文章编号： 1671-7449(2024)04-0363-08

基于过时信息年龄最小化无人机路径规划

殷　杰[1]，付　芳[2*]，李　凯[1]

（1. 山西大学 物理电子工程学院， 山西 太原 237016； 2. 海南大学 计算机科学与技术学院， 海南 海口 570228）

摘　要： 在一些对信息新鲜度要求较高的物联网(IoT)场景中，无人机凭借其高机动性和灵活性的特点可以用于辅助采集设备数据。现有的评估信息新鲜度的信息年龄存在一定局限性，无法从内容角度准确定义。该文采用过时信息年龄来表征信息新鲜度，并且提出基于好奇心驱动DQN算法的无人机辅助IoT数据采集方案，通过优化无人机轨迹，实现在满足无人机能耗约束条件下过时信息年龄的最小化。仿真结果表明，相比传统DQN算法，该文所提算法使智能体探索能力加强，收敛速度变快约45%，所得奖励值高约67%。

关键词： 物联网；过时信息年龄；无人机路径规划；好奇心驱动DQN算法
中图分类号： TP18　　**文献标识码：** A　　**doi:** 10.3969/j.issn.1671-7449.2024047
引用格式： 殷杰，付芳，李凯．基于过时信息年龄最小化无人机路径规划[J]．测试技术学报，2024，38(4)：363-370.
YIN Jie，FU Fang，LI Kai. Age of outdated information optimal UAV path planning[J]. Journal of Test and Measurement Technology，2024，38(4)：363-370.

Age of Outdated Information Optimal UAV Path Planning

YIN Jie[1]，FU Fang[2*]，LI Kai[1]

（1. College of Physics and Electronic Engineering，Shanxi University，Taiyuan 237016，China；
2. School of Computer Science and Technology，Hainan University，Haikou 570228，China）

Abstract： In some Internet of Things（IoT）scenarios that require high information freshness，UAV can be used to assist in collecting device data due to their high mobility and flexibility. The existing evaluation of information freshness has certain limitations in terms of age and cannot be accurately defined from a content perspective. This article uses the age of outdated information to characterize the freshness of information，and proposes a UAV-assisted IoT data collection scheme based on the curiosity driven DQN algorithm. By optimizing the trajectory of the UAV，the age of outdated information is minimized while meeting the constraints of UAV energy consumption. The simulation results show that compared to the traditional DQN algorithm，the proposed algorithm enhances the exploration ability of the intelligent agent，accelerates the convergence speed by about 45%，and achieves a reward value of about 67% higher.

Key words： Internet of Things；age of outdated information；UAV path planning；curiosity-driven DQN algorithm

收稿日期： 2023-08-29
作者简介： 殷　杰(1998－)，男，硕士生，主要从事强化学习、网络资源优化研究。E-mail：yinjie0919@163.com。
***通信作者：** 付　芳(1985－)，女，讲师，博士，主要从事联邦学习、区块链研究。E-mail：215780778@qq.com。

0 引 言

物联网(IoT)技术被认为是继计算机和互联网之后，信息技术领域的又一次巨大飞跃，已在工业生产、智慧交通等领域得到了广泛应用[1-2]。其中，高效采集终端数据成为提升物联网服务能力的关键，信息需要被尽可能快地传递给目的端或数据中心，以便进行数据分析并做出决策，因此，保证接收数据在目的地的新鲜度十分重要。由于无人机高机动性和高灵活性，可以用来辅助采集信息，保证信息的新鲜度[3]。

目前相关研究中，还存在以下3个不足：

1) 在评估信息新鲜度时，通常使用的评价指标是信息年龄(AoI)[4]，但其在量化信息新鲜度方面存在局限性，无法从内容角度准确定义信息的真正新鲜度。Wu等[5]提出了一个端到端的人工智能框架,并通过神经轨迹求解器规划最小AoI的无人机飞行路径；Hu等[6]联合能量传输、数据收集时间分配和无人机的轨迹规划，最小化数据的平均AoI。但是，由于AoI只考虑了信息最初生成以来经过的时间，而不是信息过时的时间。如果信息源没有新的信息内容更新，则具有较大AoI的信息仍可能被视为“新鲜”信息；Liu等[7]提出了过时信息年龄(AooI)来改进AoI。AooI的测量是从现在到信息在信息源位置第一次更改之间经过的时间。AooI非常适合用于处理非常大的信息内容状态空间。

2) 无人机的电量有限，航行过程中需要考虑电量限制，但现有的一些工作并没有考虑该限制。Zhu等[8]提出了加权A*算法来最小化无人机从地面物联网网络收集到数据的AoI，并进行无人机路径规划；Liu等[9]提出了基于逐次凸近似(SCA)方法和基于遗传算法(GA)来解决无人机环境监测系统中的无人机轨迹规划问题；Liu等[10]通过动态规划算法求解无人机轨迹和节点与信息收集点关联关系的联合优化问题，最小化传感器的最大AoI。以上这些工作没有考虑到现实情况的电量限制，不符合现实情况。

