

文章编号: 1671-7449(2022)02-0141-06

无人机中基于 C-DQN 的资源分配和 轨迹优化研究

沈 冰, 许士勋, 卢小凤, 付 芳, 张志才

(山西大学 物理工程学院, 山西 太原 030006)

摘 要: 随着无人机作为空中中继和终端技术的迅速发展, 物理层安全问题在近几年间已经成为一个研究热点. 本文将无人机发射的通信功率分为两部分(保密信息功率和人工噪声功率), 在传输保密信息时能有效防止非法窃听, 同时考虑了无人机在空中飞行时所需的动力功耗, 通过联合优化无人机的飞行轨迹和功率分配比, 以实现在固定能量下平均传输信息量最大化的目的. 将这一情景建模成马尔可夫模型(MDP), 并利用 Curiosity-Driven Deep Q-learning Network(C-DQN) 算法进行训练优化, 结果表明, 该算法具有良好的收敛效果.

关键词: 功率分配; 安全传输; 动力功耗; MDP; C-DQN 算法; 无人机

中图分类号: TP393 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2022.02.009

Research on Resource Allocation and Trajectory Optimization Based on C-DQN in UAV

SHEN Bing, XU Shixun, LU Xiaofeng, FU Fang, ZHANG Zhicai

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: With the rapid development of UAV as air relay and terminal, the issue of physical layer security has becoming a research hotspot in the recent. This article divides the communication power sent by the UAV into two parts that include power of confidential information and power of artificial noise, which can effectively prevent eavesdroppers eavesdropping illegally when UAV is transmitting confidential information. It also considers aerodynamic power consumption when UAV is flying. In order to maximize the average amount of information transferred at a fixed energy, this article optimizes the flight path and power distribution of the UAV. This scenario is modeled as a Markov model (MDP), and uses the Curiosity-Driven Deep Q-learning Network (C-DQN) algorithm to achieve optimization. The result shows that the algorithm has a good convergent effect.

Key words: aerodynamic power consumption; secure transmission; power consumption; MDP; C-DQN algorithm; UAV

0 引 言

随着移动通信技术的快速发展, 移动网络流

量需求的时空不均匀性越发明显. 无人机作为新一代的中继和终端, 具有传统固定基站所不具有的特点, 特别是在极其拥堵和偏远地区, 无人机

收稿日期: 2021-07-15

作者简介: 沈 冰(2000—), 女, 主要从事电子信息与技术研究. E-mail: shen_bing@163.com.

通信作者: 付 芳(1985—), 女, 讲师, 硕士生导师, 主要从事 5G 编码调制技术研究. E-mail: fufang0621@sxu.edu.cn

凭借其特有的移动性和视距传播的优势,作为空中基地来补充和支持现有的地面基础通信设施.结果表明,无人机可更灵活、更经济地为用户提供额外的通信和计算资源,以提高通信链路性能^[1].例如,在通信需求量大的地方充当临时通信热点,在自然灾害后快速且灵活地部署通信中继,在地势险要处起辅助勘察、搜救工作,在未覆盖无线网处提供连接等等.

无人机网络安全问题一直是不容忽视的存在,这主要是由于无线通信所固有的开放性和广播性,使得信息极易被非法截获.传统的依靠认证和加密技术虽可以部分解决无人机的安全问题,但近几年来,物理层安全凭借着其无需使用加密密钥的特点逐渐成为一个研究热点,通过设计其传输方式就可以有效阻止窃听者通过非法渠道获取信息,以实现网络安全^[2].

G. Zhang 和 X. Hong 等^[3-4]通过联合优化无人机的飞行轨迹以及合法发射功率,实现了最大化无人机与路面结点间的平均保密率;F. cheng 等^[5]通过联合优化无人机的运动轨迹和时间调度来保证具有缓存的无人机中继无线网络的安全性.然而现有的研究大多数为了保障物理层安全,在优化无人机的运动轨迹时,均未考虑到无人机在空中飞行时所需的动力功耗,这显然是不合理的.另外,在解决非法窃听问题上,H. Xing 等^[6]采用功率分裂的方法,避免了使用两架无人机,有效降低了成本.

