DOI:10.19651/j.cnki.emt.2417500

面向节点动态优先级的无人机信息收集优化算法*

韩东升[1,2] 郎宇航[1] 黄丽妍[3]
(1.华北电力大学电子与通信工程系 保定 071003;2.华北电力大学河北省电力物联网技术重点实验室 保定 071003;
3.天津市电力公司经济技术研究院 天津 300160)

摘 要:在环境监测等分布式物联网应用场景中,由于节点监测的区域重要性以及收集数据量的不同,节点往往有不同的优先级。节点优先级的动态变化会使无人机频繁更换数据采集的目标节点,造成任务完成时间延长及能量的无端浪费。因此本文针对节点具有动态优先级的分布式物联网应用场景提出了一种基于 DDQN 的无人机任务完成时间与能耗联合优化算法。训练过程中,无人机在任务完成时间、能耗及避免节点数据溢出等约束下学习产生最优策略。仿真结果表明,与最大优先级策略、贪婪策略两种现有策略相比,所提算法任务完成时间分别降低 9.2%、15.1%,能耗分别降低 10%、16.3%;与 DQN 方法相比,所提算法收敛速度更快,训练过程更稳定。

关键词:无人机;动态优先级;信息收集;DDQN
中图分类号:TN929.5 **文献标识码**:A **国家标准学科分类代码**:510.50

UAV information collection optimization algorithm for node dynamic priority

Han Dongsheng[1,2] Lang Yuhang[1] Huang Liyan[3]
(1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China;
2. Hebei Province Electric Power Internet of Things Technology Key Laboratory, North China Electric Power University, Baoding 071003, China; 3. State Grid Tianjin Electric Power Corporation Economic and Technology Research Institute, Tianjin 300160, China)

Abstract: In distributed IoT application scenarios such as environmental monitoring, nodes often have different priorities due to the different regional importance of node monitoring and the amount of data collected. The dynamic change of node priority will make the UAV frequently replace the target node of data acquisition, resulting in prolonged task completion time and unwarranted waste of energy. Therefore, we propose a joint optimization algorithm of UAV task completion time and energy consumption based on DDQN for distributed IoT application scenarios with dynamic priority of nodes. During the training process, the UAV learns the optimal strategy under the constraints of task completion time, energy consumption and avoiding node data overflow. The simulation results show that compared with the maximum priority strategy and greedy strategy, the task completion time of the proposed algorithm is reduced by 9.2% and 15.1% respectively, and the energy consumption is reduced by 10% and 16.3% respectively. Compared with the DQN method, the proposed algorithm converges faster and the training process is more stable.

Keywords: UAV; dynamic priority; information collection; DDQN

0 引 言

过去几年内,物联网(internet of things,IoT)市场规模急速增长,无线传感网络(wireless sensor network,WSN)作为一种无线连接为主的技术在环境监测、智慧交通、工业设备监测等领域大放异彩[1-2]。得益于其低成本特性与高机动性,无人机(unmanned aerial vehicle,UAV)是辅助物联网节点信息收集的理想选择[3-5]。但在自然灾害监测、工业设备报警监测等较为紧急的应用场景下,数据收集的完成时间有着极为严格的限制,这对能量与存储空间受限的无人机来说无异于是一项重大挑战。减少信息收集任务完成时间,降低能量消耗对于环境监测状态下无线传感网络信息收集尤为重要[6-9]。例如,Liu 等[10]研究了无人机辅助信息收集系统中的信息年龄最优收集问题。该文献将节点

收稿日期:2024-12-01
*基金项目:河北省省级科技计划(SZX2020034)项目资助

的信息年龄推导为节点上传时间与无人机飞行时间的加权和，并提出了一种节点关联与轨迹规划策略；Liu 等[11]考虑了数据的时间约束，无人机的能耗约束以及地面用户干扰的约束，通过联合优化无人机与地面用户的关联、无人机的飞行速度与飞行路径，最小化了无人机的任务完成时间。

上述文献通过优化无人机的轨迹及地面节点的调度完成优化目标，但并未考虑节点的优先级问题。在无线传感网络环境监测场景中，受监测区域重要性及数据量收集速率的影响，不同的节点之间有不同的优先级。引入优先级可以使关键节点优先得到服务，缩短任务完成时间，降低系统风险。现有研究大多采用固定优先级的方式。例如 Oubbati 等[12]利用指示信息相关性的参数评估设备的优先级。Johari 等[13]提出了一种两步优先级排序方法。第 1 步将信息分为安全信息、非安全信息、控制信息，并根据重要性分配固定优先级，第 2 步根据车辆类型分配任务时隙。与基于优先级的方向感知媒体访问控制（priority-based direction-aware media access control，PDMAC）等算法相比，该文献所提算法具有更低的信息年龄。

固定优先级的引入改善了优化过程，但也带来了诸多问题，例如固定的优先级会导致优先级较低的设备始终得到较迟的服务。若场景产生突发事件，某些设备亟待得到优先的服务，但固定优先级方案会导致其优先级始终不变，从而产生巨大的时延。在此情况下，动态优先级显然是更为理想的策略。例如，为解决固定优先级场景下低优先级设备无法及时得到服务，从而导致突发事件下，群组时延过高的问题，Chen 等[14]提出了一种优先级动态调整数据包调度（priority dynamic adjustment packet scheduling，PAS）算法。该算法在数据包传输过程中对优先级进行动态调整，确保低优先级的数据包能在预定时间内进行传输。Gao 等[15]引入高斯-马尔可夫移动模型描述节点的动态特征，利用初始优先级、时延参数、维持稳定数据链接的概率赋予数据包不同的优先级，并提出了一种动态优先级调度（dynamic priority scheduling scheme，DPSS）算法。与固定优先级与先入先出方案相比，该算法有更高的传输成功率与更低的传输时延。Fu 等[16]利用高斯-马尔可夫移动模型描述设备的动态过程。用设备的移动、信道变化等方面定义设备的动态优先级并提出了一种对决双深度 Q 网络（dueling double deep Q network，D3QN）算法优化无人机轨迹与数据包的调度。仿真结果表明该算法有比基线方案更高的系统利用率。Yu 等[17]利用初始数据量以及数据采集速度定义节点动态优先级，提出一种多目标深度确定性策略梯度（multi-objective deep deterministic policy gradient，MODDPG）算法，联合优化了总数据收集率、总能量收获以及无人机能量消耗 3 个目标。

上述文献通过不同的角度定义了节点或信息的动态优先级，提出了无人机辅助信息收集的场景下的任务完成时间、能耗优化的方案，并取得了显著的进展。但仍存在两方面的问题：第一，有关设备动态优先级的设置并未均衡初始优先级部分与动态变化部分的影响，导致在任务执行过程中，某一方面的影响更为显著；第二，上述文献中的无人机控制策略大多仅仅以最大优先级为目标。例如，Yu 等[17]在优化过程中，无人机始终飞向动态优先级最大的节点。但在无人机任务执行过程中，优先收集距离远但优先级最大的节点，相较于收集距离近但优先级稍小的节点，会造成任务完成时间与能耗的浪费。此外在节点动态优先级场景下无人机信息收集过程需综合考虑任务完成时间、系统能耗、节点优先级变化等诸多因素。

