DOI:10.16136/j.joel.2023.03.0254

基于MADDPG的无人机辅助通信功率分配算法

陈 剑[1],杨青青[1,2*],彭 艺[1,2]

(1.昆明理工大学 信息工程与自动化学院,云南 昆明 650500;2.昆明理工大学 云南省计算机技术应用重点实验室,云南 昆明 650500)

摘要:针对多无人机(unmanned aerial vehicle,UAV)作为空中基站辅助通信的吞吐量和公平性问题,提出了一种基于多智能体深度确定性策略梯度算法(multi-agent deep deterministic policy gradient algorithms,MADDPG)的功率分配算法,该算法通过联合优化UAV基站的功率分配和用户接入以提高系统吞吐量和公平性。本文首先构建了UAV基站为地面建立通信服务的三维场景,然后通过联合功率、用户关联和UAV位置约束,构建了吞吐量和公平性最大化的问题模型。考虑到该问题的复杂性,本文将所构建的优化问题建模为马尔科夫决策过程(Markov decision process,MDP),通过引入深度确定性策略梯度算法(deep deterministic policy gradient algorithm,DDPG)解决该问题。仿真结果表明,本文提出的基于MADDPG的UAV基站功率分配算法与其他算法相比,可以有效地提升系统的吞吐量和用户的公平性,提高通信的服务质量。

关键词:无人机(UAV);功率分配;深度强化学习(DRL);吞吐量;通信公平性

中图分类号:TN929.52 **文献标识码**:A **文章编号**:1005-0086(2023)03-0306-08

Algorithm of UAV auxiliary communication power allocation based on multi-agent deep deterministic policy gradient algorithms

CHEN Jian[1], YANG Qingqing[1,2*], PENG Yi[1,2]

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China; 2. Yunnan Key Laboratory of Computer Technologies Application, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

Abstract: Aiming at throughput and fairness issues for multi-unmanned aerial vehicle (multi-UAV) as aerial base station auxiliary communication, a power allocation algorithm based on the multi-agent deep deterministic policy gradient (MADDPG) algorithm is proposed. The algorithm improves system throughput and fairness by jointly optimizing the power allocation and user access of UAV base stations. This paper firstly constructs a 3D scene in which the UAV base station establishes communication services for the ground and then constructs a problem model of maximizing throughput and fairness by combining jonit power, user association, and UAV location constraints. Considering the complexity of the problem, this paper models the constructed optimization problem as a Markov decision process (MDP), and solves the problem by MADDPG. The simulation results show that compared with other algorithms, the UAV base station power allocation algorithm based on MADDPG proposed in this paper can effectively improve the throughput of the system and the fairness of users, to improve the service quality of communication.

Key words: unmanned aerial vehicle (UAV); power allocation; deep reinforcement learning (DRL); throughput; communication fairness

* **E-mail**:978214763@qq.com

收稿日期:2022-04-11 **修订日期**:2022-06-13

基金项目:国家自然科学基金(61761025)资助项目

0 引　言

目前,通信网络的建立主要是依靠地面基站,其灵活性受到了很大的限制。因重大自然灾害或者紧急突发事件而导致地面基站不能正常使用时,建立快速而有效的通信网络具有重要意义。携带微型基站的无人机(unmanned aerial vehicle,UAV)辅助无线通信有着广阔的应用前景,受到了人们的关注[1,2]。与地面基站相比,UAV 辅助通信有诸多优点:首先,UAV 基站在大部分时间里能够提供视距(line of sight,LOS)链路链接,通常具有更高的信道增益[3];而且 UAV 的灵活性和高度的部署适应性弥补了传统地面基站的诸多限制,能够快速部署在需要建立通信服务的场景[4]。

合理的任务调度和资源分配对系统性能极为重要。在文献[5]中,作者针对多约束条件下 UAV 辅助通信的场景,提出了一种基于惩罚对偶分解的 UAV 轨迹设计和功率分配算法,提升了地面用户(ground user,GU)的最小平均速率。文献[6]针对 UAV 和蜂窝用户的联合资源优化问题,推导了一种基于访问优先级的接收机确定方法,提高了系统的能量效率和频谱利用率。文献[7]中作者以最大化用户的最小平均速率为目标,建立了 UAV 辅助通信过程中的带宽分配和航迹规划的问题,并通过辅助变量法和交替优化法来解决,最终提升了用户的速率。上述文献均针对单 UAV 辅助通信场景,面向多 UAV 辅助通信也进行了一系列研究,文献[8]中作者研究了多个 UAV 辅助上行通信场景,提出了基于 K 均值的 UAV 部署和分组方案,同时还考虑了基于总功率最小化的资源分配问题。文献[9]联合优化了 UAV 的位置和发射功率分配的问题,通过联合凸逼近法来提高 UAV 辅助通信系统中的用户接入数量。文献[10]考虑了在非正交多址接入方式下两架 UAV 作为基站的通信场景,采用图论法进行用户分组,并利用辅助变量法将非凸的功率分配子问题转化为凸优化问题来求解。然而,文献中的优化问题大多都是高维度的离散非凸问题,以牺牲精度为代价将原非凸问题转化为凸问题并不是一个很好的解决办法[11]。

