DOI:10.19651/j.cnki.emt.2106573

一种密集无人机网络下行发射功率控制算法研究*

王　庆　黄　勇　常　晶
(公安部第一研究所 北京 100048)

摘　要：为解决密集无人机网络的下行功率控制问题，降低无人机的相互干扰、提高系统能量效率，提出了一种密集无人机网络下行发射功率控制算法。首先，将无人机网络的功率控制问题建模为高维系统状态下的平均场博弈论模型，降低无人机间的相互干扰；其次，将平均场博弈论模型转化为马尔可夫决策过程，以得到无人机密集部署下的均衡解；另外，提出了一种基于深度强化学习的平均场博弈论算法，通过使用深度神经网络获取系统的最优功率控制策略，从而最大化系统能量效率。最后，通过仿真分析将所提方法与其他3种算法进行对比分析。实验结果表明，所提方法能够使无人机与环境进行有效交互，有效降低无人机的相互干扰，增强系统网络通信性能；同时，与其他3种方法相比，所提方法的收敛速度更快、能量效率更高，具有良好的收敛性与可靠性。

关键词：无人机；功率控制；能量效率；平均场博弈论；深度强化学习
中图分类号：TN929.5　　**文献标识码**：A　　**国家标准学科分类代码**：510.50

Downlink transmit power control in dense UAV network based on mean field game and deep reinforcement learning

Wang Qing　Huang Yong　Chang Jing
(The First Research Institute of the Ministry of Public Security, Beijing 100048, China)

Abstract: To solve the problem of downlink power control of dense UAV networks, reduce the mutual interference of UAVs and improve the energy efficiency of the system, a downlink transmit power control algorithm for dense UAV networks is proposed. First, convert the power control problem of the UAV network into a mean field game theory model in a high-dimensional system state to reduce the mutual interference between UAVs. Second, convert the mean field game theory model into a Markov decision process, in order to obtain the equilibrium solution under the dense deployment of UAVs. In addition, a mean field game theory algorithm based on deep reinforcement learning is proposed, which obtains the optimal power control strategy of the system by using a deep neural network to maximize the energy efficiency of the system. Finally, the proposed method is compared with the other three algorithms through simulation analysis. The experimental results show that the proposed method can effectively interact between drones and the environment, effectively reduce the mutual interference of drones, and enhance the system network communication performance, at the same time, compared with the other three methods, the proposed method has a faster convergence faster, more energy efficient, with good convergence and reliability.

Keywords: unmanned aerial vehicle; power control; energy efficiency; mean field game; deep reinforcement learning

0　引　　言

随着移动互联网和物联网的快速发展，通信系统对网络容量和无线覆盖的要求越来越高，密集网络以其优异的基站小型化、小区密集化、节点多元化和高度协作化特性，已成为第五代移动通信系统(5G)的关键技术之一[1-2]。密集网络通过在有限的范围内密集地部署基站，能够有效地扩展网络的通信范围和系统吞吐量，从而满足当前超高流量数据信息的流通需求[3-4]。

同时，以无人机(unmanned aerial vehicles，UAV)为飞行平台的移动通信技术，可有效提升网络覆盖范围，丰富系统应用场景，是提高5G移动通信系统应对各种突发事件能力的关键技术之一[5]。因为传统的地面基站(ground base station，GBS)无法满足某些特定事件或紧急情况下，

收稿日期：2021-05-03
* 基金项目：北京市教委项目(2018Z004-005-KWY)资助

对通信系统灵活部署和快速恢复的使用需求[6]。而无人机具有体积小、便携性强、灵活性高等特点，因此在海上、自然灾害区、高海拔山区等特殊场景下，无人机网络可有效替代地面基站，实现无障碍快速通信[7-8]。

但是，单个无人机形成的网络服务范围有限，往往无法满足大规模场景中复杂任务的应用需求，因此密集的无人机网络由此诞生[9]。与传统的单无人机网络相比，密集的无人机网络可有效提升网络服务范围和数据传输速率，增加通信容量，满足大量用户接入的需求[10]。

然而，在密集无人机网络中，由于所有无人机都共享相同的下行链路通道，使得每架无人机都会受到其他无人机的干扰，导致无人机之间形成彼此竞争。其次，无人机的电量和滞空时间有限，难以保持长时间的不间断工作，导致密集无人机网络通信仍面临巨大挑战[8]。因此，降低密集无人机网络信道干扰，提高系统能量效率，是当前密集无人机网络的重要研究内容之一。

针对密集无人机网络的功率控制和能效优化问题，国内外众多专家学者进行了大量研究，文献[11]提出了一种基于无人机位置移动的动态场景资源分配算法，通过对动态网络中内容缓存的放置和内容传输的资源分配优化，可有效减小信息传输的时延。文献[12]提出了一种基于非正交多址接入的无人机空地协同联合资源分配方案，通过优先级访问方法降低能耗，提高无人机辅助非正交多址系统中的频谱效率。文献[13]提出了一种基于回声状态网络的功率控制方法，通过对用户内容请求分布和移动模式的预测，得到无人机的最佳位置，同时最小化无人机发射功率。但上述研究中的无人机数量有限，不能充分代表密集无人机网络的应用场景，在密集无人机网络的功率控制中实用性较低。文献[14]提出了一种基于无人机轨迹和任务需求的功率控制方法，通过循环迭代法和二分法进行联合求解，从而得到系统的最佳飞行半径和最佳瞬时传输功率。文献[15]提出了一种基于节能通信的多无人机覆盖模型，通过优化无人机的覆盖范围和功率来降低能量消耗，实现最优的节能覆盖部署。但上述研究中均使用了数值分析方法解决最优功率控制问题，其复杂度高，大大增加了系统的计算时间和能量消耗。