鉴于以上情况，在节点动态优先级的分布式物联网信息采集场景中，本文综合考虑节点初始优先级、区域重要性、数据采集速率，合理设置节点动态优先级。具体而言，节点监测的区域重要性越重要，节点的优先级越大。对于不同的数据类型，节点的数据采集速率不同。数据采集速率越大，节点的优先级越大。同时将最小化无人机任务完成时间与能耗的优化问题建模为马尔可夫决策过程（markov decision process，MDP），提出了一种基于双深度 Q 网络（double deep Q network，DDQN）的优化算法。该优化算法能够综合考虑节点动态优先级、任务完成时间、系统能耗、节点数据溢出次数等并作出最优的决策。

1　系统模型与问题建立

如图 1 所示，考虑一个分布式物联网环境监测信息收集场景。场景中 K 个传感器节点（sensor node，SN）随机分布在大小为 $L_x \times L_y$ m^2 的监测区域内。所有 SN 用集合 $\mathcal{K} = \{\mathrm{SN}_k, 1 \leqslant k \leqslant K\}$ 表示，SN_k 的坐标表示为 $\boldsymbol{w}_k^s = [x_k^s, y_k^s, 0]^{\mathrm{T}}$。一架旋翼无人机从任务出发点 $\mathrm{U}_{\mathrm{m}}^{\mathrm{start}}$ 出发，以固定的高度 H 飞行，收集所有 SN 存储的数据。令 T 表示 UAV 任务完成时间，在 t 时刻 UAV 的位置可表示为 $\boldsymbol{q}(t) = [x(t), y(t), z(t)]^{\mathrm{T}}$。其中，$z(t) = H$。

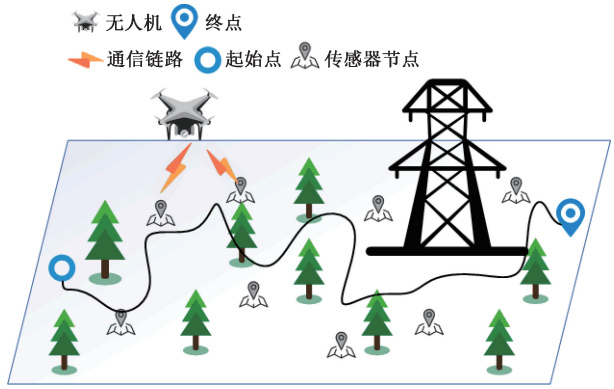

图 1　系统模型

Fig. 1　System model

1.1　通信模型

在无人机辅助信息收集系统中，UAV 与地面节点之间的链路环境较为复杂，且由于 UAV 的高速移动特性，会

出现视距链路(line of sight,LoS)与非视距链路(non-line of sight,NLoS)两种情况。因此,在本文中,UAV 与 SN 之间的信道模型采用概率 LoS 信道模型[18],t 时刻无人机与 SN_k 之间的 LoS 信道概率为:

$$p_k^{LoS}(t)=\frac{1}{1+\mathrm{a}\cdot\exp(-\mathrm{b}(\frac{180}{\pi}\arcsin(\frac{H}{d_k(t)})-\mathrm{a}))} \tag{1}$$

其中,a、b 为由环境决定的常量。$d_k(t)=\sqrt{(x(t)-x_k^s)^2+(y(t)-y_k^s)^2+H^2}$ 为 UAV 与 SN_k 之间的欧几里德距离。NLoS 信道概率为 $p_k^{NLoS}(t)=1-p_k^{LoS}(t)$。$t$ 时刻,SN_k 与 UAV 之间的路径损耗可表示为:

$$PL_k(t)=\begin{cases}(\frac{4\pi f_c d_k(t)}{c})^{\varepsilon}\eta_{LoS},\mathrm{LoS}\\(\frac{4\pi f_c d_k(t)}{c})^{\varepsilon}\eta_{NLoS},\mathrm{NLoS}\end{cases} \tag{2}$$

其中,f_c 为载波频率;c 为真空中的光速;ε 表示路径损耗常量;$\eta_{\mathrm{LoS}},\eta_{\mathrm{NLoS}}$ 分别为在 LoS 链路和 NLoS 链路下的额外路径损耗。

令 $P_{tk}(t)$ 为 SN_k 在 t 时刻的发射功率,σ^2 表示噪声功率,则在 t 时刻,SN_k 与无人机之间的信噪比可表示为:

$$\xi_k(t)=\frac{P_{tk}(t)}{PL_k(t)\sigma^2} \tag{3}$$

在 t 时刻无人机与 SN_k 的信息传输速率可表示为:

$$R_k(t)=B\log_2(1+\xi_k(t)) \tag{4}$$

1.2 节点动态优先级

本文考虑一个固定区域内的环境监测应用场景,令 $Data_k(t)$ 表示 t 时刻 SN_k 存储的数据量,且其数据增长服从泊松分布,即:

$$Data_k(t+\Delta t)=Data_k(t)+\lambda_k(t)\Delta t \tag{5}$$

其中,Δt 为时间更新间隔,$\lambda_k(t)$ 为 SN_k 在 t 时刻的数据收集率,其服从泊松分布。且因不同节点所监测区域不同,其采集环境信息的强度亦有所不同,故本文令不同节点具有不同的数据收集率。且介于节点有限的数据存储空间,节点所能储存的最大数据量有严格的限制,即 $Data_k(t)\leqslant Data_{\max}$。当数据量突破节点所能承担数据存储量的上限时,过去已存储的数据将被新收集的数据所取代,或新收集的数据立即将被放弃。不论哪一种情况都会带来数据信息的损失。因此本文设计了与数据量有关的动态优先级函数以求减少数据溢出情况。

但节点的动态优先级只与数据量有关而忽略了数据增长速度会导致初始数据量较低但数据更新较快的节点无法得到及时的采集,从而导致潜在的数据丢失的风险。综合考虑当前与未来的因素影响,本文设计节点优先级函数为:

$$Q_k(t)=\theta_k\lambda_k(t)\frac{Data_k(t)}{Data_{\max}} \tag{6}$$

其中,θ_k 为与 SN_k 任务开始时与监测区域重要性有关的系数,其数值愈大,表示所监测区域愈重要;$\lambda_k(t)$ 为 SN_k 在 t 时刻的数据收集率。根据节点采集的数据类型不同,如文字类数据、图像类数据、视频类数据等,$\lambda_k(t)$ 分别有不同的取值;$Data_k(t)$ 表示 t 时刻 SN_k 存储的数据量;$Data_{\max}$ 为节点数据最大存储量。节点存储的数据量越接近最大存储量上限则其优先级越大。且当无人机采集完节点数据时,节点的优先级归 0。