本文在优化无人机的运动轨迹时,增加了无人机的动力功耗,同时将无人机发射的通信功率分裂为两部分,一部分为发射的保密信息功率,另一部分为发射的人工干扰噪声功率,无人机通过发射人工噪声来阻止非法窃听.

1 系统架构

1.1 系统模型

本文设计了一种无人机和地面用户之间的通信模型,如图1所示,无人机(UAV)传输保密信息给合法接收者(Legitimate user),但在传输信息的途中,多个窃听者(Eavesdroppers)会以非法方式窃取信息.因此,为保护发送给合法接收者的信息不被窃取,假设无人机可以制造并发射人工干扰噪声来隐藏保密数据.同时,假设无人机和合法接收者之间的通信可以识别并排除人工噪声,

因此干扰信号只对窃听者起作用.

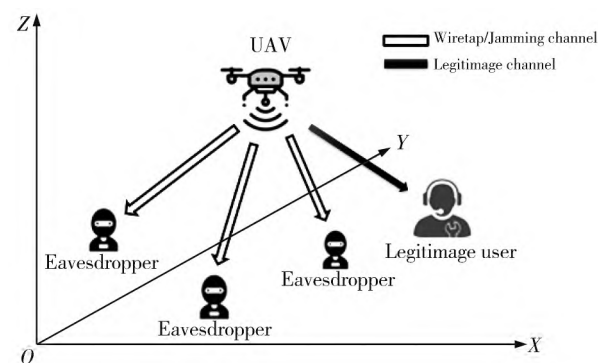


图1 系统模型

Fig. 1 System model

系统模型建立了一个三维笛卡尔坐标系,假定地面上合法接收者的位置是随机但不发生移动的,以 $(x_L, y_L, 0)$ 表示;无人机的位置是可知的,并位于固定高度 H 水平飞行,以 $(x_U[n], y_U[n], H)$ 表示;3个窃听者的位置也是随机但不发生移动的,以 $(x_{Ei}, y_{Ei}, 0)$ 表示,其中 $i = \{1, 2, 3\}$.对于这些随机位置,只要无人机上配备了光学相机或 Synthetic Aperture Radar(SAR)就可以检测到^[3].

为了简单起见,本文只关注无人机在空中飞行通信时的操作,而忽略了其起飞和着陆阶段.同时,将无人机的飞行时间 T 等间隔划分成 N 个足够小的时隙 δ_i ,可认为无人机在每个时隙 δ_i 下都处于准静态过程.

无人机与合法接收者的合法信道用 $g_{U2L}[n]$ 表示;无人机与窃听者的窃听信道和干扰信道用 $g_{U2Ei}[n]$ 表示.假定 $g_{U2L}[n]$ 和 $g_{U2Ei}[n]$ 都为视距(LoS)信道,遵循自由空间路径损耗模型,无人机到合法接收者的信道增益表示为

$$g_{U2L}[n] = \xi_0 d_{U2L}^{-2} = \frac{\xi_0}{(x_U[n] - x_L)^2 + (y_U[n] - y_L)^2 + H^2}. \quad (1)$$

无人机到某个窃听者的窃听信道和干扰信道的信道增益表示为

$$g_{U2Ei}[n] = \xi_0 d_{U2Ei}^{-2} = \frac{\xi_0}{(x_U[n] - x_{Ei})^2 + (y_U[n] - y_{Ei})^2 + H^2}, \quad i = 1, 2, 3, \quad (2)$$

式中: ξ_0 是参考距离 $d_0 = 1$ m 处的信道功率增益,取决于载波频率和收发信机的天线增益.