人们日益提高的通信需求使无线通信网络须按需进行资源分配来充分利用有限的资源,深度强化学习(deep reinforcement learning,DRL)能够让智能体直接学习动态环境规律并得到最优决策赋予网络,依据自身环境进行自我优化管理[12]。在文献[13]中,张志才等通过深度 Q 网络(deep Q-learning network,DQN)算法联合优化 UAV 的发射功率和基站的干扰功率来获得最优的功率策略。文献[14]通过 Q-learning 算法对 UAV 的能耗和飞行路线进行优化,从而实现平均通信时延最化。上述文献均是基于单 UAV 辅助通信的研究。文献[15]指出,单智能体的 DRL 算法无法刻画出环境的动态性,因此采用多智能体的强化算法很有必要,而随着智能体数量的增加,决策输出的动作维度越来越大,多智能体的深度确定性策略梯度算法(multi-agent deep deterministic policy gradient algorithms,MADDPG)在保证精度的基础上,能够解决因多智能体输出的动作维度太大而导致的算法收敛问题[16]。因此,本文在前人研究的基础之上,针对 UAV 的吞吐量和用户公平性的问题,提出了一种基于 MADDPG 的资源分配算法。

1 系统模型及优化问题构建

1.1 系统模型

本文考虑下行链路通信,构建如图 1 所示的三维通信场景,m 架 UAV 充当空中基站为 k 个 GU 通过时分多址(time division multiple access,TDMA)方式提供通信服务,UAV 可用集合表示为 $m=\{1,2,\cdots,M\}$,用户可用集合表示为 $k=\{1,2,\cdots,K\}$。将 UAV 的服务时间 T 划分为 N_E 个等长时隙 t,每个时隙的长度为 T/N_E,假设 GU 的高度为 0,第 k 个 GU 的位置可用坐标表示为 $K_k(t)=[x_k,y_k]$,UAV_m 当前时隙 t 时的三维位置可表示为 $M_m(t)=[x_m,y_m,z_m]$。其中 x_m、y_m 表示坐标中 UAV 的水平位置,z_m 为 UAV 的高度,GU 和 UAV 之间的仰角为 θ。UAV 作为空中基站为 GU 提供服务的模型图如图 1 所示。

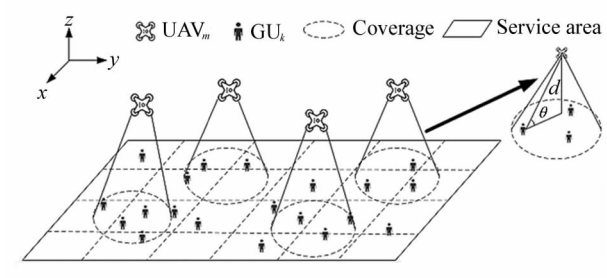


图 1　空地通信模型

Fig. 1　Air-to-ground communication model

1.2 通信模型

UAV 基站和 GU 的距离可表示为:

$$d_{m,k}(t)=\sqrt{(x_m-x_k)^2+(y_m-y_k)^2+z_m^2}。\tag{1}$$

当 UAV 和 GU 进行通信时,UAV 和用户之间

可近似为LOS传输路径。采用自由空间路径损耗模型，只需要考虑UAV和用户之间的距离即可，因此UAV_m（第m架UAV）和GU_k（第k用户）之间的信道增益可表示为：

$$g_{m,k}(t)=\rho_0 d_{m,k}^{-2}(t)=\frac{\rho_0}{(x_m-x_k)^2+(y_m-y_k)^2+z_m^2},\tag{2}$$

式中，ρ_0表示单位参考距离为1 m时的信道功率增益。

假设UAV在时隙t时的发射功率为$P_{m,k}(t)$，最大功率为$P_{\max}$，所有用户共享带宽，UAV的发射功率有如下约束：

$$0\leqslant P_{m,k}(t)\leqslant P_{\max},\forall m,k,t。\tag{3}$$

用户k接收信号的信干噪比可表示为：

$$\gamma_{m,k}(t)=\frac{P_{m,k}(t)g_{m,k}(t)}{\sum\limits_{j=1,j\neq m}^{M}P_{j,k}(t)g_{j,k}(t)+\sigma^2},\tag{4}$$

式中，σ^2是用户处的加性高斯白噪声的功率，$\sum_{j=1,j\neq m}^{M}P_{j,k}(t)g_{j,k}(t)$是时隙$t$时的传输同信道干扰。设$\gamma'$代表满足通信需求时信干噪比的阈值，则$\gamma_{m,k}(t)\geqslant\gamma'$。

在时隙t时，GU和UAV之间可实现的传输速率可表示为：

$$r_{m,k}(t)=\log_2(1+\gamma_{m,k}(t))。\tag{5}$$

单个用户的吞吐量可由传输速率表示：

$$\hat{R}_k(t)=\int_0^t r_{m,k}(t)\mathrm{d}t。\tag{6}$$

无人机的信道总容量为传输速率之和，即：

$$R_{m,k}(t)=\sum_{m=1}^{M}\sum_{k=1}^{K}r_{m,k}(t)。\tag{7}$$

因此，在UAV的服务时间T内的系统吞吐量可表示为：

$$\hat{R}_{m,k}(t)=\int_0^T R_{m,k}(t)\mathrm{d}t。\tag{8}$$

由于UAV基站采用TDMA方式为用户提供服务，为了最大化通信服务质量，每个用户在一个时隙最多被一个UAV服务，且每个GU在每个时隙至少接入一个UAV，可引入二进制变量$\alpha_{m,k}(t)$，当$\alpha_{m,k}(t)=1$时，表示接入；$\alpha_{m,k}(t)=0$，则表示未接入：

$$\begin{cases}\alpha_{m,k}(t)=1\\ \alpha_{m,k}(t)=0\end{cases}。\tag{9}$$

除此之外，UAV之间的距离和UAV的服务位置有如下约束：

$$\begin{aligned}&d(t)=M_j(t)-M_i(t)=\\&\sqrt{(x_j-x_i)^2+(y_j-y_i)^2},\\&d_{i,j}(t)\geqslant d_{\min},\\&M(t)\in D,\end{aligned}\tag{10}$$