1.3 无人机能耗模型

无人机的能量消耗包括有无人机的推进能耗、电路控制能耗以及信号收发相关的通信能耗等。且在任务执行途中,无人机的电路控制能耗与通信能耗远远低于推进能耗,故本文忽略不计。旋翼无人机的推进能耗[19]模型表示式如下:

$$P(V)=P_0\left(1+\frac{3V^2}{U_{tip}^2}\right)+P_i\left(\sqrt{1+\frac{V^4}{4v_0^2}}-\frac{V^2}{2v_0^2}\right)^{1/2}+\frac{1}{2}d_0\rho sAV^3 \tag{7}$$

其中,P_0 为无人机悬停感应功率;U_{tip} 为旋翼转动的叶尖功率;P_i 为无人机悬停的旋翼诱导功率;v_0 为无人机悬停时的旋翼平均速率;d_0 表示无人机飞行过程中的机身阻力;ρ 表示空气密度;s 表示旋翼坚固程度;A 为旋翼盘面积。

且由于无人机加减速在总任务能耗占比较少,本文忽略其造成的影响。同时,为了避免无人机能量耗尽导致任务失败的情况,规定以下约束:

$$E_{UAV}\leqslant E_{\max}^{UAV} \tag{8}$$

1.4 问题建立

本文的目标是在综合考虑节点动态优先级的情况下最小化任务完成时间与无人机能耗。无人机需要对当前环境进行预测,并规划路径。无人机的飞行决策需要综合考虑所有节点的动态优先级、自身能量消耗以及任务完成时间。为避免节点数据溢出,定义二元变量 $\beta(t)$ 为数据溢出变量,若在任务执行过程中有任一节点数据溢出则 $\beta(t)=1$,否则 $\beta(t)=0$。定义 $t_{k,k+1}$ 为无人机收集完成 SN_k 存储的数据并飞到 SN_{k+1} 上方的飞行时间。t_k 为无人机悬停收集 SN_k 存储数据的时间,则 t_k 可表示为:

$$t_k=\frac{Data_k(t)}{R_k(t)} \tag{9}$$

所以无人机总任务完成时间 T 可表示为:

$$T=\sum_{0\leqslant k<K}t_{k,k+1}+t_k \tag{10}$$

无人机的飞行能耗与采集数据时的悬停能耗可统一表示为:

$$E_{UAV}=\int_0^T P(v(t))\mathrm{d}t \tag{11}$$

综上所述本文优化问题可建立为:

$$\min_{q(t),\varphi_k(t)}\{T,\ E_{UAV}\} \tag{12}$$

$$\mathrm{s.t.}\ E_{UAV}\leqslant E_{\max}^{UAV} \tag{13}$$

$$T < T_{\max} \tag{14}$$

$$\boldsymbol{q}(t) \in M \tag{15}$$

$$\varphi_k(t) \in \{0,1\} \tag{16}$$

$$Data_k(t) < Data_{\max} \tag{17}$$

$$\beta(t) = 0 \tag{18}$$

其中，$\boldsymbol{q}(t) = [x(t), y(t), z(t)]^{\mathrm{T}}$ 为无人机在 t 时刻的位置，M 为目标区域。$\varphi_k(t)$ 为 SN_k 的选择接入状态。$\varphi_k(t) = 1$ 表示无人机在 t 时刻正在收集 SN_k 存储的信息。$\varphi_k(t) = 0$ 表示在 t 时刻，无人机未与 SN_k 通信。$Data_k(t)$ 表示 t 时刻 SN_k 存储的数据量。$Data_{\max}$ 为节点数据最大存储量。

2　算法提出

2.1　优化问题的马尔可夫决策过程建模

为解决优化问题式(12)，需要建立相应且合理的马尔可夫决策过程。首先，本文将环境模型离散化：对于面积为 $L_x \times L_y$ m^2 的监测环境，本文将其离散化为边长为 o 的单元格。在该离散环境中，无人机执行信息收集任务时间离散化为长度为 δ_t 的时间步长。在此条件下本文定义智能体的状态空间、动作空间、奖励函数及其演变过程为：

1)状态空间

本文的状态空间主要考虑无人机自身状态、传感器节点状态及任务执行状态 3 个部分。对于无人机自身状态，主要包括无人机的当前位置信息 $\boldsymbol{q}(t)$，无人机能量消耗 $E_{UAV}(t)$。对于传感器节点状态，主要包括节点接入状态 $\varphi_k(t)$；节点存储数据量 $Data_k(t)$；节点当前优先级 $pri_k(t)$；节点信息收集时间 $m_k^c(t)$；节点信息收集完成变量 $\psi_k(t)$。而任务执行状态主要包括无人机执行任务以消耗的时间 $\tau(t)$ 与节点数据溢出计数 $\beta(t)$。综上，步骤 t 时的状态 $\mathcal{S}(t)$ 为：

$$\mathcal{S}(t) = [\boldsymbol{q}(t), E_{UAV}(t), \varphi_k(t), Data_k(t), \psi(t), pri_k(t), m_k^c(t), \tau(t), \beta(t)] \tag{19}$$

式中：

(1) $\boldsymbol{q}(t) = [x(t), y(t), z(t)]^{\mathrm{T}}$ 表示无人机在步骤 t 时的位置，其中 $z(t) = H$。

(2) $E_{UAV}(t)$ 为无人机在步骤 t 时已经累计的能量消耗。在步骤 $t+1$ 时累计能量消耗演变为：

$$E_{UAV}(t+1) = \begin{cases} E_{UAV}(t) + \delta_t P(v(t)), v > 0 \\ E_{UAV}(t) + \delta_t P(0), \text{其他} \end{cases} \tag{20}$$

其中，$P(0)$ 为无人机悬停时的功率。

(3) 二元变量 $\varphi_k(t)$ 为 SN_k 的选择接入状态。$\varphi_k(t) = 1$ 表示无人机在步骤 t 时正在收集 SN_k 存储的信息。$\varphi_k(t) = 0$ 表示在步骤 t 时，无人机未与 SN_k 通信。其演变过程为：

$$\varphi_k(t+1) = \begin{cases} 1, x(t) = x_k^s \cap y(t) = y_k^s \cap \\ Data_k(t) > 0 \cap \psi_k(t) = 0 \\ 0, \text{其他} \end{cases} \tag{21}$$

(4) $Data_k(t)$ 为 SN_k 在步骤 t 时储存的信息量，其会随时间动态变化，演变过程可表示为：

$$Data_k(t+1) = \begin{cases} Data_k(t) + \delta_t \lambda_k(t), \varphi_k(t) = 0 \cap \\ \psi_k(t) = 0 \\ Data_k(t) - \delta_t R_k(t), \varphi_k(t) = 1 \end{cases} \tag{22}$$

其中，$\lambda_k(t)$ 为 SN_k 在步骤 t 时的数据收集率，其服从泊松分布。

(5) 二元变量 $\psi_k(t)$ 表示 SN_k 的数据收集完成状态。$\psi_k(t) = 1$ 表示 SN_k 存储的数据已完成收集。其演变过程为：

$$\psi_k(t+1) = \begin{cases} 1, Data_k(t) = 0 \\ 0, \text{其他} \end{cases} \tag{23}$$

