



面向可伸缩视频编码传输的 DDPG 无人机服务增强机制

闫俊杰¹, 朱亚新¹, 冯艳茹², 刘汉永³, 邓钧忆¹, 王欢¹

(1. 广西科技大学电子工程学院, 广西 柳州 545006;

2. 山西应用科技学院, 山西 太原 030062;

3. 国家电网重庆市电力公司大足供电分公司, 重庆 402360)

摘 要: 无人机 (UAV) 因其机载存储、通信和计算能力成为未来空天地一体化网络的重要组成部分。然而, 现有无人机辅助的边缘网络研究总是更多地从网络的视角出发, 缺少从用户视角出发带来的变革。为此, 从用户角度出发, 提出一种面向可伸缩视频传输的 DDPG 无人机服务增强机制。首先, 结合无人机提出一种基于可伸缩视频编码 (SVC) 的弹性视频传输方法, 提高用户的差异化体验。其次, 以最大化用户接收增强层视频的数量为目标, 提出一种基于深度确定性策略梯度 (DDPG) 算法的无人机路径规划设计, 保证 UAV 对热点区域增强覆盖的有效性。仿真结果表明, 所提出的算法与深度 Q 网络 (DQN) 算法、最短路径 (SP) 算法相比, 在不同的用户分布下得到的增强层数量平均分别可以提高 47.9%、76.4%, 该研究达到了热点区域覆盖增强和用户差异化体验的效果。

关键词: 深度确定性策略梯度; 无人机; 可伸缩视频编码; 路径规划

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2023158

UAV service enhancement mechanism deep deterministic policy gradient for scalable video coding transmission

YAN Junjie¹, ZHU Yaxin¹, FENG Yanru², LIU Hanyong³, DENG Junyi¹, WANG Huan¹

1. School of Electronic Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China

2. Shanxi College of Applied Science and Technology, Taiyuan 030062, China

3. Dazu Power Supply Company of State Grid Chongqing Electric Power Company, Chongqing 402360, China

Abstract: Unmanned aerial vehicles (UAV) are an important component of future air-space-ground integrated network due to their onboard storage, communication, and computing capabilities. However, existing research on UAV-assisted edge networks often focuses more on the network perspective and lacks consideration of user perspec-

收稿日期: 2023-05-11; 修回日期: 2023-08-09

通信作者: 闫俊杰, yanjj@gxust.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62106053); 广西自然科学基金资助项目 (No.2022GXNSFBA035642); 广西科技计划项目 (No.2021AC19167); 广西中青年教师基础能力提升项目 (No.2022KY0325, No.2021KY0364)

Foundation Items: The National Natural Science Foundation of China (No.62106053), The Natural Science Foundation of Guangxi (No.2022GXNSFBA035642), Guangxi Science and Technology Planning Project (No.2021AC19167), The Basic Ability Enhancement Program for Young and Middle-aged Teachers of Guangxi (No.2022KY0325, No.2021KY0364)



tives and their requirements. Therefore, a UAV service enhancement mechanism based on deep deterministic policy gradient (DDPG) for scalable video coding (SVC) transmission was proposed from the user's perspective. Firstly, an elastic video transmission method based on SVC was proposed in conjunction with UAV to improve differentiated user experiences. Secondly, a UAV trajectory planning design based on the DDPG algorithm was proposed to maximize the number of users receiving enhanced layer videos and ensure effective coverage enhancement by UAV in hotspot areas. Simulation results show that compared with both the deep Q network (DQN) algorithm and shortest path (SP) algorithm under different user distributions, the proposed algorithm can increase the average number of enhanced layers received by 47.9% and 76.4%, respectively. This study successfully achieves improved coverage in hotspot areas while also providing differentiated user experiences through its proposed methods.

Key words: DDPG, UAV, SVC, trajectory planning

0 引言

随着移动通信网络的飞速发展,未来网络将呈现立体化、智能化和个性化等特征,这对未来蜂窝网络的性能和效率提出更高的要求。尤其是在用户密度高的流量密集区域,单纯地依靠地面蜂窝网络难以满足日益增长的多媒体业务需求。因此,构建面向视频业务的新型网络传输架构具有重要的研究价值。其中,以无人机(unmanned aerial vehicle, UAV)为搭载平台的异构蜂窝网络因其低廉的成本、便捷的部署方式、高度的网络可扩展性等优点,成为未来 6G 网络的演进方向之一^[1-3]。

在流量需求大且相对稳定的场景下,基地站的密集部署可以实现更高的收益。然而对于时空分布不均的流量需求场景,例如,举行大型体育赛事或者游戏直播时,流量需求激增且区域分布差异性,密集部署基站并不是最佳解决方案,也无法满足运营商的经济需求。而 UAV 具备成本低和灵活部署的特点,这使得它们非常适合提供短期热点传输服务^[4-5],可以用于扩大现有陆地蜂窝网络的覆盖范围,将 UAV 应用于无线通信热点领域具有扩大覆盖范围和提高热点地区容量的巨大潜力。

近年来,在 UAV 协同的边缘网络视频传输方面已有一定的研究。文献[6]提出一种可扩展视频流的安全传输方案,其中无人机作为空中基站,以提高覆盖率。文献[7]设计了一个客户端-服务器-

地面用户框架,并提出一种用于多链路并发视频传输的拆分合并流算法,以便服务器和用户可以控制无人机根据负载、时延和丢包动态调整其传输模式。文献[8]提出一种基于多集空时移键控的多功能多输入多输出收发器,用于无人机辅助的近瞬时自适应分层视频流,该收发器能够根据估计的信道质量和无人机移动性改变吞吐量。此外,深度强化学习因其高维信息感知、不依赖环境模型、能够处理时变复杂状态空间等优势,在众多领域取得了成功的应用^[9]。目前,已有研究人员将深度强化学习与无人机通信结合,并在无人机通信应用方面取得了许多成果。文献[10]提出一种基于强化学习的无人机抗干扰视频传输方案,以选择视频压缩量化参数、信道编码率、调制和功率控制策略来对抗干扰攻击。文献[11]提出了一种深度 Q 网络(deep Q network, DQN)和深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法,解决海上无人机集群移动边缘计算网络中通信和计算的时延最小化问题。文献[12]提出了一种深度确定性策略梯度算法,该算法可以优化无人机的飞行方向和分配给每个用户的带宽,实现高效、公平的通信。在 UAV 辅助的边缘网络中利用深度强化学习算法优化 UAV 轨迹以提高网络性能是重要的研究方向之一。文献[13]研究了无人机轨迹设计和功率分配,提出一种基于深度确定性策略梯度的算法以最大化无人机的服务时间和下行链路吞吐量。文献[14]利用 DDPG

算法来优化无人机的轨迹,并找到在最小吞吐量约束下保持信息新鲜的调度策略。文献[15]设计了一种基于DDPG的算法以联合优化无人机轨迹和发射机发射功率,最大限度地提高最坏情况下的平均保密率。上述UAV协同边缘网络视频传输机制主要将网络吞吐量、时延、覆盖率和丢包率作为视频传输质量评价标准,但基于服务质量的评价指标只反映了网络的传输性能,无法体现用户对视频传输质量的差异化需求。与此不同的是,本文从用户角度出发提出一种面向可伸缩视频传输的DDPG无人机服务增强机制,保障了热点区域覆盖增强和用户差异化体验的效果。