3) 有关研究采用传统的数学优化算法，没有考虑无人机何时需要充电以及充电次数。

本文提出了基于过时信息年龄最小化的无人机辅助IoT设备数据采集路径规划的方案，并且控制无人机在电量到达一定值时进行充电。采用一种好奇心驱动的DQN算法来处理该优化问题，可以满足无人机及时充电的需求。与现有的传统深度强化学习方法仅通过环境产生的外部奖励来训练智能体不同，为了更有效地探索状态，好奇心驱动的DQN算法通过外部奖励和内部奖励来训练智能体，内部奖励被定义为智能体，预测自身行为后果的能力误差。

1 系统模型与问题建立

1.1 场景设置

如图1所示，一个城市场景，由一个无人机、M个IoT设备，N个充电站组成，$D=\{d_1,d_2,\cdots,d_m\}$表示IoT设备的集合，$C=\{c_1,c_2,\cdots,c_n\}$表示充电站的集合。无人机从设定的起点出发巡航至终点，并在飞行过程中收集尽可能新鲜的IoT设备信息，达到最小过时信息年龄的目标，并且，无人机需要在电池电量快耗尽时，寻找充电站进行充电，确保完成路径。整个时间系统分成T个等长的时隙，即$\Gamma=[0,1,\cdots,T]$，每个时隙的长度为τ秒，无人机在第一个时隙起飞。

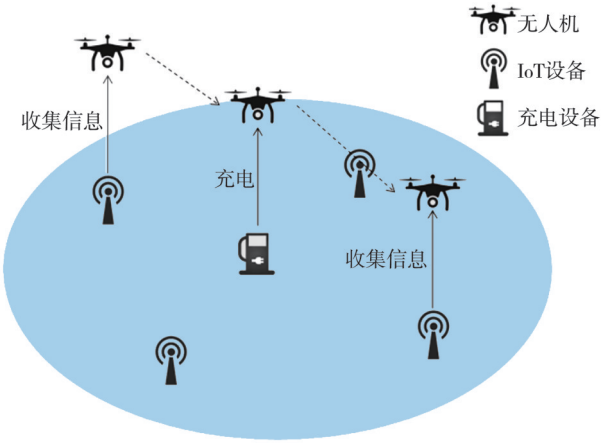


图1 场景设置

Fig. 1 Scene settings

假设无人机在固定高度H飞行，防止出现碰撞，并保持恒定的飞行速度。设$p(t)=(x,y,H)$为时隙t的无人机位置，则起点坐标$p(0)=(x_0,y_0,H)$。在调度时，一个无人机同一时刻只能访问一个IoT设备。假设$a_d(t)$为t时刻访问IoT设备$a_c(t)$，为t时刻访问充电站[11]。在t时刻访问IoT设备时$a_{d_m}(t)=1$，否则等于0，在t时刻访问充电站时$a_{c_n}(t)=1$，否则等于0，设定$a_{d_m}(t)\cdot a_{c_n}(t)=0$并且$a_{d_m}(t)+a_{c_n}(t)\neq 0$，即同一

时间无人机只能访问同个节点。

1.2　通信模型

由于在城市环境中，无人机航行过程中可能会遇到IoT设备之间有树木、房屋等阻挡，因此假设无人机和IoT设备之间的信道由LoS和NLoS混合信道控制[12]，并存在一定的概率。因此，可以得到无人机和IoT设备间的信道衰落，如下式(1)所示

$$g_s = g_{\text{free}} + \mu, \tag{1}$$

式中：$g_{\text{free}} = 20\log s + 20\log f_c + 20\log 4\pi/c$ 为自由空间路径损耗，s为无人机与IoT设备间的距离，f_c和c分别为载波频率和光速，μ是平均额外路径损失，分为LoS或NLoS两种情况，设LoS信道的概率为P_{LoS}，NLoS信道概率为P_{NLoS}，其中$P_{\text{NLoS}} = 1 - P_{\text{LoS}}$，则平均路径损耗

$$\bar{g} = g_{\text{LoS}} P_{\text{LoS}} + g_{\text{NLoS}} P_{\text{NLoS}}。 \tag{2}$$

LoS信道的概率可以简单表示为[13]

$$P_{\text{LoS}} = (1 + ae^{-b(\theta - a)})^{-1} = \left(1 + a\mathrm{e}^{-b\left(\sin^{-1}\left(\frac{H}{s}\right) - a\right)}\right)^{-1}, \tag{3}$$