(6) $pri_k(t)$ 为 SN_k 在步骤 t 时的优先级，其演变过程为：

$$pri_k(t+1) = \begin{cases} 0, \varphi_k(t) = 1 \cap Data_k(t) = 0 \\ \theta_k \lambda_k(t) \dfrac{Data_k(t)}{Data_{\max}}, \text{其他} \end{cases} \tag{24}$$

(7) $m_k^c(t)$ 为无人机收集完 SN_k 所需的悬停时间，其演变过程为：

$$m_k^c(t+1) = \begin{cases} m_k^c(t) + \delta_t, \varphi_k(t) = 1 \\ m_k^c(t), \varphi_k(t) = 0 \end{cases} \tag{25}$$

(8) $\tau(t)$ 为无人机累计的任务执行时间，其演变过程为：

$$\tau(t+1) = \begin{cases} \tau(t), \text{任务结束} \\ \tau(t) + \delta_t, \text{其他} \end{cases} \tag{26}$$

(9) 二元变量 $\beta(t)$ 表示 SN_k 在步骤 t 时是否数据溢出的标志，其演变过程为：

$$\beta(t+1) = \begin{cases} 1, Data_k(t) > Data_{\max} \\ 0, \text{其他} \end{cases} \tag{27}$$

2)动作空间

在观察系统环境之后，无人机需要根据其本地策略 π 选择合理的动作，例如在未收集传感器信息时，无人机需快速移动并选择节点，在收集完某个节点信息之后需要离开悬停模式，回归飞行轨迹继续向其他节点进行探索。鉴于在上节中本文对环境空间进行的离散化，本文令在一个飞行动作中无人机的移动距离为 $v\delta_t$。综上，本文将无人机的动作空间 $\mathcal{A}$ 定义为：

$$\mathcal{A} = \left\{ \begin{bmatrix} v\delta_t \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ v\delta_t \\ 0 \end{bmatrix}, \begin{bmatrix} -v\delta_t \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -v\delta_t \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\} \tag{28}$$

其中，$[v\delta_t, 0, 0]^{\mathrm{T}}$、$[0, v\delta_t, 0]^{\mathrm{T}}$、$[-v\delta_t, 0, 0]^{\mathrm{T}}$、$[0, -v\delta_t, 0]^{\mathrm{T}}$ 分别表示无人机向正北、正东、正南、正西移动 $v\delta_t$ 距离。$[0,0,0]^{\mathrm{T}}$ 表示无人机处于悬停状态。

3)奖励函数

本文将从确保任务完成、能量效率、防止出界、平衡优先级等方面综合设置奖励函数。

首先是确保无人机能够完成信息收集工作。本文所考

虑的场景是一个节点分布稀疏,任务完成时间周期较长的场景。若仅在收集完所有节点存储的信息之后给无人机一个较大的奖励,训练过程中会遇到稀疏奖励的问题,即因为训练过程中没有相应的奖励导致算法无法收敛的情况。因此,本文将稀疏奖励转换为密集奖励,确保每进行一次动作无人机都能获得相应的奖励。此部分奖励函数设置如下:

$$r_1(t)=\begin{cases}J_F,\sum_{k=1}^{K}Data_k(t)=0\\ J_C,\sum_{k=1}^{K}\varphi_k(t)=1\\ J_M,\sum_{k=1}^{K}\varphi_k(t)=0\cap v>0\end{cases}\tag{29}$$

其中,J_F 是无人机完成本次收集任务后获得奖励,是一个数值较大的正常数;J_C 是无人机在收集数据过程中获得的奖励;J_M 为无人机在探索过程中靠近传感器节点所获得的趋向性奖励。

其次是有关防止出界,能量效率相关的奖励。为确保无人机在目标区域内执行任务,本文设计奖励有:

$$r_2(t)=\begin{cases}P_D,\boldsymbol{q}(t)\notin M\\ 0,\text{其他}\end{cases}\tag{30}$$

其中,P_D 为一个数值较大的负常数,其含义为当无人机飞出目标区域将会得到一个较大的惩罚。

而针对能量消耗,确保无人机不会电量耗尽,本文设计奖励函数有:

$$r_3(t)=\begin{cases}F_D,E_{UAV}\geqslant E_{\max}^{UAV}\\ F_c,E_{UAV}<E_{\max}^{UAV}\cap\sum_{k=1}^{K}\varphi_k(t)=1\\ F_m,E_{UAV}<E_{\max}^{UAV}\cap\sum_{k=1}^{K}\varphi_k(t)=0\end{cases}\tag{31}$$

其中,F_D 为一个数值较大的负常数,表示在任务执行过程中无人机能量耗尽时,无人机将会受到一个较大的惩罚。而 F_c 表示无人机在数据收集过程中的悬停能量消耗。F_m 则表示无人机在探索寻找节点过程中的飞行能量消耗。

最后是有关平衡节点动态优先级的奖励函数。在本文的应用场景中,传感器节点的数据存储量与优先级都会动态变化,在任务执行的后期,由于已经执行了较长的一段时间,最后一个或几个节点的优先级和数据量都会累积到较高的程度。若直接将节点的优先级作为部分奖励函数,其过高的数值,会导致训练过程极不稳定,影响算法的收敛性。更有甚者,如果节点的优先级奖励大于无人机能量消耗导致的惩罚,甚至会出现无人机悬停不动,等待节点优先级与数据量升高之后再进行数据收集的情况。但优先考虑高优先级的节点又是本文的一个优化目标,在此情况下,本文重新设计了有关动态优先级的奖励函数,利用信息采集时间稀释节点优先级的奖励,以求在收集完优先级无人机仍能得到一个高的奖励值的同时不会影响正常的训练过程。此部分奖励函数的设置为:

$$r_4(t)=\begin{cases}\dfrac{s_k(t)pri_k(t)}{m_k^c(t)},\sum_{k=1}^{K}\varphi_k(t)=1\\ 0,\sum_{k=1}^{K}\varphi_k(t)=0\end{cases}\tag{32}$$

其中,$s_k(t)$ 为与 SN_k 存储数据量有关的系数。

综上所示,本文的奖励函数设置为:

$$r_t(s_t,a_t)=\varepsilon_1r_1(t)+\varepsilon_2r_2(t)+\varepsilon_3r_3(t)+\varepsilon_4r_4(t)\tag{33}$$