为此，针对上述研究存在的问题，提出了一种基于深度强化学习的平均场博弈论(mean filed game，MFG)方法。首先，将功率控制问题描述为一个离散的平均场博弈论模型，并将模型转化为马尔可夫决策过程求取均衡解。其次，提出了一种基于深度强化学习的多目标优化算法，通过证明平均场博弈论的解是马尔可夫决策过程的最优策略，从而获得系统的最优功率控制策略。最后，通过仿真对比分析，验证了所提方法的有效性。

1　系统模型构建

在应急通信系统中，无人机通过取代受损的地面基站，可进行用户密集区域的空地服务。但在实践中，无人机网络通信大多以计划外的方式随机部署，因此无人机的空间分布是随机的、相互独立的，通常可通过位置优化方法获得最佳位置，然后悬停在空中为地面用户提供服务。为提高无人机网络通信频带利用率，假设所有无人机都采用频率复用技术。同时，为获得实时的最优功率控制，提出了一种基于博弈论的分布式方案，无人机将根据功率控制策略调整发射功率。

如图 1 所示，为密集无人机网络与地面用户通信模型。考虑通信网络中具有 M 个完整的地面基站，并将其定义为 $M=\{m_1,m_2,\cdots,m_M\}$。此外，每架无人机在无人机网络中可为多个用户提供服务，为便于描述，假设每架无人机在特定时隙仅为一个用户提供服务，并在密集的无人机网络中设置第 $i(i\in U)$ 架无人机在时间 t 中为第 k 个($k\in K$)用户服务。其中，将网络中的无人机数量定义为 $U=\{u_1,u_2,\cdots,u_U\}$，网络用户数定义为 $K=\{k_1,k_2,\cdots,k_K\}$，发送功率集定义为 $P=(P_M,P_U)$。发射功率集中，$P_M=\{p_1,p_2,\cdots,p_M\}$表示地面基站的传输功率集；$P_U=\{p_1,p_2,\cdots,p_U\}$表示无人机的传输功率集。

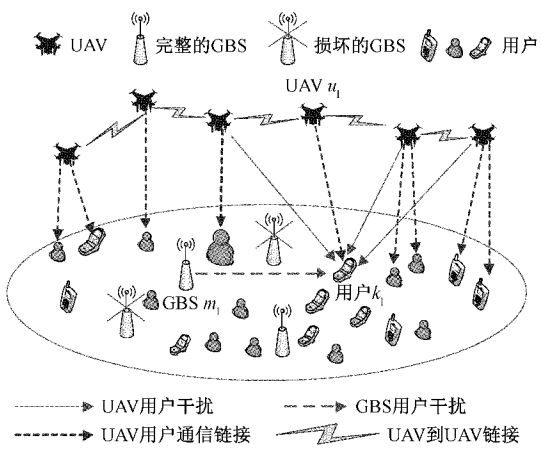

图 1　密集无人机网络与地面用户通信模型

首先，对无人机和用户之间的信道模型进行研究。当无人机悬停在地面上方时，由于地面障碍物的存在，会导致路径损耗和阴影衰落。假设在本文中无人机的高度值 H 为固定值，并将水平域中第 i 架无人机的坐标表示为 (x_i,y_i)，由第 i 架无人机服务的第 k 个用户的坐标表示为 $(x_{i,k},y_{i,k})$。因此，可将第 i 架无人机和第 k 个用户之间的距离定义为如式(1)所示。

$$|X_{i,k}(t)|=\sqrt{H^2+(x_{i,k}-x_i)^2+(y_{i,k}-y_i)^2} \tag{1}$$

随后，将第 k 个用户的接收功率表示为如式(2)所示。

$$p_r(t)=p_i(t)g_{i,k}(t)\,|X_{i,k}(t)|^{-\alpha_u} \tag{2}$$

式中：$g_{i,k}(t)=10^{\frac{\nu_{i,k}(t)}{10}}$，表示路径损耗中的阴影效应。其中，$\nu_{i,k}(t)$表示高斯随机变量，取值通常在 0～10 dB 之间，α_u 表示无人机到用户链路的路径损耗指数。

考虑到无人机的能量 $E_i(t)$ 是有限的,并将其可用能量表示为$[0,E_i(0)]$。其中,$E_i(0)$是在初始时间 $t=0$ 时的最大能量。由此可得到无人机网络能量演化的微分方程,如式(3)所示。

$$\mathrm{d}E_i(t) = -p_i(t)\mathrm{d}t \tag{3}$$

考虑到两层网络使用的是相同的频谱资源,则每个用户将分别接收来自无人机和地面基站的干扰。由于无人机使用频率复用技术,因此第 k 个用户将受到其他无人机的干扰,其数学方程如式(4)所示。

$$I_{u\to k}(t) = \sum_{j=1,j\neq i}^{U} p_j(t)g_{j,k}(t)\mid X_{j,k}\mid^{-\alpha_u} \tag{4}$$

同时,将地面基站在时间 t 引入用户的干扰表示为如式(5)所示。

$$I_{m\to k}(t) = \sum_{m=1}^{M} p_m(t)g_{m,k}(t)\mid X_{m,k}\mid^{-\alpha_m} \tag{5}$$

式中:$p_m(t)$表示第 m 个地面基站的发射功率;$g_{m,k}(t)$表示从第 m 个地面基站到第 k 个用户的信道增益;$\mid X_{m,k}\mid$表示第 m 个地面基站和第 k 个用户之间的距离;α_m 表示地面基站到用户链路中的路径损耗指数。

根据式(4)、(5),可将第 k 个用户的总干扰表示为如式(6)所示。

$$I_k(t) = I_{u\to k}(t) + I_{m\to k}(t) \tag{6}$$

由此,可将第 s 个无人机在时间 t 处服务的第 k 个用户的信噪比表示为如式(7)所示。

$$\gamma_i^k(t) = \frac{P_i(t)g_i(t)\mid X_{i,k}(t)\mid^{-\alpha_u}}{I_{p_{i(t)}} + N_0} \tag{7}$$

