

DOI:10.13232/j.cnki.jnju.2021.05.015

信息年龄受限下最小化无人机辅助无线供能网络的能耗： 一种基于 DQN 的方法

刘玲珊¹, 熊 轲^{1*}, 张 煜², 张锐晨¹, 樊平毅^{3,4}

(1. 北京交通大学计算机与信息技术学院, 北京, 100044; 2. 国网能源研究院有限公司, 北京, 102209;
3. 清华大学电子工程系, 北京, 100084; 4. 北京信息科学与技术国家研究中心, 北京, 100084)

摘 要:随着 5G/B5G 的不断发展, 无人机在实时数据采集系统中将有广泛应用. 利用无人机先给传感器节点进行无线充电, 然后传感器节点利用收集到的能量将感知的信息上传无人机, 可有效解决户外物联网节点的供电与数据采集问题. 然而, 由于无人机本身的电量受限, 如何在保证无人机充电辅助物联网系统顺利完成新鲜数据采集任务的前提下最小化无人机的能耗至关重要. 为此, 在满足信息采集新鲜度的要求下, 通过联合优化无人机的飞行时间、加速度、转角和传感器节点信息上传和能量收集调度模式, 建立无人机能耗最小化优化问题. 由于该问题含有整数变量, 大规模情况下求解较为困难. 因此, 首先将其建模为马尔科夫决策过程, 然后提出了一种基于 DQN (Deep Q Network) 的无人机能耗优化算法框架求解, 并设计相对应的状态空间、动作空间和奖励函数. 仿真结果验证了所提 DQN 算法的收敛性, 同时表明提出的 DQN 算法比传统的贪婪算法可降低 8%~30% 的无人机能耗. 当传感器个数超过八个时, 传统的贪婪算法很难求解, 而所提 DQN 算法仍然能找到最优解. 另外, 随着 AoI (Age of Information) 限制值的缩小或传感器数量的增加, 无人机的能量消耗会不断地增加, 并且由于考虑了转角约束, 所提算法优化得到的无人机飞行轨迹会更平滑.

关键词:无人机辅助无线网络, 信息年龄, 能量收集, 深度强化学习

中图分类号: TP391

文献标志码: A

Energy minimization in UAV-assisted wireless powered sensor networks with AoI constraints: A DQN-based approach

Liu Lingshan¹, Xiong Ke^{1*}, Zhang Yu², Zhang Ruichen¹, Fan Pingyi^{3,4}

(1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, China; 2. State Grid Energy Research Institute Co., Ltd., Beijing, 102209, China; 3. Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China; 4. Beijing National Research Center for Information Science and Technology, Beijing, 100084, China)

Abstract: With the development of 5G/B5G, UAV (Unmanned Aerial Vehicle) will be widely employed in real-time data collection system. UAV can wirelessly charge ground sensors, and then sensors use the collected energy to upload the perceived information to UAV, which can effectively solve the problem of power supply and data collection in outdoor Internet of Things (IoT) systems. However, due to the limited power of UAV, how to minimize the energy consumption of UAV is very important on the premise of ensuring the freshness of collected data in UAV-assisted wireless powered sensor network. Therefore, this paper investigates the optimization problem of UAV's energy consumption minimization under the Age of Information (AoI) constraint by jointly optimizing UAV's flight time, acceleration, rotation angle and scheduling of information collection and energy harvesting. As the problem is a combinational optimization problem with a set of binary

基金项目: 国家自然科学基金(62071033), 国家重点研发计划(2020YFB1806903), 国网能源研究院有限公司研究项目(526700190002)

收稿日期: 2021-06-16

* 通讯联系人, E-mail: kxiong@bjtu.edu.cn

variables, it is difficult to be solved in large-scale network. Thus, it's first modeled as a Markov decision process, and a Deep Q Network (DQN)-based UAV's energy-minimal algorithm framework is proposed to solve it, and the corresponding state spaces, action spaces and reward function are designed. Simulation results demonstrate the convergence of the proposed DQN scheme, and also show that the proposed DQN scheme can reduce the UAV's energy consumption by about 8%~30% compared with the traditional greedy scheme. When the sensors' amount is more than eight, the traditional greedy scheme becomes very difficult to solve the problem, while our presented DQN method can still find an optimal solution. Moreover, with the decrease of AoI or the increment of the number of sensors, the energy consumption of UAV increases and the trajectory of UAV becomes smoother with the rotation angle constraint.

Key words: UAV-assisted wireless network, Age of Information (AoI), energy harvesting, Deep Q Network (DQN)

随着 5G 的快速部署, 虚拟现实(Virtual Reality, VR)、增强现实(Augmented Reality, AR)、无人驾驶、智慧医疗等新型应用^[1-3]蓬勃发展. 这些应用除了要求超高可靠低时延通信, 对信息新鲜度(即信息时效性)也十分敏感, 而传统的网络指标如吞吐量、延迟等无法准确地刻画网络中的信息新鲜度. 为了准确地刻画信息的新鲜度, 学术界提出了信息年龄(Age of Information, AoI)^[4-5]的概念, 其定义为信息在被观察时刻(如在网络目的端接收时刻或被用于计算的时刻)距其从在源节点生成以来经过的时间.

AoI 已成为信息更新系统^[6-8]传感器网络中的关键性能指标之一. 为了保证网络中信息的新鲜度, 传感器节点需要频繁地采集并传输最新感知的信息, 这会消耗电池中的大量电能. 为了延长电池的续航能力, 需要频繁更换电池或充电. 在大规模应用场景中, 人工电池管理不方便且费用昂贵, 在恶劣环境下更是难以实现.