其中,ε_1、ε_2、ε_3、ε_4 分别为各部分奖励函数的平衡权重,这些可以根据优化目标的对于任务完成时间与能量消耗的侧重性而设置。

2.2 无人机任务完成时间与能量消耗联合优化算法

传统的强化学习(reinforcement learning, RL)算法利用表或线性函数近似描述 Q 函数。例如 Q-learning 就是使用表描述 Q 函数[20]。其以矩阵的形式建立一种存储每个状态下所有动作 Q 值的表格。作为一种经典的离线策略算法,其训练过程中智能体从 Q 表中选择当前状态下的最佳动作。但面对状态和动作的维度都较高的情况时,利用表格选择动作的算法显然无法胜任。而深度 Q 网络(deep Q network,DQN)利用神经网络拟合函数 Q 函数。神经网络的输入是状态 s 与动作 a ,输出的标量表示在状态 s 下采取动作 a 能获得的价值。Q 训练网络参数 θ^Q 通过最小化 Q 值与目标值的损失函数进行优化:

$$L(\theta^Q)=(Y_t-(Q^{\pi}(s,a)\mid\theta^Q))^2\tag{34}$$

Y_t 为目标函数:

$$Y_t=r(s_t,a_t)+\gamma Q(s_{t+1},a'\mid\theta^Q)\tag{35}$$

在 DQN 算法实现过程中,通过选取可以取得的最大 Q 值的动作 a 更新目标函数,但由于其为离线策略,因此在下一个状态下并不确定会选择行为 a 。因此 DQN 普遍会出现过估计问题即估计值比现实值要大,导致算法将次优动作当作最优动作。为了解决这一问题 DDQN 算法利用两个独立的神经网络估算目标 Q 值。具体而言,DDQN 利用一套神经网络 Q_ν 的输出选取价值最大的动作,但在使用该动作的价值时,用另一套神经网络 Q_{ν^-} 计算该动作的价值[21]。即使其中一套神经网络的某个动作存在较为严重的过估计的问题但由于另一套神经网络的存在,该动作最终使用的 Q 值不会存在很大的过估计问题。而 DQN 算法中本身具有目标网络、训练网络两套神经网络。可以将训练网络作为 DDQN 算法中的第一套神经网络选取动作,目标网络作为第二套神经网络计算 Q 值。相较于 DQN 算法,目标函数变化为:

$$Y_t=r(s_t,a_t)+\gamma Q_{\nu^-}(s_{t+1},\underset{a'}{\operatorname{argmax}}Q_\nu(s_{t+1},a'))\tag{36}$$

本文所提算法框架如图 2 所示。与 DQN 算法相同,本文所采取的基于 DDQN 同样为离线策略算法。本文在训练开始时使用ϵ—贪婪策略平衡探索与利用。在训练初

期无人机更多的采用随机动作探索，在训练后期无人机更倾向于按照最大的 Q 值选取动作。为更好的将 Q-learning 和深度神经网络结合，DDQN 算法采取了经验回放(experience replay)的方法，将每次无人机从环境中采样得到的四元组数据(状态、动作、奖励、下一状态)存储到经验回放缓冲区中，训练 Q 神经网络时再从回放缓冲区中随机采样若干数据进行训练，从而极大地提升样本效率。

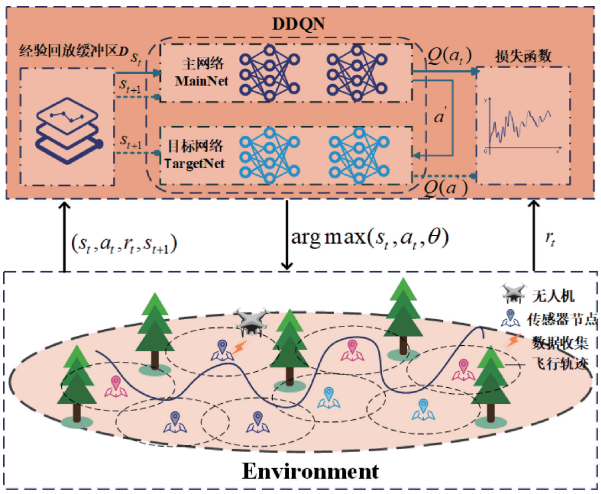

图 2　所提算法框架

Fig. 2　Framework of proposed algorithm

基于 DDQN 的无人机任务完成时间与能量消耗联合优化算法流程为：

算法 1 基于 DDQN 的无人机任务完成时间与能量消耗联合优化算法

初始化 训练集数 E，一次训练集最大迭代次数 N，ϵ 一贪婪策略参数 ϵ，目标网络参数 $\theta_{\nu^-}^{Q}$，训练网络目标参数 θ_{ν}^{Q}，批量梯度下降样本 m，奖励折扣因子 γ，经验回放缓冲区 D

1. for episodes=1 to E
2.　重置环境，初始化状态为 s_1
3.　for t =1 to N
4.　　无人机依据 ϵ 一贪婪策略选取动作，以 ϵ 概率随机选择动作，以 1− ϵ 的概率选择 a_t = argmax$Q(s_t,a)$
5.　　执行动作 a_t 并观察下一个状态 s_{t+1}，获得奖励 r_t
6.　　将 $<s_t,a_t,s_{t+1},r_t>$ 放入经验回放缓冲区 D
7.　　随机从经验回放缓冲区中选取 m 个样本
8.　　依据式(36) 计算目标函数 Y_t
9.　　最小化式(34)更新训练网络参数 θ_{ν}^{Q}
10.　　更新目标网络参数：$\theta_{\nu^-}^{Q}=\theta_{\nu}^{Q}$
11.　　更新 ϵ 一贪婪策略参数 ϵ
12.　　if 无人机飞出目标区域或无人机能量耗尽或无人机任务执行时间超过最大上限或有任一节点数据溢出或 s_{t+1} 为终止状态
13.　　终止循环
14.　end for
15. end for

3　仿真结果及分析

3.1　仿真参数设置

为验证所提算法的有效性，本文设置环境大小为 400 m×400 m。环境内随机分布有 4 个传感器节点，其初始数据量随机在[200，450] Mb 之间，数据收集率 $\lambda_k(t)$ 在{1，2，4，8}中随机选取。$\lambda_k(t)$ 取值从小到大依次代表文字类任务、图像类任务、视频类任务、目标跟踪类任务。且 SN 的重要性随上述任务依次上升。最大存储数据量 $Data_{\max}$ 为 800 Mb。在收集任务开始时无人机从出发点 U_m^{start} 出发，以固定高度 H =10 m 飞行无人机悬停收集传感器节点的数据。

当所有的数据全部被收集，无人机的飞行时间与能量消耗均未超出限制且全程无传感器节点数据溢出则视为任务完成。其他仿真参数设置[18]如表 1 所示。

表 1　部分仿真参数取值

Table 1　Simulation parameter settings

参数	取值	参数	取值
P_0 /W	99.66	P_{tk} /W	0.1
U_{tip} /(m·s^{-1})	120	σ^2 /dBm	−110
P_i /W	120.16	a，b	10，0.6
v_0	4.03	B/MHz	1
d_0	0.6	η_{LoS}/dB	3.7
ρ /(km·m^{-3})	1.22	η_{NLoS} /dB	23
s	0.05	ε	2.3
A/m^2	0.503	f_c/GHz	2

本文所提算法在训练环境参数设置由表 2 所示。在本文的布置中激活函数为线性整流函数，本文所提算法与对比的 DQN 算法参数设置相同。

表 2　训练网络参数

Table 2　Network configurations

参数	取值	含义
E	10 000	训练集数
N	100	最大迭代次数
lr	0.002	学习率
γ	0.98	奖励折扣因子
$\mid D \mid$	10 000	经验回放缓冲区大小
m	128	批量样本大小
ϵ	0.05	ϵ 一贪婪策略参数

