一种基于公平性的无人机基站通信智能资源调度方法

Intelligent Resource Allocation Based on Fairness for UAV Base Station Communications

吴官翰/WU Guanhan[1]，赵建伟/ZHAO Jianwei[1]，高飞飞/GAO Feifei[2]
（1. 火箭军工程大学，中国 西安710038；
2. 清华大学，中国 北京 100084）
(1. Rocket Force University of Engineering, Xi'an 710038, China)
(2. Tsinghua University, Beijing 100084, China)

摘要：空天地一体化网络是未来6G的关键内容。结合高精度波束赋形，无人机（UAV）的视距链路（LoS）可很好地作为空天地一体化网络的补充，但地面用户与基站间的相对运动极易造成信道容量失衡。提出一种噪声深度确定性策略梯度（Noisy-DDPG）方法。该方法以最大化通信公平性和系统容量为目标，利用DDPG优化分配方案，通过调整奖励函数策略参数来实现公平性和信道容量的平衡；通过在策略网络中利用可学习参数噪声进行扰动，得到更合理的分配方案。仿真实验验证了该算法的有效性。

关键词：无人机基站；资源调度；DDPG；公平通信；参数噪声

Abstract: The space-air-ground integrated network is an important part of the future 6G, which can be well complemented by the unmanned aerial vehicle's (UAV) line-of-sight (LoS) link combined with high-precision beamforming. However, the random channel characteristics of mobile users can easily cause channel capacity imbalance. In this paper, the Noisy-Deep Deterministic Policy Gradient (Noisy-DDPG) is proposed. To maximize communication fairness and system capacity, the Deep Deterministic Policy Gradient (DDPG) is used to optimize the allocation strategy. Besides, fairness and channel capacity are differently emphasized by adjusting the reward function policy parameters. Moreover, the learnable parameter noise is used to disturb the policy network to obtain a more reasonable allocation plan. Finally, various simulation results to verify the effectiveness of the algorithm are proposed.

Keywords: UAV base station; resource allocation; DDPG; fair communication; parameter noise

DOI：10.12142/ZTETJ.202102007
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20210401.1717.011.html

网络出版日期：2021-04-02
收稿日期：2021-02-19

研究表明，预计到2030年，以固定基站为主的5G移动通信将无法满足日益增长的数据业务需求，大量新生业务将产生海量数据资源，物理世界与数字世界之间的界限将更为模糊[1]。在此背景下，集中式的数据处理中心将承受更为巨大的压力，遥远的云端服务器也不利于满足远端用户低时延的数据处理需求。

无人机作为移动载体，可搭载5G/超5G（B5G）通信基站或边缘服务器，并结合高精度的波束赋形形成指向性强、增益高的窄波束，以减少邻居干扰，有效克服毫米波及以上频段射频信号衰减巨大这一现实问题[2-3]。将无人机基站作为未来空天地一体化网络的中间节点，卸载部分用户通信与计算任务，成为一种具有潜力的组网方式[4]。在一个无人机小区中，用户的随机移动将带来不可预知的动态拓扑结构，基站单一的带宽、功率分配策略往往会造成小区内信道容量失衡。一种合理的通信资源分配机制能够有效提升用户通信的公平性，并最大化系统平均信道容量。

近年来，人工智能在自动控制、目标识别、语义识别等领域大放异

彩，极大地推动了各行业的进步与发展。将人工智能与通信技术有机结合，是未来5G和6G的一个发展方向。深度强化学习（DRL）[5-7]具有强大的特征提取和多维决策能力，能够针对通信资源多维度的特点，做出最明智的动作决策，为无人机基站资源调度决策提供了可能[8-10]。现有研究大多集中于无人机路径规划、系统信道容量优化等方面，以满足用户最低服务质量（QoS）需求，但未考虑用户通信的公平性需求[8]和移动通信本身巨大的能量消耗[9]。