式中：a和b为环境参数；θ为IoT设备与无人机的仰角；H为垂直高度；s为通信双方距离。

无人机和IoT设备之间的上行链路吞吐量为

$$R_m = B\log_2(1 + SNR_m), \tag{4}$$

式中：B为信道带宽；$SNR_m = w_m \bar{g}/\sigma^2$ 为无人机和IoT设备间的信噪比；w_m为第m个IoT设备的发射功率；σ^2为高斯白噪声功率。

1.3　能量模型

无人机在航行期间需要考虑电量限制。无人机能耗用于无人机的悬停与飞行，与速度有关，根据文献[14]，四旋翼无人机的推进能耗为

$$E_{\text{uav}} = P_1\sqrt{\sqrt{1 + \frac{v_t^4}{4v_0^4}} - \frac{v_t^2}{2v_0^2}} + \frac{1}{2}d\rho s A_r v_t^3 + P_0\left(1 + \frac{3{v_t}^2}{U_{\text{tip}}^2}\right), \tag{5}$$

式中：P_0和P_1分别为叶型功率和推导功率；v_t为无人机在t时隙的飞行速度；U_{tip}为无人机的旋翼叶尖速度；v_0为悬停状态下的平均旋翼诱导速度；d为机身阻力比；ρ为空气密度；s为旋翼固体度，A_r为旋翼盘面积。

为了保证无人机保留足够的能量飞到充电站，需要设置一个电量约束

$$E_{\text{remain}} \geqslant \frac{\min\left\| p_{c_n} - p_{\text{uav}} \right\|}{v_t} E_{\text{uav}}。 \tag{6}$$

式中：p_{c_n}与p_{uav}为充电站与无人机的坐标。

1.4　过时信息年龄

采用文献[7]中提出的过时信息年龄(AooI)来表征采集信息的新鲜度。AooI定义为从现在到信息在信息源位置第一次更改之间经过的时间。假设$y_m(t)$表示IoT设备信息内容在时间间隔$[t_1, t_2]$中是否发生变化，$x_m(t)$为m设备t时刻的信息内容，则有

$$y_m(t) = \begin{cases} 1, & \text{if there exists } \hat{t} \in [t_1, t_2], \hat{t} \in Z^+, \\ \text{such that}, & x_m(t) \neq x_m(t_1), \\ 0, & \text{otherwize}。 \end{cases} \tag{7}$$

假设$U_m(t)$为IoT设备处自生成信息以来未发生变化的最大时间，则有

$$U_m(t) = \max_{G_m(t) \leqslant \hat{t} < t,\ \hat{t} \in Z^+} \{\hat{t}:\ y_m(G_m(t),\ \hat{t}) = 0\}, \tag{8}$$

因此，$U_m(t)+1$为变化后的第一个时隙。定义$O_m(t)$为t时刻的过时信息年龄

$$O_m(t) = t - (U_m(t) + 1)。 \tag{9}$$

每当IoT设备收集的数据与无人机收集到的内容一致时，AooI保持0。而当IoT设备收集的数据第一次改变后，AooI会随着时间线性增加。AooI考虑了收集的数据的实际信息内容，只要当前数据内容没有过时，AooI就不会随时间增加。当无人机采集了IoT设备处的信息后，IoT设备处的AooI变为1，并在下次改变前保持不变。

假设每个IoT设备处的信息内容变化或者不变化的状态遵循一个二元离散时间马尔科夫链，如图2所示。设$l_m(t) \in \{0, 1\}$为二元状态变量，表示IoT设备处的信息内容x_m在$t-1$时刻与t时刻相比是否发生变化

$$l_m(t) = \begin{cases} 1, & \text{if } x_m(t) \neq x_m(t-1); \\ 0, & \text{otherwise}。 \end{cases} \tag{10}$$

设p_m和q_m为状态转移概率，定义为

$$P(l_m(t+1) = 0 | l_m(t) = 0) = p_m, \tag{11a}$$

$$P(l_m(t+1) = 1 | l_m(t) = 0) = 1 - p_m, \tag{11b}$$

$$P(l_m(t+1) = 1 | l_m(t) = 1) = q_m, \tag{11c}$$

$$P(l_m(t+1)=0|l_m(t)=1)=1-q_m。\quad(11d)$$

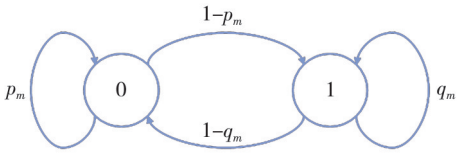


图 2 二元变量马尔科夫链

Fig. 2 Binary Markov Chain

即如果IoT设备处信息内容从t-1时刻到t时刻没有变化，那么从t时刻到$t+1$时刻，信息内容保持不变的概率为p_m，信息内容变化的概率为$1-p_m$。同样，如果IoT设备处信息内容从$t-1$时刻到t时刻发生变化，从t时刻到$t+1$时刻信息内容变化的概率为q_m，信息内容保持不变的概率为$1-q_m$。