式中，$M_j(t)$、$M_i(t)$为不同UAV的位置坐标，$d(t)$为相邻UAV之间的距离，$d_{\min}$为UAV之间的安全距离，D为UAV的服务区域。

1.3 问题描述

为了提高UAV基站和GU之间的通信服务质量，本文的目标是通过联合优化UAV基站的功率分配和用户接入来最大化系统的最大吞吐量，该问题可描述如下：

$$\begin{aligned}&P1:\max_{P_{m,k}(t),\alpha_{m,k}(t)}\ \hat{R}_{m,k}(t)\\&\text{s.t. }\hat{R}_{m,k}(t)=\int_0^T R_{m,k}(t)\mathrm{d}t,\\&C1:0\leqslant P_{m,k}(t)\leqslant P_{\max},\forall m,k,t,\\&C2:\ \gamma_{m,k}(t)\geqslant\gamma',\forall m,k,t,\\&C3:\alpha_{m,k}(t)\in[0,1],\forall m,k,t,\\&C4:d(t)\geqslant d_{\min},\\&C5:\ M(t)\in D,\end{aligned}\tag{11}$$

式中，$P1$是优化目标，$C1$表示基站的功率约束，$C2$是满足通信需求时的信干噪比阈值，$C3$表示用户接入性，$C4$、$C5$是对UAV位置的约束。

在整个服务过程中，还需要保证在每个时隙t时所有用户的通信公平性，因此引入Jain公平指数[17]，将用户的吞吐量比率定义为：$f(t)=\frac{\hat{R}_k(t)}{\hat{R}_{m,k}(t)}$，则服务时间$T$内的公平指数可表示为：

$$\hat{f}(t)=\frac{[\sum\limits_{k=1}^{K}f(t)]^2}{K[\sum\limits_{k=1}^{K}f(t)^2]}。\tag{12}$$

由此可知，$0\leqslant\hat{f}(t)\leqslant1$，且$\hat{f}(t)$越大，用户通信的公平性越高。在整个任务期间的公平吞吐量可定义为：

$$R_f(t)=\int_0^T\hat{f}(t)R_{m,k}(t)\mathrm{d}t。\tag{13}$$

2 基于MADDPG的UAV辅助通信资源分配算法

2.1 深度确定性策略梯度算法

深度确定性策略梯度算法（deep deterministic policy gradient algorithm，DDPG）算法采用行动者-评论家（actor-critic，AC）网络结构，Critic在Actor

的策略下计算动作值函数，通过深度神经网络（deep neural network，DNN）逼近行为值函数 $Q(s,a)$（Critic 网络）和 $\mu_\theta(s)$（Actor 网络）。然而 DNN 应用于 DRL 时不稳定甚至会出现发散现象，因此通常使用经验回放机制解决此问题。DRL 从经验缓冲池中随机选取一小批样本存储在训练集中，随机样本打破了序列样本之间的相关性，稳定了训练过程。DDPG 算法拥有两个独立的网络，共涉及 4 个神经网络：Critic 目标网络（target）Q' 和 Critic 当前网络（online）Q，Actor 目标网络（target）μ' 和 Actor 当前网络（online）μ。

Critic 当前网络 Q 和 Actor 当前网络 μ 具有相同的体系结构，采用时序差分（temporal-difference，TD）方式更新，损失函数为最小化均方误差：

$$L=\frac{1}{N}\sum_i[y_i-Q(s_i,a_i\mid\theta_Q)]^2\text{，}\tag{14}$$

式中，$y_i=r_i+Q'(s_{i+1},\mu'(s_{i+1}\mid\theta_{\mu'})\mid\theta_{Q'})$，$N$ 为批处理样本大小。更新 y_i 的计算过程用到了 Critic 目标网络 Q' 和 Actor 目标网络 μ'，这样做使得 Q 网络参数的学习过程更加稳定且易于收敛。通过损失函数 L 可以基于后向传播方法求得针对 θ_Q 的梯度 $\nabla_{\theta_Q}L$，对其优化更新得到 θ_Q。

Actor 当前网络 μ 的输入为当前状态，输出为当略梯度定理，式为：

$$\nabla_{\theta_\mu\mu}|s_i=\frac{1}{N}\sum_i\nabla_aQ(s,a|\theta_Q)|_{s=s_i,a=\mu(s_i)}\nabla_{\theta_\mu\mu}(s\mid\theta_\mu)|_{s=s_i}\text{。}\tag{15}$$

Critic 目标网络 Q' 和 Actor 目标网络 μ' 用于计算更新目标，采用滑动平均方式更新：

$$\theta_{Q'}\leftarrow\tau\theta_Q+(1-\tau)\theta_{Q'}\text{，}$$
$$\theta_{\mu'}\leftarrow\tau\theta_\mu+(1-\tau)\theta_{\mu'}\text{。}\tag{16}$$