1 相关工作

基于射频(Radio Frequency, RF)信号的无线能量传输(Wireless Power Transfer, WPT)技术^[9-11]能为低功率传感器节点提供稳定的电源, 延长系统运行时间, 因此基于 RF 能量收集的 AoI 网络近期逐渐得到学术界的广泛关注^[12-14]. Dabiri and Emadi^[12]分析了当无线传感器节点有容量无限的能量缓冲器和数据缓冲器时, 能量收集和到达率对系统信息平均 AoI 的影响. Wu et al^[13]在传感器电池容量受限的情况下, 通过优化在线状态更新策略降低系统长期平均 AoI. Arafa et al^[14]针对不同的能量到达模型, 分析

了不同信息更新策略的长期平均 AoI. 但这些工作均采用理想线性能量收集模型, 即认为收集的能量随接收射频信号输入功率的增加而线性增加, 却忽略了二极管等元件引起的电路非线性特性, 导致能量收集电路的输入输出显示高度的非线性特性^[12]. 因此, 采用传统理想线性能量收集模型所得的结论与实际网络偏差较大^[15-17].

近来, 为了延长户外无线传感器网络的使用寿命, 一些研究工作引入无人机(Unmanned Aerial Vehicle, UAV)技术, 即为无人机安装 RF 能量发射器(类似空中 WiFi 接入点或小基站), 通过控制无人机, 使其灵活移动到传感器节点附近与地面节点建立视距(Line-of-Sight, LoS)链路, 进而提供稳定、高质量的能量和信息传输服务. 然而, 由于大部分无人机采用电池供电, 续航时间有限, 因此节能必然是 UAV 无线网络设计的重要指标之一^[18], 所以很多工作都探讨了无人机能耗的最小化问题^[19-21]. Yang et al^[19]通过寻找最佳的传感器发射功率和无人机飞行轨迹来实现无人机推进能耗和传感器传输能量的平衡. Yang et al^[20]在满足数据传输大小的前提下, 通过联合优化无人机飞行轨迹、数据传输时间和节点传输功率来最小化无人机能耗. Zeng and Zhang^[21]研究无人机辅助网络中固定翼无人机的能耗与加速度和转角的关系, 并联合优化无人机的飞行半径和速度, 最大化系统的能量效率. 可以看出, 无人机飞行轨迹优化对降低无人机飞行能耗具有重要作用. 另外, Zeng and Zhang^[21]的工作表明, 通过优化无人机的加速度可以降低无人机的能耗.

需要说明的是, 飞行过程中无人机与传感器节点的信道时刻在变化, 能量和信息的传输也会

引起传感器节点存储能量和信息 AoI 的变化,因此,随着传感器节点个数的增加,相关优化问题的规模和维度很容易剧增,难以用动态规划等传统算法对问题进行有效求解。

近年来,深度强化学习(Deep Reinforcement Learning, DRL)算法引起了工业界和学术界的广泛关注^[22-24]。与马尔科夫决策过程(Markov Decision Process, MDP)和强化学习(Reinforcement Learning, RL)算法相比, DRL 能克服巨大的状态和动作空间,求解更复杂的优化问题。例如, Mnih et al^[24]使用神经网络作为 DQN(Deep Q Network)的函数逼近器处理巨大的状态空间,并使用经验重放和目标网络等来提高学习稳定性。目前, DRL 开始被应用于解决无人机辅助网络的优化问题^[25-27]。Yi et al^[25]基于 DQN 算法优化无人机飞行轨迹和信息传输调度策略,最小化加权 AoI,用能量受限的无人机收集传感器节点观察到的信息。Abd-ERlmagid et al^[26]基于 DQN 算法联合优化无人机的飞行轨迹和信息更新策略,最小化系统加权 AoI,无人机收集能量受限的传感器节点观察到的信息。Tong et al^[27]利用 DQN 优化无人机的飞行轨迹,最小化传感器节点信息的平均信息年龄和丢包率,能量受限的无人机收集来自传感器节点以固定或随机的速率采样的信息。需要指出,上述工作将无人机的飞行区域进行简单的离散化,不能避免无人机在飞行过程中出现大角度拐弯的情况,因此优化结果与无人机实际飞行状态不相符。另外,这些工作也只关注如何为无人机和传感器节点节能,没有引进无线射频能量收集技术为低功耗传感器节点供电。

本文以无人机能耗为指标,研究信息年龄受限的无人机辅助无线供能网络设计。另外,为了更加贴切实际情况,考虑无人机飞行转角、加速度等约束,并采用 DRL 算法对复杂优化问题进行求解。本文的主要贡献:

(1)分析无人机的能耗模型,并对系统采集信息的 AoI 进行建模。在此基础上,在信息年龄的约束下通过联合优化无人机的加速度、转角、飞行时间以及传感器节点的能量收集和信息采集策略,建立含有整数变量的组合优化问题。

(2)由于所建立优化问题非凸,而问题中的

信息年龄 AoI、能量、加速度以及转角之间存在独立性,因此将其转化为马尔科夫决策问题。由于目标问题状态空间巨大,无法使用传统的 MDP 方法求解,进一步提出基于 DQN 的无人机能耗优化算法框架并通过设计相对应的状态空间、动作空间和奖励函数求解。

(3)仿真结果证明提出的基于 DQN 的无人机控制算法的收敛性,并验证其性能优于传统的贪心策略。数值结果表明, DQN 的无人机控制框架下通过对飞行转角的优化,无人机的飞行轨迹更贴近实际,飞行过程中无人机会尽可能飞至传感器节点附近进行信息收集和能量传输,且飞行速度先增加后下降。另外,无人机的飞行能耗随着信息年龄 AoI 的减小和传感器节点个数的增加而增加。

2 系统模型

2.1 网络模型 无人机辅助无线网络场景如图 1 所示:无人机从出发点起飞,在任务时间 T 内,利用 WPT 技术为地面随机分布的 K 个传感器节点传输能量;之后,传感器节点利用收集的能量采集信息,并将采集的信息上传无人机;采集所有节点的信息后无人机飞回终点。