为解决信道容量失衡的问题，本文提出了一种基于公平性的噪声深度确定性策略梯度（Noisy-DDPG）无人机基站功率、带宽调度方法，在传统DDPG基础上结合可学习参数噪声扰动方式进行前期探索，使噪声方差依据梯度下降自适应调整；通过将训练好的策略模型用于无人机基站通信的实时部署，为任意分布的地面用户提供合理的通信资源分配方案。和传统DDPG训练方式相比，Noisy-DDPG表现出更优秀的性能：在达到相同公平指数条件下可以获得更高的系统平均信道容量。

1 系统模型与问题建模

如图1所示，在边长为D的正方形区域内，我们将单无人机基站悬停在目标区域上空，从功率和带宽两个维度，对N个运动的地面用户进行动态资源调度，在满足用户公平通信的同时最大化平均信道容量。我们定义P_{total}和B_{total}分别为无人机基站总发射功率和可用带宽，以频分复用方式对用户进行带宽分配，在保证公平性的同时最大化平均信道容量。

1.1 空地信道模型

在不同的部署环境下，搭载阵列天线的无人机在空中对用户n进行信号传输。视距（LoS）链路的概率为：

$$P_n^{\text{LoS}}(t)=\frac{1}{1+ae^{-b\left(\arcsin\left(\frac{h}{d_n(t)}\right)-a\right)}},\tag{1}$$

其中，a、b为环境相关参数，h为无人机部署高度，$d_n(t)$为t时刻无人机到用户n的距离。

由于LoS和非视距（NLoS）链路下的路径损耗存在差异，我们通常使用$L_n^{\text{LoS}}(t)$和$L_n^{\text{NLoS}}(t)$分别表示t时刻无人机到用户n的LoS与NLoS路径损耗：

$$L_n^{\text{LoS}}(t)=20\log\left(\frac{4\pi f_c d_n(t)}{c}\right)+\xi_{\text{LoS}},\tag{2}$$

$$L_n^{\text{NLoS}}(t)=20\log\left(\frac{4\pi f_c d_n(t)}{c}\right)+\xi_{\text{NLoS}},\tag{3}$$

其中，f_c为信号载频，c为光速，ξ_{LoS}和ξ_{NLoS}分别为LoS与NLoS链路下的附加损耗。

t时刻无人机到用户n的路径损耗可以表示为：

$$L_n(t)=P_n^{\text{LoS}}(t)L_n^{\text{LoS}}(t)+(1-P_n^{\text{LoS}}(t))L_n^{\text{NLoS}}(t)。\tag{4}$$

我们定义无人机基站对用户n的发射功率为$P_n(t)$，带宽分配为$B_n(t)$。由于采用了5G高指向性的波束赋形和干扰抑制技术，用户间的邻居干扰影响可以忽略。t时刻该用户的信道容量可表示为：

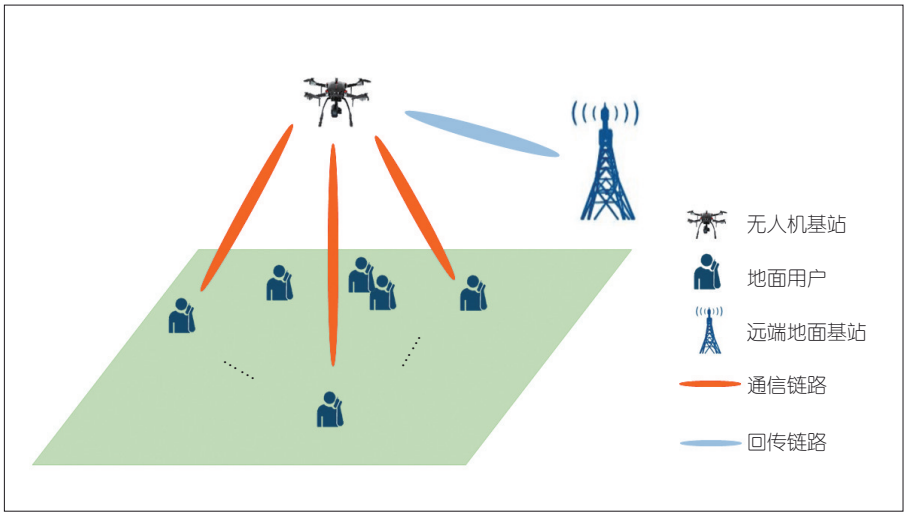
$$C_n(t)=B_n(t)\log\left(1+\frac{P_n(t)}{L_n(t)B_n(t)n_0}\right),\tag{5}$$

$$C_{\text{mean}}(t)=\frac{1}{N}\sum_{n=1}^{N}C_n(t),\tag{6}$$

其中，n_0表示高斯白噪声的功率谱密度，$C_{\text{mean}}(t)$为t时刻所有用户的平均信道容量，每个用户的信道容量为$P_n(t)$与$B_n(t)$的函数。