1.2 无人机通信功率分配模型

在保证传输信息安全性的基础上,假设无人机采用一种通信功率分割的方法,即将发射的通

信功率分裂为两部分,即保密信息功率和人工噪声功率。K. Xu 等^[7]证明,通过适当分裂无人机的发射功率,相比于没有功率分裂的情况,整个系统的性能可以得到很大提高。

本文将无人机的发射功率分为两部分,其中,一部分 p_U 用于向合法接收者发送保密信息,而另一部分 p_J 用于向窃听者发送人工噪声,这两个功率分别表示为

$$p_U(\rho) = \rho p_C, \quad (3)$$

$$p_J(\rho) = (1 - \rho) p_C, \quad (4)$$

式中: ρ 为功率分配比,需满足 $0 < \rho < 1$; p_C 为无人机的通信总功率。通过调整功率分配比 ρ ,使信息在发送过程中,既能防止被非法窃取,又可以保障在接收端接收到的信号功率仍然较高,以满足其能量采集的要求。

1.3 下行链路安全传输模型

本文模型中,由于接收端不可避免地存在高斯白噪声,因此信道传送的最大信息速率可以用香农定理表示,则合法信道的最大信息传输速率

$$C_{U2L}(\rho, n) = B \log_2 \left(\frac{1 + p_U(\rho) g_{U2L}[n]}{N_0^L} \right) = B \log_2 \left(\frac{1 + \rho p_C g_{U2L}[n]}{N_0^L} \right), \quad (5)$$

式中: B 为频道带宽; $p_U(\rho)$ 为无人机发送的保密信息功率; $g_{U2L}[n]$ 为无人机到合法接收者的信道增益; N_0^L 为信号在合法接收者处的高斯白噪声。

窃听信道和干扰信道的最大信息传输速率

$$C_{U2E, \max}(\rho, n) = \max_{i=1,2,3} B \log_2 \left(1 + \frac{p_U(\rho) g_{U2Ei}[n]}{N_0^{Ei} + p_J(\rho) g_{U2Ei}[n]} \right) = \max_{i=1,2,3} B \log_2 \left(1 + \frac{\rho p_C g_{U2Ei}[n]}{N_0^{Ei} + (1 - \rho) p_C g_{U2Ei}[n]} \right), \quad (6)$$

式中: $g_{U2Ei}[n]$ 为无人机到某个窃听者的信道增益; N_0^{Ei} 为信号在某个窃听者处的高斯白噪声; $p_J(\rho)$ 为无人机发射的人工干扰噪声功率。

根据式(5)、式(6),信息的安全容量可以根据合法信道的信息传输速率和窃听信道、干扰信道的信息传输速率表示为

$$C(\rho, n) = \max\{E[C_{U2L}(\rho, n) - C_{U2E, \max}(\rho, n)], 0\}. \quad (7)$$

1.4 无人机空中动力功耗模型

由于无人机飞行轨迹的不规则性,将飞行时

间划分成多个足够小的时隙,即将整个飞行轨迹划分成无穷小段;由于该模型建立在三维笛卡尔坐标系下,对于下一个状态,无人机只有向前、向后、向左、向右、悬停这 5 种选择。

根据旋转翼无人机的经典飞机动力学模型,处于固定高度无人机的空气动力功耗可以表示为水平飞行功率、空中悬停功率和与叶片阻力相关的剖面功率^[8]。

水平飞行功率

$$P_{\text{level}}[n] = \frac{W^2}{\sqrt{2} \rho A} \cdot \frac{1}{\sqrt{\|v_x[n], y[n]\|^2 + \sqrt{\|v_x[n], v_y[n]\|^4 + 4V_h^4}}}, \quad (8)$$

式中: $W = mg$ 表示无人机的重力; m 为无人机的质量; g 为重力加速度; ρ 表示空气密度; A 表示无人机旋翼圆盘的总面积; $\|v_x[n], v_y[n]\|$ 表示无人机的水平飞行速度,常量 $V_h = \sqrt{\frac{W}{2\rho A}}$ 代表悬停功率的一部分。

空中悬停功率

$$P_{\text{hover}}[n] = \frac{W^2}{\sqrt{2} \rho A} \cdot \frac{1}{\sqrt{4V_h^4}} = \frac{W^{\frac{3}{2}}}{\sqrt{2} \rho A}. \quad (9)$$

对比式(8)、式(9)可以发现,水平飞行比空中悬停要消耗更少的功率。

与叶片阻力相关的剖面功率

$$P_{\text{drag}}[n] = \frac{1}{8} C_{D0} \rho A \|v_x[n], v_y[n]\|^3, \quad (10)$$

