

考虑能耗中断的无人机通信中基于深度强化学习的资源管理

罗佳^{*①②} 陈前斌^② 唐伦^② 张志才^③

^①(重庆邮电大学网络空间安全与信息法学院 重庆 400065)

^②(重庆邮电大学通信与信息工程学院移动通信技术重点实验室 重庆 400065)

^③(山西大学现代通信技术研究所 太原 237016)

摘要: 最新研究表明, 高速传输导致的手机温度变化会影响相应的传输性能。针对高速传输下未考虑与手机温度有关的能耗中断而导致传输性能降低的问题, 该文提出一种基于深度强化学习的资源管理方案去考虑无人机(UAV)通信场景下的能耗中断。首先, 给出无人机通信的网络模型与智能手机热传递模型的分析; 其次, 将能耗中断的影响以约束条件的形式整合到无人机场景的优化问题中, 并通过联合考虑带宽分配、功率分配和轨迹设计优化系统吞吐量; 最后, 采用马尔可夫决策过程描述相应的优化问题并通过名为归一化优势函数的深度强化学习算法求解。仿真表明, 所提方案能有效提升系统吞吐量并得到合理的无人机飞行轨迹。

关键词: 无人机通信; 能耗中断; 深度强化学习; 带宽分配; 轨迹设计

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2023)08-2885-08

DOI: 10.11999/JEIT220907

Resource Management Based on Deep Reinforcement Learning for UAV Communication Considering Power-Consumption Outage

LUO Jia^{*①②} CHEN Qianbin^② TANG Lun^② ZHANG Zhicai^③

^①(School of Cyber Security and Information law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

^②(Key Laboratory of Mobile Communication Technology, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

^③(Institute of Modern Communication Technology, Shanxi University, Taiyuan 237016, China)

Abstract: Recent research has demonstrated that the temperature variation of smartphone caused by high data rate transmission could affect the corresponding performance on transmission. Considering the problem of performance degradation on transmission caused by the ignorance of the power-consumption outage which is related with the temperature of smartphone, a deep reinforcement learning based resource management scheme is proposed to consider the power-consumption outage for Unmanned Aerial Vehicle (UAV) communication scenario. Firstly, the analysis for the network model of UAV communication and heat transfer model in smartphone is established. Then, the influence of power-consumption outage is integrated into the optimization problem of UAV scenario in the form of constraint, and the system throughput is optimized via the joint consideration of bandwidth allocation, power allocation and trajectory design. Finally, Markov decision process is adopted to depict the problem and the optimization target is achieved by a deep reinforcement learning algorithm named normalized advantage function. Simulation results manifest that the proposed scheme can effectively enhance the system throughput and achieve appropriate trajectory of UAV.

Key words: Unmanned Aerial Vehicle (UAV) communication; Power-consumption outage; Deep reinforcement learning; Bandwidth allocation; Trajectory design

收稿日期: 2022-07-05; 改回日期: 2022-11-14; 网络出版: 2022-11-21

*通信作者: 罗佳 luojia@cqupt.edu.cn

基金项目: 国家自然科学基金(62071078), 重庆市自然科学基金(cstc2021jcyj-bsh0175), 四川省科技计划(2021YFQ0053)

Foundation Items: The National Natural Science Foundation of China (62071078), The Chongqing Municipal Natural Science Foundation (cstc2021jcyj-bsh0175), The Sichuan Science and Technology Program (2021YFQ0053)

1 引言

作为无线通信的一个分支,无人机通信与蜂窝网络结合的范式正受到业界的持续关注,这一结合使无人机具有空中基站的功能,从而进一步辅助和增强蜂窝通信网络的应用范围^[1]。另外,为支撑巨大的网络流量需求,将无线传输速率提高到千兆比特每秒级别是当前及未来移动通信发展的主要方向之一。需注意的是,面对如此高的传输速率,接收端需对海量接收数据进行数据处理,从而导致智能手机等移动终端中计算资源的巨大消耗,这一现象使得未来移动通信网络将面临新的问题——能耗中断^[2]。

能耗中断与智能手机的散热以及相应计算资源的消耗有关。具体来说,智能手机芯片中的计算单元,尤其是基带处理器(Baseband Processor, BP)和应用处理器(Application Processor, AP)是对接收数据进行处理的核心模块。芯片执行计算时产生的热量可能会提高智能手机背板的温度,从而导致用户体验变差^[3]。为保护用户免受低温烫伤,相关的热力学设计将45 °C设定为手机背板温度的阈值^[4]。一旦手机背板温度超过该阈值,智能手机内部的温度调节器需降低芯片工作频率以降低手机背板温度^[5]。对于传统移动通信系统,由于无线传输速率相对较低,手机端处理无线接收数据对计算资源的需求量较低从而使其对手机温度的影响较小,手机背板过热则通常与本地应用中的计算密集型任务有关^[6]。然而,作为未来移动通信网络的主要特点,千兆比特每秒级别的高速无线传输使得无线通信的数据处理或将成为手机背板过热的导火索之一^[7]。文献^[8]首先指出,高速传输下基站侧的下行发送速率可能大于用户侧的最大接收速率。最大接收速率表示智能手机芯片能够对接收数据进行稳定的数据处理的速率。基于这种发送速率和接收速率不匹配的现象,文献^[2]指出并定义了能耗中断这一新的问题来描述与无线通信相关的计算过载,这种计算过载正是由于对无线接收数据处理的计算需求激增导致芯片过热时频率降低而引起的。另外,无人机通信在未来移动通信网络中逐渐占据重要地位^[9]。文献^[10]研究了无人机作为空中基站的自适应视频传输,通过联合考虑通信资源和无人机轨迹优化相应的用户体验。文献^[11]则针对无人机作为中继节点辅助传统基站传输的视频业务场景,联合考虑传输带宽和功率分配并通过李雅普诺夫优化方法实现长期用户体验优化。文献^[12]研究了无人机作为终端用户与传统基站通信的场景,提出一种优化的功率控制算法从而最大化上行传输的能量效率。