1.2 问题建模

当无人机基站对用户分配带宽、发射功率时，均等的分配方案会使边缘用户长期处于较低信道容量；而最大化系统总吞吐量的方式又可能会导致资源分配不公平，即基站只向最近的用户分配绝大部分带宽和功率以换取较高信道容量，而其他用户则陷入较低信道容量的困扰。公平通信是指在保证每个用户满足最低通



▲图1 无人机基站辅助通信场景

信需求下，尽量减少因远近效应带来的信道容量差异。我们通常用公平指数来衡量这种差异：

$$f(t)=\frac{\left(\sum_{n=1}^{N}f_n(t)\right)^2}{N\left(\sum_{n=1}^{N}f_n^2(t)\right)},\tag{7}$$

$$f_n(t)=\frac{C_n(t)}{\sum_{n=1}^{N}C_n(t)},\tag{8}$$

其中，$f(t)$表示t时刻通信系统的公平指数，$f(t)\in[0,1]$，$f(t)$越大代表该分配方案下用户的理论信道容量差异越小。在公平性准则下，最大化系统信道容量可定义优化目标为：

$$\max f(t),C_{\text{mean}}(t),\tag{9}$$

$$\text{s.t.}\sum_{n=1}^{N}P_n(t)=P_{\text{total}},\tag{10}$$

$$\sum_{n=1}^{N}B_n(t)=B_{\text{total}},\tag{11}$$

$$B_n(t)\geqslant B_{\min}。\tag{12}$$

公式（10）、（11）为总资源的约束，公式（12）为每个用户的最小带宽需求。值得注意的是，由于基站到每个用户的路径损耗不同，无人机在保证公平通信时势必会在资源分配上略微向边缘用户倾斜，从而提升其信道容量；但是从整个通信系统来看，这可能会影响平均信道容量的提升。因此，公平指数$f(t)$和平均信道容量C_{mean}无法同时达到最大化，在不同场景下便需要考虑不同的侧重。

2 基于Noisy-DDPG的资源调度算法

传统方法可采用遗传算法、粒子群算法、模拟退火算法等启发式算法解决以上的问题，但这一类算法一般用于通信资源实时调度，在每个时隙间隔均需要针对不同拓扑进行迭代。这不仅需要较多的计算开销，还需要大量的时间成本，不利于有实时性需求的优化。DRL是利用训练有素的神经网络模型，完成当前状态到最佳决策动作的直接映射，在实时控制决策方面具有优良特性。利用DRL的泛化能力能够处理未训练过的类似状态。

2.1 资源调度的MDP模型

强化学习是建立在马尔可夫决策过程（MDP）基础之上，通过优化(s_t,a_t,r_t,s_t')轨迹、最大化Bellman方程得到的累积奖励。其中，s_t为t时刻状态，a_t为决策动作，r_t为采取动作后的单步奖励，s_t'为采取动作后转移到的下一个状态。MDP通常由$(\boldsymbol{S},\boldsymbol{A},\boldsymbol{P},\boldsymbol{R},\gamma)$进行定义，$\boldsymbol{S}$为状态空间，$\boldsymbol{A}$为动作空间，$\boldsymbol{P}$为状态转移矩阵，$\boldsymbol{R}$为奖励空间，$\gamma$为折扣因子（代表智能体对未来奖励的重视程度）。