可伸缩视频编码(scalable video coding, SVC)技术可以在保证高质量视觉体验的同时支持多种分辨率、比特率和质量级别选择。将SVC技术引入无线网络中来实现用户差异化需求是目前研究的一个热点^[16]。基于SVC的视频技术可以根据信道反馈灵活地调整视频层的速率。用户可以根据不同的信道条件接收增强层,信道条件好的用户可以接收基础层和更多的增强层。然而,将SVC视频流的分层结构特征引入UAV无线边缘网络中尚未得到深入研究。因此,如何结合SVC技术,构建符合UAV边缘的弹性视频传输架构,并依据用户对视频质量的差异化需求,建立面向SVC视频传输的资源分配和UAV轨迹联合优化问题,如协调UAV轨迹以最大化用户整体接收的视频质量,也是一个非常有意义的课题。

为此,本文从用户角度出发,提出一种基于DDPG算法的弹性视频传输方法来实现用户差异化需求。具体来说,本文结合UAV引入SVC技术构建了一个热点区域用户差异化观看视频的场景,针对这种场景,提出优化UAV轨迹和悬停位置的问题,以最大化用户接收增强层视频数量为目标。由于目标函数是多约束且非凸的,传统的优化方法难以直接解决,深度强化学习方法在处理动态复杂的优化问题时有很大的优势。为了解

决这个挑战,本文通过将问题建模为马尔可夫决策过程(Markov decision process, MDP),将路径规划问题转化为序列决策问题,即UAV根据当前状态决策下一个飞行的方向。并通过DDPG算法将用户接收增强层视频的数量作为奖励项,UAV通过优化决策使用户获得更多的增强层数量。

1 系统模型

本文的研究面向突发热点区域,如大型体育赛事、网络直播、演唱会等场景。在一个大小为 $L \times L$ 的区域,存在 Z 个用户有视频服务需求。在缓存使能的UAV辅助的异构蜂窝网络中,有一个宏基站和多个小基站(small base station, SBS)及一架UAV通过优化轨迹及悬停位置辅助地面基站。基于SVC的UAV轨迹优化系统模型如图1所示。在一场足球比赛中,体育场附近的小基站过载,不能满足高峰时段用户的需求;在这种情况下,UAV可以利用空地链路高概率的视距传输特性增加链路吞吐量,实现热点区域的流量增强覆盖需求,缓解地面基站的压力。

视频文件库中有 G 个内容项,该视频集表示为 $\mathbf{G} = \{1, 2, \dots, G\}$ 。每个视频文件被编码为 J 层文件。编码的层元素表示为 $\mathbf{J} = \{1, 2, \dots, J\}$,其中元素 $\{1\}$ 是基本层文件,其他元素 $\{2, \dots, J\}$ 是增强层文件。由于每个视频都被编码为 J 层,因此每个视频也有 J 个质量级别。在这个场景中,UAV和SBS缓存流行视频的基本层和增强层的副本。为了充分利用有限的缓存空间和UAV通信的优势,假设UAV只缓存增强层文件,SBS可以同时缓存基本层文件和增强层文件。在流量高峰需求时段,用户的增强层视频需求过多导致小基站负载,此时UAV通过路径规划为用户提供服务,减轻基站的压力。假设UAV u 以固定高度 H 飞行,最大速度为 $v_{\max}$,需要为 M 个用户提供增强层服务。小基站 $f \in F$ 为 N_f 个用户提供基本层服务。因此,所有小基站服务的基本层用户 Z 可以表示为:

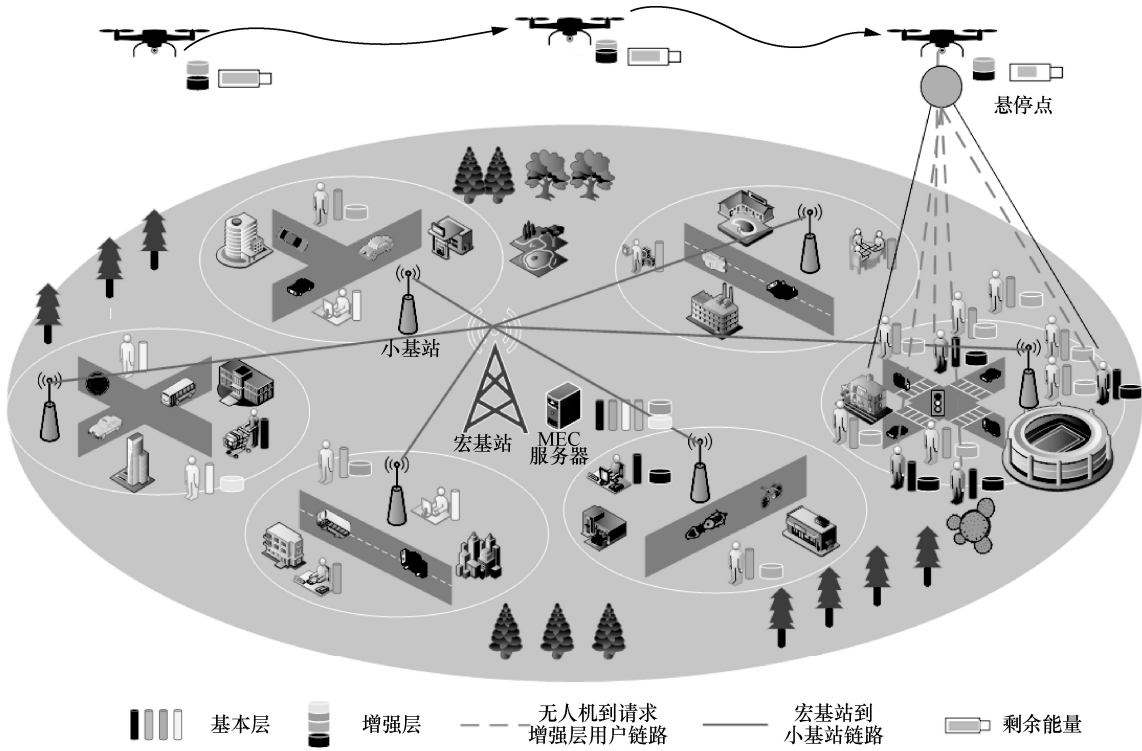


图1 基于 SVC 的 UAV 轨迹优化系统模型

$$Z = \sum_{f=1}^F N_f \quad (1)$$

其中, 所有具有增强层需求的用户一定有基本层需求, 所以 $M \in Z$ 。

1.1 通信模型

UAV 将服务周期分成若干时隙 $\mathcal{T} = \{1, 2, \dots, T\}$, 在第 t 个时隙投射到地面的位置为 $L_u = (x_u^t, y_u^t)$, 其服务用户 m 的水平坐标为 $L_m = (x_m^t, y_m^t)$ 。SBS 服务用户 n 的水平坐标为 $L_n = (x_n^t, y_n^t)$ 。