对于当前的研究,一方面,已有文献给出了能耗中断的基本原理^[2],然而其在具体场景中对资源管理和系统性能的影响以及相应的优化机制尚缺乏进一步的研究。另一方面,无人机可利用毫米波频段的大规模MIMO(Multiple-Input Multiple-Output)技术实现高速下行传输,该场景需考虑无线传输对终端温度的影响以避免终端手机用户发生低温烫伤事故,当前无人机通信的相关研究中并未考虑这一因素的影响。基于此,本文针对无人机通信场景,将能耗中断的影响引入高速传输的无人机通信模型,研究考虑能耗中断的无人机通信中多用户资源管理的基本范式。本文指出,为减轻能耗中断对过热用户的不利影响,作为空中基站的无人机侧可主动降低相应的传输速率,以避免手机终端侧的计算过载。因此,无人机可主动释放在于过热手机用户的部分通信资源,从而增强其他用户的无线数据传输,进而提高资源利用率。另外,由于手机背板温度随时间变化,且与无人机轨迹、传输带宽和发射功率等决定下行传输速率的因素有关,因此,为实现无人机轨迹设计以及带宽和功能分配的联合优化,本文采用马尔可夫决策过程建模考虑能耗中断的优化问题,并利用深度强化学习实现相应的优化目标。此外,为了方便地执行深度强化学习,本文基于相应问题的优化变量重新设计了动作向量以满足无人机的速度限制,并构造惩罚函数以考虑相应的约束条件。

2 系统模型和问题

2.1 网络模型

网络模型针对开阔的农村地区,考虑单个无人机在固定高度 H_u 飞行并作为空中基站为 N 个终端用户提供服务。为支持高速传输,无人机与用户间的下行传输采用毫米波频段的大规模MIMO技术,其中,无人机的发送天线数为 N_{Tx} ,用户侧的接收天线数为 N_{Rx} 。无人机利用不同的波束服务多个用户,即针对不同用户的无线传输使用不同的带宽。无人机在时隙 t 的水平坐标为 $\mathbf{q}_u(t) = [x_u(t), y_u(t)]$,用户 $n(n \in \{1, 2, \dots, N\})$ 的坐标固定为 $\mathbf{q}_n = [x_n, y_n]$ 。作为空中基站,无人机在每个时隙传输时悬停在空中以提供稳定的无线链路,并在下一个时隙更新其位置。由于其速度限制,无人机在两个相邻时隙间的最大移动距离为 D_{max} ,因此,可得约束条件

$$\|\mathbf{q}_u(t+1) - \mathbf{q}_u(t)\| \leq D_{max}, \forall t \quad (1)$$

本文在网络模型中采用文献^[13]中描述的设定,即无人机的能量供应由激光充电提供,这使得无人机能不间断地在空中飞行以提供持续的无线传

输服务。另外, 基于通信场景(农村地区)和无人机的高机动性, 可合理假设无人机和用户间的下行链路主要为视距(Line of Sight, LoS)传输。路径损耗的计算则采用文献[14]中的标准对数正态阴影模型, 因此, 无人机和用户 n 之间在时隙 t 视距传输的路径损耗(单位: dB)可表示为

$$l_{n,\text{LoS}}(t) = L_{\text{FS}}(d_0) + 10\chi_{\text{LoS}} \lg(d_n(t)) + \psi_{\text{LoS}} - G \quad (2)$$

其中, G 是无人机的发射天线增益。 χ_{LoS} 和 ψ_{LoS} 是路径损耗指数和视距传输的阴影变量。 $L_{\text{FS}}(d_0) = 20\lg(4\pi d_0 f_{\text{CF}}/c)$ 是自由空间路径损耗。 c 是光速。 d_0 代表自由空间的参考距离。 f_{CF} 是载波频率。 $d_n(t)$ 表示无人机与用户 n 之间的距离, 其计算公式为

$$d_n(t) = \sqrt{H_u^2 + [x_u(t) - x_n]^2 + [y_u(t) - y_n]^2} \quad (3)$$

类似地, 无人机和用户 n 之间在时隙 t 非视距传输的路径损耗(单位: dB)可表示为

$$l_{n,\text{NLoS}}(t) = L_{\text{FS}}(d_0) + 10\chi_{\text{NLoS}} \lg(d_n(t)) + \psi_{\text{NLoS}} - G \quad (4)$$

用复数矩阵 $\mathbf{H} \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{Tx}}}$ 表示信道矩阵。考虑小尺度衰落时, 由于存在视距传输, 本文采用莱斯衰落信道模型。根据莱斯衰落的性质[15], 对于复矩阵 \mathbf{H} 中的任一元素 h , 其实部和虚部分别为相互独立且方差相同、均值非零的正态随机变量, $\mathbb{E}[|h|^2] = b^2 + \sigma^2$, 其中, $b^2 = 10^{-l_{n,\text{LoS}}(t)/10}$, $\sigma^2 = 10^{-l_{n,\text{NLoS}}(t)/10}$ 。基于此, h 实部和虚部的方差均为 $\sigma^2/2$, 均值分别为 $b\cos\theta$ 和 $b\sin\theta$, 其中, θ 为任意实数。用 $p_n(t)$ 和 $\varpi_n(t)$ 分别表示分配给用户 n 的发射功率和传输带宽。因此, 对于时隙 t , 用户 n 的下行传输速率由式(5)表示

$$R_n(t) = \varpi_n(t) \log_2 \det \left(\mathbf{I}_{N_{\text{Rx}}} + \frac{p_n(t) \mathbf{H} \mathbf{H}^H}{\varpi_n(t) N_0} \right) \quad (5)$$