在无人机基站辅助通信的场景中，我们将一个回合内所有用户随机运动的过程建模为MDP模型。这样可以保证在公平指数较高的情况下，系统的平均信道容量可以实现最大化。具体来说，用户终端将与无人机基站进行关联，并将自身位置信息通过上行链路导频或全球定位系统（GPS）上传给无人机，无人机再依据自身位置计算出到用户n的直线距离。状态的定义需要包含此时区域内的拓扑情况，于是状态s_t可定义为：

$$s_t=(d_1,d_2,\cdots,d_N)。\tag{13}$$

$d_i,i\in[1,2,\cdots,N]$为用户i到无人机的直线距离，动作a_t表示在t时刻针对s_t状态下的通信资源分配策略：

$$a_t=(P_n,B_n),n\in[1,2,\cdots,N],\tag{14}$$

$$\text{s.t.}\sum_{n=1}^{N}P_n=P_{\text{total}},\tag{15}$$

$$\sum_{n=1}^{N}B_n=B_{\text{total}},B_n\geqslant B_{\min}。\tag{16}$$

为了满足公平通信条件下最大化平均信道容量的需求，同时遵循总功率和带宽的分配条件，输出的动作需要满足公式（15）、（16）的约束。我们用r_t来描述无人机基站对当前拓扑状态下资源分配策略的奖励得分。由于$f(t)$和$C_{\text{mean}}(t)$这两个指标无法同时达到最大化，为满足不同场景下对公平通信的需求，在训练模型时我们用参数λ表示对公平指数不同程度的侧重：

$$r_t=f(t)C_{\text{mean}}(t)+\lambda f(t)。\tag{17}$$

2.2 DDPG算法

深度Q网络（DQN）算法开创了DRL先例，即用神经网络解决无限维的状态映射问题。传统DQN这类基于价值的强化学习算法只能处理离散有限的动作空间，而DDPG算法利用Actor-Critic模式和确定性策略梯度的方式解决了连续动作空间输出的问题。

DDPG算法中定义了4个神经网络结构，Actor现实网络和Actor目标网络的结构相同，Critic现实网络和Critic目标网络的结构相同。作为策略网络，Actor网络用来为当前状态输出决策动作；作为评估网络，Critic网络用来评价Actor输出的策略，拟合$Q(s_t,a_t)$函数，并用θ^{μ}、$\theta^{\mu'}$分别表示Actor现实网络参数和目标网络参数，θ^{Q}和$\theta^{Q'}$分别表示Critic现实网络参数和目标网络参数。对于Critic网络的更新，我们定义其损失函数为：

$$J(\theta^Q)=\frac{1}{N_s}\sum_i[r_i+\gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})-Q(s_i,a_i|\theta^Q)]^2, \quad (18)$$

其中，N_s是从经验回放池中采样的数据批次大小，$\mu'(s_{i+1}|\theta^{\mu'})$为Actor目标网络对下一个状态的策略估计。以梯度下降的方式来更新Critic可以使其对Q值的评估更加准确。而Actor网络则采用梯度上升的方式更新如下的梯度目标函数：

$$\nabla J(\theta^\mu)=\frac{1}{N_s}\sum_i\nabla_a Q(s,a|\theta^Q)\big|_{s=s_i,a=\mu(s_i)}\nabla_{\theta^\mu}\mu(s_i|\theta^\mu)。 \quad (19)$$

2.3 Noisy-DDPG算法

在DRL的训练过程当中，通常需要在训练前期添加一定的不确定性来丰富经验，探索更充分的样本空间，以使智能体学习更加全面。当前，DDPG通常采用给输出的动作施加衰减的高斯噪声来进行探索。相比于在策略网络的输出中添加噪声的方式，在神经网络权重中添加参数化噪声能够实现更加全面的探索[5]。

我们将一种可学习的策略噪声添加到Actor的全连接层以实现探索与利用，结合这种策略噪声的算法被称为Noisy-DDPG。具体来说，不同于文献[5]中通过衡量添加噪声前后策略的差异来调整噪声方差，在策略网络中添加自适应参数噪声方法中的参数可以通过梯度下降进行学习，从而实现端到端调整。简而言之，Actor网络不仅需要学习网络参数，还需要学习生成噪声的方差：

$$\chi_k=\omega\chi_s+b。 \quad (20)$$

对于神经网络中的线性单元部分如公式(20)所示，χ_s和χ_k分别表示输入输出，ω为权重矩阵，b为偏置向量，施加参数化噪声后可表示为：

$$\chi_k=(\mu_\omega+\sigma_\omega\odot\varepsilon_\omega)\chi_s+\mu_b+\sigma_b\odot\varepsilon_b。 \quad (21)$$

