作价值估计和梯度信号，可以有效减少训练过程中的不稳定性，使得算法更加稳定和

可靠，故 Dueling-DQN 算法相比于 LOPA 和 DQN 算法，可达速率较为稳定和集中，且获得了最高的用户可达速率。

图 7 展示了不同地面数量下的服务公平性，随着地面用户数量的增加，5 种算法的公平性都不断降低，其中最大功率算法和随机功率算法的公平性最低且始终在一个较低的水平；相较于 LOPA 算法，DQN 算法在公平性上有所提高，但是在地面用户数量较多时优势不明显；Dueling-DQN 算法采用速率补偿机制，不但提高了较高的服务公平性还有效应对了地面用户数量增大的影响，避免了公平性的快速下降。但是随着节点数量的增加，信道状态和干扰关系变得更加复杂，公平性不断降低，这种情况下如何保持住较高的服务公平性，是一个值得继续研究的问题。

图 8 展示了不同用户数量下的饥饿节点数量。本文中饥饿节点的数量定义为一定的仿真时间内，没有获得 UAV-BS 通信服务的节点，反映了算法的通信覆盖能力，这是由 UAV-BS 的选择性服务偏好导致的。从图 8 中可以看出，随着地面用户数量的增加，5 种算法中饥饿节点的数量不断上升，其中 LOPA 算法、最大功率算法、随机功率算法等传统算法饥饿节点数量增加较快，严重影响了用户的服务体验和系统的公平性；Dueling-DQN 算法采取动态速率补偿机制，随着 UAV-BS 与环境的交互，通过奖励函数最大化的策略逐渐满足用户的最低速率需求，可以显著减少饥饿节点的数量，增强了对地面用户的服务保障。