其中, $\mathbf{I}_{N_{\text{Rx}}} \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{Rx}}}$ 为单位矩阵, N_0 表示噪声谱密度。无人机的最大发射功率和总带宽分别用 P_{tx} 和 B 表示, 因此, 应满足以下约束

$$\sum_{n=1}^N \varpi_n(t) \leq B \quad (6)$$

$$\sum_{n=1}^N p_n(t) \leq P_{\text{tx}} \quad (7)$$

2.2 热传递模型

能耗中断与智能手机的热传递有关, 因此, 需明确其热传递模型。首先, 介绍热传递模型相关的参数。 c_{chip} (单位: J/(kg·K))和 m_c (单位: g)分别表示手机芯片的比热容和质量。手机的散热器将芯片产生的热量由芯片传递到手机背板。 k_1 (单位: W/(m·K))为芯片的导热系数。 A (单位: cm²)和 L (单位: mm)

分别表示散热片的面积和长度。此外, D (单位: mm)为手机背板的厚度, k_2 (单位: W/(m·K))表示其导热系数。空气的对流换热系数记为 h_{air} (单位: W/(m²·K))。对于用户 n 的芯片, 其总发热功率可表示为[2]

$$Q_n^{\text{Total}}(t) = P_n^{\text{BP}}(t) + P_n^{\text{AP}}(t) + P_n^{\text{sys}}(t) + \lambda Q_n^{\text{LNA}}(t) \quad (8)$$

其中, $P_n^{\text{AP}}(t)$ 和 $P_n^{\text{BP}}(t)$ 分别表示应用处理器和基带处理器的功耗。 $P_n^{\text{sys}}(t)$ 表示手机维持基本的系统运作所产生的功耗。 $Q_n^{\text{LNA}}(t)$ 表示低噪声放大器(Low Noise Amplifier, LNA)的发热功率。此外, λ 表示热量从低噪声放大器传递到手机芯片的比率。对于时隙 t , 用户 n 的基带处理器功耗可计算为[7]

$$P_n^{\text{BP}}(t) = K_{\text{BP}} F_{\text{BP}} \alpha_1 I_c N_0 \ln 2 R_n(t) \quad (9)$$

其中, K_{BP} 表示基带处理器处理单位比特的接收数据所需的逻辑操作数。 F_{BP} 是基带处理器中加载的逻辑门数量。 α_1 表示基带处理器晶体管的活跃因子。 I_c 是一个和半导体技术有关的不成熟因子。类似地, 应用处理器功耗可由式(10)给出

$$P_n^{\text{AP}}(t) = K_{\text{AP}} F_{\text{AP}} \alpha_2 I_c N_0 \ln 2 R_n(t) \quad (10)$$

其中, 应用处理器的 K_{AP} , F_{AP} 和 α_2 对应于基带处理器的 K_{BP} , F_{BP} 和 α_1 。此外, 手机维持基本系统运作的功耗可由式(11)给出[2]

$$P_n^{\text{sys}}(t) = h_{\text{air}} A (T_0^{\text{sur}} - T_{\text{env}}) \quad (11)$$

其中, T_0^{sur} 表示手机和周围环境处于热平衡状态时的手机背板温度, T_{env} 是环境温度。进一步地, 低噪声放大器的发热功率可计算为

$$Q_n^{\text{LNA}}(t) = N_{\text{Rx}} P_{\text{LNA}} (1 - \eta) \quad (12)$$

其中, P_{LNA} 和 η 是低噪声放大器的功耗和效率。基于上述模型, 可计算得出, 手机芯片在发热功率为 $Q_n^{\text{Total}}(t)$ 的强度下工作一段时间 τ 之后, 其手机背板温度为[2]

$$T_{\text{end}} = \frac{Q_n^{\text{Total}}(t)}{h_{\text{air}} A} \left(1 - e^{-\frac{z\tau}{c_{\text{chip}} m_c}} \right) + (T_{\text{start}} - T_{\text{env}}) \cdot e^{-\frac{z\tau}{c_{\text{chip}} m_c} + T_{\text{env}}} \quad (13)$$

其中, $z = 1 / \left(\frac{L}{k_1 A} + \frac{D}{k_2 A} + \frac{1}{h_{\text{air}} A} \right)$, T_{start} 表示该时间周期 τ 开始时的手机背板温度。

2.3 问题建模

令 $T_n^{\text{sur}}(t)$ 表示用户 n 在时隙 t 的初始手机背板温度。通常来说, 为保护用户免受低温烫伤, 需预设一个安全温度阈值 T_{safe} 。当手机背板温度达到或超过该安全温度时($T_n^{\text{sur}}(t) \geq T_{\text{safe}}$), 手机需降低其芯片工作频率以降低手机背板温度。因此, 可采用式(14)表示用户 n 的状态

$$H(\Delta_n(t)) = \begin{cases} 1, & \Delta_n(t) \geq 0 \\ 0, & \Delta_n(t) < 0 \end{cases} \quad (14)$$

其中, $\Delta_n(t) = T_n^{\text{sur}}(t) - T_{\text{safe}}$ 。 $H(\Delta_n(t)) = 1$ 表示时隙 t 用户 n 的初始手机背板温度达到或超过安全阈值 T_{safe} , 若未达到安全阈值, 则 $H(\Delta_n(t)) = 0$ 。