UAV 和请求增强层视频的用户之间的数据传输信道假设为视距 (line of sight, LOS) 链路。由于 UAV 的机动性, 假设多普勒效应得到完美补偿。在任意时隙 t , UAV 与地面用户 m 的距离为 $d_m(t) = \sqrt{H^2 + \|L_u - L_m\|^2}$, $\|\cdot\|$ 表示欧几里得范数。UAV 可以使用正交频分多址 (orthogonal frequency division multiple access, OFDMA) 同时与多个强求增强层视频的用户通信, 即为每个用

户分配总带宽/传输功率的一小部分^[17]。因此, 每个请求增强层视频的用户不受其他用户的干扰。UAV 到位于 L_m 的用户的瞬时信道功率增益可以表示为:

$$h_{um} = \rho_u d_m^{-2}(t) = \frac{\rho_u}{H^2 + \|L_u - L_m\|^2}, 0 \leq t \leq T \quad (2)$$

其中, ρ_u 表示 UAV 在参考距离 $d_0 = 1$ m 处的信道功率增益。同理 SBS 与请求基本层视频的用户信道增益为:

$$h_{fn} = \rho_f d_n^{-2}(t) = \frac{\rho_f}{H^2 + \|L_f - L_n\|^2}, 0 \leq t \leq T \quad (3)$$

其中, ρ_f 表示 SBS 在参考距离 $d_0 = 1$ m 处的信道功率增益, $d_n = \|L_f - L_n\|$ 为第 n 个用户到 SBS 的距离。

假设 UAV 与 SBS 复用相同的频带, 因此 UAV 与请求增强层视频的用户 m 之间产生的信干噪比 (signal to interference plus noise ratio, SINR) 可以表示为:

$$\text{SINR}_{um} = \frac{P_{um} h_{um}}{P_{fn} h_{fn} + \sigma} \quad (4)$$

其中, P_{um} 表示 UAV 对请求增强层视频的用户传输功率, P_{fn} 表示 UAV 复用的 SBS 的传输功率, h_{um} 代表 UAV 与请求增强层视频的用户信道增益, h_{fn} 代表 SBS 与请求基本层视频的用户信道增益, 用 B 表示复用的带宽大小, σ 表示加性白高斯噪声 (additive white Gaussian noise, AWGN)。UAV 与请求增强层视频的用户 m 的传输速率可以表示为:

$$R_{um} = B \log(1 + \text{SINR}_{um}) \quad (5)$$

1.2 视频质量需求模型

为了满足用户对视频质量的需求, 将 SVC 技术引入无线视频传输中, SVC 视频流可以编码成一个基本层和多个增强层, 用户能够解调的视频层数越多, 获得的视频质量也就越高。用户可以根据自身的信道条件接收增强层, 信道条件好的用户可以接收更多的增强层。在热点区域流量需求高峰时段, SBS 向所有用户提供基本层服务, 同时由于缓存容量限制, 只能为一部分用户提供增强层服务, 因而通过 UAV 辅助为具有增强层需求的用户提供服务。无人机协助下基于 SVC 的视频层传输模型如图 2 所示。不同用户能接收的 SVC 视频层数也会根据数据速率的不同而有所差异。为了确保请求视频的用户都能接收到视频信息, SBS 向具有基本层需求的用户传输基本层数据的传输速率不小于 R_B ^[18], 请求增强层视频的用户接收的增强层数量与 UAV 之间的传输速率关系可以表示为:

$$\beta_{u,m} = \begin{cases} 0, R_{um} < R_{E1} \\ 1, R_{E1} < R_{um} < R_{E2} \\ 2, R_{um} > R_{E2} \end{cases} \quad (6)$$

其中, R_{E1} 为 UAV 与请求增强层 1 用户之间的速率阈值, R_{E2} 为 UAV 与请求增强层 2 用户之间的速率阈值。

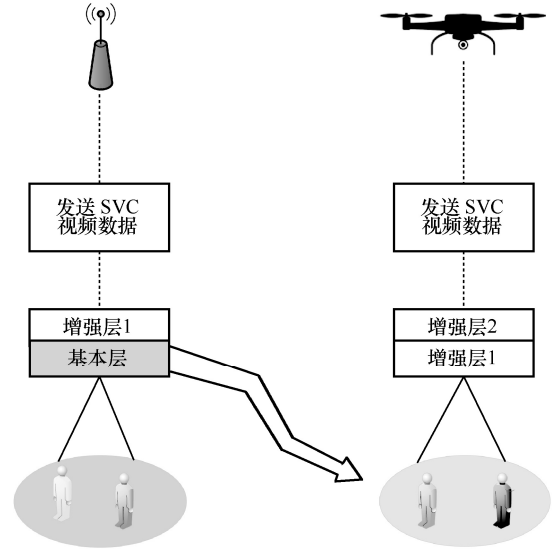


图2 无人机协助下基于 SVC 的视频层传输模型

1.3 UAV 能耗模型

尽管 UAV 具有视距信道和高移动性的优势, 但是 UAV 无法提供长时间的续航能力, 因此在研究 UAV 通信时需要考虑其能耗问题。通信 UAV 服务能耗主要包括飞行能耗、悬停能耗和传输能耗。 E_u^{fly} 、 E_u^{hover} 和 E_{u-m}^{Trans} 分别表示 UAV 的飞行能耗、悬停能耗和传输能耗, E_{UAV} 表示三者之和。

$$E_{\text{UAV}} = E_u^{\text{fly}} + E_u^{\text{hover}} + E_{u-m}^{\text{Trans}} \quad (7)$$

(1) 飞行能耗

旋翼 UAV 需要消耗推进能量来保持 UAV 在空中飞行并支持其运动, 速度 v 下的飞行功率可表示为^[19]:

$$P_u^{\text{fly}}(v) = P_0 \left(1 + \frac{3v^2}{U_{\text{tip}}^2} \right) + P_i \left(\sqrt{1 + \frac{v^4}{4v_0^4}} - \frac{v^2}{2v_0^2} \right)^{\frac{1}{2}} + \frac{1}{2} d_0 \rho s A v^3 \quad (8)$$

其中, P_0 、 U_{tip} 、 P_i 、 v_0 、 d_0 、 ρ 、 s 和 A 是与 UAV 质量、机翼面积、空气密度等因素相关的常数。UAV 在任意时隙 t 的飞行能耗表示为:

$$E_u^{\text{fly}}(t) = P_u^{\text{fly}}(\|v(t)\|) \quad (9)$$



(2) 悬停能耗

基于 DDPG 算法获得 UAV 的最优悬停位置, UAV 在时间段 $[t_0, t_0 + t]$ 的悬停能耗为:

$$E_u^{\text{hover}} = \int_{t_0}^{t_0+t} P_u^{\text{hover}} dt \quad (10)$$

其中, P_u^{hover} 为 UAV 在悬停时间的功率。

(3) 传输能耗

无线传输能量消耗与通过无线回程链路传输的数据位总数成正比。 E_{u-m}^{Trans} 为 UAV 到用户 m 的传输能耗:

$$E_{u-m}^{\text{Trans}}[t] = \sum_{m=1}^{m'} P_{u,m} \frac{\mu_{u,m}}{R_{u,m}} \quad (11)$$