DDPG 算法通过 Critic 网络评价 Actor 网络动作的好坏来指导值函数和策略函数进行梯度更新，使用贝尔曼方程可将系统确定性策略的值函数表示为：

$$Q_\mu(s_t,a_t)=E_{s_{t+1}\sim E}[r(s_t,a_t)+\gamma Q_\mu(s_{t+1},\mu(s_{t+1}))]\text{，}\tag{17}$$

式中，$\gamma\in[0,1]$ 是折扣因子，用于衡量当前奖励和未来奖励的重要性，$Q_\mu(s_{t+1},\mu(s_{t+1}))$ 为 Critic 目标网络对下一时刻状态 s_{t+1} 的评价和 Actor 目标网络对下一时刻所选动作 $\mu(s_{t+1})$ 的评价，$r(s_t,a_t)$ 是系统的奖励函数。由于动作－状态函数的 Q 值与环境有关，因此通过 Q 值迭代可以求解最优策略为：

$$\nabla_{\theta_\mu\mu}=E_{\mu'}[\nabla_aQ(s,a\mid\theta_Q)|_{s=s_t,a=\mu(s_t)}\nabla_{\theta_\mu\mu}(s\mid\theta_\mu)|_{s=s_t}]\text{。}\tag{18}$$

DDPG 的算法框架如图 2 所示。

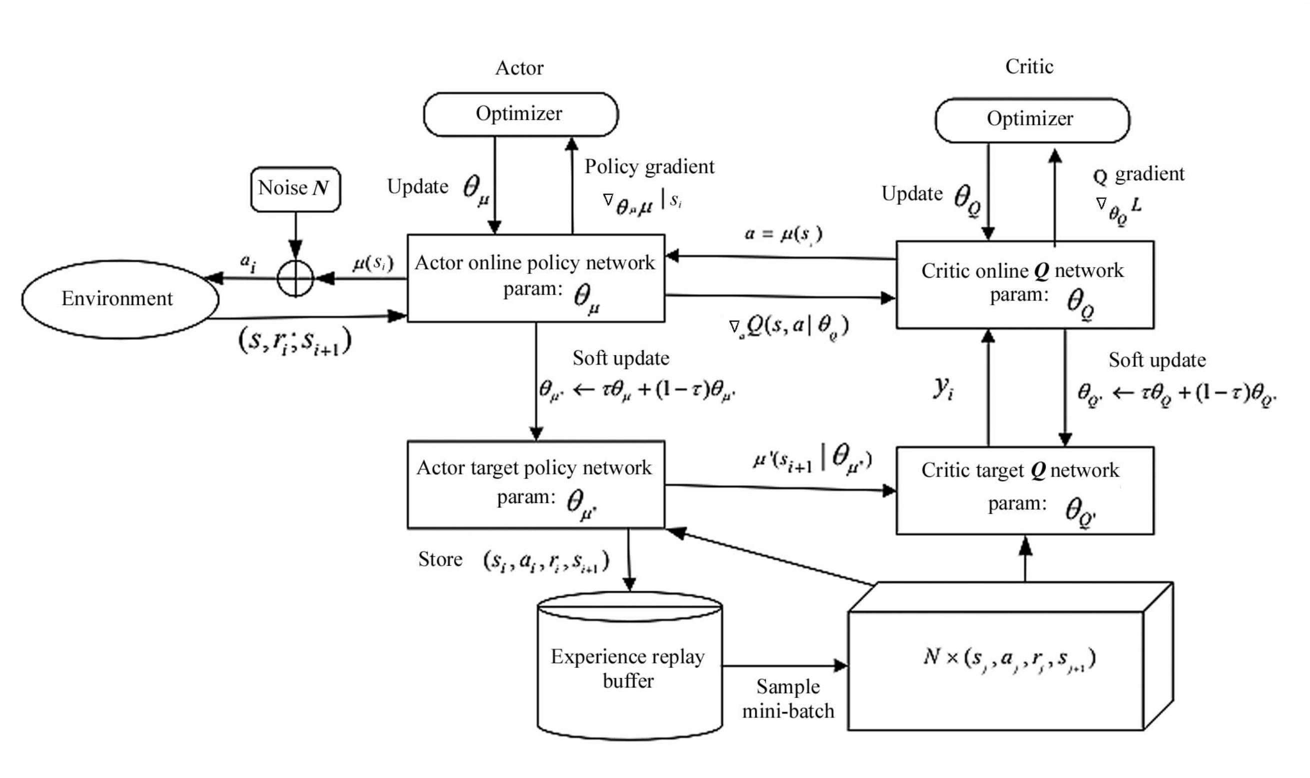


图 2　DDPG 算法框架

Fig. 2　Diagram of DDPG

2.2 状态空间

UAV通过下行链路对GU提供通信服务，因此UAV和GU之间的吞吐量只与它们之间的路径损耗有关，本文中算法的状态空间由GUs、UAVs和环境组成，在时隙 t 时的系统状态可定义为 $s_t=\{M(t),K(t),D,d\}$，$M(t)$ 代表UAV的位置，$K(t)$ 表示用户的位置，D 和 d 为UAV的轨迹信息，其中 D 为UAV的工作区域，d 为UAV之间的距离。UAV的距离必须大于最小安全距离，否则可能会发生碰撞。