式中: C_{D0} 是取决于无人机旋翼叶片几何形状的阻力系数,并且可以发现剖面功率与水平飞行速度的三次方成正比关系。

根据式(8)~式(10),无人机的空中动力功耗模型可以表示为

$$P_{\text{fly}}[n] = P_{\text{level}}[n] + P_{\text{hover}}[n] + P_{\text{drag}}[n] = \frac{W^2}{\sqrt{2} \rho A} \cdot \frac{1}{\sqrt{\|v_x[n], v_y[n]\|^2 + \sqrt{\|v_x[n], v_y[n]\|^4 + 4V_h^4}}} + \frac{W^{\frac{3}{2}}}{\sqrt{2} \rho A} + \frac{C_{D0} \rho A \|v_x[n], v_y[n]\|^3}{8}. \quad (11)$$

1.5 优化目标

本文的目的是通过不断优化无人机的运动轨迹和功率分配比 ρ ,即动作 a ,以达到在固定能量

下实现平均传输信息量最大化,即

$$\max_{a \in A} \frac{C(\rho, n)}{\alpha P_c + \beta P_{\text{fly}}[n]}, \quad (12)$$

$$\text{s. t. } C1: 0 < \rho < 1, \quad (13)$$

$$C2: x_{\min} \leq x_U[n] \leq x_{\max}, \\ y_{\min} \leq y_U[n] \leq y_{\max}, \forall n, \quad (14)$$

$$C3: z_U[n] = H, \forall n, \quad (15)$$

$$C4: P_c = P. \quad (16)$$

C1 限制了保密信息功率和人工噪声功率在总通信功率中的比例大小;C2 限制了无人机的水平飞行范围;C3 固定了无人机的飞行高度,在该高度下可以认为无人机躲避了所有建筑物的阻挡,处于无障碍自由飞行状态;C4 表示了无人机的通信功率为一定值。

1.6 马尔可夫决策模型(MDP)

由于无人机飞行的下一状态只与当前状态有关,而与更早之前的状态无关,具有马尔可夫性。因此,本文将上述优化问题和限制条件式(12)~式(16)建模为马尔可夫决策模型(MDP),用一个元组表示 $\langle S, A, P, R \rangle$ 。

S 表示状态空间,包括无人机与合法接收者的距离 d_{U2L} 、无人机与窃听者的距离 d_{U2E} 、保密信息功率 p_U 和噪声功率 p_J ,即 $S = \{d_{U2L}, d_{U2E1}, d_{U2E2}, d_{U2E3}, p_J, p_U\}$ 。

A 表示动作空间,包括无人机的飞行方向(向前 a 、向后 b 、向左 l 、向右 f 、悬停 h)和功率分配比 ρ ,即 $A = \{a, b, l, f, h, \rho\}$ 。

P 表示状态转移概率,即在状态 s 下采用动作 a 后转移到下一状态 s' 的概率,但在本模型中,下一个状态完全由无人机的飞行方向以及当前状态所决定,不存在其他不确定因素,因此转移到下一状态 s' 的概率为 100%,所以不存在状态转移概率。

r 表示及时奖励,可以表示为

$$r_\pi = \max_{a \in A} \frac{C(\rho, n)}{\alpha P_c + \beta P_{\text{fly}}[n]}, \quad (17)$$

式中: α 和 β 分别表示无人机通信功率和飞行功率的折扣因子,需满足 $0 < \alpha < 1, 0 < \beta < 1$ 。

通过不断地训练学习参数化的状态动作价值函数 $Q(s, a)$,找到最优的状态动作价值函数 $Q^*(s, a)$,间接找到最优的策略 π^* 。根据贝尔曼方程,动作状态价值函数

$$Q_\pi(s, a) =$$

$$E_\pi[r_{t+1} + \gamma Q_\pi(S_{t+1}, A_{t+1}) | S_t = s, A_t = a], \quad (18)$$

式中: γ 为后继状态奖励的折扣因子,需满足 $0 < \gamma < 1$,价值函数由及时奖励和后继奖励组成,且下一步对当前策略的影响最大,时间越远,影响越小。

