面向公平性数据采集和能量补充的无人机路径规划算法研究

高思华,李军辉,李建伏*,刘宝煜
(中国民航大学计算机科学与技术学院,天津 300300)

摘 要: 针对无人机(Unmanned Aerial Vehicle,UAV)辅助WSN(Wireless Sensor Networks)数据采集和能量补充工作中存在的数据来源单一和能量补充不均衡现象,本文首先提出数据采集和能量补充公平性问题并进行数学建模. 其次,本文设计一种DPDQN(Double Parametrized Deep Q-Networks)强化学习算法,规划无人机的飞行路线和悬停位置,优化数据采集和能量补充效果. DPDQN学习离散动作与多种连续动作相混合的动作选择策略,算法网络模型包括离散动作网络和连续动作网络两部分. 前者规划无人机访问数据采集节点的顺序,后者优化无人机在数据采集节点周围的悬停位置. 仿真实验结果显示,本文算法在数据采集公平性、能量补充公平性、飞行距离和四种影响公平性的因素比较中均优于三种现有对比算法,并具有良好的鲁棒性和稳定性.

关键词: 公平性数据采集和能量补充;无人机路径规划;深度强化学习;无线传感器网络
基金项目: 国家自然科学基金(No.62173332);中央高校基本科研业务费专项资金(No.3122019118)
中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112(2024)11-3699-12
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230299

Research on UAV Path Planning Algorithm for Fairness Data Collection and Energy Supplement

GAO Si-hua, LI Jun-hui, LI Jian-fu*, LIU Bao-yu
(*School of Computer Science and Technology*, *Civil Aviation University of China*, *Tianjin* 300300, *China*)

Abstract: UAV (Unmanned Aerial Vehicle)-assisted WSN (Wireless Sensor Networks) suffers from single-source data collection and uneven energy supplement. In this article, we first investigate and develop a mathematical model for the problem of fairness for data collection and energy supplement. Then, a novel deep reinforcement learning algorithm, named DPDQN (Double Parametrized Deep Q-Networks), is designed to resolve the proposed problem. The DPDQN algorithm incorporates a hybrid discrete-continuous action strategy, which consists of two components, namely, discrete action network and continuous action network. The former schedules the UAV's visiting order to sensors in WSN, and the latter optimizes the UAV's hover position around each visited sensor. Numerical results demonstrate that the DPDQN algorithm outperforms three existing solutions in data collection fairness, energy replenishment fairness, flying distance, and four factors that influence fairness. Furthermore, the results validate our algorithm is robust and stable.

Key words: fairness data collection and energy supplement; unmanned aerial vehicle path planning; deep reinforcement learning; wireless sensor networks

Foundation Item(s): General Program of National Natural Science Foundation of China (No.62173332); The Fundamental Research Funds for the Central Universities (No.3122019118)

1 引言

无线传感器网络(Wireless Sensor Networks,WSN)[1]以其自组织性强、部署简单等特点,在国防军事、工业生产等方面得到广泛应用[2~4]. 然而,复杂环境下WSN易出现"信息孤岛"[5]和"能量空洞"[6],造成数据传递受阻,节点能量补充困难等问题. 近些年,无人

收稿日期:2023-04-04;修回日期:2024-05-12;责任编辑:王天慧
*通讯作者:李建伏

机(Unmanned Aerial Vehicle,UAV)凭借成本低、速度快、续航长等优势[7~9],辅助WSN解决无线覆盖[10]、数据分发[11]和信息传播[12]等问题,扩展了WSN的应用场景和工作效果.