同时,令任何满足服务质量约束的无人机符合式(8)所示条件。

$$\frac{P_i(t)g_i(t)\mid X_{i,k}(t)\mid^{-\alpha_u}}{I_{k(t)} + N_0} \geqslant \tau_k \tag{8}$$

式中:τ_k 表示第 k 个用户的目标阈值信噪比。

因此,系统在时间 t 处的吞吐量,可表示为如式(9)所示。

$$R_i^k(t) = W\log(1+\gamma_i^k(t)) \tag{9}$$

式中:W 表示系统带宽。

随后,根据式(7)定义了在时间间隔 T 期间无人机能量消耗的传输效率,如式(10)所示。

$$\eta = \frac{R_i^k}{p_e(t)} = \frac{W\log(1+\gamma_i^k)}{p_i(t)+p_c} = \frac{W\log\left(1+\dfrac{P_i(t)g_i(t)\mid X_{i,k}(t)\mid^{-\alpha_u}}{I_k(t)+N_0}\right)}{p_i(t)+p_c} \tag{10}$$

式中:p_c 表示固定电路功耗,当 $p_c=0$ 时,即当不考虑固定电路功耗时,能量效率是发射功率的严格递减函数;当$p_c\neq 0$时,能量效率随发射功率的增大,先增后减,即能量效率相对于发射功率呈凹函数。

综上所述,可将无人机网络的下行功率控制问题表示为当无人机向用户传输数据时,每架无人机都需要考虑其他无人机和其他基站对用户的干扰。同时,无人机的最优功率控制可根据无人机在时间间隔 T 内消耗的能量传输效率确定。

2　问题表述和博弈论分析

2.1　平均场博弈

本文目标是通过控制每架无人机的下行发射功率,从而使每架无人机的能量效率最大化。但是,随着无人机发射功率的增加,每个用户的吞吐量和接收到的用户干扰也随之增加。同时,为保证用户的服务质量,须限制无人机的发射功率。然而,每架无人机在最大化其能量效率的同时,也会影响其他无人机的发射功率。因此,所提能量效率最大化问题为非合作博弈问题。

考虑到密集网络无人机的数量较多,建立了一种多用户接入模型,从而解决密集无人机网络中的下行功率控制问题。此处,将第 i 架无人机对其他用户造成的干扰定义为如式(11)所示。

$$I_{i\to K} = \sum_{f=1,f\neq k}^{K} p_i(t)g_{i,f}(t)\mid X_{i,f}(t)\mid^{-\alpha_m} \tag{11}$$

根据系统模型,干扰状态的数学方程如式(12)所示。

$$\mathrm{d}I_{i\to K} = \sum_{f=1,f\neq k}^{K} p_i(t)g_{i,f}(t)\mid X_{i,f}(t)\mid^{-\alpha_m}\mathrm{d}p_i(t) \tag{12}$$

随后,将平均场的状态空间定义为如式(13)所示。

$$\pi_i^k(t) = [E_i(t), I_{i\to K}], i\in U \tag{13}$$

由此,状态空间的分布可表示为如式(14)所示。

$$\pi^k(t) = [\pi_1^k(t),\pi_2^k(t),\cdots,\pi_n^k(t)] \tag{14}$$

式中:n 为状态空间的第 n 个分布,并将状态空间集合的分布大小定义为 N。

在拥有大量无人机的平均场博弈论模型中,由于其他无人机通过整体网络的动作对另一个无人机产生影响,使得无人机行为的波动被平均化。因此,当无人机密集部署时,单个无人机对平均场的影响可以忽略。因此,在讨论平均场博弈论函数时,须首先考虑密集无人机网络的平均场。令 $\alpha_i(a^d,t)$表示在时间 t 处系统的全部动作,其中 $d=1,2,\cdots,D$,D 为动作空间集合的大小,其数学方程如式(15)所示。

$$\alpha_i(a^d,t) = \lim_{U\to\infty}\frac{1}{U}\sum_{i,j\in U,j\neq i} R(a_{j,t}=a^d) \tag{15}$$

式中:$R(a_{j,t}=a^d)$表示指示函数;$a_{j,t}=a^d$ 表示第 j 架无人机在时间 t 采取的行为为a^d。

为方便起见,使用 α 表示$\alpha_i(a^d,t)$,并做如下假设:

1)无人机数量 U 足够大;

2)每架无人机都是相同且可互相替代的;

3)无人机在有限平均场中相互作用。

由于每架无人机都是独立的,因此每架无人机的状态

空间是相同的，则第 i 架无人机可根据当前时刻的剩余能量和第 k 个用户的干扰确定奖励函数，然后从一种状态演化到另一种状态。

在平均场博弈论模型中，可通过哈密顿-雅可比-贝尔曼方程（Hamilton-Jacobi-Bellman，HJB）和福克-普朗克-可尔莫哥洛夫方程（Fokker-Planck-Kolmogorov，FPK）描述整个系统的相互作用及演变。其中，FPK 方程表示平均场的状态变化，如式(16)所示。

$$\pi^k(t+1)=\sum_{i=1,i\neq j}^{i,j\in U}P_{i,j}(\alpha,t)\pi^k(t) \tag{16}$$

式中：$P_{ij}(\alpha,t)$ 表示转移概率，是第 k 个用户在平均场 α，i，$j\in U$ 的影响下，在时间 t 从 i 到 j 的状态概率。

因此，将第 i 架无人机的能量效率定义为模型的奖励函数，如式(17)所示。

$$\begin{aligned}
r(\iota)&=\max\eta=\max\frac{W\log\left(1+\dfrac{P_i(t)g_i(t)\mid X_{i,k}(t)\mid^{-a_u}}{I_k(t)+N_0}\right)}{p_i(t)+p_c}\\
\text{s.t.}\quad &\mathrm{d}E_i(t)=-p_i(t)\mathrm{d}t\\
&\mathrm{d}I_{i\to K}=\sum_{j-1,j\neq k}^{K}g_{i,j}(t)\mid X_{i,j}(t)\mid^{-a_m}\mathrm{d}p_i(t)\\
&0\leqslant p_i(\iota)\leqslant p_{\max}\\
&\gamma_i^k\geqslant\tau_k
\end{aligned} \tag{17}$$