记传感器节点集合为 $V = \{v_1, \dots, v_K\}$ 。为方便将时间 T 分成 N 个时间片,即 $T = N \cdot \delta$, δ 代表一个足够小的时间间隙。假设无人机飞行的固定高度为 H ,无人机的出发点的水平坐标为 $q_0 = [x_0, y_0]$,终点的水平坐标为 $q_F = [x_F, y_F]$,在第 n 个时刻,无人机的飞行水平坐标为 $q(n) = [x(n), y(n)]$,其中, $n \in [1, 2, \dots, N]$ 。

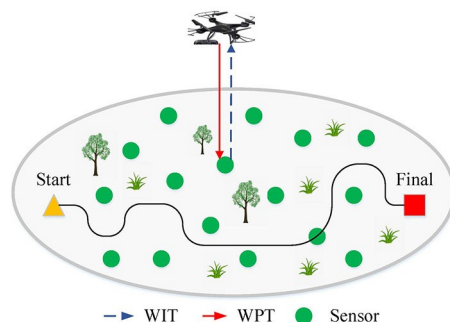


图1 系统模型

Fig. 1 The system model

2.2 信道模型 无人机飞行时从空中与地面传感器节点建立 LoS 链路,同时考虑树木、小草、建筑物等影响,小规模尺度衰落也不容忽视,因此,基于莱斯衰落信道模型^[28],在第 n 个时刻无人机与传感器节点 v_k 的信道增益可表示为:

$$g_{u,k}(n) = \sqrt{\frac{\beta_0}{d_{u,k}^2}} g_k = \sqrt{\frac{\beta_0}{H^2 + \|q_u(n) - q_k\|^2}} g_k \quad (1)$$

其中, $d_{u,k}$ 是无人机到传感器节点 v_k 的距离, β_0 表示在参考距离 $d_0 = 1$ 米的信道增益, $\|\cdot\|$ 为欧氏距离, g_k 为小尺度衰落系数,即:

$$g_k = \sqrt{\frac{K_R}{K_R + 1}} \bar{g} + \sqrt{\frac{1}{K_R + 1}} \bar{g} \quad (2)$$

其中, K_R 是无人机与传感器节点信道的莱斯因子, \bar{g} 是视距分量, \bar{g} 是散射分量,表示零均值单位方差圆对称复高斯(Circularly Symmetric Complex Gaussian, CSCG)随机变量.

2.3 信息和能量采集模型 为了避免干扰,无人机通过时分多址(Time Division Multiple Access, TDMA)的方式与传感器节点进行能量传播或信息传输. 因此,在每一时刻,传感器节点只能进行能量收集或信息传输,且在信息传输时最多只能有一个传感器节点传输更新信息. 若第 n 时刻,第 k 个传感器被调度进行信息传输,根据香农公式,传感器节点 v_k 传输信息所需的能量可表示为:

$$E_k^T(n) = \frac{\sigma_0^2}{g_{u,k}(n)} \left(2^{\frac{S}{\delta B}} - 1 \right) \delta \quad (3)$$

其中, S 和 B 分别为数据包大小和信道带宽.

无人机从 q_0 出发,在飞行过程中通过广播方式给每个传感器节点充电. 传感器节点 v_k 装配了能量收集器,其射频电路接收的输入功率可表示为:

$$e_k(n) = p_u g_{u,k}(n) \quad (4)$$

其中, p_u 表示无人机的发射功率.

本文采用非线性能量收集模型^[15]来刻画射频到直流的转换. 因此,在第 n 时刻 v_k 收集的能量可建模为:

$$E_k^H(n) = M \frac{1 - ae_k(n)}{1 + e^{-a(e_k(n) - b)}} \delta \quad (5)$$

其中, M 表示能量收集电路饱和时在能量收集接收器处的最大收集功率, a 和 b 分别表示与实际电

路灵敏度,电阻等有关的常数. 那么, v_k 的电池容量变化过程可表示为:

$$B_k(n) = \begin{cases} \min(B_k(n-1) + E_k^H(n), B_{\max}), & c(n) = 0 \\ B_k(n-1) - E_k^T(n), & c(n) = k \text{ and } B_k(n-1) \geq E_k^T(n) \\ B_k(n-1), & \text{otherwise} \end{cases} \quad (6)$$

其中, B_{\max} 表示节点电池最大容量, $c(n)$ 表示信息传输和能量采集的策略,即 $c(n) = \{0, 1, 2, \dots, K\}$, $c(n) = 0$ 时表示所有节点进行能量采集, $c(n) = k$ ($k \in \{0, 1, 2, \dots, K\}$) 时表示 v_k 被调度上传信息,需要满足其电池存储的能量大于传输信息消耗的能量.

2.4 无人机能耗模型 无人机系统的能耗主要由无人机的推进能耗和无人机通信相关能耗(如信号处理和天线发射等所需能量)两部分组成,其中通信相关能耗远小于推进能耗,因此,与 Yang et al^[20]相似,本文忽略通信相关能耗. 依据 Zeng and Zhang^[21],固定翼无人机的飞行能耗可表示为:

$$E_u^{\text{fly}} = \sum_{n=1}^N \left(c_1 \|v(n)\|^3 + \frac{c_2}{\|v(n)\|} \left(1 + \frac{\|a(n)\|^2}{g^2} \right) \right) \delta + \frac{1}{2} m (\|v(T)\|^2 - \|v(0)\|^2) \quad (7)$$

其中, c_1 和 c_2 是两个常数,与飞机重量、机翼面积、空气密度等有关, g 表示重力加速度, m 是无人机的质量, $v(n)$ 和 $a(n)$ 分别表示无人机的速度和加速度. 从式(7)可看出,无人机的推进能耗主要与其速度和加速度有关. 另外,无人机的飞行轨迹可表示为:

$$v(n) = v(n-1) + \delta a(n) \quad (8)$$

$$p(n) = p(n-1) + o(n) \quad (9)$$

$$x(n) = x(n-1) + \delta v(n) \cos(p(n)) \quad (10)$$

$$y(n) = y(n-1) + \delta v(n) \sin(p(n)) \quad (11)$$

其中, $p(n)$ 和 $o(n)$ 分别表示第 n 时刻无人机的飞行角度及变化角度.