3.2　仿真结果及分析

首先验证所提算法的有效性。在无人机飞行速度为

20 m/s 时智能体训练过程中的学习曲线如图 3 所示,为方便观察,本文绘制了以 100 个数据为周期的奖励平均移动曲线。可以看出在训练过程中无人机迅速学习并获得了较高的期望奖励,并在经历了四次波动之后,平稳在较高的水平。在最初的 50 次训练中,智能体累计奖励的水平非常低。这是由于缺乏足够的训练,无人机的动作大多为随机选取。而当经验回放缓冲区填满之后,无人机开始从训练网络中储存积累经验元组。之后无人机在 200 次训练之后累计奖励升高到了一个较高的水平。而奖励函数在 1 000、2 000、3 000 和 4 000 次训练时有 4 次不同程度的波动,与环境中收集 4 个传感器节点的过程相吻合,最终奖励在 195 上下波动。

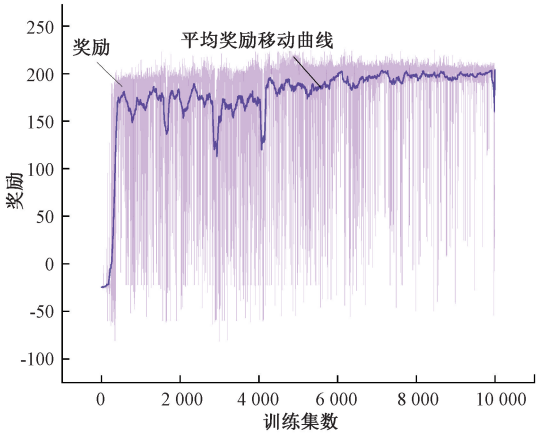

图 3 $v=20$ m/s 时所提算法的累积训练奖励

Fig. 3 Accumulate reward when $v=20$ m/s

本文所提算法与 DQN 算法的训练平均奖励移动曲线对比如图 4 所示。本文所提算法的优势体现在 3 点。其一,相较于 DQN 算法,本文所提算法的收敛速度更快。在 100 次训练之后,本文所提算法就有着较高的奖励积累,而 DQN 算法须训练 500 次之后才有明显的积累奖励的提升。其二,本文所提算法的训练过程更加稳定。所提算法分别在 1 000、2 000、3 000 和 4 000 次经历了 4 次波动,之后便稳定在一个较高的水平。而 DQN 算法分别在 1 000、2 000、3 000、3 500、4 700、6 000、7 000、7 500 等训练次数存在幅度更大的波动,同样也可以看到在 DQN 算法的性能提升后,它仍会持续出现一定程度的震荡,这主要是神经网络过拟合到一些局部经验数据后由 argmax 运算所持续带来的影响。其三,本文所提算法累计奖励更高。在训练过程的后期,本文所提算法的累计奖励平稳在 195 附近,而 DQN 算法则稳定在 170 上下。其主要是在训练过程中,受神经网络过拟合到一些局部经验数据的影响,无人机过高的估计了悬停动作所能够带来的奖励,所以在地面设备已经完成数据收集之后仍然格外进行了悬停动作。具体而言,在无人机飞行速度为 20 m/s 的条件下,所提算法进行了 10 次飞行动作,25 次悬停动作。而 DQN 算法同样进行 10 次飞行动作,却进行了 30 次悬停动作,导致了额外的能量消耗,累计奖励同样受到影响。

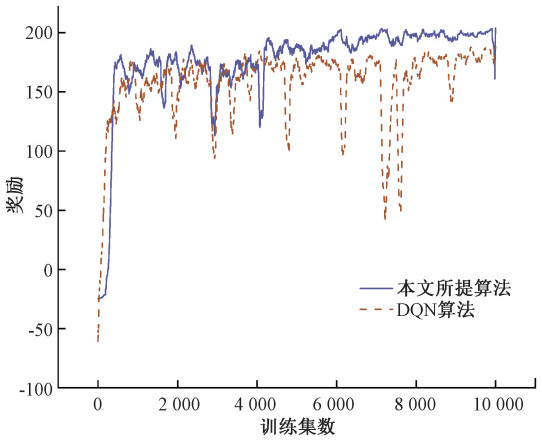

图 4 $v=20$ m/s 时所提算法与 DQN 算法训练奖励移动曲线对比

Fig. 4 Training reward movement curves of different algorithms when $v=20$ m/s

而 DQN 算法训练过程不平稳,存在过估计的问题同样也可以从训练中最大 Q 值变化曲线中直观的观察。如图 5 所示,前 4 000 次训练过程中,DQN 算法存在较为严重的过估计问题,而且 4 000 次训练之后,当 4 700、6 000、7 000、7 500 左右存在最大 Q 值的波动时,算法训练过程中的平均奖励移动曲线也会存在明显的波动变化。

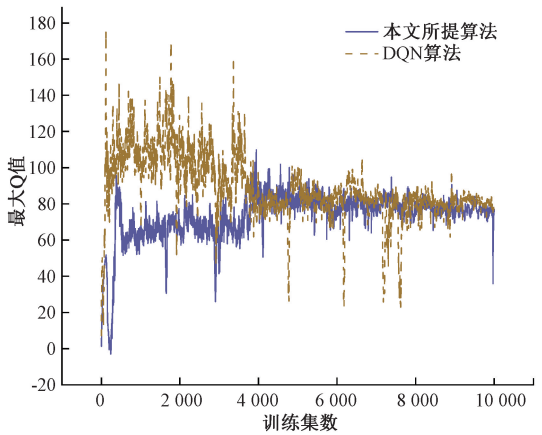

图 5 $v=20$ m/s 时两种算法训练过程最大 Q 值对比

Fig. 5 Maximum Q value curve of different algorithms when $v=20$ m/s

本文所提算法与 DQN 算法在不同飞行速度条件情景下的累计奖励值对比如图 6 所示。可以看出,在不同速率条件下本文所提算法累计奖励值均高于 DQN 算法。在无人机的速率达到 35 m/s 之前,两种算法的累计奖励值均不断增加,且两种算法之间的差距不断缩小。这主要是由于随着无人机飞行速率的上升,无人机的飞行时间变短,传感器节点的缓存的数据量变小,无人机进行动作选择的频率

也放缓。这就导致了两种算法性能差距不断缩小。但在无人机飞行速率为 15 m/s 时，无人机飞行时间延长，节点缓存的数据量变大，无人机动作选择的频率提升。在此情况下本文所提算法性能优势更加明显。而当无人机速率上升至 35 m/s 时，由于无人机飞行能耗的膨胀，其飞行时间缩短不足以抵消无人机飞行功耗过高所导致的负面影响，两种算法的累计奖励均有所回落。但本文所提算法累计奖励值仍高于 DQN 算法。