由此得到系统的值函数，如式(18)所示。

$$V_i^k=\max_{p_i^t}\{r(t,\pi_i^k(t),p_i(\alpha,t))+\sum_{i=1,i\neq j}^{i,j\in u}P_{ij}(t)V_i(t+1)\} \tag{18}$$

通过状态方程和值函数，可推导出满足最优控制问题的 HJB 方程，如式(19)所示。

$$\frac{\partial V_i^k(t)}{\partial t}-\max_{p_i(t)}\left\{r(\iota,\pi_i^k(\iota),p_i(\iota))+p_i(\iota)\frac{\partial V_i^k(t)}{\partial\pi_i^k(t)}\right\}=0 \tag{19}$$

式中：HJB 方程模拟了单个用户和平均场之间的相互作用关系。HJB 函数表明函数的最终值是已知的，同时可确定在[0，T]时刻 $V_i(\iota)$ 的值。因此，HJB 方程总是在时间顺序上向后求解，即从 $t=T$ 时刻开始，到 $t=0$ 时刻结束。当求解整个状态空间时，HJB 方程是最优解的充要条件，而 FPK 方程则随时间向前发展。因此，这种相互作用的演化最终导致了平均场均衡（mean field equilibrium，MFE）。

2.2　基于 MFG 的深度强化学习

通常，平均场均衡可通过有限差分方法计算获得，该方法将求解域划分为差分网格，并用有限数量的网格节点替换连续求解域，但是该方法的复杂度较高。考虑到优化问题的本质为随机控制问题，当状态空间离散化时，可表示为有限时域的马尔可夫决策过程，如式(20)所示。

$$\begin{aligned}
&V^*(\pi^k)=:\sum_{i\ 1}^{u}\pi_i^tV_i^k=\\
&\sum_{i=1}^{u}\pi_i^t\max\Big\{\sum_j p_{ij}(\pi^k,P_i)+\sum_j p_{ij}V_j^{t+1}\Big\}=\\
&\max\Big\{\sum_{i=1}^{u}\pi_i^k\sum_j p_{ij}(\pi^k,P_i)+\sum_{i=1}^{u}\pi_i^t\sum_j p_{ij}V_j^{t+1}\Big\}=\\
&\max\{R(\pi^k,P)+\sum_{j=1}^{}\pi_j^{t+1}V_j^{t-1}\}=\max\{R(\pi^k,P)+V^*(\pi^{k+1})\}
\end{aligned} \tag{20}$$

通常，马尔可夫决策过程可被用于解决最优化问题。因此，密集无人机网络的下行功率控制问题，可以转化为马尔可夫决策过程的优化问题，从而可减少平均场解的约束，简化求解过程。然后，通过应用线性回归方法得到马尔可夫决策过程最优控制策略。如图 2 所示为强化学习的示意图。由图 2 可知，强化学习方法通过构建一个深度神经网络（deep neural network，DNN）来学习功率优化策略，其隐含层是多层神经网络结构的模型，并将系统状态和动作作为神经网络的输入，奖励函数作为神经网络的输出。

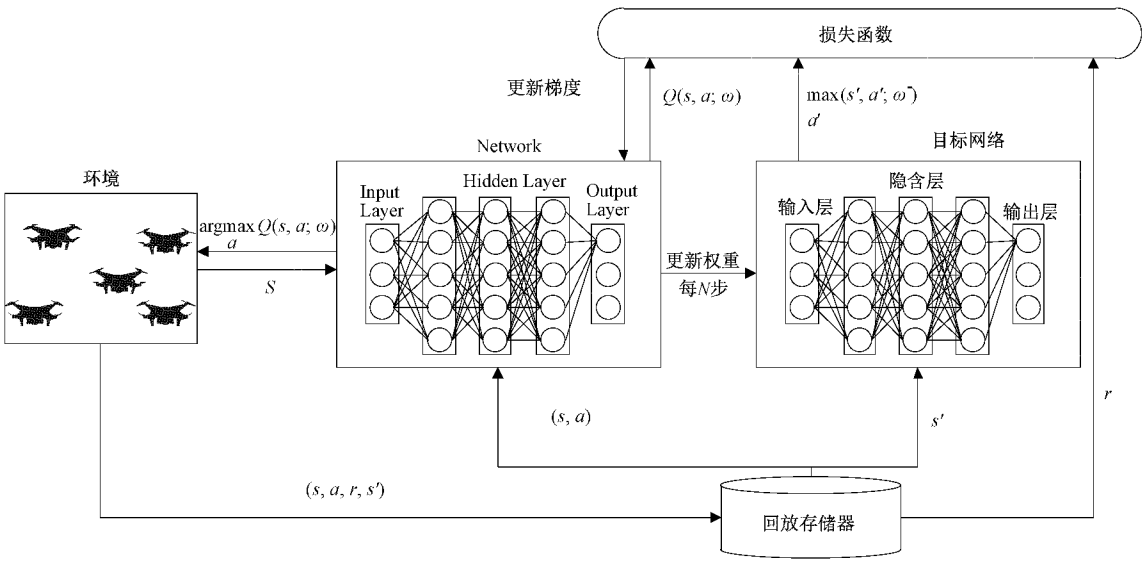

图 2　强化学习中主体和环境的相互作用

由于强化学习在平均场博弈论模型中的应用促进了所有无人机之间的相互作用,即单个无人机的最优功率控制策略可基于无人机网络的动态演化获取,而无人机网络的动态演化也可根据单个无人机的功率控制策略进行更新。

当无人机向用户传输数据时,无人机与当前环境进行交互。首先,无人机及时调整发射功率,使用户干扰和无人机剩余能量发生变化。随后,在智能体和环境之间进行不断交互,以学习更多的数据,直到获得最佳策略。