2.5 AoI 模型 信息年龄 AoI 是刻画采集信息新鲜度的一个重要指标,这里定义为无人机最近一次采集的信息产生经过的时间. 假设 $U_k(n)$ 表示在第 n 时刻无人机最近一次从 v_k 采集信息时的

时间,该信息的信息年龄 AoI 可表示为:

$$A_k(n) = (n - U_k(n))\delta \quad (12)$$

不失一般性,假设信息产生时刻的信息年龄 AoI 为 1 个归一化单位的时间,每经过一个时刻,其 AoI 增加 1. 一旦产生新的信息,将覆盖原有信息,信息年龄也将减少为 1. 图 2 展示无人机采集 v_k 信息的信息年龄 AoI 的变化过程,在 t_1 和 t_2 时刻 v_k 上传更新信息,因此其信息 AoI 减少为 1,其他时刻的信息 AoI 增加 1,故 v_k 信息的 AoI 在 T 的变化过程可表示为:

$$A_k(n) = \begin{cases} \delta, & c(n) = k \text{ and } B_k(n) \geq E_k^T(n) \\ A_k(n-1) + \delta, & \text{otherwise} \end{cases} \quad (13)$$

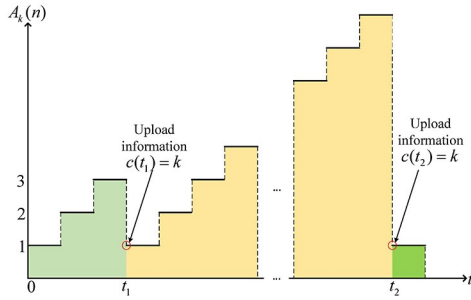


图 2 AoI 模型

Fig. 2 The AoI model

3 系统模型

系统的设计目标是通过联合优化无人机的飞行轨迹、飞行时间以及信息采集和能量收集的策略来最小化无人机能耗. 因此,问题的数学描述如下:

$$(P_0) \min_{\{a, o, c, T\}} E_u^{\text{fly}} \quad (14)$$

$$\text{s.t. } A_k(N) < (N+1)\delta, k \in \{1, 2, \dots, K\} \quad (15)$$

$$\frac{1}{N} \frac{1}{K} \sum_{n=1}^N \sum_{k=1}^K A_k(n) \leq A_r \quad (16)$$

$$q(0) = q_0, q(N) = q_F \quad (17)$$

$$a(n) \in [a_{\min}, a_{\max}] \quad (18)$$

$$o(n) \in [o_{\min}, o_{\max}] \quad (19)$$

$$v(n) \in [v_{\min}, v_{\max}] \quad (20)$$

其中, A_r 表示系统接受的有效信息的最大信息年龄 AoI. 式(15)表示无人机收集的信息至少更新一次. 式(16)保证无人机收集的信息 AoI 满足信息新鲜度的限制. 式(17)表示无人机从初始地点

出发,经过 T 完成任务并到达目的地. 式(18)至式(20)是对无人机的加速度、速度和转角的约束限制,保证正常飞行. 由于 c 是离散组合变量,并与变量 a, o 和 T 相耦合,可看出问题 (P_0) 属非线性整数的非凸问题. 由于控制无人机飞行的加速度、转角、传感器节点储存的电量和收集信息的 AoI 之间的独立性,因此可以将问题建模为具有有限状态和动作空间的马尔科夫决策问题. 然而,状态空间巨大使该问题难以使用传统标准的马尔科夫算法解决,而深度强化学习中的神经网络擅长提取高维度数据特征,因此下节将基于深度强化学习方法来解决该问题.

4 基于 DQN 的求解算法

如图 3 所示,基于 DQN 的无人机控制框架包含环境和智能体. 本文所提 DQN 算法将无人机视为一个智能体,将在空中的飞行运动、与传感器节点的交互如信息能量传输等视为环境. 在每个训练周期中,智能体需要感知周边环境状态 s_n , 包括传感器节点的电量、上传信息的 AoI 等,根据当前环境决定下一时刻的动作 a_n ; 执行动作后,智能体获得环境相应的反馈奖励 r_n , 并继续观察下一时刻状态 s_{n+1} . 智能体在训练过程中,通过不断地与环境进行交互,观察执行动作后状态的变化和环境反馈的奖励,调整动作策略,通过反复迭代学习最大化累积回报可以获得更好的动作策略.

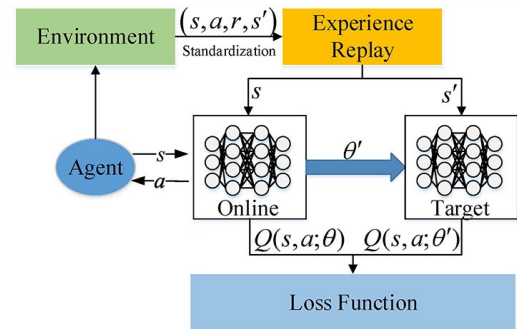


图 3 无人机能耗优化算法框架

Fig. 3 The framework of the UAV's energy-minimal algorithm

智能体在每次迭代过程中需要对给定的策略计算其值函数并根据值函数给出策略,即通过神经网络非线性逼近值函数 $Q(s_n, a_n | \theta_n)$ (θ 是 ANN

的权重)评估策略 π 下状态 s_n 采取动作 a_n 的代价, 值函数 $Q(s_n, a_n | \theta_n)$ 的更新规则如下:

$$Q_{n+1}(s_n, a_n) = (1 - \beta)Q(s_n, a_n | \theta_n) + \beta(r_n + \gamma \max_a Q(s_{n+1}, a)) \quad (21)$$

其中, β 为学习率, γ 是折扣因子.