芯片的功耗与其工作频率相关^[16], 可以看出, 在计算相应处理器功耗的式(9)和式(10)中, $K_{\text{BP}}R_n(t)$ 和 $K_{\text{AP}}R_n(t)$ 即是与芯片工作频率相关的单位时间逻辑操作数。因此, 当芯片频率满足数据处理需求从而支撑 $R_n(t)$ 的传输速率时, 式(9)和式(10)可用于计算相应处理器的功耗。然而, 持续的高速传输会造成巨大的数据处理需求, 从而导致手机过热, 该情况下需降低芯片工作频率以降低手机背板温度, 从而降低芯片的逻辑运算能力, 这将导致手机端没有足够的计算资源可用于处理接收数据(即计算过载)以维持当前的高传输速率 $R_n(t)$, 以上这一系列连锁反应即是本文关注的能耗中断问题。可以看出, 若用户 n 在时隙 t 触发能耗中断, 其用户侧的最大接收速率 $R_n^{\text{rx}}(t)$ 实际上应低于 $R_n(t)$, 当采用式(9)和式(10)计算相应处理器功耗时应使用实际接收速率 $R_n^{\text{rx}}(t)$ 而非 $R_n(t)$, 因此, 无人机分配给用户 n 的通信资源中对应 $R_n(t)$ 比 $R_n^{\text{rx}}(t)$ 高出的部分是被浪费掉的。另外, 从式(13)可以看出手机背板温度与芯片发热功率 $Q_n^{\text{Total}}(t)$ 相关, 进一步地, $Q_n^{\text{Total}}(t)$ 则与速率 $R_n(t)$ 有关。受上述相关性的启发, 为了减轻能耗中断的不利影响, 本文设计在能耗中断发生的情况下在无人机侧主动降低 $R_n(t)$ 以适配用户侧的接收速率 $R_n^{\text{rx}}(t)$, 从而降低相应的手机背板温度。根据式(13)可得, 如果用户 n 需在时隙 t 降低其手机背板温度($T_{\text{end}} < T_n^{\text{sur}}(t)$), 应满足不等式

$$Q_n^{\text{Total}}(t) < h_{\text{air}}A(T_n^{\text{sur}}(t) - T_{\text{env}}) \quad (15)$$

因此, 对于任意用户 n , 应满足不等式

$$H(\Delta_n(t))Q_n^{\text{Total}}(t) < h_{\text{air}}A(T_n^{\text{sur}}(t) - T_{\text{env}}), \forall n \quad (16)$$

上述不等式表明, 应限制过热用户的传输速率以满足芯片的总发热功率约束。因此, 在涉及手机背板温度变化的场景中, 不等式(16)至关重要, 可利用该不等式在传统资源管理中去考虑能耗中断的影响。基于此, 相应的优化问题可表述为

$$\begin{aligned} \mathcal{P}_1: & \max_{\mathbf{q}_n(t+1), \boldsymbol{\omega}(t), \mathbf{p}(t)} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^N R_n(t) \\ \text{s.t. C1:} & \|\mathbf{q}_n(t+1) - \mathbf{q}_n(t)\| \leq D_{\text{max}}, \forall t \\ & \text{C2: } \sum_{n=1}^N \varpi_n(t) \leq B \end{aligned}$$

$$\text{C3: } \sum_{n=1}^N p_n(t) \leq P_{\text{tx}}$$

$$\text{C4: } H(\Delta_n(t))Q_n^{\text{Total}}(t) < h_{\text{air}}A(T_n^{\text{sur}}(t) - T_{\text{env}}), \forall n$$

在多用户环境下, 约束条件C4可促使无人机释放过热用户的部分带宽资源。另外, 无人机的物理位置也会影响下行传输速率。因此, 联合考虑带宽和功率分配以及无人机轨迹设计可提升低速非过热用户的传输性能从而减轻能耗中断对系统吞吐量的不利影响。基于此, 问题 \mathcal{P}_1 旨在最大化系统中多个用户的系统吞吐量。另外, 无人机轨迹以及带宽和功率分配不仅会影响当前时隙的系统吞吐量, 还会影响手机背板的温度变化, 从而影响未来的系统吞吐量。因此, 上述问题的目标函数是时间平均的形式。

3 基于深度强化学习的优化框架

由于手机背板温度和无人机轨迹的动态性, 可采用马尔可夫决策过程来描述问题 \mathcal{P}_1 。由于马尔可夫决策过程的动态变化性质, 当前时隙的动作会影响下一时隙的状态从而对将来时隙中的即时奖励产生影响。因此, 对马尔可夫决策过程的求解不应像静态优化一样仅关注当前时隙的即时奖励, 而应寻求多个时隙累计奖励的最大化。在已知马尔可夫决策过程模型(即转移概率已知)的情况下, 动态规划中的值迭代和策略迭代可用于求解马尔可夫决策过程。然而, 实际情况中大多数马尔可夫决策过程模型的转移概率均无法获得, 深度强化学习可在转移概率未知的情况下通过与环境的交互和学习求解马尔可夫决策过程。因此, 相应的问题求解可采用深度强化学习。时隙 t 的状态可表示为 $\mathbf{s}_t = [\mathbf{q}_n(t), \mathbf{T}^{\text{sur}}(t)]$, 其中 $\mathbf{T}^{\text{sur}}(t) = [T_1^{\text{sur}}(t), T_2^{\text{sur}}(t), \dots, T_N^{\text{sur}}(t)]$ 。根据问题 \mathcal{P}_1 , 深度强化学习智能体位于无人机侧, 其需在每个时隙 t 做出关于 $\mathbf{q}_n(t+1)$, $\boldsymbol{\omega}(t) = [\varpi_1(t), \varpi_2(t), \dots, \varpi_N(t)]$ 和 $\mathbf{p}(t) = [p_1(t), p_2(t), \dots, p_N(t)]$ 的决策, 因此动作向量应为 $\mathbf{a}_t = [\mathbf{q}_n(t+1), \boldsymbol{\omega}(t), \mathbf{p}(t)]$ 。然而, 值得指出的是, 该类问题中的约束条件可促使动作向量的重新设计从而以更直接的方式去满足相应的约束条件。基于此, 本文针对约束条件C1, 将动作向量重新设计为 $\mathbf{a}_t = [d(t), \xi(t), \boldsymbol{\omega}(t), \mathbf{p}(t)]$, 其中, $d(t)(0 \leq d(t) \leq D_{\text{max}})$ 是无人机在时隙 t 和 $t+1$ 之间的水平移动距离。 $\xi(t)(0 < \xi(t) \leq 2\pi)$ 为水平面的移动角。因此, $\mathbf{q}_n(t+1)$ 可计算为