针对平均场博弈论的纳什均衡解,将每架无人机的最优策略 π^* 定义为如式(21)所示。

$$\pi^* = \operatorname{argmax} V_i^k(\pi_i) \tag{21}$$

由此,对于所有策略 π,其数学关系如式(22)所示。

$$V_i^k(\pi_i) \leqslant V_i^k(\pi^*) \tag{22}$$

因此,根据最优策略 π^*,可得到满足式(16)、(18)平均场博弈论模型的纳什均衡解。

通常,平均场博弈论只有 1 个唯一的平衡解。令 $v_i(p_i)=\partial V(\pi_i)/\partial p_i$,若值函数是凸的,则 $v_i(p_i)$ 为严格对角凸函数,此时满足式(23)所示关系。

$$\sum_{i=1}^{N}(p_i^1 - p_i^2)(v_i(\pi^1) - v_i(\pi^2)) > 0 \tag{23}$$

式中:p_i^1 和 p_i^2 分别表示策略 π^1 和 π^2 在第 i 个状态下所选择执行的动作。基于纳什均衡最大化的唯一性,令奖励函数为单调分布,如式(24)所示。

$$\sum_{i-1}^{N}(\pi^2 - \pi^1)(r(\pi^2) - r(\pi^1)) \geqslant 0 \tag{24}$$

然而,由于大量博弈个体的存在,使得模型的复杂性较高,系统需处理的信息量较大,包括接收到的用户干扰和剩余能量。因此,基于上述分析,提出了一种用于求解马尔可夫决策过程最优策略的基于平均场博弈论的深度强化学习(deep reinforcement learning-based MFG,DRL-MFG)算法,其中共包含如下重要元素。

1)智能体:$U=\{u_1,u_2,\cdots,u_i,\cdots,u_U\}$ 表示密集无人机网络的智能体集。其中,无人机数量 U 为任意大。

2)状态:将智能体的状态定义为用户接收到的干扰和无人机电池剩余能量的组合。因此,系统状态由用户当前的干扰 $\mathrm{d}I_{i\to K}$ 和无人机当前的剩余能量 $E_i(t)$ 两部分组成。

同时,令状态空间为 S,则系统在时间 t 的状态 $S_t\in S$ 被定义为如式(25)所示。

$$s_t = [\mathrm{d}I_{1\to K},\mathrm{d}I_{2\to K},\cdots,\mathrm{d}I_{U\to K};E_1(t),E_2(t),\cdots,E_i(U)] \tag{25}$$

由此可得到系统状态的演变过程,如式(26)所示。

$$\pi_i(t+1) = \sum P_{ij}(\alpha,t)\pi_i(t) \tag{26}$$

3)动作:无人机的可选动作是一组可能离散的发射功率值,如式(27)所示。

$$A_t = \{a_1^i,a_2^i,\cdots,a_d^i,\cdots,a_D^i\} \tag{27}$$

式中:d 表示第 i 架无人机的潜在发射功率。

4)控制策略:功率控制策略由 $Q^*(t)$ 决定,从而使时间间隔 T 内的平均奖励函数最大化。

5)奖励函数:第 k 架无人机的奖励函数为自身的能量效率。

如算法 1 所示,为所提基于平均场博弈论的深度强化学习算法(DRL-MFG)。其中,平均场博弈论和强化学习之间的相互作用过程如图 3 所示。

算法 1 基于 MFG 的 DRL(DRL-MFG)算法

初始化 UAV 和 GBS 的数量
用缓冲容量初始化回放存储器
使用随机权重 ω 初始化函数 Q 并使用权重 ω $=\omega$ 初始化目标 DNN
通过式(11)获取 UAV 的干扰,剩余能量作为 UAV 的当前状态 $s^{(1)}$
for episode $:=1,\cdots,M$ **do**
　重置密集 UAV 网络环境
　随机获取每个 UAV 的初始状态
　for epoch $t=1,\cdots,T$ **do**
　　for each $i\in U$ **do**
　　通过密集 UAV 网络更新无人机的功率 p_i
　　根据 ε 贪婪策略选择 UAV 的随机动作
　　否则选择 $a_t=\operatorname{argmax}_a Q(s_t,a;\omega)$
　　通过选择 UAV 获得 s_{t+1} 并奖励 r_t
　　将 s_t,a_t,r_t,s_{t+1} 存储在回放存储器中;
　　从回放存储器中随机抽样;
　　如果步数为 $t+1$,设置 $y_t=r_t$;
　　否则 y_t 设置为式(34)
　　通过最小化函数式(30)来更新 ω;
　　设置 $\omega_0=\operatorname{argmin}L(\omega)$;
　　每 C 步更新一次 DNN 权重值
　　end for
　end for
end for
通过训练网络来获得密集 UAV 网络中的最佳功率控制策略

在强化学习过程中,值函数可表示为如式(28)所示。

$$Q(s,a) = E[r_{t+1} + \gamma r_{t-2} + \gamma^2 r_{t+3} + \cdots \mid s,a] = E_{s'}[r + \lambda Q(s',a') \mid s,a] \tag{28}$$