DQN算法包含两个神经网络, 即在线网络和目标网络. 这两个网络初始权重相同, 在线网络权重每次迭代都会更新, 而目标网络权重是每隔一段时间才会更新. 为了打破数据间的关联性, DQN算法使用经验池来存放历史数据, 每次训练时从中随机抽取数据. 考虑到状态空间中能量、AoI和位置量纲的不同, 在存放数据前需对历史数据进行标准化, 即 $s = (\bar{s} - \mu) / \sigma$, 其中, μ 为均值, σ 为方差. 对于任意一组 (s_n, a_n, r_n, s_{n+1}) 历史数据, 在线网络基于最小化损失函数准则进行训练, 损失函数定义如下:

$$L(\theta_{j+1}) = [r_n + \gamma \max_a Q(s_{n+1}, a | \theta_k) - Q(s_n, a_n | \theta_j)]^2 \quad (22)$$

其中, j 为更新权重的轮数. 使用梯度下降法更新在线网络, 其梯度为:

$$\nabla_{\theta_{j+1}} L(\theta_{j+1}) = [r_n + \gamma \min_a Q(s_{n+1}, a_n | \theta_j) - Q(s_n, a_n | \theta_{j+1})] \times \nabla_{\theta_{j+1}} Q(s_n, a_n | \theta_{j+1}) \quad (23)$$

基于DQN的无人机能耗优化算法是将无人机能耗最小化优化问题转化为马尔科夫决策过程, 由三元组 $\{s, a, r\}$ 来描述, 其中, s 表示智能体的状态, a 表示智能体执行的动作, r 表示智能体执行动作后环境反馈的奖励. 具体设计如下:

(1) 状态: 在第 n 个时刻, 智能体状态 s_n 包含两部分: 传感器节点和无人机在第 n 时刻的状态信息, 包括传感器节点的当前储存能量 B 、无人机收集信息的新鲜度 A 、无人机的地理位置、飞行速度、飞行角度、能耗、飞行时间以及距离终点的距离. 因此, 状态 s_n 表示为:

$$s_n = \{B, A, q(n), v(n), p(n), E(n), n, d(n)\} \quad (24)$$

其中 $B = [B_1, B_2, \dots, B_K]$, $A = [A_1, A_2, \dots, A_K]$, B_k, A_k 分别表示为第 k 个节点存储的电量和采集信息的AoI; $d(n) = \|q(n) - q_F\|$ 为无人机在第 n

时刻与终点的位距离. 因此, 系统的状态空间为 $S \subseteq \mathbb{R}^K \times \mathbb{R}^K \times \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{N} \times \mathbb{R}$, 其中 $s_n \in S$.

(2) 动作: 在第 n 个时刻, 动作 a_n 包含三部分: 无人机的加速度、转角以及节点信息传输和能量收集策略, 可以表示为:

$$a_n = \{a(n), o(n), c(n)\} \quad (25)$$

将无人机的转角和加速度进行离散化分为 k_o 和 k_a 个等级, 即:

$$o(n) \subseteq \left[o_{\min}, o_{\min} + \frac{o_{\max} - o_{\min}}{k_o}, \dots, o_{\max} \right] \quad (26)$$

$$a(n) \subseteq \left[a_{\min}, a_{\min} + \frac{a_{\max} - a_{\min}}{k_a}, \dots, a_{\max} \right] \quad (27)$$

(3) 奖励函数: 在第 n 个时刻的奖励函数 r_n 的设计主要包含两部分, 即无人机与终点的距离和无人机消耗的能量. 若当前状态满足问题 (P_0) 中的限制条件(式(15)至式(20)), 即无人机到达终点且收集信息AoI满足新鲜度, 则给予一定的奖励; 若违反约束条件, 则应受到惩罚. 因此, 奖励函数可设计为:

$$r_n = \begin{cases} k_1 - \frac{E_u^{\text{fly}}}{k_2}, & d(n) = 0 \text{ and } \frac{1}{N} \frac{1}{K} \sum_{k=1}^K A_k(N) \leq A_r \\ -\frac{d(n)}{k_3} - \frac{E(n)}{k_2}, & \text{otherwise} \end{cases} \quad (28)$$

其中:

$$E(n) = \sum_{k=1}^n \left(c_1 \|v(n)\|^3 + \frac{c_2}{\|v(n)\|} \left(1 + \frac{\|a(n)\|^2}{g^2} \right) \right) \delta + \frac{1}{2} m (\|v(n)\|^2 - \|v(0)\|^2)$$

表示无人机从出发到第 n 个时刻以来的能耗, k_1, k_2, k_3 为正常数.

为清晰起见, 将基于DQN的无人机能耗优化算法的流程归纳为算法1.

算法1 基于DQN的无人机能耗优化算法

输入: 经验池 D , 在线网络 Q 的随机权重 θ , 目标网络 Q' 的随机权重 $\theta' = \theta$;

输出: θ' 和 θ ;

1. For each episode Do
2. 初始化状态 $s(0)$;
3. For each time slot Do
4. 将状态 s_n 进行标准化;
5. 使用 ϵ -greedy策略选择动作, 即以 ϵ 概率选择

随机动作作为 a_n , $1 - \epsilon$ 概率选择动作 $\arg\max_a Q(s_n, a | \theta_k)$

作为 a_n ;

6. 在当前状态 s_n 下, 执行动作 a_n 并获得对应的奖励 r_n 和下一状态 s_{n+1} ;

7. 将 $\{s_n, a_n, r_n, s_{n+1}\}$ 标准化后作为训练数据存储在经验池 D 中;

8. $s_n = s_{n+1}$;

9. 如果任务完成, 则退出当前循环;

10. 随机从经验池中选取 b 组训练数据 $\{s_n, a_n, r_n, s_{n+1}\}$;

11. 根据式(23), 更新梯度;

12. 每隔 J 步, 更新目标网络的权重 $\theta' = \theta$.