2 C-DQN 算法的资源分配

采用 C-DQN 算法求解马尔可夫模型中的及时奖励 r ,该算法由 D. Pathak 等^[9]提出。如图 2 所示,该算法由两部分组成:一部分是传统的 DQN 模块;另一部分是基于好奇心的内在驱动模块。该算法相较于 DQN 算法(仅能由与环境交互的外部奖励驱动)的最大区别在于智能体可以由自身的内在好奇奖励所驱动,使算法的收敛速度更快、收敛值更大。

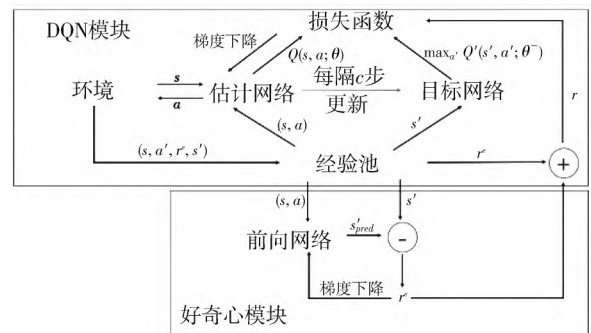


图 2 C-DQN 算法

Fig. 2 C-DQN algorithm

2.1 DQN 模块

智能体与环境进行交互,在 s_t 状态下根据 ϵ 策略随机选择 a_t 动作,得到奖励 r_t^e 并转移到下一个状态 s_{t+1} ,将大量的经验 $(s_t, a_t, r_t^e, s_{t+1})$ 存储到经验池中,用于神经网络的训练。训练时,从经验池中随机选取一小片经验 (s_j, a_j, r_j, s_{j+1}) ,根据贝叶斯公式得到标签

$$y_j = \begin{cases} r_j, & j+1 \text{ 步结束,} \\ r_j + \gamma \max_{a_{j+1}} Q'(s_{j+1}, a_{j+1}; \theta^-), & \text{其它,} \end{cases} \quad (19)$$

式中: r_j 不同于传统的 DQN 网络,包括环境反馈的奖励 r_j^e 和内生刺激产生的奖励 r_j^i , γ 表示折扣因子,需满足 $0 < \gamma < 1$, Q' 是参数为 θ 的目标网络的状态动作价值函数。

损失函数

$$\text{Loss}(\theta) = E(y_j - Q(s_j, a_j; \theta))^2, \quad (20)$$

式中: Q 是参数为 θ 的估计网络的状态动作价值函数. DQN 算法通过梯度下降的策略最小化损失函数来优化 DQN 神经网络的参数 θ , 该策略可以表示为

$$\theta \leftarrow \theta - \alpha_q (y - Q(s, a; \theta)) \cdot \nabla_{\theta} Q(s, a; \theta), \quad (21)$$

式中: α_q 是 DQN 模块的学习率, 需满足 $0 < \alpha_q < 1$.

估计网络每隔 C 步对目标网络进行参数的更新, 可以有效消除时间相关性.

2.2 基于好奇心的内在驱动模块

根据原始的好奇心模型, 本文删除了其反向模型, 只保留了前向模型, 这是因为反向模型的作用是预测智能体从 s_t 转移到 s_{t+1} 所采取的动作 a_t , 但在本文的模型中, 已对状态空间进行了筛选, 去除了对智能体无影响的状态. 通过前向模型对下一状态进行预测, 将状态的预测值与下一状态的实际值相比较, 即可得到内在奖励.

从经验池中选取 a_t 和 s_t , 先将原始状态 s_t 编码为特征向量 $\phi(s_t)$, 再利用前向模型对这两个输入进行预测, 得到含参数为 ϑ 的下一状态的预测值

$$\phi_{pred}(s_{t+1}) = f(a_t, \phi(s_t); \vartheta). \quad (22)$$

损失函数

$$Loss(\vartheta) = \frac{1}{2} \|\phi_{pred}(s_{t+1}) - \phi(s_{t+1})\|_2^2. \quad (23)$$

内在驱动模块通过梯度下降策略最小化损失函数来优化更新好奇心神经网络的参数 ϑ , 该策略可以表示为

$$\vartheta \leftarrow \vartheta - \alpha_i (y - \phi_{pred}(s_{t+1})) \cdot \nabla_{\vartheta} \phi_{pred}(s_{t+1}), \quad (24)$$

式中: α_i 是内在驱动模块的学习率, 需满足 $0 < \alpha_i < 1$.