2.3 动作空间

基于系统的当前状态和环境，UAV的动作空间可表示为 $a_t=\{P_{m,k}(t),\alpha_{m,k}(t)\}$，包括功率分配和用户的接入。UAV基站需要基于系统的当前状态和观察到的环境选择动作。

2.4 奖励函数设置

在DRL中，智能体的行为是基于奖励函数的，适当的奖励函数对智能体的动作起着至关重要的作用，本文针对UAV基站的功率分配和用户关联性进行优化，实现最大化系统吞吐量的目标，因此奖励函数可构建如下：

$$r_t=r(s_t,a_t)=k_t\hat{R}_{m,k}(t)+r'_t, \tag{19}$$

式中，k_t 是一个常数，用来调整使吞吐量最大化部分的报酬，r'_t 代表惩罚，如果UAV违反了约束条件（如飞出任务范围或者发生碰撞等），则累积奖励会收到一个负回报作为惩罚，通过获取累积奖励得到优化目标。

2.5 算法伪代码

在DDPG算法中，训练行为策略的Critic网络参数和Actor网络参数通过迭代更新。Actor网络通过DNN的权重 θ_μ 生成动作，然后从环境中观察当前状态 s_t，得到奖励值 r_t，并更新环境状态为 s_{t+1}。算法的计算复杂度取决于DNN的大小（输入、输出和隐藏层）以及UAV和GUs的数量。本文基于MADDPG的UAV功率分配算法如算法1所示：

1) 输入：训练片段长度 E、服务长度 T、Critic网络 $Q(s,a\mid\theta_Q)$ 的参数 θ_Q、Actor网络 $\mu(s\mid\theta_\mu)$ 的参数 θ_μ、折扣因子 γ、软件更新因子 τ、经验缓冲池 B、最小批处理大小 N、高斯分布噪声 $\boldsymbol{N}$

2) 输出：最优网络参数 θ_Q 和 θ_μ 以及最优策略

3) 随机初始化Actor网络的权重 θ_μ 和Critic网络 θ_Q 的权重

4) 将Critic、Actor的参数拷贝给对应的目标网络 Q' 和 μ' 的参数：$\theta_{Q'}\leftarrow\theta_Q$，$\theta_{\mu'}\leftarrow\theta_\mu$

5) 清空经验缓冲池 B

6) for each episode $e=1,2,\cdots,E$：

7) 初始化GUs和UAV的位置，得到初始状态 s_1

8) for $t=1,2,\cdots,T$ do

9) 根据噪声和当前策略得到行为 $a_t=\mu(s_t\mid\theta_\mu)+\boldsymbol{N}$

10) UAV执行动作 a_t 得到回报 r_t 和状态 s_{t+1}

11) if 缓冲池未满

12) 则得到转换序列 (s_t,a_t,r_t,s_{t+1}) 存储在 B 中

13) else

14) 将得到的转换序列随机替代缓冲池内的一组序列

15) 从缓冲池中随机抽取一批序列作为Critic和Actor online策略网络的训练数据

16) 令 $y_i=r_i+Q'(s_{i+1},\mu'(s_{i+1}\mid\theta_{\mu'})\mid\theta_{Q'})$

17) 通过最小化损失函数 L 更新online Critic的网络参数 θ_Q

18) $L=\frac{1}{N}\sum_i(y_i-Q(s_i,a_i\mid\theta_Q))^2$

19) 通过计算样本策略梯度 $\nabla_{\theta_\mu\mu}\mid s_i$，更新online Actor网络参数 θ_μ：

20) $\nabla_{\theta_{\mu\mu}}\mid s_i=\frac{1}{N}\sum_i\nabla_a Q(s,a\mid\theta_Q)\mid_{s=s_i,a=\mu(s_i)}$
$\nabla_{\theta_{\mu\mu}}(s\mid\theta_\mu)\mid_{s=s_i}$

21) 通过滑动平均更新目标网络参数 $\theta_{Q'}$，$\theta_{\mu'}$

22) $\theta_{Q'}\leftarrow\tau\theta_Q+(1-\tau)\theta_{Q'}$
$\theta_{\mu'}\leftarrow\tau\theta_\mu+(1-\tau)\theta_{\mu'}$

23) end for

24) end for

3 仿　真

为了验证算法的有效性，本文通过Python3.8和Tensor Flow2.0环境进行仿真以评估算法可行性及效率。本节首先设置了仿真参数并确定了最佳算法参数，然后将本文提出的算法与异步优势行动者—评论家算法（asynchronous advantage actor-critic，A3C），DQN算法和Greedy算法进行仿真对比，并对结果进行了分析。

3.1 仿真参数设置

本文考虑了一个1000 m×1000 m×100 m的三维场景，GU随机分布在水平高度为0的地面上，UAV的飞行速度固定为20 m/s，随机出现在高度固

定为 100 m 的空中,通过仿真测试了 UAV 和用户数量对系统公平性的影响,并在 UAV 和用户数量一定的情况下对系统的总吞吐量和公平吞吐量进行了仿真测试,具体的仿真数值如表 1 所示。

表 1　仿真参数设置

Tab. 1　Simulation parameter setting

Parameter	Meaning	Value
V	Flight speed	20 m/s
$P_{\max}$	Maximum power	1 W
ρ_0	Channel gain	−30dB
σ^2	Gaussian noise	−100 dB
$d_{\min}$	Safety distance	100 m
γ	SINR threshold	18 dB
B	Buffer size	1 000
E	Episode	1 000
N	Batch size	64