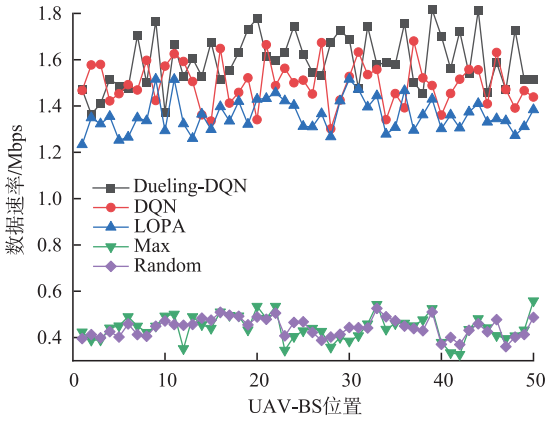


图 6 不同位置下的数据速率

Fig. 6 Data rate at different locations

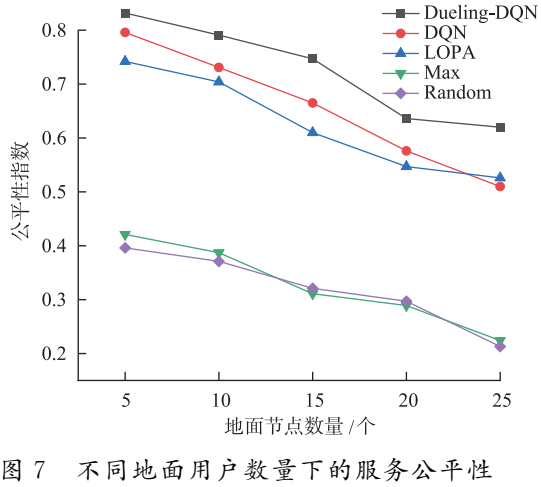


图 7 不同地面用户数量下的服务公平性

Fig. 7 Fairness of service with different numbers of ground users

4 结 语

研究了 UAV-BS 的下行链路多用户资源管理和分配问题，提出了一种基于深度强化学习的 Dueling-DQN 算法，算法通过 DQN 学习并挖掘 UAV-BS 对用户的信道状态和服务信息，通过决斗网络克服神经网络中动态环境的可观测性问题，通过联合优化 UAV-BS 的空中位置与功率分配，在保证地面用户 QoS 的基础上，最大化了地面用户的总数据速率。采用了一种动态的速率补偿机制有效减少了饥饿用户节点的数量，在不同地面用户数量和 UAV-BS 位置下检验了 Dueling-DQN 算法的性能，并将其与其他基准方案（DQN 算法、LOPA 算法、最大功率算法、随机功率算法）进行了对比，证明了 Dueling-DQN 算法可以部署更优的位置，为用户提供更高的服务速率，并保证了服务公平性，同时网络的规模越大 Dueling-DQN 算法有着更好的优越性。随着网络规模的增大，饥饿节点数量不断上升且系统的服务公平性越来越差，加强多个 UAV-BS 之间的协作和交流可能是一种有效的方式，同时研究地面用户不同的 QoS 需求（如不同用户有不同的数据速率要求）与动态性将是下一阶段的工作重点。

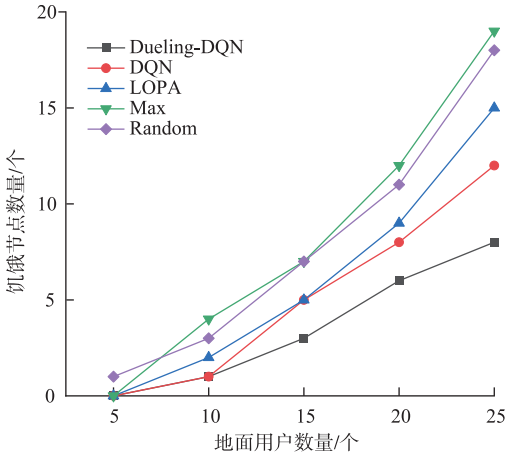


图 8 不同地面用户数量下的饥饿节点数量

Fig. 8 Number of starved nodes for different number of ground users

参考文献/References：

[1] NA Zhenyu，ZHANG Mengshu，WANG Jun，et al. UAV-assisted wireless powered Internet of Things：Joint trajectory optimization and resource allocation[J]. Ad Hoc Networks，2020，98. DOI：10.1016/j.adhoc.2019.102052.

[2] MKIRAMWENI M E，YANG Chungang，LI Jiandong，et al. A survey of game theory in unmanned aerial vehicles communications[J]. IEEE Communications Surveys & Tutorials，2019，21(4)：3386-3416.

[3] WANG Haijun，ZHAO Haitao，ZHANG Jiao，et al. Survey on unmanned aerial vehicle networks：A cyber physical system perspective[J]. IEEE Communications Surveys & Tutorials，2020，22(2)：1027-1070.

[4] ZENG Yong，ZHANG Rui，LIM T J. Wireless communications with unmanned aerial vehicles：Opportunities and challenges[J]. IEEE Communications Magazine，2016，54(5)：36-42.

[5] 尹雅楠，甄然，武晓晶，等. 自适应多启发蚁群算法的无人机路径规划[J]. 河北科技大学学报，2021，42(1)：38-47.
YIN Yanan，ZHEN Ran，WU Xiaojing，et al. Research on UAV route planning based on adaptive multi heuristic ant colony algorithm[J]. Journal of Hebei University of Science and Technology，2021，42(1)：38-47.

[6] GU Xiaohui，ZHANG Guoan. A survey on UAV-assisted wireless communications：Recent advances and future trends[J]. Computer Communications，2023，208：44-78.

[7] 张先玉，陈勇，张余，等. 无人机通信中的资源分配及部署位置联合优化[J/OL]. 西南交通大学学报.（2023-11-28）[2023-12-21]. http://kns.cnki.net/kcms/detail/51.1277.U.20231127.1035.002.html.
ZHANG Xianyu，CHEN Yong，ZHANG Yu，et al. Joint optimization of resource allocation and deployment location in UAV communication[J/OL]. Journal of Southwest Jiaotong University：1-9.（2023-11-28）[2023-12-21]. http://kns.cnki.net/kcms/detail/51.1277.U.20231127.1035.002.html.