1.5 优化问题

本文的优化目标是在满足能量约束的情况下，通过优化无人机的轨迹$p(t)$，使整个巡航周期内的AooI长期加权和最小。

$$\min_{p(t)}\sum_{t=1}^{T}\sum_{m=1}^{M}\theta_m O_m^t,\ \text{s.t.}p(0)=(x_0,y_0,H),\quad(12a)$$

$$\|v_t\|=1\quad\forall t\in[0,T],\quad(12b)$$

$$a_{d_m}(t)\in[0,1],\ \forall d_m\in D,\ t\in T,\quad(12c)$$

$$a_{c_n}(t)\in[0,1],\ \forall c_n\in C,\ t\in T,\quad(12d)$$

$$a_{d_m}(t)\cdot a_{c_n}(t)=0,\quad(12e)$$

$$a_{d_m}(t)+a_{c_n}(t)\neq 0,\quad(12f)$$

$$E_{\text{remain}}\geqslant\frac{\min\|p_{c_n}-p_{uav}\|}{v_t}E_{uav},\quad(12g)$$

式中：θ_m为设备的权重系数，表示设备间的相对重要性；式(12a)为无人机的初始位置，式(12b)为设置恒定的飞行速度；式(12c)、(12d)、(12e)、(12f)为无人机的调度策略，即在同一时间只能访问一个设备或进行充电；式(12g)为设置的能量约束，使无人机剩余的能量满足足够的电量在接下来的航程中飞到最近的充电站。

2 问题求解

2.1 马尔可夫决策过程

在本文中，把无人机采集数据这一过程用一个由元组(S,A,R)组成的马尔科夫决策过程建模，将无人机看做智能体。描述如下：

1) 状态空间S：时刻t的状态定义为

$$S(t)=\{O_m(t),p(t),E_{\text{remain}}(t)\},\quad(13)$$

式中：$O_m(t)=(O_1(t),O_2(t),\cdots,O_m(t))$，为$m$个不同的IoT设备处的过时信息年龄AooI；$p(t)=(x,y,H)$为无人机当前的位置；$E_{\text{remain}}(t)\triangleq[0,E_{\max}]$为无人机当前的剩余电量。将所有可能状态的空间用$S$表示，为了使状态空间有限，假设AooI最大可以达到$\Delta_{\max}$，其是有限的，但可以任意大。

2) 动作空间A

设置动作空间是恒定速度并且满足式(12c)，(12d)，(12e)，(12f)的调度策略的无人机轨迹。

3) 奖励函数R

设置的MDP中，无人机奖励函数为即时奖励函数。将时刻t时的AooI加权和的负值以及当前电量与最大电量的比值作为瞬时奖励，表示为

$$R=-\sum_{m=1}^{M}\theta_m O_m^t+k\frac{E_{\text{remain}}}{E_{\max}},\quad(14)$$

式中：k为电量消耗权重系数。

马尔可夫决策过程的目标是，通过折扣因子$\gamma\in[0,1]$最大化长期累积折现奖励，得到最优策略π^*。折扣因子反映的是即时和未来回报的重要性。$\gamma=0$表示只考虑下一个奖励，而$\gamma=1$表示所有奖励都同样重要，无论这些事件发生的距离有多远。因此得到最优策略

$$\pi^*=\arg\max_{\pi}\lim_{T\to\infty}E\left[\sum_{t=1}^{T}\gamma^{t-1}R(S,A)|S\right]。\quad(15)$$

2.2 算法设计

传统的强化学习算法旨在通过最大化环境提供的奖励来学习策略，以实现目标任务。然而，这些现有的强化学习方法大多缺乏探索能力，并且强化学习是基于奖励机制的，这就面临两个问题，其一是稀疏奖励，在复杂的环境下，很难做到每一个行动得到即时奖励。其二是在强化学习的环境中，设置的外部奖励是固定的，无法扩展。Pathak等[15]提出好奇心驱动学习这一解决方法，即建立一个拥有内在奖励函数的智能体。好奇心其实是一种内部奖励，是智能体在其当前状态下预测其自身行为的后果的误差(也就是在给定当前状态和采取的行动的情况下预测下一个状态)。为了计算这一误差，文中介绍了内在好奇心模块(ICM)，分为前向模型模块、反向模型模块、编码器模块来训练智能体。