$$x_u(t+1) = x_u(t) + d(t) \cos(\xi(t)) \quad (17)$$

$$y_u(t+1) = y_u(t) + d(t) \sin(\xi(t)) \quad (18)$$

为优化时间平均的总吞吐量, 奖励函数应包括

$\sum_{n=1}^N R_n(t)$ 。此外，还需满足约束条件C2, C3和C4，对此，本文引入惩罚函数

$$\zeta_1(t) = \max\left(\sum_{n=1}^N \varpi_n(t) - B, 0\right) \quad (19)$$

$$\zeta_2(t) = \max\left(\sum_{n=1}^N p_n(t) - P_{tx}, 0\right) \quad (20)$$

$$\zeta_3(t) = \sum_{n=1}^N \max\left(H(\Delta_n(t)) Q_n^{\text{Total}}(t) - h_{\text{air}} A(T_n^{\text{sur}}(t) - T_{\text{env}}), 0\right) \quad (21)$$

上述惩罚函数用于量化时隙 t 的动作决策违反约束条件C2, C3和C4的程度，如果满足约束条件C2, C3和C4，上述惩罚函数值均为零，如果违反约束，则相应的惩罚函数值为正数。基于此，本文将上述惩罚函数加入奖励函数，使得智能体可在训练过程中学习如何不违反相应的约束条件。因此，问题 \mathcal{P}_1 的即时奖励可表示为

$$r_t = \sum_{n=1}^N R_n(t) - \lambda_1 \zeta_1(t) - \lambda_2 \zeta_2(t) - \lambda_3 \zeta_3(t) \quad (22)$$

其中， λ_1, λ_2 和 λ_3 是即时奖励中的惩罚权重系数。另外，深度Q网络(Deep Q Network, DQN)算法是通过深度强化学习求解马尔可夫决策过程的经典算法^[17]，然而，DQN不适合具有连续动作空间的问题 \mathcal{P}_1 。为克服该缺点，本文采用归一化优势函数(Normalized Advantage Function, NAF)算法^[18]求解上述问题。为更好地解释NAF，首先给出两个函数

$$V(\mathbf{s}_t) = \mathbb{E}\left\{\sum_{\tau=0}^{\infty} \gamma^\tau r_{t+\tau} | \mathbf{s}_t\right\} \quad (23)$$

$$A(\mathbf{s}_t, \mathbf{a}_t) = Q(\mathbf{s}_t, \mathbf{a}_t) - V(\mathbf{s}_t) \quad (24)$$

其中， $V(\mathbf{s}_t)$ 是状态值函数， $A(\mathbf{s}_t, \mathbf{a}_t)$ 是优势函数。

对于基于值函数的深度强化学习，需计算状态动作值函数(即Q函数)并选择最大Q值对应的动作向量，然而，当动作空间连续时，无法计算所有的Q值。对此，NAF通过其巧妙的设计使得神经网络可直接获得当前状态对应的最大Q值并输出相应的动作向量。具体来说，NAF使用 $V(\mathbf{s}_t)$ 和 $A(\mathbf{s}_t, \mathbf{a}_t)$ 计算Q值，并将优势函数 $A(\mathbf{s}_t, \mathbf{a}_t)$ 构造为二次函数

$$Q(\mathbf{s}, \mathbf{a}|v) = A(\mathbf{s}, \mathbf{a}|v) + V(\mathbf{s}|v) \quad (25)$$

$$A(\mathbf{s}, \mathbf{a}|v) = -\frac{1}{2}(\mathbf{a} - \alpha'(\mathbf{s}|v))^T P(\mathbf{s}|v) (\mathbf{a} - \alpha'(\mathbf{s}|v)) \quad (26)$$

其中，参数集 v 是神经网络中各神经元权重和偏差的集合， $P(\mathbf{s}|v)$ 为正定方阵。NAF利用神经网络的部分输出项构建一个下三角矩阵 $\mathcal{L}(\mathbf{s}|v)$ ，并通过计算 $P(\mathbf{s}|v) = \mathcal{L}(\mathbf{s}|v) \mathcal{L}^T(\mathbf{s}|v)$ 得到式(26)中的正定方阵。神经网络的输出项由以下3部分组成：(1)状态值函数 $V(\mathbf{s}|v)$ ；(2)与动作相关的向量 $\alpha'(\mathbf{s}|v)$ ；(3)用于构建下三角矩阵 $\mathcal{L}(\mathbf{s}|v)$ 的向量 $\mathcal{L}_0(\mathbf{s}|v)$ 。由式(25)和式(26)可知，Q函数是关于动作向量 \mathbf{a} 的二次函数，因此，当状态为 \mathbf{s} 时，可选择 $\alpha'(\mathbf{s}|v)$ 作为动作向量，其对应的Q值为当前状态 \mathbf{s} 的最大Q值。基于此，为实现上述计算，NAF下的神经网络结构可由图1表示。

算法1描述了求解问题 \mathcal{P}_1 的NAF算法的具体步骤，其中， T_{max} 和 \tilde{T} 分别表示最大的训练回合(Episode)数以及单位训练回合包含的时隙数。为鼓励智能体充分探索动作空间，在神经网络的输出 $\alpha'(\mathbf{s}|v)$ 中加入噪声。

4 仿真结果

仿真设定单位时隙长度为10 s，单用户的最大带宽为 $B_s = 100$ MHz。所有用户的物理位置都是在1个500 m×500 m的区域内随机生成的，无人机的初始位置为(250, 250, H_{fl})。神经网络为具有3个隐