13. End For

14. End For

首先, 初始化经验池来存储训练数据并随机初始化在线网络参数, 以与在线网络相同的网络结构引入目标网络, 并赋值与在线网络相同的神经网络参数; 然后, 在 E 训练周期内, 初始化无人机的初始状态, 并标准化处理状态, 在无人机最大允许任务时间 T 内使用 ϵ -greedy 贪心策略探索环境即在选择动作 a_n 时有一定的探索概率 ϵ 随机选择动作; 接着, 在执行动作后, 智能体根据计算得到环境反馈的奖励以及下一时刻状态 s_{n+1} , 同时将 $\{s_n, a_n, r_n, s_{n+1}\}$ 训练数据标准化后存储于经验池 D 中; 若无人机任务完成, 则退出当前飞行任务, 进入下一轮训练; 每次飞行任务结束, 采用小批量方法从经验池中随机选取 b 组训练数据, 以打破数据之间的相关性, 并根据式(23), 更新梯度; 最后, 每隔 J 步, 将在线网络的权重同步更新目标网络。

5 仿真结果与分析

为验证提出的算法的有效性, 基于 Pytorch 和 Python3.5 搭建了一个四层完全连接的神经网络, 其中包括两个完全连接的隐藏层, 每层分别有 128 和 64 个神经元. 输入层大小与状态空间大小相同即 $2K + 7$, 输出层大小等于动作总数即 $k_a k_o (K + 1)$. 详细的仿真数值如表 1 所示.

5.1 算法收敛性 为验证无人机能耗优化算法的收敛性, 图 4 显示了传感器节点个数分别为 10,

表 1 仿真参数

Table 1 The simulation parameters

变量	取值	变量	取值
β_0 (dB)	-50	M (mJ)	20
S (bits)	800	a	6400
B (MHz)	1	b	0.003
σ_0^2 (dBm)	-100	v_{\min} (m·s ⁻¹)	5
δ (s)	0.1	v_{\max} (m·s ⁻¹)	20
P_u (w)	0.1	α_{\min}	-60°
c_1	0.002	α_{\max}	60°
c_2	70.698	a_{\min} (m·s ⁻²)	-5
g (m·s ⁻²)	9.8	a_{\max} (m·s ⁻²)	5
m (kg)	9.65	β	0.001
ϵ	0.1	γ	0.8

12, 14 的累计回报变化情况, 其中横坐标为训练迭代次数, 纵坐标为累计回报. 图 4a 至图 4c 中 $A_r = 5$, 图 4d 至图 4f 中 $A_r = 6$. 可以看出, 虽然奖励值随着训练周期不断增大, 但由于采用了贪心策略, 累计回报在训练周期发生了抖动. 由图 4a 至图 4c 可知, 当节点个数增加时, 智能体的动作和状态空间的增加会使训练至收敛状态的次数增加. 另外, 随着信息年龄 AoI 的减小, 算法收敛的训练次数会增加.

5.2 无人机飞行轨迹与速度 图 5 显示了 $A_r = 6$ 时传感器节点不同分布下对应的无人机的飞行轨迹和速度变化情况. 图 5a 至图 5c 为无人机的飞行轨迹情况, 可以看出, 无人机会尽可能飞到传感器节点附近传输信息和能量. 与图 5c 无转角约束相比, 有转角约束的图 5a 和图 5b 中的无人机飞行轨迹更平滑, 更贴近实际. 图 5d 至图 5f 显示无人机的飞行速度变化情况, 可以观察到无人机在出发后速度先上升, 因为速度快能减少任务执行时间, 减少无人机能耗, 在飞行后期的飞行速度会下降, 是因为无人机的能耗与其到达终点的速度有关, 降低速度能减少无人机能耗.

5.3 信息新鲜度对无人机能耗的影响 图 6 展示了 AoI 对无人机能耗的影响情况, 其中横坐标为传感器节点个数, 纵坐标为无人机能耗. 从图中可看出节点个数的增加使无人机的飞行能耗增加. 与此同时, 图 6 的结果也表明无人机的能耗会随着系统的信息年龄 AoI 的减小而增加.