数据采集和能量补充是无人机辅助WSN工作中最基础、最重要的工作之一.文献[13]中,作者通过拉格朗日乘子法和几何方法选择数据采集目标并优化无人机悬停位置.该算法可有效补充传感器能量,但忽略传感器缓存区的数据量,造成无人机采集的数据量受限.文献[14]通过带宽分配策略增加无人机数据采集的并行性,但忽略了获得低带宽的传感器节点只能通过消耗额外能量提升数据上传功率的事实,影响后续采集任务的进行.文献[15]中,作者使用KKT(Karush Kuhn Tucker)条件分配采集和充能的时间,并通过动态规划方法指导无人机的飞行路线.该算法加快了无人机采集数据的进度,但无人机不能及时为低能量传感器节点进行能量补充,影响了无人机采集的数据量.文献[16]中,作者通过整数规划方法对数据采集和能量补充同时进行场景下的无人机最小飞行时间问题进行建模,并使用遗传算法求解.该方法缩短了无人机悬停时长,但无法平衡无人机悬停阶段用于数据采集和能量补充的时间,造成能量补充效果较差.文献[17]提出一种基于交替优化和黄金分割的线性搜索算法规划无人机飞行路线.该算法增加了被充能范围覆盖的节点数量,但每次悬停无人机只为节点补充部分已损失能量.文献[18]中,作者提出一种基于信誉值的无人机数据采集算法.该算法控制数据传输数量降低无人机群能耗,但通信质量较差的传感器节点因信誉值过低而失去上传数据的机会.

随着人工智能的发展,越来越多的学者将数据采集和能量补充问题抽象为序列决策问题,通过强化学习方法规划无人机的飞行路线.文献[19]中,作者将WSN区域栅格化,以飞行能耗和数据采集量作为奖励值,使用Q-learning算法指导无人机的数据采集路线.然而,受限于Q表的表示能力,无人机只能悬停在少数网格中心采集数据和提供能量补充服务.为求解更为复杂的路径规划问题,强化学习将自身的决策能力与深度学习的感知能力相结合,形成深度强化学习方法[20].文献[21]中,作者通过DQN(Deep Q-learning Network)算法选择无人机的通信节点、通信方式以及飞行速度,减少通信中的丢包率.该算法增加通信质量,但无人机仅与飞行路线附近节点建立通信,忽略了距离较远通信质量较差的节点.文献[22]将无人机的飞行速度和方向离散化处理,通过DQN算法规划其飞行路线,旨在减小各传感器节点缓存区平均数据量和延长网络生命周期.该算法在小规模网络中表现较好,当WSN规模增加后,无人机仅往返于WSN中节点分布密集的区域.文献[23]中,作者使用TD3(Twin-delayed Deep Deterministic)算法优化无人机的飞行速度和飞行方向,最小化采集数据的AOI(Age Of Information).该算法增强了数据的时效性,但无人机会忽略距离基站较远且AOI较大的传感器节点.文献[24]研究发现无人机在数据采集节点通信范围内自由悬停能为更多节点补充能量.基于此想法,作者通过DDPG(Deep Deterministic Policy Gradient)算法指导无人机的飞行路线.该算法虽然增加被充能节点个数,但无法主动为低能量节点补充能量,造成WSN中能量分配不均衡.文献[25]中,作者提出一种强化学习算法MODDPG(Multi-Objective Deep Deterministic Policy Gradient)规划无人机的运动速度和方向.无人机在被采集节点通信范围内悬停,悬停时采集数据并为充能范围内的其他传感器补充能量.该算法令无人机在缓存区内数据较多的节点间飞行以增加数据采集量,但存在无人机飞行能耗增加和数据来源不丰富等问题.

从研究内容上看,现有成果主要以采集数据量和WSN中传感器节点能量补充效果为优化目标,忽略了数据来源的多样性、能量补充的均衡性,以及数据采集和能量补充的内在联系.从解决方法上看,现有的强化学习算法无法合理地规划无人机的飞行路线和悬停位置.Q-learning、DQN等离散动作输出型强化学习算法搜索空间不足,无人机的悬停位置被限制在WSN中某些特定区域.DDPG等连续动作输出型强化学习算法扩大了搜索空间,但无法很好地学习数据采集节点和能量补充节点选择策略.针对上述问题,本文的主要贡献如下:

(1)提出数据采集和能量补充公平性问题,并进行数学建模.采集数据量和数据来源多样性描述数据采集的公平性;能量补充的公平性则通过WSN获得的能量补充和WSN中传感器节点的能量分布共同决定.无人机需要在能量限制条件下规划飞行路线和悬停位置,最大化数据采集和能量补充的公平性.