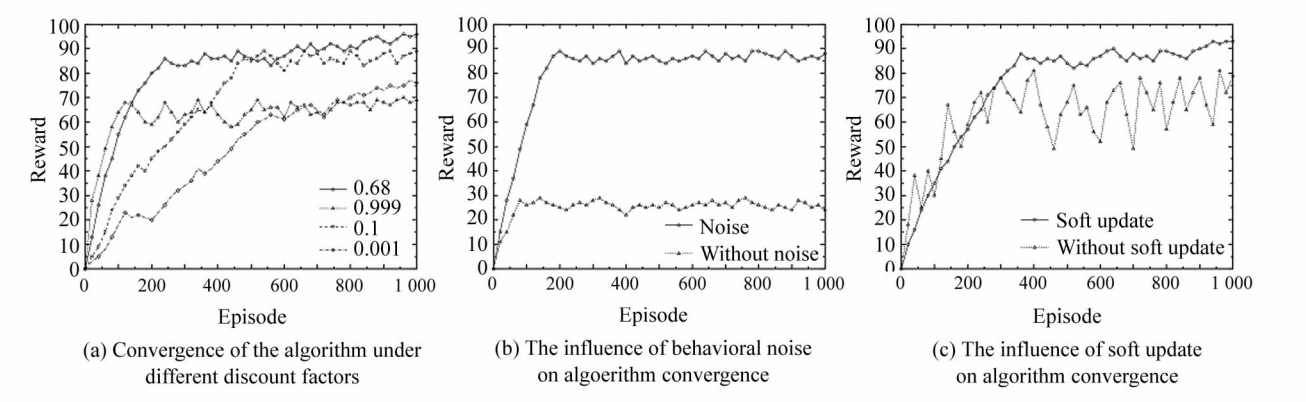
3.2　算法参数分析

在强化学习中,算法的参数对性能的影响很大,本文在不同折扣因子、行为噪声函数和软更新参数情况下对 DDPG 算法的收敛性进行了仿真,以确定最佳的算法参数。

图 3 显示了本文所使用的算法在不同参数下的收敛性能,图 3(a)所示为多组不同的折扣因子对算法累计期望回报和收敛性的影响,由图可知,当折扣因子较小(0.001)时,累计期望回报达到收敛的时间太长,学习效率太低;当折扣因子较大(0.99)时,可能会使得累计期望陷入局部最优,无法达到最大值。因此,本文在测试了多组折扣因子之后,折扣因子的值最终选择 0.68。

图 3(b)所示为探索噪声对算法训练的影响,由于 DDPG 算法所采用的是确定性策略,输出的动作是确定性动作,因此智能体探索能力较低,通过给确定性策略添加噪声构建行为网络,可以保证算法的高效探索性;由图可知,通过增加探索噪声,算法的收敛速度大大提高。如果没有行为噪声,算法的探索性较低,累计期望达不到最大值,得不到最优的行为策略。

图 3(c)所示的是有无软更新对累计期望回报的影响,由图可知,如果没有软更新,直接将参数复制给目标网络的参数,会引起算法的剧烈变化,不够稳定,加入了软更新之后目标网络只能缓慢变化,提高了学算法的稳定性。



(a) Convergence of the algorithm under different discount factors　(b) The influence of behavioral noise on algoerithm convergence　(c) The influence of soft update on algorithm convergence

图 3　不同参数下算法的性能比较

Fig. 3　Performance comparison of algorithms under different parameters

3.3　不同算法公平性对比

在本节中,通过设置不同数量的 GU 和 UAV 基站,对比了不同算法的公平性。

如图 4(a)所示,UAV 数量固定为 2 时,随着用户数量的增加,所有算法的公平性都在逐渐降低,因为随着用户的增加,UAV 基站不能很好地兼顾所有的 GU,导致用户的公平性是逐渐降低的,其中本文基于 MADDPG 的优化算法与其他几种基准算法相比,虽然公平性也呈下降趋势,但是明显可以看出公平指数下降幅度最小,能稳定在 0.95 左右。

由图 4(b)所示,当固定 GU 数量为 10,随着 UAV 基站数量的增加,所有算法的公平性都在逐步上升,较多的基站意味着用户可接入的选择变多,本文基于 MADDPG 的优化算法公平指数最高,虽然增长速度最慢,但是能够稳定保持在 0.95 左右,其他几种算法的增幅随着 UAV 数量的增加都逐渐趋于稳定,没有本文所采用的基于 DDPG 的优化算法公平性指数高。