公式(21)将ω和b分别用噪声数值来参数化，$\omega\sim\mathcal{N}(\mu_\omega,\sigma_\omega)$，$b\sim\mathcal{N}(\mu_b,\sigma_b)$，$\varepsilon_\omega$和$\varepsilon_b$为采样的高斯噪声$\varepsilon_{\omega,b}\sim\mathcal{N}(0,I)$。此时，神经网络需要通过学习噪声的均值和方差来自适应调整，并且重参数化的方式保证网络可微。

在采样高斯噪声方面，如果将每个神经网络参数作为独立高斯噪声进行采样，计算开销会随着网络规模的增大而迅速增加。相比于独立噪声，这种分解高斯噪声的方式不仅减少了噪声采样数量，还能在神经元数目较多时减少计算开销。具体来说，将上一层神经元数量设为s，下一层的神经元数量设为k，每个神经元分别生成一个独立的单位高斯噪声，即为$\varepsilon_i, i\in[1,2,\cdots,s]$，$\varepsilon_j, j\in[1,2,\cdots,k]$。此时，神经网络参数中添加的噪声可表示为：

$$\varepsilon_\omega=f(\varepsilon_j)f(\varepsilon_i^T), \quad (22)$$

$$\varepsilon_b=f(\varepsilon_j), \quad (23)$$

其中，$f(x)=\mathrm{sgn}(x)\sqrt{x}$。在定义Actor网络结构时，我们将含噪声方式的全连接层放在靠近输出层的位置，以便输出含噪动作。

3 仿真实验及对比分析

本次实验中，动作噪声是在输出的动作上直接加入不断衰减的高斯噪声，并使其满足公式(15)、(16)的约束，以形成合法决策。我们通过在训练前期引入动作的不确定性来获取更加多样性的经验样本。参数噪声是直接扰动神经网络参数，使网络通过学习自适应调整噪声参数。

3.1 实验参数设置

本次实验仿真模拟的是城市环境，主要参数设置为：$D=2$ km，$N=10$，$a=9.61$，$b=0.28$，$\xi_{\mathrm{LoS}}=1$ dB，$\xi_{\mathrm{NLoS}}=20$ dB，$n_0=10^{-17}$ W/Hz，$P_{\mathrm{total}}=1$ W，$B_{\mathrm{total}}=50$ MHz，$B_{\mathrm{min}}=1$ MHz，$f_c=2$ GHz。基于该算法训练策略模型时，我们将无人机基站设置在坐标为(1 000，1 000，500)的目标区域上空，进行$L=1\,000$个回合的迭代训练。每个训练回合包含$T=300$个阶段，即在每个阶段所有用户都能随机运动到一个新的位置，并将自身位置上传给无人机。在训练过程中，当每个回合开始时，我们都初始化环境，并在目标区域内随机生成N个用户分布。无人机计算每个用户到自己的距离并生成初始状态s_0，然后通过Actor网络生成的分配策略，与环境交互得到单步奖励r_t。由于分配策略需要包含功率分配和带宽分配信息，这里采用了不同方式输出动作分量：

$$P_n=\mathrm{softmax}(\chi_p)P_{\mathrm{total}}, \quad (24)$$

$$B_n=(B_{\mathrm{total}}-NB_{\mathrm{min}})\mathrm{softmax}(\chi_b)+B_{\mathrm{min}}, \quad (25)$$

其中，χ_p和χ_b分别表示功率分配输出层和带宽分配输出层的预激活变量，并由softmax激活函数归一化输出。公式(25)保证了在带宽分配时，每个用户的最小带宽需求B_{min}能得到满足。