(2)提出一种基于深度强化学习的DPDQN(Double Parametrized Deep Q-Networks)算法规划无人机的飞行路线和悬停位置.无人机飞行路线和悬停位置由一个离散变量和两个连续变量共同决定,DQN、DDPG等传统强化学习算法并不适用.DPDQN算法首次将PDQN(Parametrized Deep Q-Networks)[26]算法思想引入无人机辅助WSN数据采集和能量补充的工作中并加以改进,实现离散动作与多种连续动作相混合的智能体训练模式.在网络结构上,DPDQN与PDQN均由连续动作网络和离散动作网络组成.不同点在于DPDQN将连续动作网络扩展为双分支结构,共同决定了无人机悬停位置与数据采集节点位

置的相对关系. 同时,连续网络设计最小化无人机所有动作的动作-价值函数之和的倒数为损失函数,用于学习无人机悬停位置的选择策略.

本文结构如下:第一部分介绍无人机辅助无线传感器网络中数据采集和能量补充的研究现状与当前存在的问题;第二部分介绍系统模型,建模数据采集和能量补充公平性问题;第三部分详细介绍DPDQN算法;第四部分通过仿真实验对比DPDQN算法与其他算法在数据采集公平性、能量补充公平性、飞行距离以及其他影响公平性的指标方面的性能,并通过超参数分析实验验证了DPDQN算法的可行性和鲁棒性;最后,第五部分对全文进行总结.

2　系统模型和问题描述

2.1　传感器模型

本文假定在大小为$L\times L$的二维平面区域A内随机部署n个传感器$Q=\{q_1,q_2,\cdots,q_n\}$,$\left(x_q^i,y_q^i\right)$为传感器节点q_i的坐标. q_i装载最大容量为$L_{\max}$的缓冲区用于存储数据,同时配备最大能量为$J_{\max}$的可充电电池以保障正常工作. q_i处于工作状态时,以固定速率v_q^i感知数据,填充缓冲区直至容量上限. q_i可通过配备的唯一天线与连通范围内处于悬停状态的无人机建立一条连通链路,用于将缓冲区数据传递给无人机,或从无人机获取能量补充.

2.2　无人机模型

当无人机处于巡航状态时,在固定高度H,以恒定速度V在WSN上方飞行,此时无法与传感器节点建立连通链路. 处于悬停状态时,无人机选择连通范围内的单个传感器q_i为数据采集节点,与之建立唯一的数据上传链路,并以功率P_d采集传感器q_i缓冲区内的所有数据;与此同时,无人机分别与连通范围内的其他所有传感器建立充能下行链路,并以功率P_c为其补充能量至$J_{\max}$. 完成上述两种服务后,无人机断开所有链路,巡航至下一悬停位置. 在第k次悬停过程中,无人机与最大连通半径D范围内的传感器建立链路,如公式(1)所示:

$$d_i(k)=\sqrt{\left(x_u(k)-x_q^i\right)^2+\left(y_u(k)-y_q^i\right)^2}\leqslant D \quad (1)$$

式中,$\left(x_u(k),y_u(k)\right)$表示无人机的水平悬停位置,$d_i(k)$为传感器$q_i$与无人机的水平距离,$D$为无人机的水平最大连通半径. $Q(k)$表示处于无人机最大连通范围内的传感器节点集合,$Q'(k)$为$Q(k)$中与无人机建立充能下行链路的传感器集合. 无人机与传感器节点建立链路过程如图1所示. 蓝色圆形区域为无人机水平连通范围,黄色虚线和绿色虚线分别表示数据上传链路和充能下行链路. 无人机在第k次悬停中与黄色节点q_1建立数据上传链路并采集数据. 在同一时刻,无人机与绿色节点q_2、q_3和q_4同时建立充能下行链路并提供能量补充服务. 因此,有$Q(k)=\{q_1,q_2,q_3,q_4\}$,$Q'(k)=\{q_2,q_3,q_4\}$. 无人机悬停的时间$t_u^h(k)$为数据上传时间$t_u^d(k)$和充能时间$t_i^c(k)$中的较大者,如公式(2)所示:

$$t_u^h(k)=\max\left\{t_u^d(k),\ \max_{i\in Q'(k)} t_i^c(k)\right\} \quad (2)$$

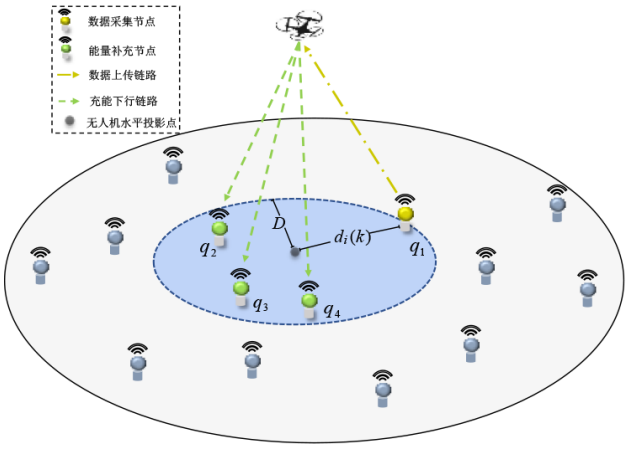


图1　数据采集与充能

2.3　信道模型

信道功率增益决定数据和能量传输的速率和质量. 本文信道模型与文献[25]类似,考虑了信道可视情况和通径损失. 无人机在第k次悬停过程中与传感器q_i建立视距信道的概率为

$$P_i^{\mathrm{LoS}}(k)=\frac{1}{1+a\exp\left[-b\left(\rho_i(k)-a\right)\right]} \quad (3)$$

式中,a、b为视距信道LoS(Line of Sight)和非视距信道NLoS(Non Line of Sight)的环境依赖常数. $\rho_i(k)$为传感器q_i与无人机之间的仰角. 建立非视距信道概率为$P_i^{\mathrm{NLoS}}(k)=1-P_i^{\mathrm{LoS}}(k)$. 两种信道下的通径损失分别是:

$$\mathrm{PL}_i(k)=\begin{cases}\zeta\sqrt{d_i(k)^2+H^2}^{-\alpha}, & \mathrm{LoS}\\ \mu\zeta\sqrt{d_i(k)^2+H^2}^{-\alpha}, & \mathrm{NLoS}\end{cases} \quad (4)$$

式中,ζ为单位距离下的信道功率增益;α,μ分别表示通径损失指数和非视距信道的额外衰减系数. 综上所述,无人机在第k次悬停过程中与传感器q_i建立的上行信道功率增益$g_i(k)$和下行信道功率增益$h_i(k)$为

$$g_i(k)\approx h_i(k)=\left(P_i^{\mathrm{LoS}}(k)+\mu P_i^{\mathrm{NLoS}}(k)\right)\zeta\sqrt{d_i(k)^2+H^2}^{-\alpha} \quad (5)$$

令W和σ分别表示信道的带宽和噪声功率,则q_i与无人机之间的数据传输速率$v_u^i(k)$计算方式如下:

$$v_u^i(k)=W\log_2\left(1+\frac{P_d\left|g_i(k)\right|^2}{\sigma^2}\right) \tag{6}$$

2.4 能耗模型

2.4.1 传感器能耗模型

传感器能耗分为感知能耗和数据传输能耗. 感知能耗与传感器节点的工作时间线性相关,$E_i^s(k)$表示无人机第$k-1$次悬停结束至第k次悬停结束时传感器q_i的感知能耗,计算方法如式(7)所示:

$$E_i^s(k)=\begin{cases}\mathrm{c}\cdot t_u^f(k), & d_i(k)\leqslant D\\ \mathrm{c}\cdot\left(t_u^h(k)+t_u^f(k)\right), & d_i(k)>D\end{cases} \tag{7}$$