内在奖励与预测的下一个状态有关, 可以表示为

$$r_t^i = \frac{\eta}{2} \|\phi_{pred}(s_{t+1}) - \phi(s_{t+1})\|_2^2, \quad (25)$$

式中: η 是缩放因子, 需满足 $\eta > 0$. 当内在奖励趋于 0 时, 就表明已不存在好奇心, 该模块已完成参数的训练更新.

3 算法仿真与数据分析

仿真模拟一个固定高度飞行的无人机对地面

上一个随机初始化位置且固定不动用户的防窃听通信情景. 对该情景采用 python 仿真器进行仿真, 具体参数设置见表 1.

表 1 系统参数

Tab. 1 System parameter

参数	值
无人机的水平飞行范围/m	$0 \leq x \leq 160,$ $0 \leq y \leq 160$
无人机的飞行高度 H /m	100
单位信道功率增益 ξ_0	1.42×10^{-4}
频道带宽 B /MHz	20
背景噪声功率 N_0 /W	1.99×10^{-13}
无人机的通信总功率 p_C /W	10
无人机的质量 m /kg	4
重力加速度 g /m \cdot s $^{-2}$	9.8
空气密度 ρ /kg \cdot m 3 [8]	1.225
无人机旋翼圆盘的总面积 A /m 2 [8]	0.18
无人机的水平飞行速度 $\ (v_x[n], v_y[n])\ $ /m \cdot s $^{-1}$	6
无人机旋翼叶片几何形状 的阻力系数 C_{D0} [8]	0.08

图 3 是不同学习率下 DQN 算法的平均外部奖励收敛图. 由图 3 可知, $\alpha_q = 0.001$ 曲线收敛速度比 $\alpha_q = 0.015$ 曲线更快, 且两者在收敛后的奖励值趋于相同, 虽然 $\alpha_q = 0.005$ 曲线的收敛速度比上述两者更快, 但其收敛值明显更小. 因此, 综合考虑, $\alpha_q = 0.001$ 曲线仿真效果更好.

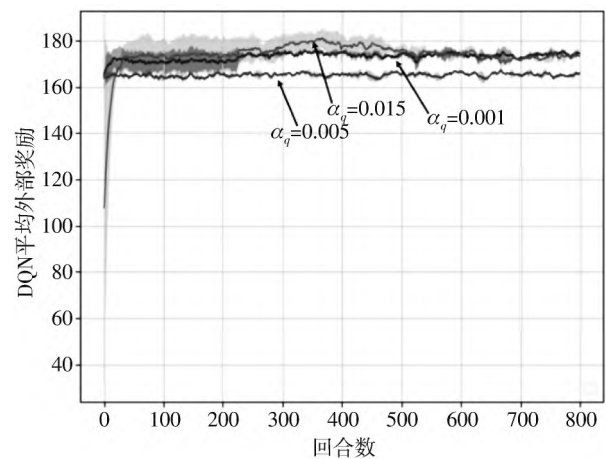


图 3 不同 DQN 学习率下收敛图

Fig. 3 Convergence graph under different DQN learning rates

图 4 是不同学习率下 C-DQN 算法的平均外部奖励收敛图. 由图 4 可知, 这 3 种学习率的曲线收敛速度相似, 但 $\alpha_q = 0.004$ 曲线的收敛值要略大于其它两种学习率的曲线. 因此, 综合考虑, $\alpha_q = 0.004$ 曲线仿真效果更好.