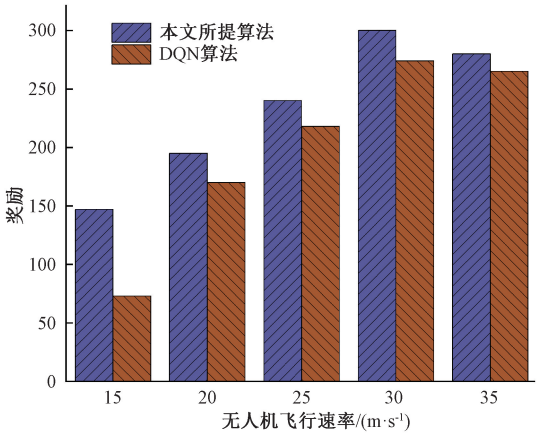

图 6　不同飞行速率条件下两种算法累计奖励值对比

Fig. 6　Comparison of cumulative reward values of two algorithms under different flight rates

为更好的评估所提算法的性能，本文选取两种现有的控制策略进行对比。第 1 种策略为最大优先级策略[17]，无人机始终飞向当前优先级最大的传感器节点，并悬停在其上方收集数据；第 2 种为贪婪策略[22]，无人机始终飞向距离最近的传感器节点，并悬停在其上方收集数据。

3 种不同策略的轨迹如图 7 所示。从中可以看出，3 种策略控制下的轨迹各不相同。在最大优先级策略控制下，无人机首先飞向距离起点最近，优先级最高的设备 2，并在采集完数据之后飞向当前优先级最高的设备 4。而在贪婪策略控制下，无人机始终飞向距离当前距离最近的设备。在本文所提策略的控制下，无人机优先采集当前优先级不是最高，但是信息增长速度更快的设备 1，并重新规划了采集 IoT 设备数据的路径，避免了出现重复采集同一设备的错误。

图 8 对比了在 5 种不同飞行速度下的 3 种策略的任务完成时间。可以看出，在所有飞行速度的条件下，所提算法的任务完成时间均为最小，贪婪策略最大，最大优先级策略次之。且随着飞行速度的不断提升，3 种不同策略的任务完成时间均呈现下降趋势，但三者之间的差距却不断缩小，这是因为随着飞行速度的提升，无人机在不同设备之间移动的时间逐渐变小，则设备数据增长的时间也愈短，无人机采集数据量也越靠近设备的初始数据量。

3 种不同策略的能量消耗随速度变化曲线如图 9 所示。与任务完成时间相同的是，所提算法的能量消耗始终

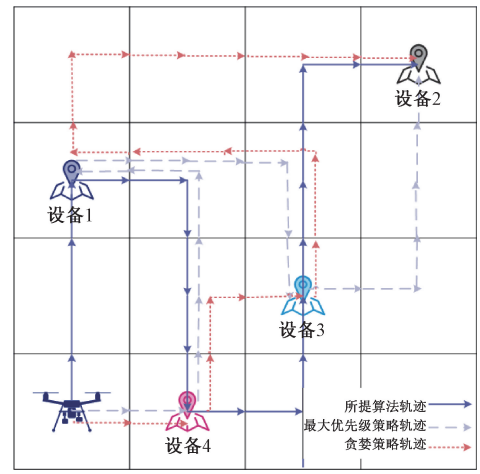

图 7　无人机轨迹

Fig. 7　UAV flight trajectory

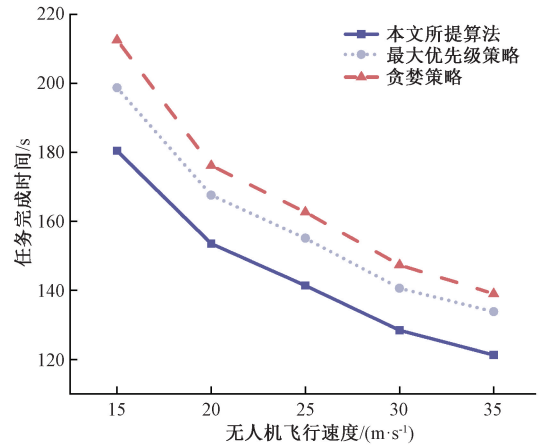

图 8　3 种策略任务完成时间对比

Fig. 8　Task completion time comparison curve of the three strategies

是最小的，贪婪策略最大，最大优先级策略次之。不同的是，3 种策略的能量消耗并未和任务完成时间一样呈现单独的上升或下降趋势，而是一种先下降后上升，再下降再上升的过程。结合无人机飞行推进功率与速度相关图像来看，其主要原因为，当无人机以较低的速度飞行时，飞行功率较低，但在设备与设备之间的飞行时间较长，导致设备的数据增长的较多，而无人机悬停时的功率高于无人机飞行的功率。从而导致了更高的能量消耗。而在 20、25、30 m/s 飞行速度下，无人机因速度升高所少收集的数据量，从而节省下来的悬停能耗大致与飞行速度加快而多消耗的飞行能耗相同，所以能量消耗并未产生明显的波动。如图 10 所示，当飞行速度升高到 35 m/s 时，无人机的飞行功率膨胀到了一个较大的数值，远高于无人机的因飞行速度升高所节省的悬停能耗。所以无人机的能量消耗有着明显的提高。