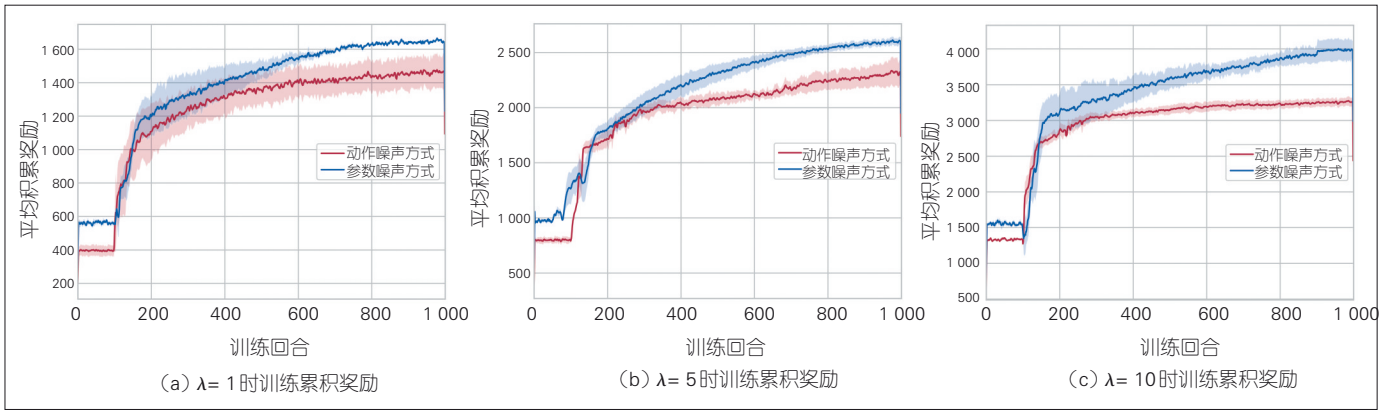
3.2 实验结果及对比分析

为满足不同场景对公平指数的要求，可在训练前设置不同的参数λ。λ越大，通信的公平性越能得到重视，用户之间的理论信道容量差异则越

小。但过大的λ则会导致无人机基站减少对通信系统信道容量的考虑，在追求高公平指数的同时在一定程度上忽略了系统内部资源合理调度对平均信道容量的影响。因此，在设置λ时需要调整好对信道容量公平性与系统平均信道容量的侧重关系。

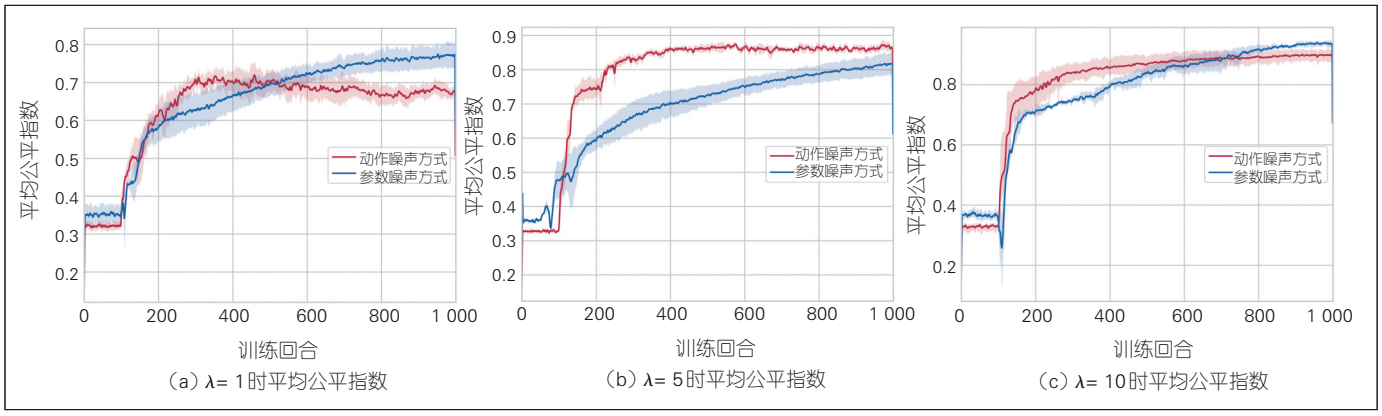
图2为训练累积奖励变化情况，图3与图4分别为在不同λ时收敛到的平均公平指数的变化情况和平均信道容量的变化情况。图2—4都反映了在λ=1、5、10时，无人机基站对通信系统公平性和平均信道容量的不同侧重。从图2可以发现，在所有λ下，参数噪声训练方式相对于传统动作噪声方式能收敛到更高的累积奖励，从而验证了所提方法的优越性。