但是,DRL-MFG 算法中仅得到值函数是不够的,还需要通过值函数来求解最优策略,得到最优值函数,如式(29)所示。

$$Q^*(s,a) = \max Q(s,a) \tag{29}$$

式中:$Q^*(s,a)$ 表示最优值函数,是所有策略下的最大值函数。通过最优值函数这个确定解的唯一性,从而求解整个马尔可夫决策过程。

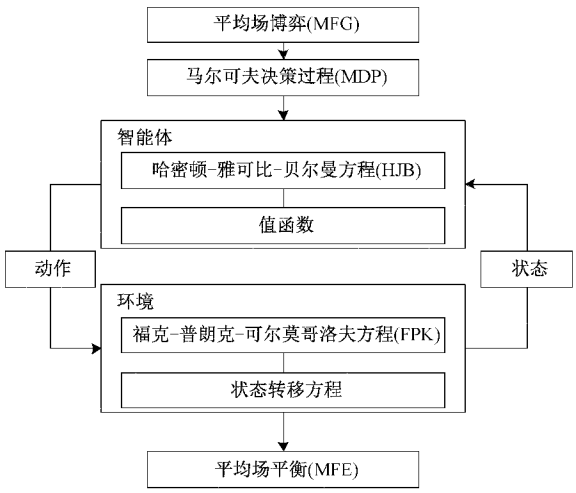

图 3　DRL-MFG 的基本程序

在博弈过程中，所有的计算结果均存储在可检索的内存回放$<s,a,r,s'>$中。当训练神经网络时，随机的小批量内存片段均取自于内存回放，而不是直接使用最近的计算结果。该方式切断了相似训练的连续性，使相似性倾向于将整个网络限制在一个小的状态区域内。此外，内存回放使训练任务更趋近于正常的监督学习，简化了算法的调试和测试过程，且能够直接收集智能体的经验，从而进行继续迭代学习。

在强化学习中，由于所有用户都不知道马尔可夫决策过程的传递模型，所以使用近似的 Q 值函数代替值函数。然后，无论状态空间的维数有多大，都可通过降低矩阵的维数来计算 Q 值函数，并将其作为单个值输出。通过所提 DRL-MFG 算法学习 $Q(s,a)$，可找到最优功率控制策略。DRL-MFG 算法通过将深度神经网络与 Q 学习算法相结合，直接从高维原始数据中学习控制策略。由于深度神经网络的输入为状态空间 S，输出为对应于每个动作的值评估，因此可通过 Q 学习算法获得大量训练样本，并对样本进行标记以训练神经网络。

由于神经网络训练的本质是一个优化问题，神经网络通过使标签和网络输出之间的偏差最小化得到最小化损失函数。在所提算法中，对通过 Q 学习算法训练得到的目标 Q 值进行标记，使其接近目标 Q 值。因此，将损失函数定义为如式(30)所示。

$$L(\omega)=E[(r+\max_{a'}Q(s',a',\omega)-Q(s,a,\omega))^2] \tag{30}$$

随后，通过反向传播更新神经网络参数，在神经网络中平均场的状态是神经网络的输入，其输出状态如式(31)所示。

$$Q_t(s,a)=\omega_t^{\mathrm{T}}x+b \tag{31}$$

式中：w_t 表示权重，可利用式(32)进行权重更新：

$$\omega_{t+1}=\omega_t+\lambda(y_t-Q_t(s,a)) \tag{32}$$

式中：λ 表示学习率；y_t 表示最优 Q 值，如式(33)所示。

$$y_t=(1-\beta)Q_t(s,a)+\beta(r(s,a)+\gamma\max\omega_t^{\mathrm{T}}x+b) \tag{33}$$

式中：β 表示折扣因子。

由此可得到最优动作函数，如式(34)所示。

$$a(s)=\max Q(s,a) \tag{34}$$

由于动作空间和状态空间的复杂性，在学习过程中网络结构会不断深化，无人机会根据每一步的当前状态进行动作选择，形成训练数据集。大量的无人机和训练迭代次数使得训练数据的样本量较大，使神经网络具有良好的收敛性，进而使用户能够有效地学习最优策略。

3　仿真实验与分析

3.1　仿真参数设置

实验中，在 5 km×5 km 的区域内，随机分布有 60 架无人机和 3 个地面基站。其中，无人机的最大发射功率设置为 $p_{\max}=0.5$ W，且无人机的发射功率为预定义的。令无人机到用户的路径损耗指数 $\alpha_{U-K}=3$，地面基站到用户的路径损耗指数 $\alpha_{M-K}=4$，地面基站到用户的距离为 200 m。其中，每架无人机均通过 ε 贪婪方法进行动作选择，即无人机的动作选择为 ε 的随机行为，并通过自适应神经网络逼近值函数。最后，将噪声功率设置为 $N_0=0.001$ W。系统仿真模型具体参数设置如表 1 所示。

表 1　仿真模型参数值

参数	值	参数	值
S	60	α_u	3
M	3	α_m	4
$P_{\max}$/W	0.5	p_c/W	0.01
W/kHz	20	N_0/W	0.001
β	0.08	γ	0.01

本文采用 Adam 算法更新权重，其迭代数量设置为 256。通常，用于逼近作用值函数的深度神经网络函数由 3 个完全连通的前馈隐藏层组成，所提仿真模型中将 3 个隐藏层的神经元数量分别设置为 256、256 和 512 个。此外，在学习过程中，将学习率设置为 0.01，折扣系数设置为 0.08。

为验证所提 DRL-MFG 算法的有效性，将所提算法与 Q 学习、随机策略和有限差分 3 种算法进行了对比分析。其中，Q 学习是强化学习的非典型算法，随机策略是自由模型搜索策略的主要方法之一，有限差分法是经典的平均场博弈论算法。

3.2　结果分析

如图 4 所示，为奖励函数收敛时能量效率、用户干扰和发射功率之间的关系。由图可知，当无人机的能量效率处于相对稳定的状态时，其对用户的干扰较小，表明无人机已获得了最优功率控制策略，使其能够在与用户交互的