图 11 为无人机飞行速度为 20 m/s 时不同策略采集数据量对比图。

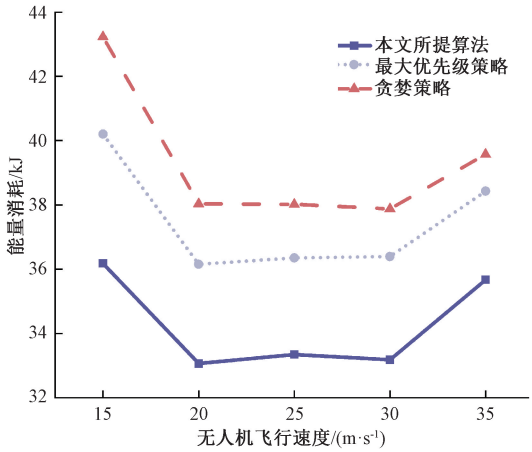

图 9 3 种策略能量消耗对比

Fig. 9 Energy consumption comparison curve of the three strategies

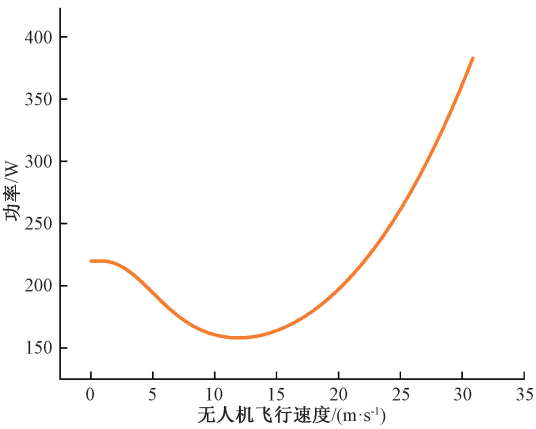

图 10 无人机飞行功率与速度变化曲线

Fig. 10 UAV flight power and speed change curve

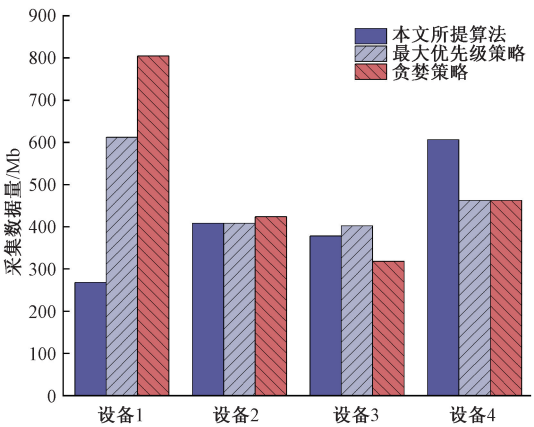

图 11 3 种策略采集数据量对比

Fig. 11 Data collection amount comparison curve of the three strategies

本文定义无人机采集设备存储数据时,该设备的数据量便不再变化,同时遍历完所有节点之后则视为一轮任务结束。本文所提算法相较于贪婪策略和最大优先级策略有着更小的任务完成时间与更低的能量消耗,但在一次的任务完成过程中,对于某一个或者某几个数据增长速度较快的节点,无人机收集其数据较少。

4 结 论

在节点具有动态优先级的无人机辅助信息收集场景中,本文研究了联合优化任务完成时间与能耗问题。考虑到场景的不确定性与动态性,本文提出了一种基于 DDQN 的算法,将无人机决策问题转化为马尔可夫决策过程,在满足无人机能量限制、任务时间限制、数据溢出次数限制的条件下训练无人机的动作以解决节点选择、路径规划问题。并探究了不同飞行速度对于任务完成时间与无人机能耗的影响。实验结果表明,所提算法相较于两种现有策略任务完成时间更短、无人机能耗更低。此外,该算法有比 DQN 算法更快的收敛速度与更高的训练稳定度。

参考文献

[1] LATA S, MEHFUZ S, UROOJ S. Secure and reliable WSN for internet of things: challenges and enabling technologies[J]. IEEE Access, 2021, 9:161103-161128.

[2] KHOSRAVI M R, SAMADI S. BL-ALM: A blind scalable edge-guided reconstruction filter for smart environmental monitoring through green IoMT-UAV networks [J]. IEEE Transactions on Cognitive Communications and Networking, 2021, 5(2):727-736.

[3] XU X B, ZHAO H, YAO H P, et al. A blockchain-enabled energy-efficient data collection system for UAV-assisted IoT [J]. IEEE Internet of Things Journal, 2021, 8(4): 2431-2443.

[4] ZHAN CH, ZENG Y, ZHANG R. Energy-efficient data collection in UAV enabled wireless sensor network[J]. IEEE Wireless Communication Letters, 2018, 7(3): 328-331.

[5] ZHU M Y, WEI ZH Q, QIU CH, et al. Joint data collection and sensor positioning in multi-UAV-assisted wireless sensor network[J]. IEEE Sensors Journal, 2023, 23(19): 23664-23675.

[6] GHORBEL M B, RODRIGUEZ-DUARTE D, GHAZZAI H, et al. Joint position and travel path optimization for energy efficient wireless data gathering using unmanned aerial vehicles[J]. IEEE Transactions on Vehicular Technology, 2019, 68(3): 2165-2175.

[7] YANG SH ZH, DENG Y SH, TANG X X, et al. Energy efficiency optimization for UAV-assisted backscatter communications[J]. IEEE Communications Letters, 2019, 23(11): 2041-2045.

[8] LI M, HE SH SH, LI H. Minimizing mission

completion time of UAVs by jointly optimizing the flight and data collect trajectory in UAV-enabled WSNs[J]. IEEE Internet of Things Journal, 2022, 9(15): 13498-13510.

[9] 孙淑光, 孙涛. 基于融合 A* 算法的无人机路径规划研究[J]. 电子测量技术,2022,45(9):82-91.
SUN SH G, SUN T. Research UAV path planning based on fusion A* algorithm [J]. Electronic Measurement Technology, 2022, 45(9): 82-91.

[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] LIU K, ZHENG J. UAV trajectory planning with interference awareness in UAV-enabled time-constrained data collection systems [J]. IEEE Transactions on Vehicular Technology, 2024, 73(2): 2799-2815.

[12] OUBBATI O S, ATIQUZZAMAN M, LAKAS A, et al. Multi-UAV-enabled AoI-aware WPCN: A multi-agent reinforcement learning strategy[C]. IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops(INFOCOM WKSHPS). IEEE, 2021: 1-6.

[13] JOHARI S, KRISHNA M B. Prioritization for time slot allocation and message transmission in TDMA MAC for VANETs[C]. 2022 IEEE 11th International Conference on Communication Systems and Network Technologies(CSNT). IEEE, 2022: 515-520.

[14] CHEN X, SHI X S, WANG Y Q. Packet scheduling algorithm based on priority adjustment in wireless sensor networks [C]. 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference(ITAIC). IEEE,2022: 1632-1639.

[15] GAO M D, ZHANG B L, WANG L. A dynamic priority packet scheduling scheme for post-disaster UAV-assisted mobile ad hoc network[C]. 2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE,2021: 1-6.

[16] FU X Y, MIAO J S, YAO Y SH, et al. A dynamic priority packet scheduling for UAV assisted AOI-aware network: A deep reinforcement learning approach[C]. 2024 IEEE 99th Vehicular Technology Conference(VTC2024-Spring). IEEE, 2024: 1-5.

[17] YU Y, TANG J, HUANG J Y, et al. Multi-objective optimization for UAV-assisted wireless powered IOT networks based on extended DDPG algorithm [J]. IEEE Transactions on Communications, 2021, 69(9): 6361-6374.

[18] WANG X J, YI M J, LIU J, et al. Cooperative data collection with multiple UAVs for information freshness in the internet of things [J]. IEEE Transactions on Communications, 2023, 71(5): 2740-2755.

[19] OUBBATI O S, LAKAS A, GUIZANI M. Multiagent deep reinforcement learning for wireless-powered UAV networks[J]. IEEE Internet of Things Journal, 2022, 9(17): 16044-16059.

[20] SUTTON R S. Dyna an integrated architecture for learning, planning, and reacting [J]. ACM Sigart Bulletin, 1991, 2(4): 160-163.

[21] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double q-learning [C]. AAAI conference on artificial intelligence,2016.

[22] 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.

作者简介

韩东升,博士,教授,博士生导师,主要研究方向为无线通信新技术、电力系统通信。

E-mail:handongsheng@ncepu.end.cn

郎宇航(通信作者),硕士研究生,主要研究方向为无线通信新技术。

E-mail:nexus_9527@163.com

黄丽妍,硕士,高级工程师,主要研究方向为通信网规划。

E-mail:12979850@qq.com