其中,c为常数,$t_u^f(k)$为无人机在第k次悬停前最近一次飞行的时长,表示为

$$t_u^f(k)=\frac{\sqrt{\left(x_u(k)-x_u(k-1)\right)^2+\left(y_u(k)-y_u(k-1)\right)^2}}{V} \tag{8}$$

$L_i(k)$表示无人机第k次悬停结束时传感器q_i缓冲区内的数据量:

$$L_i(k)=\begin{cases}0, & q_i\in Q(k)-Q'(k)\\ \min\left\{L_i(k-1)+v_q^i\cdot t_u^f(k),L_{\max}\right\}, & q_i\in Q'(k)\\ \min\left\{L_i(k-1)+v_q^i\cdot\left[t_u^f(k)+t_u^h(k)\right],L_{\max}\right\}, & \text{其他}\end{cases} \tag{9}$$

传感器节点的传输能耗与数据传递量和传输距离相关. 令无人机在第k次悬停过程中与传感器q_i建立数据上传链路,则上行链路保持时间$t_u^d(k)$为

$$t_u^d(k)=\frac{\min\left\{L_i(k-1)+v_q^i\cdot t_u^f(k),L_{\max}\right\}}{v_u^i(k)} \tag{10}$$

所需能耗$E_i^d(k)$计算如式(11)所示:

$$E_i^d(k)=\min\left\{L_i(k-1)+v_q^i\cdot t_u^f(k),L_{\max}\right\}\left(\epsilon_{\mathrm{el}}+\epsilon_{\mathrm{amp}}d_i(k)\right) \tag{11}$$

其中,ϵ_{el}为q_i传输1 bit数据所需最小能耗,ϵ_{amp}为传输能耗随距离增加的额外值.

2.4.2 无人机能耗模型

本文假设无人机的能耗仅发生在巡航、悬停、采集数据和提供能量补充四个过程中,其他忽略不计. 无人机的初始能量为E_u,飞行过程中的牵引功率为$P(V)$[27]. 第k次悬停前无人机最近一次飞行的能耗$E_u^f(k)$计算如下:

$$E_u^f(k)=t_u^f(k)\cdot P(V) \tag{12}$$

无人机悬停时的牵引功率为$P(0)$,能耗计算式如下:

$$E_u^h(k)=t_u^h(k)\cdot P(0) \tag{13}$$

在第k次悬停过程中,无人机维持上行链路的能耗$E_u^d(k)$和维持下行链路的能耗$E_u^c(k)$计算式如下:

$$E_u^d(k)=P_d\cdot t_u^d(k) \tag{14}$$

$$E_u^c(k)=\sum_{i\in Q'(k)}P_c\cdot t_i^c(k) \tag{15}$$

第k次悬停结束后,无人机的剩余能量$E_u(k)$计算式如下:

$$E_u(k)=E_u-\sum_{m=1}^{k}\left[E_u^f(m)+E_u^h(m)+E_u^d(m)+E_u^c(m)\right] \tag{16}$$

2.5 能量补充模型

无人机在悬停过程中为连通范围内的所有能量补充节点同时提供服务,保障每个节点能量充至$J_{\max}$. 由于各链路的信道功率增益不同,无人机为各节点充能的功率存在差异. 假设无人机在第k次悬停过程中与传感器q_i建立充能下行链路,q_i的充能功率为

$$P_i^c(k)=\left|h_i(k)\right|^2P_c \tag{17}$$

q_i通过无人机补充的能量为$E_i^c(k)=J_{\max}-\left(J_i(k-1)-E_i^s(k)\right)$,其中,$J_i(k)$为无人机第$k$次悬停结束时传感器节点$q_i$的剩余能量:

$$J_i(k)=\begin{cases}\max\left\{J_i(k-1)-E_i^s(k)-E_i^d(k),0\right\}, & q_i\in Q(k)-Q'(k)\\ J_{\max}, & q_i\in Q'(k)\\ \max\left\{J_i(k-1)-E_i^s(k),0\right\}, & \text{其他}\end{cases} \tag{18}$$