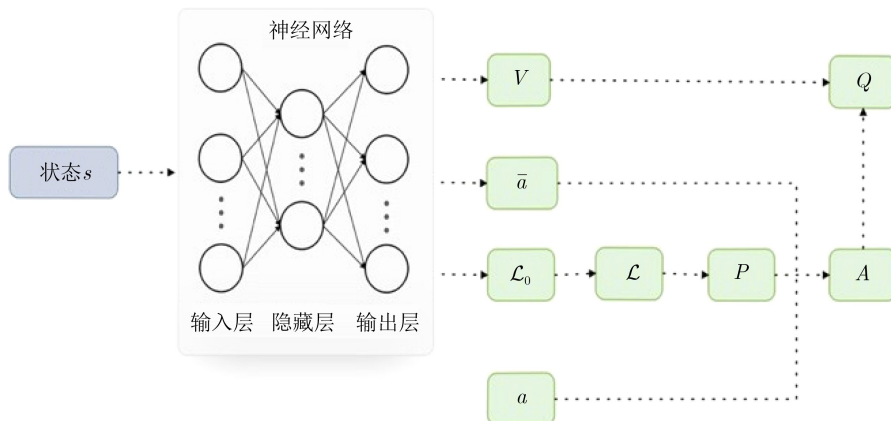


图1 NAF下的神经网络结构

算法1 求解问题 \mathcal{P}_1 的NAF算法

输入: 主网络 Q 的参数集 \mathbf{v} , 目标网络 \widehat{Q} 的参数集 $\mathbf{v}^- = \mathbf{v}$, 经验池 $X = \emptyset$, 计数器 $t = 0$, $T = 0$

输出: 动作向量 $\mathbf{a}(t)$

Repeat:

 获得当前时隙状态 \mathbf{s}_t , $t_{\text{start}} = t$

 Repeat:

 产生随机噪声向量 \mathcal{N}_t

 选择当前时隙动作 $\mathbf{a}_t = \alpha'(\mathbf{s}_t|\mathbf{v}_t) + \mathcal{N}_t$

 执行动作 $\mathbf{a}(t)$ 并获得即时奖励 r_t 和下一时隙状态 \mathbf{s}_{t+1}

 将经验 $(\mathbf{s}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1})$ 存储到经验池 X

 从经验池 X 中随机抽样包含 M 条经验的Mini-batch

 对于经验 m 计算: $y_m = r_m + \gamma \widehat{V}(\mathbf{s}_{m+1}|\mathbf{v}_t^-)$

 计算损失函数:

$$L(\mathbf{v}_t) = \frac{1}{M} \sum_{m=1}^M (y_m - Q(\mathbf{s}_m, \mathbf{a}_m|\mathbf{v}_t))^2$$

 使用梯度下降法对主网络进行更新:

$$\mathbf{v}_t := \mathbf{v}_t - \alpha \nabla L(\mathbf{v}_t)$$

$t := t + 1$, 每隔 Y 个时隙更新目标网络 $\mathbf{v}_t^- = \mathbf{v}_t$

 Until $t == t_{\text{start}} + \widetilde{T}$

$T := T + 1$

Until $T > T_{\text{max}}$

藏层且每个隐藏层神经元数量为128的全连接神经网络, 其余仿真参数则在表1给出。

仿真主要将本文提出的算法方案与其他两种方案进行比较。方案1表示不考虑能耗中断影响的传统资源管理。方案2则是本文提出的方案, 其中, $T_{\text{safe}}=318$ K (45 °C)。方案3类似文献[2]分析单用户场景下能耗中断所建议的处理方式, 一旦某个用户过热, 手机端将停止对无线接收数据的处理, 直到手机背板温度降至预定阈值 $T_{\text{th}}=310$ K (37 °C)。

图2给出了上述3种方案下系统吞吐量随用户数量变化的结果。如图所示, 考虑能耗中断时, 本文所提方案2的性能优于方案3。其原因在于能耗中断发生时, 方案3仅在手机端停止对无线接收数据的处理, 从而造成实际接收速率与发送速率不匹配, 导致无人机分配给能耗中断用户的资源被浪费, 而本文所提的方案2在无人机侧主动降低分配给能耗中断用户的资源, 从而实现对手机背板温度变化的自适应调整, 因此能实现更高效的资源利用并获得更高的系统吞吐量。另一方面, 由于传统资源管理(方案1)忽略了可导致过热问题的能耗中断影响, 因此, 当手机背板温度达到或超过安全温度 T_{safe} 时, 手机芯片并不会降低工作频率以降低对接收数据的处理能力, 这将导致手机背板温度持续维持在较高的温度, 而手机端实际接收速率将与发送速率相同, 即手机背板温度并不会限制无线传输速率。因此, 方案1相比受制于手机背板温度变化的方案2能获得更高的系统吞吐量。然而, 方案1对能耗中断的忽略却使终端用户可能面临手机过热引起的低温烫伤, 从而导致用户体验变差。并且, 这种持续维持在较高温度的状态也不利于手机及其零部件的使用寿命。此外, 单用户的带宽限制和手机背板温度会限制系统总带宽的利用率, 导致系统吞吐量相对较低。当用户数量增加时, 往往会使更多的系统带宽得到合理的使用, 从而提高系统吞吐量。而当用户数量超过10之后, 系统带宽已几乎被充分利用, 加上手机背板温度的限制, 系统吞吐量并没有显著提高。

为展示考虑手机背板温度状态变化下不同方案对无人机轨迹的影响, 图3以4个用户和6个连续时隙为例, 给出了各方案下的无人机轨迹。图中用户

表1 仿真参数

参数	值	参数	值	参数	值
\widetilde{T} (回合)	30	I_c	10^2	G (dBi)	10
M (经验)	128	η	0.59	f_{CF} (GHz)	40
Y (时隙)	300	F_{BP}	3	H_{th} (m)	70
N_{Tx} (天线)	64	F_{AP}	4	T_{env} (K)	298
N_{Rx} (天线)	4	B (GHz)	1	T_0^{sur} (K)	303
θ	$\pi/2$	m_c (g)	1	K_{BP}	8×10^7
χ_{NLoS}	2.4	P_{tx} (W)	5	K_{AP}	6×10^7
ψ_{NLoS}	5.27	d_0 (m)	5	P_{LNA} (mW)	24.3
χ_{LoS}	2	L (mm)	2	N_0 (dBm/Hz)	-174
ψ_{LoS}	5.3	D (mm)	1	k_1 (W/(m·K))	401
α_1	0.1	A (cm ²)	1	k_2 (W/(m·K))	130
α_2	0.2	D_{max} (m)	90	h_{air} (W/(m ² ·K))	26.3
λ	0.3	T_{max} (回合)	600	c_{chip} (J/(kg·K))	1030