基础上快速调整发射功率。另外,能量效率虽有一定波动,但波动中心与理论值基本相等,表明所提算法是合理、可行的。

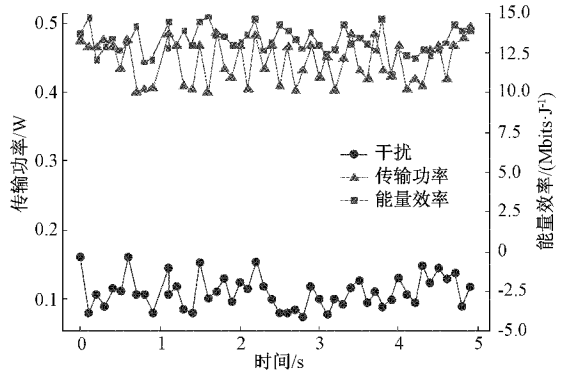

图 4　发射功率、干扰和能量效率随迭代次数的变化趋势

如图 5 所示,为所提算法与其他 3 种算法的收敛性比较。由图可知,所提算法的收敛速度明显高于其他 3 种算法。其中,随机算法由于无人机不能根据自身状态及时选择合理的行为,其收敛性较差。而 Q 学习算法由于必须建立一个庞大的询问表,会减缓系统在密集网络环境下的收敛速度。此外,有限差分算法虽获得了较高的能量效率和平滑收敛曲线,但其动作空间和状态空间的维数相对较大,使得有限差分算法的计算复杂度较高,导致收敛速度较慢。同时,根据图 5 可知,迭代次数对系统奖励函数会产生显著影响,迭代次数越少,智能体越难学到足够的经验获得有效的功率控制策略。因此,为获得更好的学习效果,设置合理的迭代次数是非常重要的。

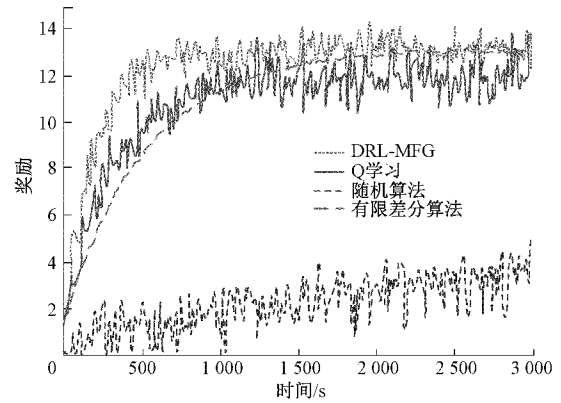

图 5　4 种不同算法的计算结果

如图 6 所示,为所提算法在不同学习速率下的收敛性能。由图可知,随着迭代次数的增加,DRL-MFG 方法是收敛的,且系统的能量效率趋于稳定,表明无人机学会了有效的功率控制策略。此外,还比较了不同学习率下奖励函数的收敛性,为避免出现较大振荡,学习率设置须适中,若学习率太小,奖励函数收敛太慢,学习率过大,函数波动较大难以收敛。因此,本文选择学习率 γ 为 0.001,0.01,0.1 进行对比。其中,当 $\gamma=0.001$ 时,收敛速度最快;当 $\gamma=0.1$ 时算法收敛速度最慢。

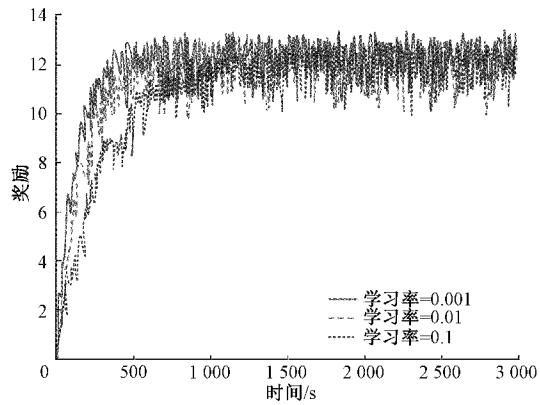

图 6　收敛性分析

如图 7 所示,为不同数量无人机在不同算法中的能量效率。由图可知,MFG-DRL 算法得到的能量效率非常接近于有限差分法,表明所提算法具有较好的正确性,且能得到平均场博弈论的平衡解。同时,当无人机数量越少,4 种方法的能效就越高。然而,随着无人机数量的增加,Q 学习算法的性能明显低于 DRL-MFG 算法,且随机算法的能量效率呈快速下降趋势。

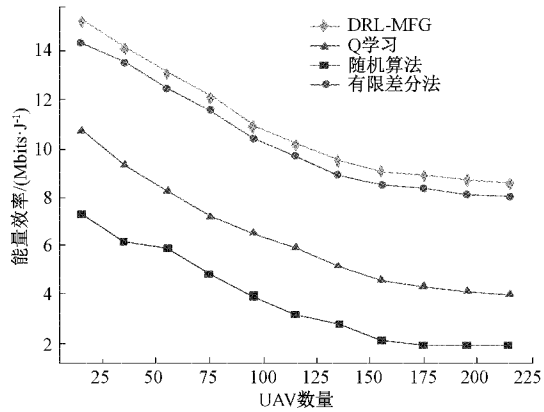

图 7　能源效率与无人机数量的对比

如图 8 所示,为 4 种算法能量效率与发射功率对比结果。由图可知,所提算法得到的能量效率接近于有限差分法,且在相同发射功率下明显优于 Q 学习算法和随机算法。此外,随发射功率的增加,能量效率呈先增后减的趋势。为防止无人机无限增大发射功率以达到目标信噪比,系统能量效率逐渐降低,有效降低了系统的功耗。

如图 9 所示,为能源效率与具有不同能源约束的无人机数量的关系。由图可知,由于用户的干扰随单位面积无人机密度的增加而增加,使得系统能量效率随无人机数量的增加而降低。同时,由图 8 可知,当无人机的能量在一定范围内增加时,其能力效率会降低,因为当无人机选择较大发射功率时,必然产生一定的资源浪费。