无人机与q_i之间的下行链路保持时间$t_i^c(k)$为

$$t_i^c(k)=\frac{E_i^c(k)}{P_i^c(k)} \tag{19}$$

2.6 数据采集和能量补充公平性问题

数据采集公平性考虑无人机数据采集总量和数据来源多样性两个因素. 无人机采集数据量越大、来源越广泛,则数据采集公平性越高. 因此,无人机在规划飞行路径时应尽量选择缓冲区内数据多且被采集次数少的传感器节点. 令K为无人机悬停的总次数,q_i为无人机第k次悬停时的数据采集节点,该节点已上传$m_i(k)$次数据. 数据采集公平性F_{data}计算方法如下:

$$F_{\mathrm{data}}=\sum_{k=1}^{K}\frac{L_i(k)}{m_i(k)+1},i\in Q(k)-Q'(k) \tag{20}$$

能量补充公平性考虑无人机的补充能量和获得能量补充的传感器节点数量两个因素. 无人机提供的能量越多,并且获得能量补充的传感器节点数量越多,则

能量补充公平性越高,WSN中传感器节点的能量分布$\mathrm{NUM}_v(k)$也更均衡. 本文通过统计无人机第k次悬停后WSN中剩余能量大于$v\cdot J_{\max}$的传感器节点数量来表示$\mathrm{NUM}_v(k)$,计算方法如下:

$$\begin{aligned}&\mathrm{NUM}_v(k)=\mathrm{COUNT}\left(v\cdot J_{\max}\leqslant J_i(k)\right),\\&i\in\{1,2,\cdots,n\},v\in[0,1]\end{aligned}\tag{21}$$

因此,无人机在悬停时应尽量选择能够服务更多传感器节点,满足更大能量补充需求的位置. 能量补充公平性F_{charge}表示如下:

$$F_{\mathrm{charge}}=\sum_{k=1}^{K}\sum_{i\in Q^t(k)}E_i^c(k)\cdot\mathrm{NUM}_v(k)\tag{22}$$

飞行路线和悬停位置的选择直接影响无人机的数据采集和能量补充公平性. 合理的飞行路线减少无人机飞行能耗,将更多能量用于数据采集和能量补充. 合理的悬停位置兼顾无人机采集数据和传感器节点补充能量,使无人机采集更多数据的同时,延长网络内传感器节点的工作时间,提升传递数据能力. 综上所述,本文目标是在能量约束下规划无人机飞行路线和悬停位置,最大化数据采集和能量补充的公平性,具体描述为

$$\max\left(F_{\mathrm{data}},F_{\mathrm{charge}}\right)\tag{23}$$

$$\text{s.t.}\begin{cases}C1\colon E_u(k)\geqslant 0\\C2\colon K\geqslant 1\\C3\colon 0\leqslant\mathrm{NUM}_v(k)\leqslant n\\C4\colon m_i(k)\in[0,k],i\in\{1,2,\cdots,n\}\\C5\colon J_i(k)\in\left(0,J_{\max}\right],i\in\{1,2,\cdots,n\}\\C6\colon L_i(k)\in\left[0,L_{\max}\right],i\in\{1,2,\cdots,n\}\end{cases}\tag{24}$$

约束条件式(24)中,$C1$保证无人机的能量消耗不高于初始能量;$C2$为无人机的悬停次数限制;$C3$表示满足剩余能量在某一阈值之上的传感器数量不大于WSN中传感器个数;$C4$表示任意传感器节点上传数据次数不大于无人机的悬停次数;$C5$和$C6$表示任意传感器缓存区中的数据量和剩余能量均不溢出.