其中, m' 表示 t 时隙 UAV 服务的用户总数, $\mu_{u,m}$ 表示用户 m 所需视频文件大小。

UAV 飞行时受到电池容量的限制。因此, 在时刻 t , UAV 的剩余能量为最大能量减去 UAV 飞行能耗、传输能耗与悬停能耗, 可表示为:

$$E_u^{\text{rem}}(t) = \left[E_u^{\text{max}} - \int_0^t P_u^{\text{fly}}(t) dt - E_{u-m}^{\text{Trans}}[t] - E_u^{\text{hover}} \right]^+ \quad (12)$$

其中, $[x]^+ = \max\{x, 0\}$, 即在任意时隙 t , UAV 的剩余能量都大于或等于 0。

1.4 视频质量感知的优化问题构建

为了最大化系统中用户整体接收的视频质量, 需要优化 UAV 的轨迹, 使覆盖范围内的用户整体收到的增强层数最多。为便于实验和性能评估, 本文考虑 SVC 方式, 保证所有请求视频资源的用户都收到基本层, 在此基础上, 用户整体接收到的增强层越多, 其体验感越佳。UAV 在为请求增强层用户提供服务时, 通过考虑能耗约束来优化 UAV 轨迹, 以优化视频质量感知为目标来最大化用户接收增强层数量。对应的优化问题表述如下。

$$P1: \max \sum_{m \in M} \beta_{u,m}$$

$$C1: x_u^{\min} \leq x_u^t \leq x_u^{\max}, y_u^{\min} \leq y_u^t \leq y_u^{\max}$$

$$C2: \|L_u[t] - L_u[t+1]\|^2 \leq (v_{\max} \Delta t)^2, \forall t \in 1, 2, \dots, T$$

$$C3: E_u^{\text{rem}}(t) \geq 0, \forall t \in 1, 2, \dots, T$$

$$C4: \beta_{u,m}(t) \in (0, 1, 2), \forall m, t \quad (13)$$

$$C5: \sum_{m=1}^M \beta_{u,m}(t) \leq 1, \forall m, t$$

其中, 在约束 C1 中 x_u^t, y_u^t 为 UAV 位置投射地面的水平坐标; 约束 C2 代表 UAV 飞行距离不能超过最大飞行速度在一个时隙内能够飞行的距离; 约束 C3 表示 UAV 的能量限制; 约束 C4 中 $\beta_{u,m} = 1$ 表示用户 m 可以接收到来自 UAV 的增强层 1, $\beta_{u,m} = 2$ 表示用户 m 可以接收到来自 UAV 的增强层 2,

$$\beta_{u,m} = 0 \text{ 则未收到, 即 } \beta_{u,m} = \begin{cases} 0, R_{um} < R_{E1} \\ 1, R_{E1} < R_{um} < R_{E2} \\ 2, R_{um} > R_{E2} \end{cases}; \text{ 约}$$

束 C5 表示在任意时隙 t 用户最多选择接收一个增强层。

2 基于 DDPG 的 UAV 轨迹优化设计

由于上述优化问题是多约束且非凸的, 传统的优化方法难以直接解决。因此本文将该问题中动态的无线环境有效地建模为 MDP, 它允许强化学习中的智能体通过与未知环境交互来学习, 以最大限度地提高回报。为此, 本文将用户接收增强层的数量最大化问题表述为 MDP 并随后采用一种基于连续动作控制的 DDPG 深度强化学习算法, 通过训练深度神经网络, 获得累积用户接收增强层数量的奖励, 以最大化用户接收增强层的数量。本文所采用的算法将 UAV 辅助的异构蜂窝网络系统建模为环境, UAV 本身被视为智能体。利用 DDPG 深度强化学习算法能够成功地优化 UAV 轨迹, 相比于深度 Q 网络这种只能处理离散动作空间的算法和传统的最短路径 (shortest path, SP) 算法, DDPG 深度强化学习算法能最大限度地增加用户接收到的增强层的数量。本节展示了将问题转换为 MDP 的过程和本文提出的 UAV 路径规划算法。

2.1 算法设计

在这种动态网络环境中, 本文采用 DDPG 算法来学习和估计状态动作值函数(称为 Q 值函数)。首先构建环境模型, 在该模型中 UAV 的位置、用户的位置和用户接收的增强层数量都具有马尔可夫性质, 可将其映射到 MDP 的交互环境, 基于 UAV 的 DDPG 网络结构如图 3 所示。在数学上 MDP 被定义为四元组 $\langle S, A, r, p \rangle$, 即状态空间、动作空间、奖励和状态转移概率。在每个时隙 t 中, 智能体从环境中获取当前状态 s_t , 执行动作 a_t , 与环境交互后对应地到达下一个状态 s_{t+1} , 并得到奖励值 r_t 。UAV 将这种迭代转换作为经验存储在回放缓冲池中, 并从回放缓冲池中随机采样以训练两个神经网络, 即在线策略网络 and 在线 Q 网络, 从而分别近似动作函数和动作值函数。UAV 的状态空间、动作空间和奖励设计如下。

(1) 状态空间

S 为智能体状态的集合, s_t 表示智能体在时隙 t 的状态, $s_t \in S$, 由当前 UAV 的坐标、用户坐标和用户接收增强层的数量组成, 即 $s_t = \{L_u^t, L_m^t, F_m^t\} \in L \times L \times F$ 。

(2) 动作空间

A 为智能体动作的集合, a_t 表示智能体在时

隙 t 的动作, $a_t \in A$, 智能体的动作由 3 个部分组成: UAV 的水平坐标 x_u^t 、UAV 的竖直方向坐标 y_u^t 、UAV 角度 $\theta \in (0, \pi/2)$ 。

(3) 奖励设计

DDPG 的目标是最大化奖励的累积期望。为了实现 P1 中的目标, 应该在奖励函数中包含用户接收增强层的数量。对于一个时间步长 t , 用户接收增强层的数量 $\beta_{u,m}$ 可以根据 UAV 的位置计算得出。本文将用户接收增强层的数量作为奖励。这样, UAV 将其状态转向用户接收增强层数量最大的最优状态, 从而获得最大的回报, 然后维持在此最优状态, 使累积期望奖励最大化。此外, 在达到最大时间步长 $t_{\max}$ 之前, UAV 不会停止状态转换。如果 UAV 在时间步长 $t < t_{\max}$ 时已转向最优状态, 则根据一个非零值的动作继续将状态转向次优状态, 因此, 需要一个惩罚奖励来限制行为, 即位移 d_t , 以提高收敛性能。在时间步长 t 处的奖励函数可以定义为:

$$r_t = \kappa \beta_{u,m}^t - \lambda \|d_t\|^2 \quad (14)$$

其中, κ 、 λ 为正常数, 可将 $\beta_{u,m}^t$ 和 $\|d_t\|^2$ 重新调整, 将奖励调整到一个合理的数量级。在训练开始时, $\beta_{u,m}^t$ 支配着奖励, 随着训练的不停迭代, 奖励的增加变得缓慢, 然后位移惩罚的第二项生效,