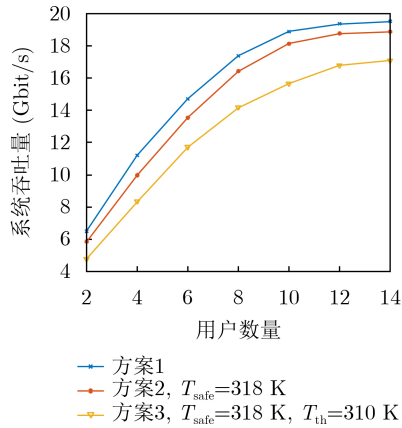


图 2 用户数量对系统吞吐量的影响

的状态变化(正常或过热)为方案2实施的结果, 方案3下的用户状态变化类似。方案2由于在无人机侧

考虑了手机背板温度变化的影响并自适应地对手机背板温度过高导致的能耗中断做出响应, 因此, 其在无人机轨迹方面能对用户状态变化做出更及时的优化调整。而方案3仅在手机端对用户状态变化做出调整, 无人机则缺少相应对手机背板温度变化的考虑, 导致其在无人机轨迹方面对温度变化的适应度较低。可以看到, 与方案3相比, 方案2使得无人机远离过热用户, 避免了资源的浪费, 从而为正常用户提供更好的服务。对于方案1, 其在无人机和手机端均未考虑手机背板温度变化的影响, 因此, 无人机轨迹与温度变化无关, 其主要受到无线信道随机衰落的影响。可以看到, 由于方案1忽略了能耗中断的影响, 相应的无人机轨迹变化相对较小, 从而为每个用户提供了可观的吞吐量。

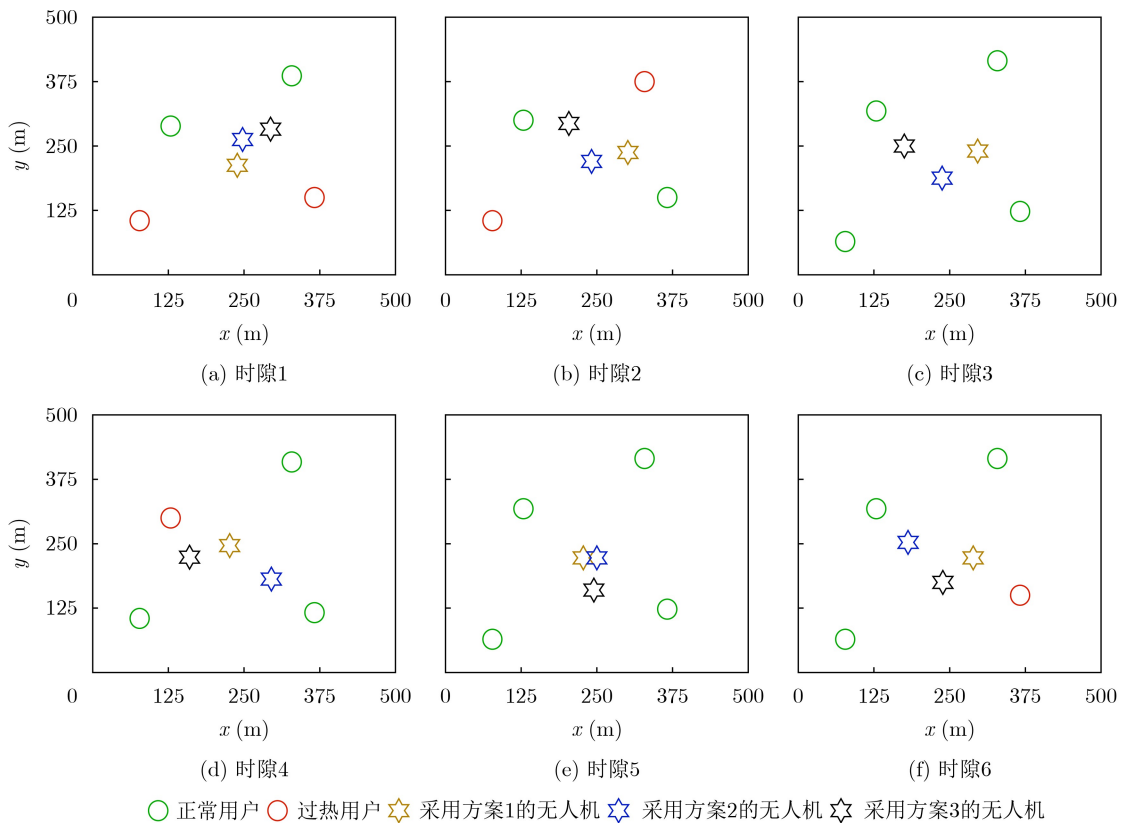


图 3 不同方案下的无人机轨迹对比

5 结束语

针对能耗中断对传统资源管理的影响, 本文在考虑能耗中断的同时, 联合优化无人机高速传输场景中的轨迹设计以及带宽和功率分配。基于仿真结果讨论说明了本文所提方案在系统吞吐量提高和轨迹设计适应性方面的有效性, 为进一步研究能耗中断在无人机通信场景中的影响提供参考。

参考文献

[1] GERACI G, GARCIA-RODRIGUE A, AZARI M M, *et al.* What will the future of UAV cellular communications be? A flight from 5G to 6G[J]. *IEEE Communications Surveys & Tutorials*, 2022, 24(3): 1304–1335. doi: 10.1109/COMST.2022.3171135.