[8] BEKMEZCI I，SAHINGOZ O K，TEMEL S. Flying Ad-Hoc networks（FANETs）：A survey[J]. Ad Hoc Networks，2013，11(3)：1254-1270.

[9] 曾晓婉，王海军，马东堂，等. 面向最大化服务用户数的无人机基站 3D 部署方法[J]. 电讯技术，2023，63(11)：1732-1741.
ZENG Xiaowan，WANG Haijun，MA Dongtang，et al. A UAV base station 3D deployment method for maximum users served[J]. Telecommunication Engineering，2023，63(11)：1732-1741.

[10] NGUYEN T M，AJIB W，ASSI C. A novel cooperative NOMA for designing UAV-Assisted wireless backhaul networks[J]. IEEE Journal on Selected Areas in Communications，2018，36(11)：2497-2507.

[11] ZENG Fanzi，HU Zhenzhen，XIAO Zhu，et al. Resource allocation and trajectory optimization for QoE provisioning in Energy-Efficient UAV-Enabled wireless networks[J]. IEEE Transactions on Vehicular Technology，2020，69(7)：7634-7647.

[12] LIU Xiaonan，WANG Jingjing，ZHAO Nan，et al. Placement and power allocation for NOMA-UAV networks[J]. IEEE Wireless Communications Letters，2019，8(3)：965-968.

[13] 靳晓洁，石建迈，伍国华，等. 无人机基站部署问题综述：模型与算法[J]. 控制理论与应用，2022，39(12)：2219-2232.
JIN Xiaojie，SHI Jianmai，WU Guohua，et al. A review of the UAV base station deployment problem：Models and algorithms[J]. Control Theory & Applications，2022，39(12)：2219-2232.

[14] CHEN Binqiang，LIU Dong，HANZO L. Decentralized trajectory and power control based on Multi-Agent deep reinforcement learning in UAV networks[C]//ICC 2022-IEEE International Conference on Communications，Seoul：IEEE，2022：3983-3988.

[15] ZHOU Shiyang，CHEN Yufan，LEI Xia. Multi-agent model-based reinforcement learning for trajectory design and power control in UAV-Enabled networks[C]//2022 3rd Information Communication Technologies Conference（ICTC）. Nanjing：IEEE，2022：33-38.

[16] BENFAID A，ADEM N，KHALFI B. AdaptSky：A DRL based resource allocation framework in NOMA-UAV networks[C]//2021 IEEE Global Communications Conference（GLOBECOM）. Madrid，Spain：IEEE，2021：1-7.

[17] 郭少艾. 基于深度强化学习的认知物联网资源分配研究[D]. 长春：吉林大学，2023.
GUO Shaoai. Research on Cognitive Internet of Things Resource Allocation Based on Deep Reinforcement Learning[D]. Changchun：Jilin University，2023.

[18] 牛海斌. 无人机基站部署与资源分配策略研究[D]. 北京：北京邮电大学，2022.
NIU Haibin. Research on Deployment and Resource Allocation Strategy of UAV Base Station[D]. Beijing：Beijing University of Posts and Telecommunications，2022.

[19] ZHOU Li，SHENG Zhengguo，WEI Lei，et al. Green cell planning and deployment for small cell networks in smart cities[J]. Ad Hoc Networks，2016，43：30-42.

[20] TACHIBANA K，OTSUKA K. Wind prediction performance of complex neural network with ReLU activation function[C]//2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan（SICE）. Nara：IEEE，2018：1029-1034.

[21] KINGMA D P，BA J. Adam：A Method for Stochastic Optimization[EB/OL].（2014-12-22）[2023-05-10]. https://arxiv.org/abs/1412.6980.