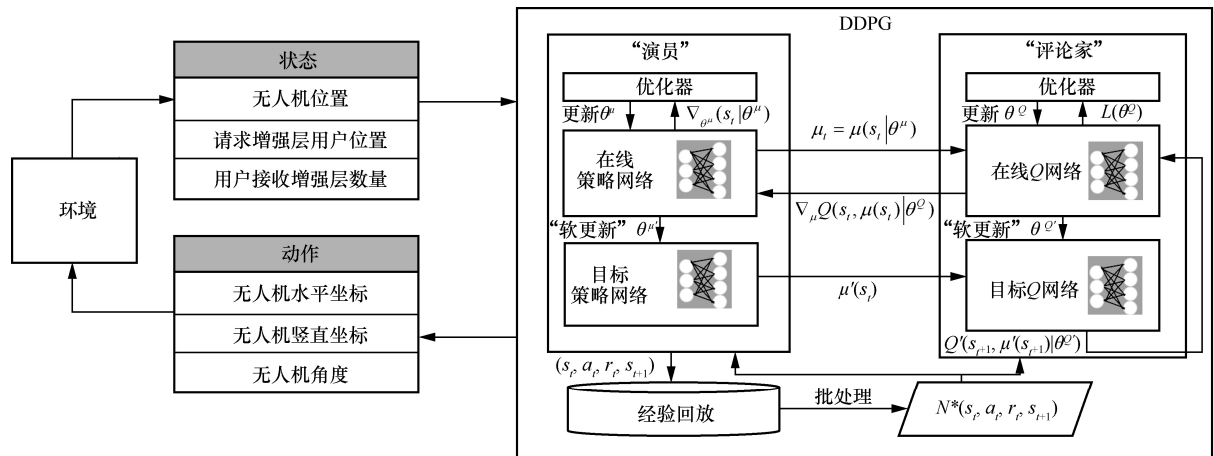


图3 基于 UAV 的 DDPG 网络结构



以阻止 UAV 大规模探索,从而使 UAV 能够更稳定地悬停在最佳位置。

2.2 算法流程

本文针对 UAV 辅助热点区域可伸缩视频传输场景,提出一种基于 DDPG 的 UAV 路径规划设计算法,以最大化用户接收增强层视频的数量。在该算法中,UAV 通过 DDPG 找出最优的策略,即对于每一个状态 s_t ,找出使未来累积奖励值 r 最大的动作,定义 $\mu(s_t|\theta^\mu)$ 为动作函数,实现了在时间步长 t 上从状态 $s_t \in S$ 到动作 $a_t \in A$ 的映射。动作-价值函数定义为 $Q(s_t, a_t|\theta^Q)$, 输出一个 Q 值来计算在 s_t 处的 a_t 。其中, θ^μ 和 θ^Q 分别是在线策略网络 and 在线 Q 网络的参数, θ^μ 通过最小化基于 Q 值的损耗来更新:

$$L(\theta^\mu) = \frac{1}{N} \sum_i -Q(s_i, \mu(s_i|\theta^\mu)|\theta^Q) \quad (15)$$

其中, N 是用于训练的经验样本的大小。

为了稳定训练过程,在线策略网络 and 在线 Q 网络中分别使用了具有相同结构的子网络,即目标神经网络。通过最小化基于时序差分(temporal difference, TD)误差的损失来更新在线 Q 网络的参数 θ^Q :

$$L(\theta^Q) = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i|\theta^Q))^2 \quad (16)$$

其中, $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$ 由目标策略网络 $\mu'(s|\theta^{\mu'})$ 和目标 Q 网络 $Q'(s, a|\theta^{Q'})$ 计算得出, $\theta^{\mu'}$ 和 $\theta^{Q'}$ 分别是目标策略网络 and 目标 Q 网络的参数, γ 是奖励折扣因子。 $\theta^{\mu'}$ 和 $\theta^{Q'}$ 更新通过“软更新”实现:

$$\theta^{\mu'} \leftarrow \tau \theta^\mu + (1-\tau) \theta^{\mu'}, \theta^{Q'} \leftarrow \tau \theta^Q + (1-\tau) \theta^{Q'} \quad (17)$$

其中, τ 为“软更新”系数,且 $\tau \ll 1$ 。

本文基于 DDPG 算法来优化 UAV 的轨迹,通过介绍该算法的结构,给出了优化 UAV 轨迹的过程。基于 DDPG 的 UAV 轨迹优化算法如算法 1 所示。

算法 1 基于 DDPG 的 UAV 轨迹优化算法

输入 请求增强层用户的位置 $\{L_1, L_2, \dots, L_M\}$, UAV 的传输功率 P_{um} , 经验大小 N 和学习率

初始化 $\mu(s|\theta^\mu)$ 、 $\mu'(s|\theta^{\mu'})$ 和 $Q(s, a|\theta^Q)$ 、 $Q'(s, a|\theta^{Q'})$, 其中, $\theta^{\mu'} = \theta^\mu$, $\theta^{Q'} = \theta^Q$

初始化经验缓冲池

for 训练回合 = 1: E **do**

 获得 s_1

for step = 1: T **do**

 将 UAV 观测到 UAV 的位置、用户的位置、用户接收的增强层数量作为状态 s_t

 得到动作 $a_t = \mu(s_t|\theta^\mu) + \text{noise}$, 其中 noise 为添加的随机噪声

 UAV 执行动作 a_t , 更新 UAV 位置, 得到 r_t 和 s_{t+1}

 存储 (s_t, a_t, r_t, s_{t+1}) 到经验缓冲池中

if 回放缓冲池已满 **then**

 从缓冲池中随机取样 N 个经验

 通过式 (15) 和式 (16) 计算损失, 并更新在线策略网络 and 在线 Q 网络

 通过式 (17) “软更新”目标神经网络

end if

end for

end for

返回 UAV 最优悬停位置和接收的最大增强层数量

3 仿真结果

本节通过仿真实验验证了基于 DDPG 的 UAV 路径规划算法的性能,并与其他算法进行了对比。本文共设置 4 种不同用户分布仿真场景,每个场景均经过 2 000 次训练回合以验证算法的鲁棒性。场景参数设置为: UAV 飞行速度 $v_{\max} = 20$ m/s, 区域长度 $L = 400$ m, 用户个数 $Z = 100$ 。使用 Python 进行仿真设计,模型仿真参数见表 1。Actor 神经网络和 Critic 神经网络的

结构都由一个输入层、两个隐藏层和一个输出层组成。采用 Adam 优化器对神经网络进行训练, 经验大小 $N = 256$, DDPG 算法参数为: “软更新”系数 $\tau = 0.001$, 奖励折扣因子 $\gamma = 0.9$, 动作-价值网络的学习率 $l_r = 0.001$, 动作噪声方差 $\sigma = 0.05$ 。由于用户位置的状态空间太大, 本文取 10 次训练的平均值作为训练回合值, 同样最终收敛值也有一定的波动, DQN 算法参数的参数设置与 DDPG 基本相同。