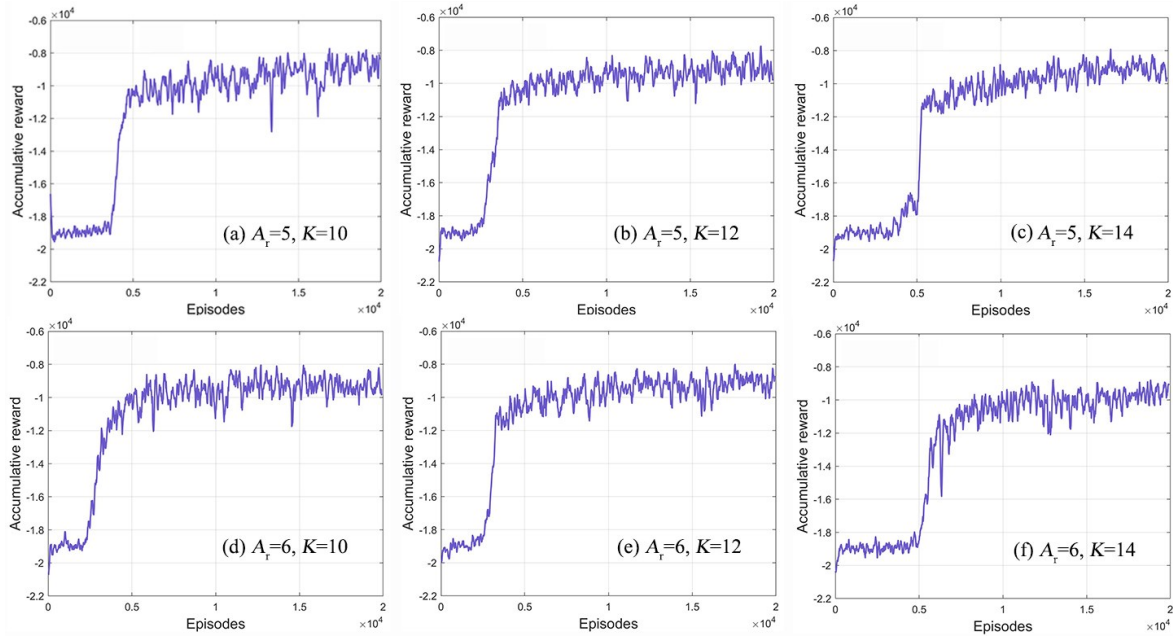


图 4 算法收敛性

Fig. 4 Convergence behavior of our algorithm

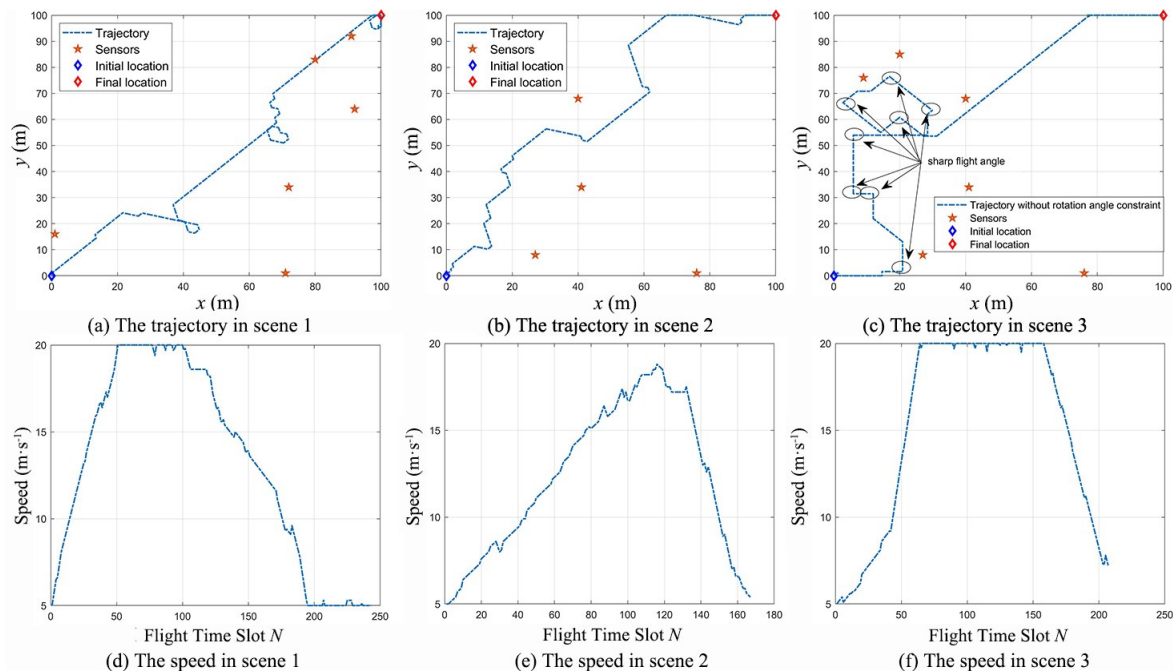


图 5 不同场景下无人机飞行轨迹和速度变化过程

Fig.5 The trajectory and speed of UAV in different scenes

5.4 算法比较 为了验证无人机能耗优化算法的有效性,图7将本文所提算法与贪婪算法进行了比较,图中横坐标为传感器节点个数,纵坐标为无人机能耗。从图7可看出,提出的DQN算法和传统的贪婪算法相比,可降低无人机能耗8%~

30%。贪心算法是根据当前状态选择环境反馈奖励中最大的对应动作,没有考虑信息长期AoI,因此,当传感器节点个数较大(如大于8)时,贪心算法得到的动作策略因无法满足信息AoI的要求而无法求解问题。

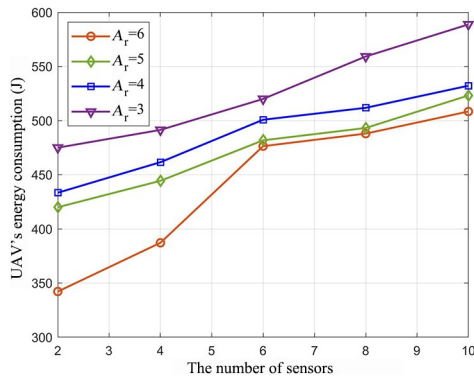


图6 AoI对无人机能耗的影响

Fig. 6 Performance of UAV with different AoI

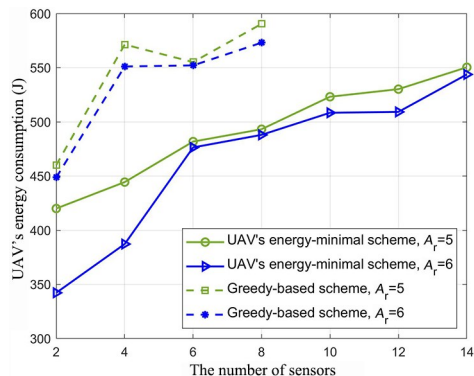


图7 本文算法与传统贪心算法的比较

Fig. 7 UAV's energy consumption of our algorithm and other traditional greedy schemes

6 结 论

本文研究了信息年龄受限的无人机辅助无线供能网络设计,其中无人机为传感器节点广播能量,收集传感器节点感知的信息,并通过联合优化无人机的飞行加速度、转角、飞行时间以及能量收集和信息更新调度策略以最小化无人机能耗.为了求解该非凸问题,提出了基于深度强化学习DQN的无人机能耗优化算法框架,它克服了状态空间巨大而带来的收敛速度慢的问题.仿真结果验证了提出的DQN算法的收敛性,且取得的无人机飞行能耗小于传统的贪心策略.数值结果表明,无人机会尽可能飞至传感器节点附近,飞行过程中速度先增大后减小,能够减少飞行能耗,并在飞行转角的约束下使无人机的飞行轨迹更平滑.另外,随着信息年龄的减小和传感器节点个数的增加,无人机的能耗也会增加.