图 4(c)所示为当固定用户数量为 20,基站数量为 4 时,随着迭代次数的增加算法的公平性比较。

由图可知，随着迭代次数的增加，所有算法的公平性都在缓慢上升然后趋于稳定，本文采用的基于 MADDPG 的优化算法公平性最高，优于其他的几种算法。

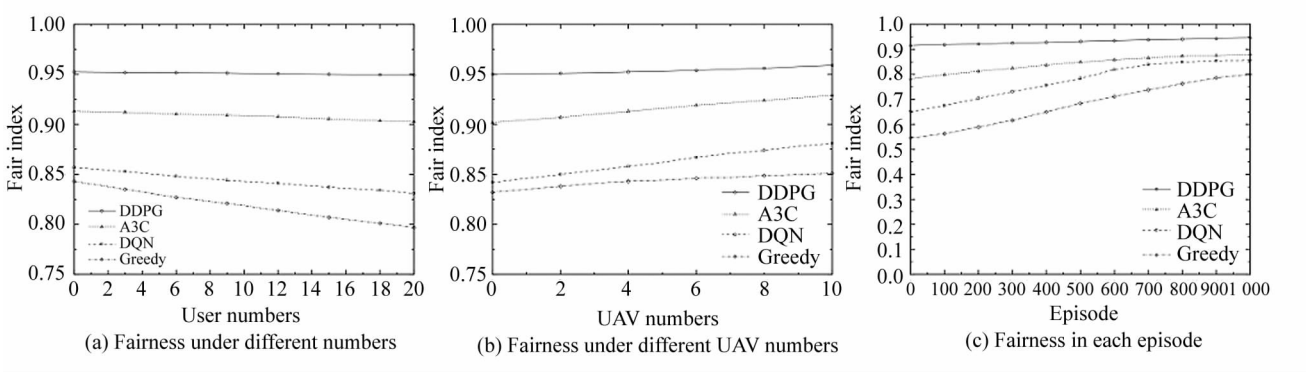


图 4　不同算法的公平性比较

Fig. 4　Comparison of fairness under different algorithms

3.4　不同算法系统吞吐量对比

本文考虑 4 个 UAV 基站，20 个 GU 的场景下，使用不同算法得到系统吞吐量和公平吞吐量的对比。

图 5 所示为不同算法的系统吞吐量对比，由图可以看出，随着迭代次数的增加，所有算法的吞吐量都是呈先快速上升至慢慢变缓的趋势，本文所提出的基于 MADDPG 算法的优化算法在迭代次数为 400 左右时逐渐开始收敛，且吞吐量高于其他几种算法；A3C、DQN、Greedy 算法分别在迭代次数为 550、650、600 左右的时候才逐渐收敛，其中 Greedy 虽然收敛速度比 DQN 算法快，但是其吞吐量略微弱于 DQN 算法。DDPG 算法结合了 DQN 算法和 AC 算法的优势，通过确定性策略更新动作，算法速率高，收敛较快，但是确定性策略使得智能体不能很好地遍历所有状态空间，容易陷入局部最优。因此通过增加探索噪声和采用 AC 框架可以让智能体跳出局部最优、朝着全局最优的方向收敛。

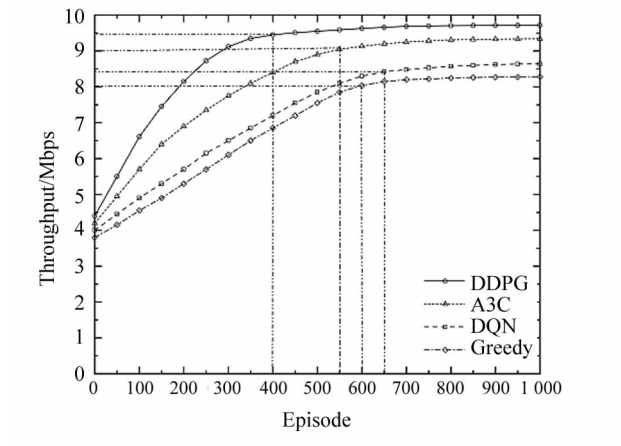


图 5　不同算法吞吐量对比

Fig. 5　Comparison of throughput under different algorithms

图 6 所示为不同算法的公平吞吐量对比，由图可以看出，随着迭代次数的增加，所有优化算法的公平吞吐量都是呈先上升后变缓的趋势，其中本文基于 MADDPG 的优化算法公平吞吐量最高，A3C 算法次之，接下来是 DQN 算法，Greedy 算法的公平性最低。

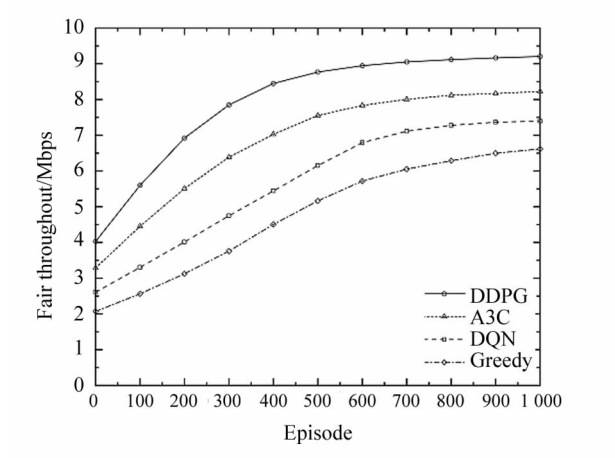


图 6　不同算法公平吞吐量对比

Fig. 6　Comparison of fair throughput under different algorithms

4　结　论

本文针对多 UAV 辅助通信过程中功率分配的问题，提出了一种基于 MADDPG 的功率分配算法，该算法联合 UAV 基站的功率分配和用户关联性，优化了 UAV 辅助通信过程中的吞吐量和公平性。通过对比现有的几种算法验证了该算法的有效性。

但是本文仍有几个需要后续研究的地方，首先

本文的仿真结果只考虑了 UAV 和用户之间只有自由空间路径损耗的场景,在实际场景中还需要考虑到其他的衰落。除此之外,还需要考虑 UAV 本身的能耗和 UAV 移动过程中的障碍物,这对整个系统的性能有很大的影响。在未来的工作中,还将考虑这些影响因素并进一步尝试其他的最新算法来解决此问题。

参考文献:

[1] ZHANG Z,XIONG T B,CHEN J Q,et al. Research on the spatial characterization of a 3D UAV air-to-ground channel model[J]. Journal on Communications,2020,41(2):123-130.
张治,熊天波,陈建侨,等.无人机三维空地信道模型的空间特性研究[J]. 通信学报, 2020, 41(2):123-130.