为了使智能体更有效地探索，采用基于好奇心的DQN算法来最优化过时信息年龄的长期加

权和。将好奇心这一思想结合DQN来解决最优化问题，具体流程如图 3 所示。

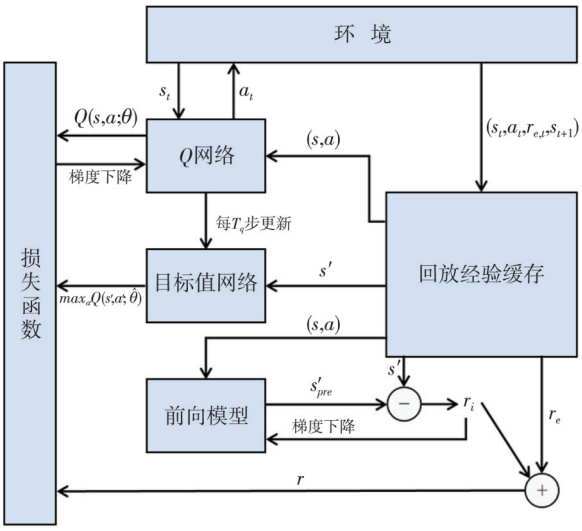


图 3　好奇心驱动DQN算法

Fig. 3　Curiosity-driven DQN algorithm

算法伪代码如表 1 所示，算法主体为DQN算法，在此基础上加入一个好奇心网络，网络由一个前向模型组成。算法开始时，清空经验缓冲区，初始化Q网络参数，初始化环境状态。算法进入循环，每个回合开始，随机选择一个状态s_t输入Q网络，根据ε-贪心算法选择a_t，得到下一状态s_{t+1}和奖励r_e。将这一元组(s_t, a_t, r_e, s_{t+1})储存到经验缓冲区。

表 1　算法伪代码

Tab. 1　Algorithm pseudocode

算法：好奇心驱动的DQN算法
1）初始化Q网络学习率l_q、好奇心网络学习率l_{cur}、折现因子γ、权重参数ϕ、φ、和时间T。
2）初始化经验缓存区M。
3）for episode z= 1, Z do
4）初始化状态s_1。
5）for t=1, T do
6）采用$\varepsilon-greedy$策略选择一个动作a_t。
7）根据动作a_t得到下一状态s_{t+1}和奖励值$r_{e,t}$。
8）更新经验缓存区：
9）$M \leftarrow M \cup \{s_t, a_t, r_{e,t}, s_{t+1}\}$。
10）训练：从M中随机抽取小批次经验元组，$(s_m, a_m, r_{e,m}, s'_m)_{m=1}^{M}$。
11）根据式(19)更新好奇心网络参数ϕ。
12）根据式(17)计算内部奖励r_i。
13）根据式(22)更新Q网络参数φ。
14）更新目标网络参数$\hat{\varphi}$。
15）end for
16）end for

在训练阶段，从经验缓冲区取出小批次的(s, a)对和s'分别输入到Q网络、目标值网络以及好奇心网络中。在Q网络中得到$Q(s,a)$值，在目标值网络中得到$\max_{a'} Q(s', a')$值，在好奇心网络中，(s, a)输入前向模型，得到s'_{pre}，即根据前向模型函数和状态动作对得到的下一状态预测值

$$s'_{\text{pre}} = f(s, a, \phi)。\tag{16}$$

与经验池中输入的s'实际下一状态做比，得到了二者的误差，即是内部奖励

$$r_i = \frac{\delta}{2}\|s' - s'_{pre}\|_2^2。\tag{17}$$

$\delta > 0$为比例系数。根据最小化前向模型损失函数梯度更新参数ϕ，

$$\text{Loss}(\phi) = \frac{1}{2}\|s' - f(s, a, \phi)\|_2^2,\tag{18}$$

$$\phi \leftarrow \phi - l_{\text{cur}}(Q_{\text{tar}} - f(s, a; \phi))\nabla_\phi f(s, a; \phi),\tag{19}$$

式中：l_{cur}为好奇心网络的学习率，参数经过数个步长后更新。内部奖励r_i与外部奖励r_e相加，得到总的奖励值，即$r = r_e + r_i$。使用$Q(s,a)$，$\max_{a'} Q(s', a')$以及总奖励值计算Loss值，

$$\text{Loss}(\varphi) = E[Q_{\text{tar}} - Q(s, a; \varphi)]^2,\tag{20}$$

式中：Q_{tar}为目标Q值。

$$Q_{tar} = \begin{cases} r, \\ r + \gamma \max_{a'} \hat{Q}(s', a'; \hat{\varphi}), \end{cases}\tag{21}$$

式中：γ为折现因子；$\hat{Q}$是参数为$\hat{\varphi}$的目标网络的近似值，数个步长后更新$\hat{\varphi} = \varphi$。参数φ根据上述损失函数随机梯度下降进行更新，由下式得到，

$$\varphi \leftarrow \varphi - l_q(Q_{\text{tar}} - Q(s, a; \varphi))\nabla_\varphi Q(s, a; \varphi),\tag{22}$$