参考文献

- [1] Ullah Z, Al-Turjman F, Mostarda L. Cognition in UAV - aided 5G and beyond communications: A survey. *IEEE Transactions on Cognitive Communications and Networking*, 2020, 6(3): 872–891.
- [2] Villarruel J E G, Corona B T. Proposal for a remote surgery system based on wireless communications, electromyography and robotics//*Proceedings of Electronics, Robotics and Automotive Mechanics Conference*. Cuernavaca, Mexico: IEEE, 2008: 93–98.
- [3] Kamenov S S, Todorova V D. Some challenges in creating of tactile sensor network in internet environment//*2018 IEEE 27th International Scientific Conference: Electronics*. Sozopol, Bulgaria: IEEE, 2018:1–3.
- [4] Yates R D, Kaul S. Real - time status updating: Multiple sources//*2012 IEEE International Symposium on Information Theory*. Cambridge, MA, USA: IEEE, 2012:2666–2670.
- [5] Kosta A, Pappas N, Angelakis V. Age of information: A new concept, metric, and tool. *Foundations and Trends® in Networking*, 2017, 12 (3):162–259.
- [6] Wu X W, Yang J, Wu J X. Optimal status updating to minimize age of information with an energy harvesting source//*2017 IEEE International Conference on Communications*. Paris, France: IEEE, 2017:1–6.
- [7] Hu L M, Chen Z C, Dong Y Q, et al. Status update in IoT networks: Age of information violation probability and optimal update rate. *IEEE Internet of Things Journal*, 2021, doi: 10.1109/JIOT.2021.3051722.
- [8] Farazi S, Klein A G, Brown D R. Average age of information for status update systems with an energy harvesting server//*IEEE Conference on Computer Communications Workshops*. Honolulu, HI, USA: IEEE, 2018:112–117.
- [9] Yang L, Zeng Y, Zhang R. Wireless power transfer with hybrid beamforming: How many RF chains do we need? *IEEE Transactions on Wireless Communications*, 2018, 17(10):6972–6984.

- [10] Zhang R C, Xiong K, Guo W, et al. Q-Learning-based adaptive power control in wireless RF energy harvesting heterogeneous networks. *IEEE Systems Journal*, 2021, 15(2): 1861—1872.
- [11] Di X F, Xiong K, Fan P Y, et al. Simultaneous wireless information and power transfer in cooperative relay networks with rateless codes. *IEEE Transactions on Vehicular Technology*, 2017, 66(4): 2981—2996.
- [12] Dabiri M, Emadi M J. Average age of information minimization in an energy harvesting wireless sensor node//2018 9th International Symposium on Telecommunications. Tehran, Iran; IEEE, 2018: 123—126.
- [13] Wu X W, Yang J, Wu J X. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Transactions on Green Communications and Networking*, 2018, 2(1): 193—204.
- [14] Arafa A, Yang J, Ulukus S, et al. Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies. *IEEE Transactions on Information Theory*, 2020, 66(1): 534—556.
- [15] Li L H, Cai R T, Jiang H, et al. Rate-energy tradeoff for SWIPT systems with multi-user interference channels under non-linear energy harvesting model//2019 IEEE 89th Vehicular Technology Conference. Kuala Lumpur, Malaysia; IEEE, 2019: 1—6.
- [16] Lu Y, Xiong K, Fan P Y, et al. Global energy efficiency in secure MISO SWIPT systems with non-linear power-splitting EH model. *IEEE Journal on Selected Areas in Communications*, 2019, 37(1): 216—232.
- [17] Boshkovska E, Ng D W K, Zlatanov N, et al. Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model. *IEEE Transactions on Communications*, 2017, 65(5): 1984—1999.
- [18] Wu Q Q, Li G Y, Ng D W K, et al. An overview of sustainable green 5G networks. *IEEE Wireless Communications*, 2017, 24(4): 72—80.
- [19] Yang D C, Wu Q Q, Zeng Y, et al. Energy tradeoff in ground-to-UAV communication via trajectory design. *IEEE Transactions on Vehicular Technology*, 2018, 67(7): 6721—6726.
- [20] Yang Z H, Xu W, Shikh-Bahaei M. Energy efficient UAV communication with energy harvesting. *IEEE Transactions on Vehicular Technology*, 2020, 69(2): 1913—1927.
- [21] Zeng Y, Zhang R. Energy-efficient UAV communication with trajectory optimization. *IEEE Transactions on Wireless Communications*, 2017, 16(6): 3747—3760.
- [22] Tan R J, Zhou J, Du H B, et al. An modeling processing method for video games based on deep reinforcement learning//2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference. Chongqing, China; IEEE, 2019: 939—942.
- [23] Moreno-Vera F. Performing deep recurrent double Q-learning for Atari games//2019 IEEE Latin American Conference on Computational Intelligence. Guayaquil, Ecuador; IEEE, 2019: 1—4.
- [24] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*, 2019, 518(7540): 529—533.
- [25] Yi M J, Wang X J, Liu J, et al. Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks//IEEE Conference on Computer Communications Workshops. Toronto, Canada; IEEE, 2020: 716—721.
- [26] Abd-Elmagid M A, Ferdowsi A, Dhillon H S, et al. Deep reinforcement learning for minimizing age-of-information in UAV-assisted Networks//2019 IEEE Global Communications Conference. Waikoloa, HI, USA; IEEE, 2019: 1—6.
- [27] Tong P, Liu J, Wang X J, et al. Deep reinforcement learning for efficient data collection in UAV-Aided internet of things//2020 IEEE International Conference on Communications Workshops. Dublin, Ireland; IEEE, 2020: 1—6.
- [28] You C S, Zhang R. 3D trajectory optimization in Rician fading for UAV-enabled data harvesting. *IEEE Transactions on Wireless Communications*, 2019, 18(6): 3192—3207.

(责任编辑 杨可盛)