[2] YANG Jing, GE Xiaohu, THOMPSON J, *et al.* Power-

- consumption outage in beyond fifth generation mobile communication systems[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(2): 897–910. doi: [10.1109/TWC.2020.3029051](https://doi.org/10.1109/TWC.2020.3029051).
- [3] GARIMELLA S V, PERSONS T, WEIBEL J A, *et al.* Electronics thermal management in information and communications technologies: Challenges and future directions[J]. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2017, 7(8): 1191–1205. doi: [10.1109/TCPMT.2016.2603600](https://doi.org/10.1109/TCPMT.2016.2603600).
- [4] CHIRIAC V, MOLLOY S, ANDERSON J, *et al.* A figure of merit for mobile device thermal management[C]. The 15th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), Las Vegas, USA, 2016: 1393–1397. doi: [10.1109/ITHERM.2016.7517712](https://doi.org/10.1109/ITHERM.2016.7517712).
- [5] BHAT G, GUMUSSOY S, and OGRAS U Y. Power and thermal analysis of commercial mobile platforms: Experiments and case studies[C]. 2019 Design, Automation & Test in Europe Conference & Exhibition (DATE), Florence, Italy, 2019: 144–149. doi: [10.23919/DATE.2019.8714831](https://doi.org/10.23919/DATE.2019.8714831).
- [6] ARNOMO S A, SIMANJUNTAK P, and NUR SADIKAN S F. Overheating analysis of mobile phone temperature based on multitasking process[C]. 2021 International Conference on Computer Science and Engineering (IC2SE), Padang, Indonesia, 2021: 1–6. doi: [10.1109/IC2SE52832.2021.9792125](https://doi.org/10.1109/IC2SE52832.2021.9792125).
- [7] MAMMELA A and ANTTONEN A. Why will computing power need particular attention in future wireless devices[J]. *IEEE Circuits and Systems Magazine*, 2017, 17(1): 12–26. doi: [10.1109/MCAS.2016.2642679](https://doi.org/10.1109/MCAS.2016.2642679).
- [8] YANG Jing, GE Xiaohu, and ZHONG Yi. How much of wireless rates can smartphones support in 5G networks?[J]. *IEEE Network*, 2019, 33(3): 122–129. doi: [10.1109/MNET.2018.1800025](https://doi.org/10.1109/MNET.2018.1800025).
- [9] 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. doi: [10.11999/JEIT210789](https://doi.org/10.11999/JEIT210789).
CHEN Xinying, SHENG Min, LI Bo, *et al.* Survey on unmanned aerial vehicle communications for 6G[J]. *Journal of Electronics & Information Technology*, 2022, 44(3): 781–789. doi: [10.11999/JEIT210789](https://doi.org/10.11999/JEIT210789).
- [10] ZHAN Cheng and HUANG Renjie. Energy efficient adaptive video streaming with rotary-wing UAV[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(7): 8040–8044. doi: [10.1109/TVT.2020.2993303](https://doi.org/10.1109/TVT.2020.2993303).
- [11] CHEN Yan, ZHANG Hangjing, and HU Yang. Optimal power and bandwidth allocation for multiuser video streaming in UAV relay networks[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(6): 6644–6655. doi: [10.1109/TVT.2020.2985061](https://doi.org/10.1109/TVT.2020.2985061).
- [12] FU Xiuhua, DING Tian, KADOCH M, *et al.* Uplink performance analysis of UAV cellular communications with power control[C]. 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020: 676–679. doi: [10.1109/IWCMC48107.2020.9148412](https://doi.org/10.1109/IWCMC48107.2020.9148412).
- [13] LIU Xiao, LIU Yuanwei, CHEN Yue, *et al.* Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(8): 7957–7969. doi: [10.1109/TVT.2019.2920284](https://doi.org/10.1109/TVT.2019.2920284).
- [14] CHEN Mingzhe, SAAD W, and YIN Changchuan. Echo-liquid state deep learning for 360 content transmission and caching in wireless VR networks with cellular-connected UAVs[J]. *IEEE Transactions on Communications*, 2019, 67(9): 6386–6400. doi: [10.1109/TCOMM.2019.2917440](https://doi.org/10.1109/TCOMM.2019.2917440).
- [15] GOLDSMITH A. *Wireless Communications*[M]. Cambridge, USA: Cambridge University Press, 2005: 78–79.
- [16] ZHAO Pengtao, TIAN Hui, CHEN K C, *et al.* Context-aware TDD configuration and resource allocation for mobile edge computing[J]. *IEEE Transactions on Communications*, 2020, 68(2): 1118–1131. doi: [10.1109/tcomm.2019.2952580](https://doi.org/10.1109/tcomm.2019.2952580).
- [17] MNIH V, KAVUKCUOGLU K, SILVER D, *et al.* Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533. doi: [10.1038/nature14236](https://doi.org/10.1038/nature14236).
- [18] GU Shixiang, LILLICRAP T, SUTSKEVER I, *et al.* Continuous deep Q-learning with model-based acceleration[C]. The 33rd International Conference on International Conference on Machine Learning, New York, USA, 2016: 2829–2838.
- 罗佳: 男, 讲师, 博士, 研究方向为下一代无线通信网络、人工智能、区块链等。
- 陈前斌: 男, 教授, 博士生导师, 研究方向为个人通信、多媒体信息处理与传输、异构蜂窝网络等。
- 唐伦: 男, 教授, 博士生导师, 研究方向为下一代无线通信网络、异构蜂窝网络、图像处理等。
- 张志才: 男, 讲师, 博士, 研究方向为移动边缘计算、无人机、车联网和机器学习等。

责任编辑: 马秀强