[2] CHEN X Y,SHENG M,LI B,et al. Survey on unmanned aerial vehicle communications for 6G[J]. Journal of Electronics & Information Technology,2022,44(3):781-789.
陈新颖,盛敏,李博,等.面向 6G 的无人机通信综述[J].电子与信息学报,2022,44(3):781-789.

[3] WU Q,MEI W,ZHANG R. Safeguarding wireless network with UAVs:a physical layer security perspective[J]. IEEE Wireless Communications,2019,26(5):12-18.

[4] YOU X H,WANG C X,HUANG J,et al. Towards 6G wireless communication networks:Vision, enabling technologies, and new paradigm shifts[J]. Science China (Information Science),2021,64(1):5-78.

[5] CUI F Y,CAI Y L,ZHAO J M. Joint trajectory design and power allocation for NOMA-based mobile-UAV communication networks[J]. Journal of Hangzhou Dianzi University (Natural Sciences),2020,40(1):14-20.
崔方宇, 蔡云龙, 赵民建.基于 NOMA 的无人机轨迹与功率联合优化[J]. 杭州电子科技大学学报(自然科学版), 2020,40(1):14-20.

[6] LIU M,GUI G,ZHAO N,et al. UAV-aided air-to-ground cooperative non-orthogonal multiple access[J]. IEEE Internet of Things Journal,2019,7(4):2704-2715.

[7] LANG L,WANG J N,WANG Y,et al. Radio resource and trajectory optimization for UAV assisted communication based on user route[J]. Journal on Communications, 2022,43(3):225-232.
郎磊,王荆宁,王一,等.无人机辅助通信中基于用户轨迹的无线资源和航迹优化[J].通信学报,2022,43(3):225-232.

[8] WANG J,LIU M,SUN J,et al. Multiple unmanned aerial vehicles deployment and user pairing for non-orthogonal multiple access schemes[J]. IEEE Internet of Things Journal,2020,8(3):1883-1895.

[9] HU D,ZHANG Q,LI Q,et al. Joint position,decoding order,and power allocation optimization in UAV based NOMA downlink communications[J]. IEEE Systems Journal, 2019,14(2):2949-2960.

[10] LI G Q,LIN J Z,XU Y J,et al. User grouping and power allocation algorithm for UAV-aided NOMA network[J]. Journal on Communications,2020,41(9):21-28.
李国权,林金朝,徐勇军,等.无人机辅助的 NOMA 网络用户分组与功率分配算法[J]. 通信学报,2020,41(9):21-28.

[11] WU G H,JIA W M,ZHAO J W,et al. MARL-based design of multi-unmanned aerial vehicle assisted communication system with hybrid gaming mode[J]. Journal of Electronics & Information Technology,2021,43(3):940-950.
吴官翰, 贾维敏, 赵建伟,等.基于多智能体强化学习的混合博弈模式下多无人机辅助通信系统设计[J].电子与信息学报,2021,43(3):940-950.

[12] LIANG Y C,TAN J J,NIYATO D,et al. Overview on intelligent wireless communication technology[J]. Journal on Communications,2020,41(7):1-17.
梁应敞, 谭俊杰,NIYATO D,等.智能无线通信技术研究概况[J].通信学报,2020,41(7):1-17.

[13] ZHANG Z C,FU F,YI Z H. Research on resource allocation based on energy efficiency in UAV system[J]. Journal of Test and Measurement Technology, 2021, 35(6):503-507.
张志才, 付芳, 尹振华.无人机系统中基于能量效率的资源分配研究[J].测试技术学报, 2021, 35(6):503-507.

[14] ZHANG G C,YAN Y L,CUI M, et al. Online trajectory optimization for the UAV-enabled base station multicasting system based on reinforcement learning[J]. Journal of Electronics & Information Technology,2022,44(3):969-975.
张广驰,严雨琳,崔苗,等.基于强化学习的无人机基站多播通信系统的飞行路线在线优化[J].电子与信息学报,2022,44(3):969-975.

[15] SUN C Y,MU C X. Important scientific problems of multi-agent deep reinforcement learning[J]. Acta Automatica Sinica,2020,46(7):1301-1312.
孙长银, 穆朝絮.多智能体深度强化学习的若干关键科学问题[J].自动化学报, 2020, 46(7):1301-1312.

[16] LI B,YUE K Q,GAN Z G,et al. Multi-UAV cooperative autonomous navigation based on multigent deep deterministic policy gradient[J]. Journal of Astronautics,2021,42(6):757-765.
李波, 越凯强, 甘志刚,等.基于 MADDPG 的多无人机协同任务决策[J].宇航学报, 2021, 42(6):757-765.

[17] CHI H L,CHEN Z,TANG J,et al. Energy-efficient UAV control for effective and fair communication coverage:a deep reinforcement learning approach[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(9):2059-2070.

作者简介:

杨青青　(1981—),女,博士,讲师,硕士生导师,主要从事信息处理、应急通信方面的研究.