表 1 模型仿真参数

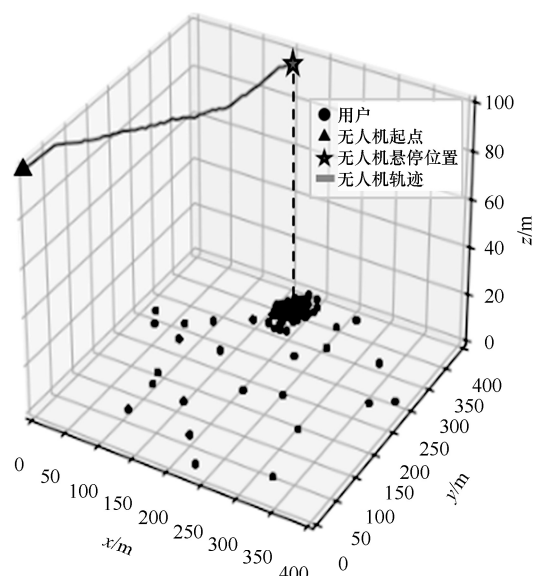
参数	值
用户数 Z	100
UAV 高度 H	100 m
用户范围和 UAV 轨迹范围 $\{x_u^{\max}, y_u^{\max}\}$	[400 m, 400 m]
1 m 参考距离的信道增益 ρ_u 、 ρ_f	-30 dB
路径损耗指数 α	2
带宽 B	10 MHz
UAV 的传输功率 P_{um}	1 W
UAV 的速度 $v_{\max}$	20 m/s
奖励函数中的缩放常数 κ	1×10^{-2}
奖励函数中的缩放常数 λ	1×10^{-3}
SBS 与基本层用户的速率阈值 R_B	0.442 Mbit/s
UAV 与增强层 1 的用户速率阈值 R_{E1}	0.827 Mbit/s
UAV 与增强层 2 的用户速率阈值 R_{E2}	1.114 Mbit/s

本文选取 UAV 一个飞行周期内服务的用户接收增强层的数量作为衡量不同算法性能指标。在仿真实验中, 本文通过 DDPG 算法优化 UAV 最优悬停位置并与另外两种算法进行比较。第一种为 SP 算法, 即 UAV 沿着一条最短的路径飞行, 直到最大飞行时间 (因为该算法为确定性算法, 其服务用户接收增强层的数量不随训练回合变化)。第二种为 DQN 算法, DQN 只能在离散的行动空间中搜索策略。因此, 将连续的动作空间划分为几个等长的离散动作, 从而用 DQN 求解有约束的 MDP。具体来说, 将 UAV 连续的 x 轴飞行速度 $[0, v_{\max}]$ 划分为离散的 x 轴速度集

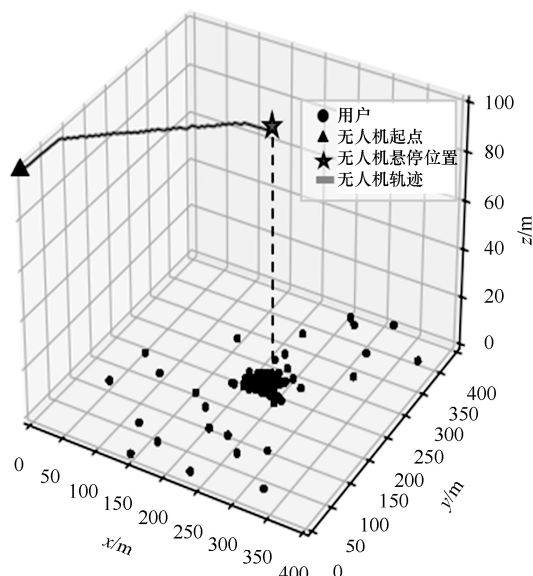
$v_x = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\} \times v_{\max}$, 将连续的 y 轴飞行速度划分为离散的 y 轴速度集 $v_y = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\} \times v_{\max}$ 。连续发射功率被划分为离散发射功率集 $P = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\} \times P_{um}$ 。

在最终收敛的情况下, 不同用户分布下 UAV 的轨迹和最优悬停位置如图 4 所示。可以观察到, UAV 在不同的用户分布下总是在较多用户分布的热点上方悬停, 并为这些用户建立良好的信道便于传输增强层视频。因为算法在设计奖励项时与目标函数紧密相关, 该算法可以有效指导 UAV 做出最优决策, 因此本文算法可以有效地提升用户差异化体验。DDPG 在最大化用户接收增强层数量上优于 DQN, 由于行为空间离散, DQN 可能会错过最优策略。而 DDPG 探索的是一个连续的动作空间, 采取的是准确的动作, 因此可以找到最优的策略。

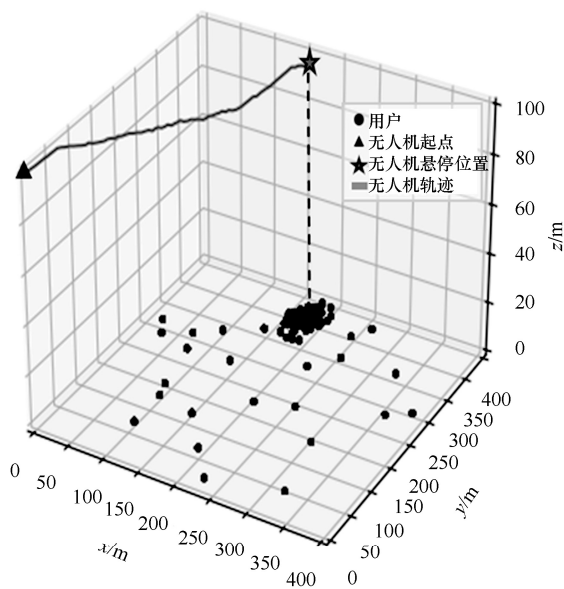
不同用户分布下用户接收的增强层数量如图 5 所示, 可以观察到 DQN 算法总是比 DDPG 算法性能更差, 并且差距逐渐扩大。这是因为离散化引起的偏差随着动作空间维数的增加而增大。而 SP 算法总是以最短路径悬停在用户热点分布上方, 没有奖励指导 UAV 做出决策, 所以用户接收到的增强层数量几乎不会变化。因此基于 DDPG 这种可以处理连续动作空间的算法可以获得比 DQN、SP 更优的性能。仿真结果表明, 在用户热点分布为(160,360)时, 与 DQN 算法、SP 算法相比, DDPG 算法用户接收增强层的数量分别提高了 34.8%、71.4%; 在用户热点分布为(210,270)时, 与 DQN 算法、SP 算法相比, 用户接收增强层的数量分别提高了 57.8%、75.7%; 在用户热点分布为(260,340)时, 与 DQN 算法、SP 算法相比, 用户接收增强层的数量分别提高了 45.3%、78.6%; 在用户热点分布为(320,190)时, 与 DQN 算法、SP 算法相比, 用户接收增强层的数量分别提高了 53.7%、80%。