式中：l_q为Q网络的学习率。

3　实验结果与分析

提出了基于好奇心驱动DQN算法的无人机最小化过时信息年龄路径规划的结果，并与DQN算法进行比较。模拟过程基于Python实现算法代码。

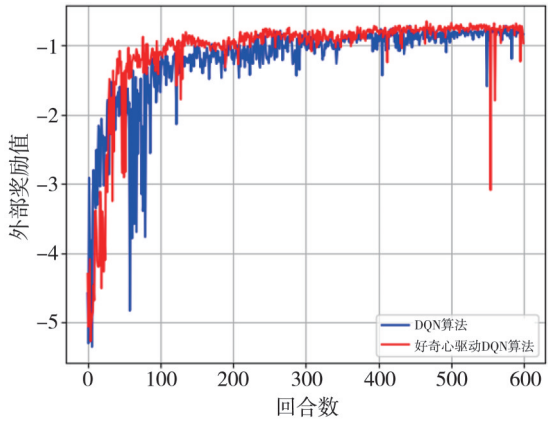
考虑一个城市场景，一架无人机在一片固定区域巡航，无人机从设定的起点出发，在飞行过程中收集尽可能新鲜的IoT设备信息，达到最小过时信息年龄的目标。无人机的飞行区域为[−200 m, 200 m]×[−50 m, 300 m]。无人机的出发点在原点位置，飞行高度设置为 20 m 或 50 m，飞行速度固定为 10 m/s。为了方便无人机收集IoT设备的信息，在无人机所在高度规定以各设备为圆心的通信范围，当无人机在该范围内时即可收集设备数据。无人机的初始电量为2×10^4 J，系统带宽为 100 kHz，载波频率为 5 GHz。本文中考虑所有物联网设备具有相同的权重因子，即$\theta_m = 1/m$。参数如表 2 所示。

表 2 参数设置

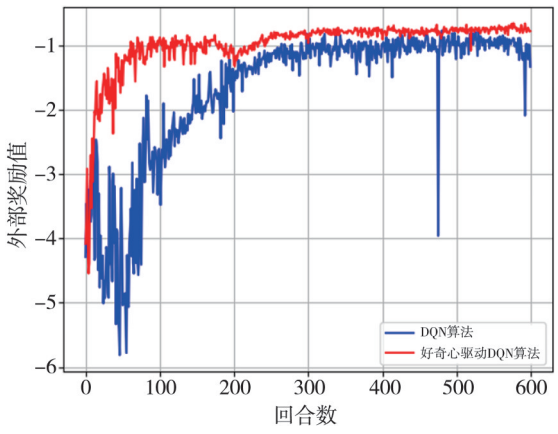
Tab. 2 Parameter settings

参数	描述	值
f_c	载波频率	5 GHz
$(a,b,\eta_{LoS},\eta_{NLoS})$	城市环境的环境参数	(9.61,0.16,1,20)
B	带宽	100 kHz
w_m	发射功率	0.1 W
σ^2	高斯白噪声功率	1.99×10^{-10} mW
(p_m,q_m)	转移概率	(0.2,0.3)

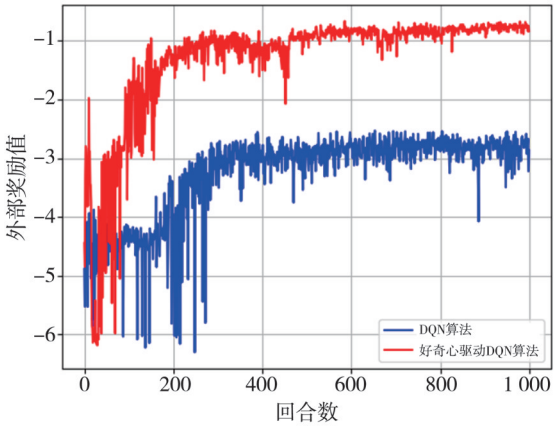
图 4(a)、图 4(b)和图 4(c)分别展示的是无人机高度设置为 20 m 和 50 m 以及通信范围半径为 10 m 和 20 m 时，好奇心驱动 DQN 算法和 DQN 基准算法的收敛性比较，其中 $l_{cur}=0.001$，$l_q=0.001$，$\gamma=0.95$。从图中可以看出，本文使用的好奇心 DQN 算法在 3 种不同的场景下都可以实现稳定的收敛效果，而 DQN 算法只有在无人机高度较低，通信范围较大时才能实现效果比较好的收敛，例如高度在 20 m 处，通信范围半径为 20 m 时，其性能和好奇心 DQN 的性能基本相同。图 4(b)显示当高度不变，通信范围缩小时，DQN 算法性能下降。