在图3和图4中，λ=1时参数噪声方式的公平指数最终收敛到0.79，平均信道容量收敛到0.65 Mbit/s，动作噪声DDPG方式的公平指数收敛到0.68，平均信道容量收敛到0.64 Mbit/s。分析实验后我们发现：在训练后期，



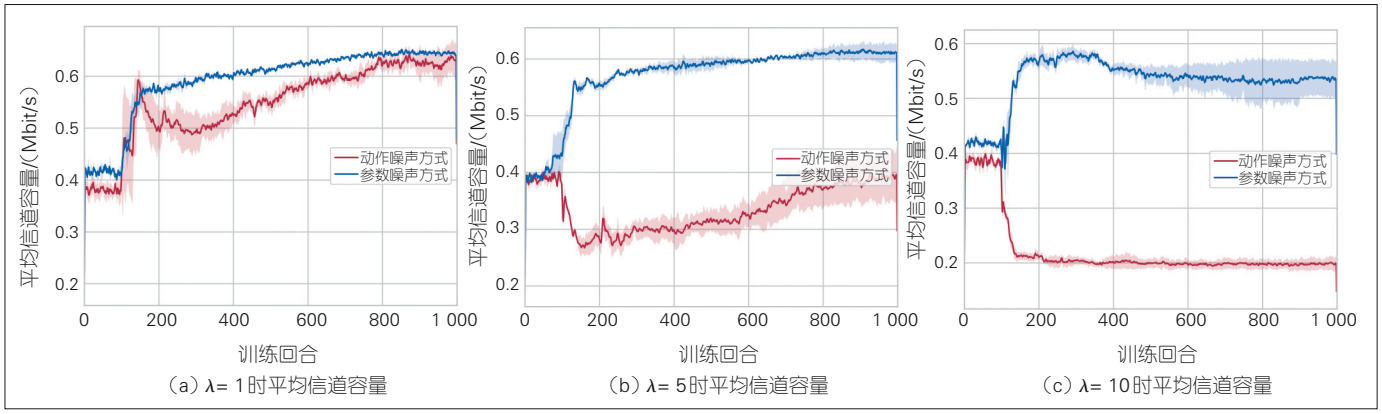
（a）λ=1时训练累积奖励　（b）λ=5时训练累积奖励　（c）λ=10时训练累积奖励

▲图2 训练过程回合累积奖励对比



（a）λ=1时平均公平指数　（b）λ=5时平均公平指数　（c）λ=10时平均公平指数

▲图3 训练过程回合平均公平指数对比



（a）λ=1时平均信道容量　（b）λ=5时平均信道容量　（c）λ=10时平均信道容量

▲图4 训练过程平均信道容量对比

动作噪声方式的公平指数有所下降，而平均信道容量有所回升。这是因为在训练后期，该方式放松了对公平指数的追求，将通信资源倾向于距离较近的用户，以公平性的下降换取平均信道容量的提升。反观参数噪声DDPG方式，各项指标提升均较为稳定。λ= 5时参数噪声DDPG方式的公平指数最终收敛到0.84，平均信道容量收敛到0.62 Mbit/s，而动作噪声DDPG方式的公平指数最终收敛到0.87，平均信道容量收敛到0.43 Mbit/s。可以发现，此时动作噪声的方式更倾向于追求公平指数带来的奖励，而忽略了系统内部资源合理调度对平均信道容量的影响。这导致动作噪声方式所收敛的累积奖励低于参数噪声方式，而参数噪声方式却能在保持公平指数的同时保证了平均信道容量的大小，很好地平衡了两者的关系。在λ=10时，两种方式均倾向于公平指数带来的高回报：参数噪声方式公平指数最终收敛到0.94，平均信道容量收敛到0.55 Mbit/s；动作噪声方式公平指数收敛到0.90，平均信道容量收敛到0.21 Mbit/s。因为追求高公平指数，两种方式在平均信道容量上均会有所降低。但是参数噪声方式显然在分配方面更加合理，在达到相同公平指数的前提下能够保证更好的平均信道容量，从而可以带来更高的收益。