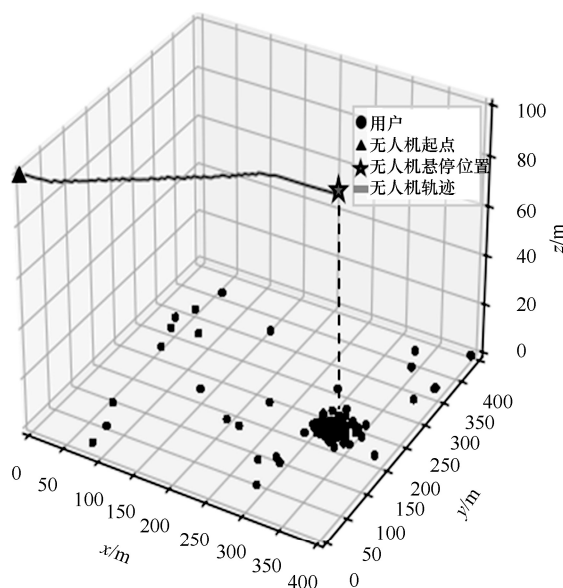
(a) 用户热点分布为(160,360)



(b) 用户热点分布为(210,270)



(c) 用户热点分布为(260,340)



(d) 用户热点分布为(320,190)

图4 不同用户分布下 UAV 的轨迹和最优悬停位置

3 种算法的奖励回报值对比如图 6 所示, 随着迭代次数的增加, DDPG 算法和 DQN 算法的动作价值函数逐渐收敛。从图 6 可以看出, 随着训练的进行, 智能体很快学会获得更高的预期总奖励, 然后使回报值在高水平上稳定收敛。在前 40 个训练回合, 奖励的回报值在非常低的水平上波动, 这是因为 UAV 处于完全探索阶段, 在没有足够经验可供学习的情况下, 行动是随机选

择的。当回放缓冲池满时, UAV 开始对存储的经验元组进行采样, 以训练网络。可以看到, 在前 40 个训练回合, DQN 的奖励回报值高于 DDPG, 但随着迭代次数的增加, DDPG 展示出较好的回报值。SP 几乎不受训练次数的影响, 始终维持在一个固定值。DDPG 随训练迭代的损失值变化如图 7 所示, 可以看到随着损失函数的振荡收敛, 优化目标基本稳定。仿真结果表明,

本文所提出的基于 DDPG 的优化方案可以使用户接收到更多的增强层数量。

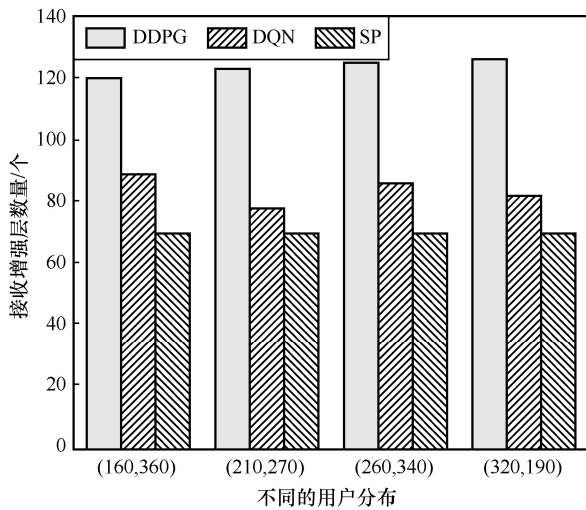


图5 不同用户分布下用户接收的增强层数量

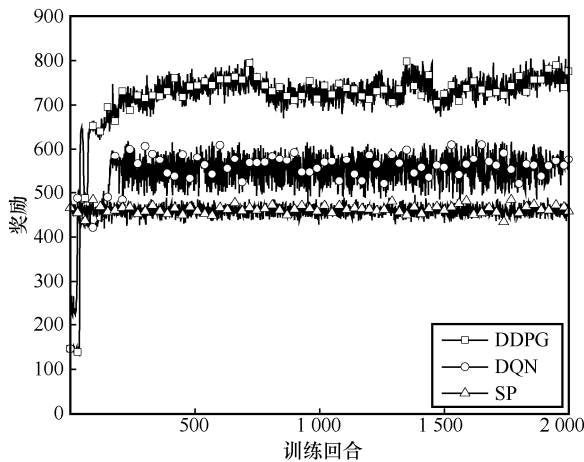


图6 3种算法的奖励回报值对比

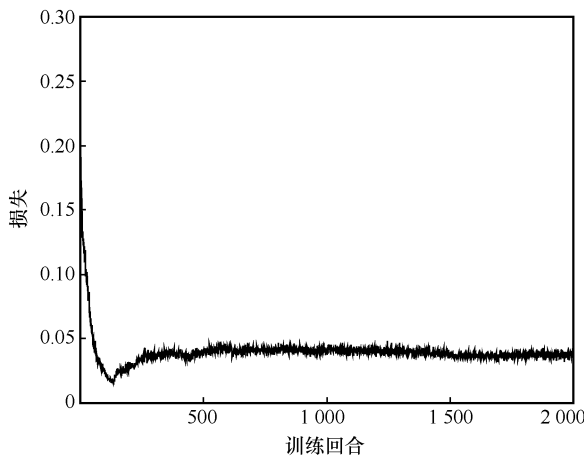


图7 DDPG 随训练迭代的损失值变化

3种方案的能耗变化如图8所示。设置观察时间为100个时隙，在用户数量固定的情况下，随着UAV从起点飞行到悬停位置，UAV为越来越多的用户提供增强层视频传输服务，UAV的能耗呈增长趋势。可以观察到，与其他算法相比，所提出的基于DDPG的算法可以实现最低的能耗。这是由于DDPG探索的是一个连续的行动空间，采取的是准确的行动，因此可以找到最优的策略以节省能耗。当时隙为100时，DDPG算法的能耗分别比DQN和SP算法的能耗低16.4%、53.1%。

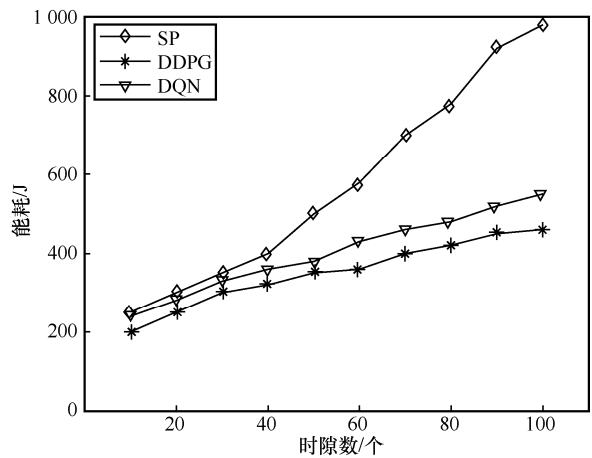


图8 3种方案的能耗变化

在仿真环境下用不同模型运行2000次迭代后，统计用户接收视频图像的峰值信噪比（peak signal to noise ratio, PSNR）。在基本层数据率为0.442 Mbit/s、增强层1数据率为0.827 Mbit/s和增强层2数据率为1.114 Mbit/s的情况下，PSNR的累积分布函数（cumulative distribution function, CDF）如图9所示。从图9可以看出，DDPG算法的PSNR主要分布于36.5~37.4 dB，DQN算法的PSNR主要分布于35.5~36.3 dB，SP算法的PSNR主要分布于32.5~33.5 dB。DDPG算法在用户接收视频质量的分布上优于DQN和SP算法。这是由于DDPG在估值^[20]方面虽然类似于DQN，但是DQN只能处理离散的动作空间，这导致了最佳