(a) h=20 m，r=20 m



(b) h=20 m，r=10 m



(c) h=50 m，r=20 m

图 4 不同条件下算法性能图

Fig. 4 Algorithm performance graphs under different conditions

在训练大约 200 个回合时，DQN 算法收敛慢于好奇心驱动 DQN 算法，在收敛后的外部奖励值约是好奇心驱动 DQN 算法的 50%。

从图 4(c)可以看出，在高度上升而通信范围不变时，好奇心驱动 DQN 算法和 DQN 算法都在训练大约 300 个回合后趋于收敛，而与 DQN 算法相比，本文所提出的好奇心驱动 DQN 算法更稳定，没有明显的波动。此外，在达到收敛后，好奇心驱动 DQN 算法奖励值比传统 DQN 算法高约 67%。

图 5 是在高度 50 m，通信范围半径为 20 m 的环境下，采用好奇心驱动 DQN 算法与 DQN 算法的无人机轨迹图。

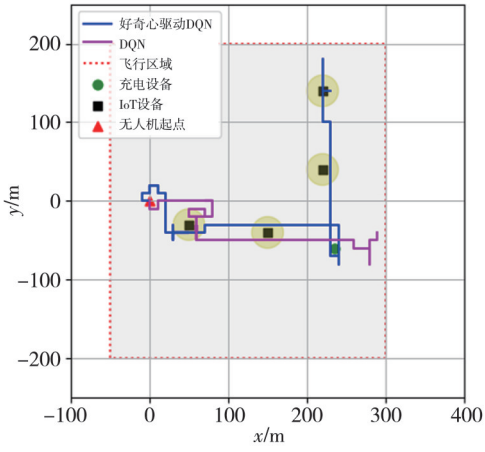


图 5 不同算法下无人机的轨迹图

Fig. 5 Trajectory maps of UAV under different algorithms

从图 5 中可以看出好奇心驱动 DQN 算法明显优于 DQN 算法。首先，在坐标为[50，−30]的 IoT 设备处，采用 DQN 算法的无人机在此处进行了反复兜圈，而好奇心驱动 DQN 算法的无人机不

多做停留，迅速前往下一个节点；其次，采用DQN算法的无人机忽略了两个IoT设备没有采集，而好奇心驱动DQN算法全部采集到了信息；最后，在充电设备右边，采用DQN算法的无人机航行了一些无意义的路径，而好奇心驱动DQN算法全程没有过多的无用路径。

图6是无人机采用不同电量消耗权重k时的剩余电量图。图中横轴表示时间步长，纵轴表示当前剩余电量占初始电量的百分比。可以看出，当k=0.5时，最低剩余电量百分比为22%。随着k的增加，无人机的充电频率也在增加，剩余电量可以维持在一个更高的水平。当k=1时，最低无人机剩余电量最低百分比为27%。当k=1.5时，最低无人机剩余电量最低百分比为30%。但频繁充电可能会造成收集到的IoT设备信息不够新鲜。

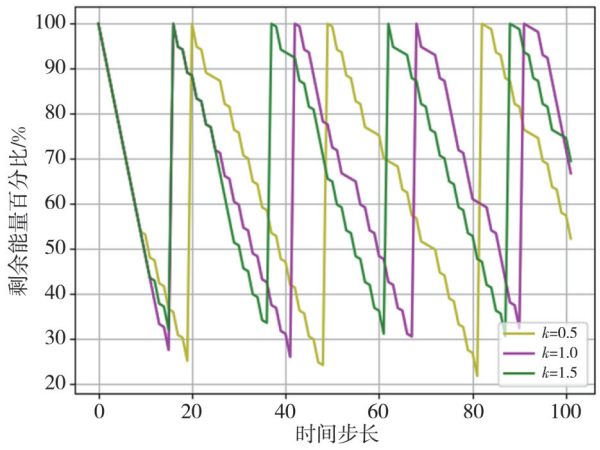


图6　无人机剩余电量变化图

Fig. 6　Change chart of UAV's remaining power

图7展示的是在一个周期内过时信息年龄加权和随时间步长的变化图，在时间步长40以前，无人机处于探索阶段，过时信息年龄持续增长。

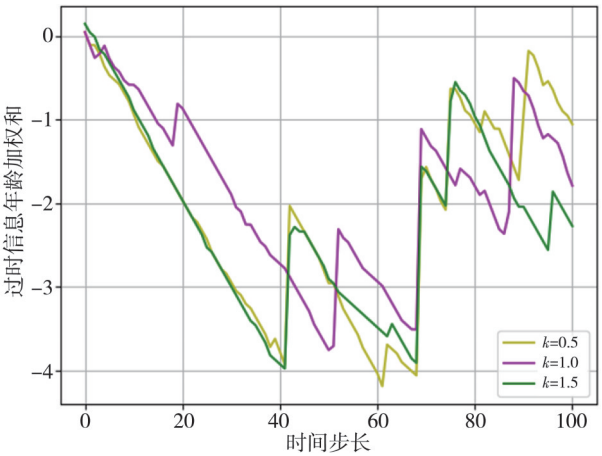


图7　过时信息年龄加权和变化图

Fig. 7　Weighted sum variation chart of AoI

在时间步长40以后，无人机开始训练，并在时间步长60以后开始收敛。当k=0.5时，无人机收集到的信息相对最新鲜，在周期结束时过时信息年龄加权和值为−1，当k=1时，加权和值为−1.8，而当k=1.5时，加权和值为−2.3。随着k值的上升，信息新鲜度逐渐降低。因此，可以根据需求设置k值，达到无人机充电频率与信息新鲜度的平衡。