4 结束语

针对未来空天地一体化网络中无人机辅助通信的多维资源调度公平性问题，本文提出了一种名为Noisy-DDPG的资源分配策略模型训练方法。这种方法适用于无人机搭载5G大规模天线阵列辅助地面移动通信的场景。在不同公平性需求下，通过调节奖励函数参数λ来实现公平指数与平均信道容量不同程度侧重，以使多维通信资源分配得更加合理高效。在模型训练时，采用一种可学习的自适应分解高斯噪声对输出策略进行扰动，使DDPG算法能够在训练中进行更深层次的探索。相比于传统动作噪声的探索方式，本文所提的方法能够获得更好的效果，仿真实验也进一步验证了方法的有效性。

致谢

本文的部分研究成果和撰写指导得到火箭军工程大学贾维敏教授的帮助与鼓励，在此谨致谢意！

参考文献

[1] 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 sciences, 2021, 64(1): 110301. DOI: 10.1007/s11432-020-2955-6
[2] MOZAFFARI M, SAAD W, BENNIS M, et al. Drone small cells in the clouds: design, deployment and performance analysis [C]//2015 IEEE Global Communications Conference (GLOBECOM). San Diego, CA, USA: IEEE, 2015: 1-6. DOI: 10.1109/GLOCOM.2015.7417609
[3] LI B, FEI Z S, ZHANG Y. UAV communications for 5G and beyond: recent advances and future trends [EB/OL]. (2019-06-11) [2021-01-22]. https: //arxiv.org/abs/1901.06637
[4] MISHRA D, NATALIZIO E. A survey on cellular-connected UAVs: design challenges, enabling 5G/B5G innovations, and experimental advancements [EB/OL]. (2020-03-14) [2021-01-23]. https: //arxiv.org/abs/2005.00781
[5] PLAPPERT M, HOUTHOOFT R, DHARIWAL P, et al. Parameter space noise for exploration [C]//Proceedings of International Conference on Learning Representations (ICLR). Vancouver, BC, Canada: ICLR, 2018
[6] FORTUNATO M, AZAR M G, PIOT B, et al. Noisy networks for exploration [C]//Proceedings of International Conference on Learning Representations (ICLR). Vancouver, BC, Canada: ICLR, 2018
[7] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing Atari with deep reinforcement learning [C]//27st Conference on Neural Information Processing Systems (NIPS). Lake Tahoe, Nevada, USA: NIPS, 2013: 1-9
[8] GHANAVI R, KALANTARI E, SABBAGHIAN M, et al. Efficient 3D aerial base station placement considering users mobility by reinforcement learning [C]//2018 IEEE Wireless Communications and Networking Conference (WCNC). Barcelona, Spain: IEEE, 2018: 1-6. DOI: 10.1109/WCNC.2018.8377340
[9] LIU C H, 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. DOI: 10.1109/jsac.2018.2864373
[10] ZHANG Y, MOU Z Y, GAO F F, et al. UAV-enabled secure communications by multi-agent deep reinforcement learning [J]. IEEE transactions on vehicular technology, 2020, 69(10): 11599-11611. DOI: 10.1109/TVT.2020.3014788

作 者 简 介

吴官翰，火箭军工程大学在读硕士研究生、酒泉卫星发射中心助理工程师；主要研究方向为深度强化学习、无人机通信组网等。

赵建伟，火箭军工程大学讲师；主要研究方向为5G/B5G、无人机通信组网、深度强化学习等。

高飞飞，清华大学自动化系副教授、IEEE Fellow、国家自然科学基金委优秀青年项目获得者，担任多本知名刊物的编委；主要从事通信原理和智能信号处理技术在无线通信中的应用研究；获2018年中国通信学会青年科技奖、2017年中国通信学会自然科学奖二等奖（排名第1）；发表论文160余篇。