解的丢失。而 DDPG 能够处理连续动作空间，决策 UAV 的动作更加准确，使 UAV 的轨迹和悬停位置更加合理。

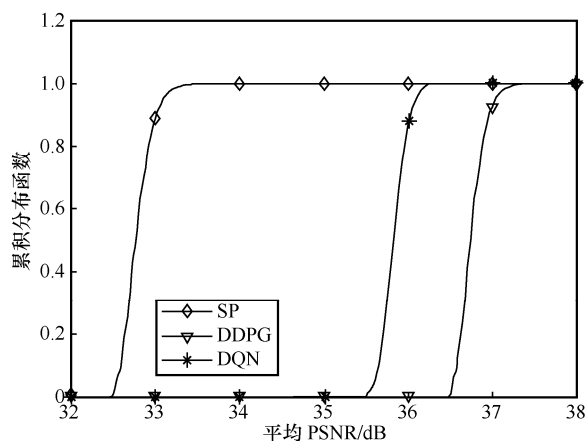


图9 PSNR 的 CDF

4 结束语

针对热点区域小基站负载的场景，本文基于深度强化学习中的 DDPG 算法提出优化 UAV 的轨迹和最佳悬停位置的算法，以最大化用户接收增强层视频的数量来实现热点区域覆盖增强和用户差异化体验。仿真结果表明，本文所提的算法与 DQN 算法、SP 算法相比，在不同的用户分布下得到的增强层数量均有所提高，验证了所提算法的可行性。在今后的工作中，随着无人机作业的任务复杂度不断升级，单个无人机往往难以完成复杂的作业任务，笔者会对多 UAV 组网的部署展开一定的研究。

参考文献：

- [1] WU Q Q, ZENG Y, ZHANG R. Joint trajectory and communication design for multi-UAV enabled wireless networks[J]. IEEE Transactions on Wireless Communications, 2018, 17(3): 2109-2121.
- [2] FU Y J, MEI H B, WANG K Z, et al. Joint optimization of 3D trajectory and scheduling for solar-powered UAV systems[J]. IEEE Transactions on Vehicular Technology, 2021, 70(4): 3972-3977.
- [3] 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.

- 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.
- [4] ZHOU H, HU F H, JURAS M, et al. Real-time video streaming and control of cellular-connected UAV system: prototype and performance evaluation[J]. IEEE Wireless Communications Letters, 2021, 10(8): 1657-1661.
- [5] ZHAN C, HUANG R J. Energy efficient adaptive video streaming with rotary-wing UAV[J]. IEEE Transactions on Vehicular Technology, 2020, 69(7): 8040-8044.
- [6] ZHAO N, CHENG F, YU F R, et al. Caching UAV assisted secure transmission in hyper-dense networks based on interference alignment[J]. IEEE Transactions on Communications, 2018, 66(5): 2281-2294.
- [7] LIU Z C, JIANG Y. Cross-layer design for UAV-based streaming media transmission[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 4710-4723.
- [8] ZHANG Y Q, XU C, HEMADEH I A, et al. Near-instantaneously adaptive multi-set space-time shift keying for UAV-aided video surveillance[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 12843-12856.
- [9] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [10] XIAO L, DING Y Z, HUANG J H, et al. UAV anti-jamming video transmissions with QoE guarantee: a reinforcement learning-based approach[J]. IEEE Transactions on Communications, 2021, 69(9): 5933-5947.
- [11] LIU Y, YAN J J, ZHAO X H. Deep reinforcement learning based latency minimization for mobile edge computing with virtualization in maritime UAV communication network[J]. IEEE Transactions on Vehicular Technology, 2022, 71(4): 4225-4236.
- [12] DING R J, GAO F F, SHEN X S. 3D UAV trajectory design and frequency band allocation for energy-efficient and fair communication: a deep reinforcement learning approach[J]. IEEE Transactions on Wireless Communications, 2020, 19(12): 7796-7809.
- [13] CUI Y L, DENG D H, WANG C W, et al. Joint trajectory and power optimization for energy efficient UAV communication using deep reinforcement learning[C]//Proceedings of IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2021: 1-6.
- [14] SAMIR M, ASSI C, SHARAFEDDINE S, et al. Age of infor-

mation aware trajectory planning of UAVs in intelligent transportation systems: a deep learning approach[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 12382-12395.

- [15] KANG H Y, CHANG X L, MIŠIĆ J, et al. Improving dual-UAV aided ground-UAV Bi-directional communication security: joint UAV trajectory and transmit power optimization[J]. IEEE Transactions on Vehicular Technology, 2022, 71(10): 10570-10583.
- [16] WU D P, LIU Q R, WANG H G, et al. Cache less for more: exploiting cooperative video caching and delivery in D2D communications[J]. IEEE Transactions on Multimedia, 2019, 21(7): 1788-1798.
- [17] SONG D, ZHAI X B, LIU X, et al. Jointly optimal fair data collection and trajectory design algorithms in UAV-aided cellular networks[C]//Proceedings of 2021 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2021: 1-6.
- [18] CONDOLUCI M, ARANITI G, MOLINARO A, et al. Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks[J]. IEEE Transactions on Vehicular Technology, 2016, 65(5): 2907-2921.
- [19] ZENG Y, XU J, ZHANG R. Energy minimization for wireless communication with rotary-wing UAV[J]. IEEE Transactions on Wireless Communications, 2019, 18(4): 2329-2345.
- [20] TANG F X, HOFNER H, KATO N, et al. A deep reinforcement learning-based dynamic traffic offloading in space-air-ground integrated networks (SAGIN)[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(1): 276-289.

[作者简介]



闫俊杰 (1990-), 男, 博士, 广西科技大学电子工程学院讲师, 主要研究方向为 MEC、UAV 通信、D2D 通信。



朱亚新 (1998-), 女, 广西科技大学电子工程学院硕士生, 主要研究方向为 MEC、UAV 通信、D2D 通信。



冯艳如 (1990-), 女, 山西应用科技学院讲师, 主要研究方向为信息传输及信息安全。



刘汉永 (1992-), 女, 国家电网重庆市电力公司大足供电分公司工程师, 主要研究方向为区块链、网络安全、密码学、自动化、电网营销。



邓钧忆 (1985-), 男, 博士, 广西科技大学电子工程学院助理研究员, 主要研究方向为车联网、MEC。



王欢 (1987-), 男, 博士, 广西科技大学电子工程学院副教授, 主要研究方向为智能网络与信息安全、高可用软件。