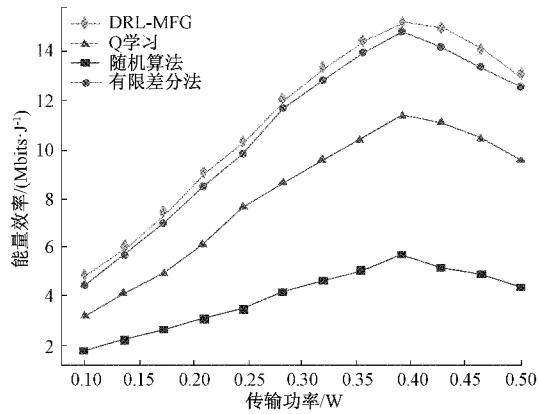

图 8　能源效率与发射功率的对比

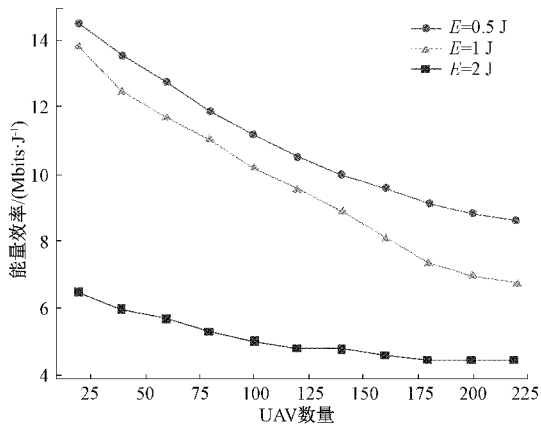

图 9　能源效率与不同能源约束的无人机数量的关系

最后，分析了无人机高度对系统能量效率的影响。如图 10 所示，为不同高度下无人机的发射功率能量效率的变化趋势。由图可知，随无人机发射功率的增加，系统能量效率先增后减。同时，由图可知，无人机的高度与能量效率的关系更大。与 60 和 300 m 高度相比，当无人机高度为 150 m 时其能量效率最大，主要因为当无人机高度较低时，会受到较大的地面环境影响。但是，当无人机高度增加时，为保证用户的通信质量，无人机需消耗更多能量，并增加发射功率，这同时伴随着较大的路径损耗。

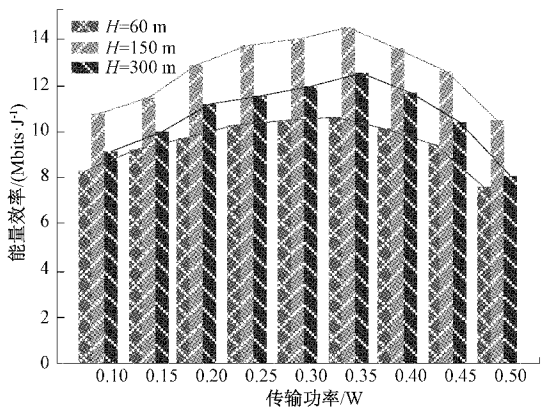

图 10　不同高度下能量效率与发射功率的关系

4　结　　论

针对密集无人机网络中的下行功率控制问题，提出了一种基于深度强化学习的平均场博弈论算法，并通过仿真对比实验得出如下结论。

1)在密集无人机网络中，所提方法可有效降低无人机之间的相互干扰，并提高系统能量效率、增强网路通信性能。

2)在不同学习速率、不同迭代次数、不同无人机数量下，所提方法仍趋于收敛，能量效率仍趋于稳定，并能够获得最优功率控制策略。

3)4 种算法均能获得较高的能量效率和平滑的收敛曲线，但与其他 3 种经典方法相比，所提方法的收敛速度更快，系统能量效率更高，综合性能更优。

参考文献

[1]　陈安民，张春元，张泽林. 基于长短期记忆网络的无人机认知无线电频谱预测[J]. 国外电子测量技术，2021，40(1)：37-43.

[2]　GE X，TU S，MAO G，et al. 5G ultra-dense cellular networks[J]. IEEE Wireless Communications，2016，23(1)：72-79.

[3]　罗霄，薛亚洲，张乐. 一种无人机飞控计算机硬件平台的设计实现[J]. 电子测量技术，2021，44(1)：50-54.

[4]　KAMEL M，HAMOUDA W. Ultra-dense networks：A survey [J]. IEEE Communications Surveys & Tutorials，2016，18(4)：2522-2545.

[5]　贺子健，艾元，闫实，等. 无人机通信网络的容量与覆盖性能[J]. 电信科学，2017，33(10)：65-70.

[6]　周剑，贾金岩，张震，等. 面向应急保障的 5G 网联无人机关键技术[J]. 重庆邮电大学学报(自然科学版)，2020，32(4)：511-518.

[7]　刘炜. 基于多维谱峰联合搜索的无人机控制抗扰动算法[J]. 电子测量技术，2019，42(9)：19-23.

[8]　SHARMA V，BENNIS M，KUMAR R. UAV-assisted heterogeneous networks for capacity enhancement [J]. IEEE Communications Letters，2016，20(6)：1207-1210.

[9]　邓耀华，姚可星，孙成，等. 无人机航测高精度 RTK 接收机信号捕获与跟踪仿真方法研究[J]. 电子测量与仪器学报，2020，34(12)：43-48.

[10]　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]　王子端，张天魁，许文俊，等. 无人机应急通信网络中的动态资源分配算法[J]. 北京邮电大学学报，2020，43(6)：42-50.

[12]　LIU M，GUI G，ZHAO N，et al. UAV-aided air-to-ground cooperative nonorthogonal multiple access[J].

IEEE Internet of Things,2020,7(4):2704-2715.

[13] CHEN M Z, MOHAMMAD S, WALID S, et al. Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience[J]. IEEE Journal on Selected Areas in Communications,2017,35(5):1046-1061.

[14] 马松辉,芦永超,刘可佳,等.基于小型无人机的高精度天线测试方法研究[J].仪器仪表学报,2019,40(5):36-42.

[15] RUAN L, WANG J L, CHEN J, et al. Energy-efficient multi-UAV coverage deployment in UAV networks: A game-theoretic framework[J]. China Communications, 2018, 15(10):194-209.

作者简介

王庆,硕士,工程师,主要研究方向为无人机检测技术。
E-mail:njnnf51@163.com

黄勇,硕士,工程师,主要研究方向为警用特种车辆、无人机检测技术研究。
E-mail:huanyong@163.com

常晶,大专,高级技工,主要研究方向为机械制造加工、无人机检测技术研究。
E-mail:45379768@163.com