4　结　语

本文使用过时信息年龄表征信息新鲜度，通过对过时信息年龄加权和与无人机电量消耗联合优化，实现最小化过时信息年龄目标的无人机路径规划。针对传统DQN算法探索性不足的缺点，采用好奇心驱动的DQN算法来处理该优化问题，通过设置内部奖励鼓励无人机进行探索。相比传统DQN算法，在改变通信范围半径时，好奇心驱动的DQN算法收敛速度加快约1.5倍，所得奖励值即过时信息年龄高约20%；在改变无人机高度时，好奇心驱动的DQN算法收敛速度加快约45%，所得奖励值即过时信息年龄高约67%。这些优点表明了该算法相比DQN算法的优势，并且可以使收集到的信息更新鲜。下一步研究将考虑在该系统中部署多架无人机来完成过时信息年龄最小化的目标。

参考文献：

[1] COMPARE M, BARALDI P, ZIO E. Challenges to IoT-enabled predictive maintenance for industry 4.0 [J]. IEEE Internet of Things Journal, 2020, 7(5): 4585-4597.

[2] QI X Y, MEI G, PICCIALLI F. Resilience evaluation of urban bus-subway traffic networks for potential applications in IoT-based smart transportation [J]. IEEE Sensors Journal, 2021, 21(22): 25061-25074.

[3] AHANI G, YUAN D, ZHAO Y X. Age-optimal uav scheduling for data collection with battery recharging [J]. IEEE Communications Letters, 2021, 25(4): 1254-1258.

[4] HU L M, CHEN Z C, JIA Y J, et al. Asymptotically Optimal arrival rate for IoT networks with AoI and peak AoI constraints [J]. IEEE Communications Letters, 2021, 25(12): 3853-3857.

[5] WU T H, LIU J F, LIU J, et al. A novel AI-based framework for AoI-Optimal trajectory planning in

UAV-assisted wireless sensor networks [J]. IEEE Transactions on Wireless Communications, 2022, 21(4): 2462-2475.

[6] HU H M, XIONG K, QU G, et al. AoI-minimal trajectory planning and data collection in UAV-assisted wireless powered IoT networks [J]. IEEE Internet of Things Journal, 2021, 8(2): 1211-1223.

[7] LIU Q Y, LI C Z, HOU Y, et al. AooI: Minimizing age of outdated information to improve freshness in data collection [C]//IEEE INFOCOM 2022-IEEE Conference on Computer Communications, IEEE, 2022: 1359-1368.

[8] ZHU B T, BEDEER E, NGUYEN H, et al. UAV trajectory planning for AoI-minimal data collection in UAV-Aided IoT networks by transformer [J]. IEEE Transactions on Wireless Communications, 2023, 22(2): 1343-1358.

[9] LIU K, ZHENG J. UAV trajectory optimization for time-constrained data collection in UAV-enabled environmental monitoring systems [J]. IEEE Internet of Things Journal, 2022, 9(23): 24300-24314.

[10] LIU J, TONG P, WANG X J, et al. UAV-aided data collection for information freshness in wireless sensor networks [J]. IEEE Transactions on Wireless Communications, 2021, 20(4): 2368-2382.

[11] WU M J, CHI H J, GAN S Y, et al. AoI optimal UAV Trajectory Planning: A Deep Recurrent Reinforcement Learning Approach [C]//2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), IEEE, 2021:1-6.

[12] LI Z M, TONG P, LIU J, et al. Learning-Based data gathering for information freshness in UAV-assisted IoT networks [J]. IEEE Internet of Things Journal, 2023, 10(3): 2557-2573.

[13] 牟治宇, 张煜, 范典, 等. 基于深度强化学习的无人机数据采集和路径规划研究[J]. 物联网学报, 2020, 4(3): 42-51.
MOU Zhiyu, ZHANG Yu, FAN Dian, et al. Research on the UAV-aided data collection and trajectory design based on the deep reinforcement learning [J]. Chinese Journal on Internet of Things, 2020, 4(3): 42-51. (in Chinese)

[14] SUN M Y, XU X D, QIN X Q, et al. AoI-energy-aware UAV-assisted data collection for IoT networks: A deep reinforcement learning method [J]. IEEE Internet of Things Journal, 2021, 8(24): 17275-17289.

[15] PATHAK 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), IEEE, 2017: 488-489.