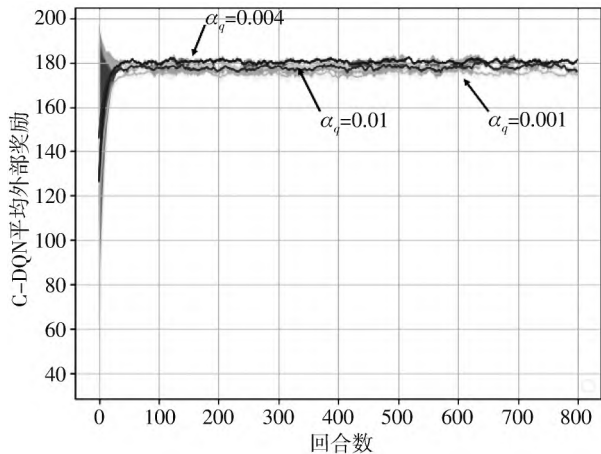


图4 不同C-DQN学习率下收敛图

Fig. 4 Convergence graph under different C-DQN learning rates

图5是在本文的情景下DQN算法和C-DQN算法的性能比较图。其中，DQN算法的学习率为0.001；C-DQN算法的外部学习率为0.004，内部学习率为0.01。由图5可知，当C-DQN算法在内部奖励趋于0时，外部奖励也将趋于稳定。对比DQN和C-DQN算法的外部奖励曲线，可以发现，两者的收敛速率近似相同，但后者只需在大约50迭代次数处就可达到收敛而前者要在230迭代次数左右才达到收敛，后者的收敛速率更快，且后者的外部奖励在收敛时明显大于前者的外部奖励。因此，本文所采用的C-DQN算法性能优于DQN算法。

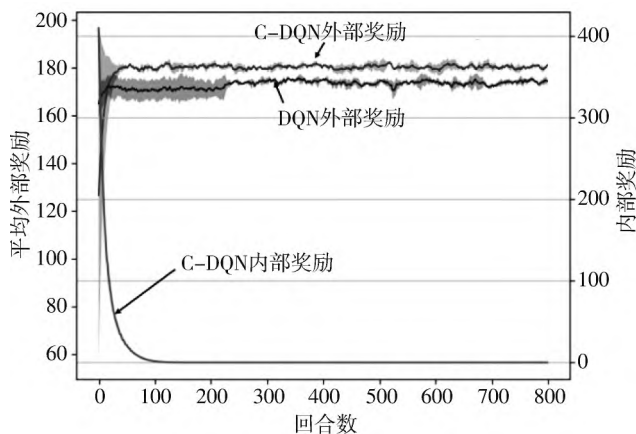


图5 C-DQN和DQN算法的性能比较

Fig. 5 Performance comparison of C-DQN and DQN algorithms

4 总结

本文提出了一种在无人机系统中实现物理层安全的方法，通过联合优化无人机的功率分配比

和飞行轨迹，实现在固定能量下平均传输信息量最大化。由于本文所优化的物理量具有后无效性，可将上述问题建模成马尔可夫模型，并利用C-DQN算法来解决该问题。仿真结果表明，该算法相比于DQN算法的学习效果更佳。

参考文献：

- [1] LIU X, LIU Y, CHEN Y, 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.
- [2] ZHOU X X, WU Q, YAN S, et al. UAV-enabled secure communications: joint trajectory and transmit power optimization[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 4069-4073.
- [3] ZHANG X G, WU Q, CUI M, et al. Securing UAV communications via joint trajectory and power control [J]. IEEE Transactions on Wireless Communications, 2019, 18(2): 1376-1389.
- [4] HONG X X, LIU P, ZHOU F, et al. Resource allocation for secure UAV-assisted SWIPT systems[J]. IEEE Access, 2019, 7: 24248-24257.
- [5] CHENG X F, GUI G, ZHAO N, et al. UAV-relaying-assisted secure transmission with caching [J]. IEEE Transactions on Communications, 2019, 67(5): 3140-3153.
- [6] XING X H, LIU L, ZHANG R. Secrecy wireless information and power transfer in fading wiretap channel [J]. IEEE Transactions on Vehicular Technology, 2016, 65(1): 180-190.
- [7] XU X K, ZHAO M M, CAI Y, et al. Low-complexity joint power allocation and trajectory design for UAV-enabled secure communications with power splitting[J]. IEEE Transactions on Communications, 2021, 69(3): 1896-1911.
- [8] SUN X Y, XU D, NG D W K, et al. Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems[J]. IEEE Transactions on Communications, 2019, 67(6): 4281-4298.
- [9] PATHAK X D, AGRAWAL P, EFROS A A, et al. Curiosity-driven exploration by self-supervised prediction[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017: 488-489.