3　基于DPDQN的路径规划算法

3.1　环境建模

3.1.1　状态

s_k为无人机第k次悬停开始前的环境状态,由无人机信息和WSN的信息组成. 无人机信息包括位置$\left(x_u(k-1),y_u(k-1)\right)$和自身剩余能量$E_u(k-1)$;WSN信息包括网络中所有传感器节点的状态. 任一传感器节点q_i的状态描述为位置$\left(x_q^i,y_q^i\right)$、剩余能量$J_i(k-1)$、上传数据次数$m_i(k-1)$和数据缓存量$L_i(k-1)$. 状态集S包括无人机所有悬停开始前的环境状态,表示如下:

$$\begin{aligned}S=\Big\{s_k|s_k=\Big\{&\left(x_u(k-1),y_u(k-1)\right),\left(x_q^i,y_q^i\right),L_i(k-1),\\&m_i(k-1),J_i(k-1),E_u(k-1)\Big\},i\in\{1,2,\cdots,n\}\Big\}\end{aligned}\tag{25}$$

3.1.2　动作

动作a_k由离散动作和连续动作混合而成,表示无人机的悬停位置. 离散动作i表示数据采集节点q_i,并将无人机悬停位置限制在以q_i为圆心,半径为D的圆形区域内. 连续动作$\left(\delta_i(k),\theta_i(k)\right)$表示无人机与$q_i$的相对位置关系. $\delta_i(k)$为无人机第k次悬停位置与q_i的水平距离,$\theta_i(k)$为无人机悬停位置的水平投影与q_i位置连线的方位角,如图2所示. 无人机的动作空间A如式(26)所示:

$$\begin{aligned}A=\Big\{&a_k|a_k=\left\{i,\left(\delta_i(k),\theta_i(k)\right)\right\},\\&i\in\{1,2,\cdots,n\},\delta_i(k)\in[0,D],\theta_i(k)\in[0,2\pi)\Big\}\end{aligned}\tag{26}$$

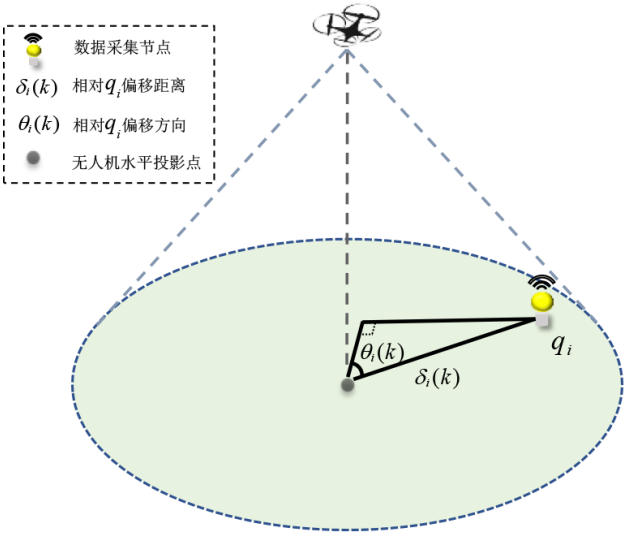


图2　无人机动作表示

3.1.3　奖励

无人机第k次悬停得到的奖励r_k包括公平性数据采集奖励$r_d(k)$、公平性能量补充奖励$r_c(k)$和能耗代价$r_e(k)$. $r_d(k)$对应无人机在传感器节点q_i采集数据的公平性,计算方法如下:

$$r_d(k)=\frac{L_i(k)}{m_i(k)+1}\tag{27}$$

$r_c(k)$对应q_i周围节点补充能量的公平性,计算方法如下:

$$r_c(k)=\sum_{i\in Q^t(k)}E_i^c(k)\cdot\mathrm{NUM}_v(k)\tag{28}$$

式中,$r_e(k)$为无人机进行公平性数据采集和能量补充

的飞行能耗代价 $E_u^f(k)$. 综上所述，r_k 鼓励无人机飞行较少距离，完成更多公平性数据采集和能量补充，计算方法如下：

$$r_k=r_d(k)+r_c(k)-r_e(k) \tag{29}$$

3.2 DPDQN 算法

3.2.1 网络结构

DPDQN 由连续动作网络 $\psi(s_k;\omega)$ 和离散动作网络 $\chi(s_k,\psi(s_k;\omega);\phi)$ 两部分组成. 与传统的 PDQN 不同，DPDQN 中 $\psi(s_k;\omega)$ 采用双分支结构，两个分支共享两层全连接层提取的状态信息，分别输出长度为 n 的序列 $\delta(k)$ 和 $\theta(k)$. $\psi(s_k;\omega)=(\delta(k),\theta(k))$ 记录各候选悬停点与对应传感器节点的相对位置关系. 状态信息 s_k 与 $\psi(s_k;\omega)$ 的拼接结果 $(s_k,\psi(s_k;\omega))$ 作为离散网络的输入. $\chi(s_k,\psi(s_k;\omega);\phi)$ 由两层全连接层组成，选择数据采集节点 q_i，并结合 $\psi(s_k;\omega)$ 输出无人机的混合动作 $a_k=\{i,\psi_i(s_k;\omega)\}$. 其中，$\psi_i(s_k;\omega)=(\delta_i(k),\theta_i(k))$. DPDQN 的网络结构如图 3 所示.

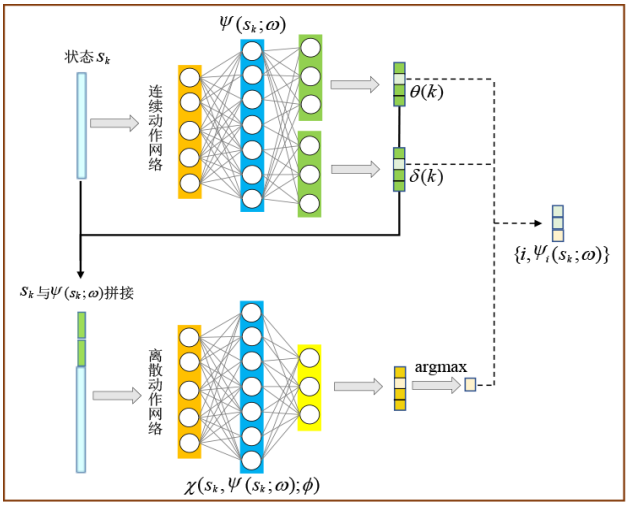


图 3 DPDQN 网络结构

3.2.2 算法执行流程

状态 s_k 输入 DPDQN 网络后，离散网络计算所有动作的动作价值 $Q(s_k,i,\psi_i(s_k;\omega);\phi)$，并根据 ξ-greedy 行为策略选择动作 a_k：

$$a_k=\begin{cases}\arg\max\limits_{i\in\{1,2,\cdots,n\}} Q(s_k,i,\psi_i(s_k;\omega);\phi), & \text{以概率}\xi\\ \text{均匀抽取}A\text{中一个动作}, & \text{以概率}1-\xi\end{cases} \tag{30}$$

无人机执行动作 a_k 后，得到环境反馈的奖励 r_k 并进入下一个状态 s_{k+1}. 与 DQN 和 PDQN 类似，DPDQN 算法收集轨迹 (s_k,a_k,r_k,s_{k+1}) 加入经验池（memory pool），通过经验回放（experience replay）技术加快训练速度；与此同时，DPDQN 算法创建目标连续网络 $\psi'(s_k;\omega')$ 和目标离散网络 $(\chi'(s_k,\psi'(s_k;\omega');\phi')$ 缓解训练出现的高估问题. 训练流程如图 4 所示.

在训练过程中，DPDQN 算法随机选取适量批次的经验，通过最小化损失函数值训练网络. 离散网络的损失函数 $l_k^\chi(\phi)$ 设计如下：

$$l_k^\chi(\phi)=\frac{1}{2}\left[Q(s_k,i,\psi_i(s_k;\omega);\phi)-y_k\right]^2 \tag{31}$$

$$y_k=r_k+\gamma\max_{i\in\{1,2,\cdots,n\}} Q(s_{k+1},i,\psi_i(s_{k+1};\omega');\phi') \tag{32}$$

连续网络的训练目的是在无人机确定数据采集节点的前提下，优化悬停位置获得更多的奖励. 在离散网络参数和状态输入固定时，若所有动作的动作-价值函数之和 $\sum_{i=1}^{n}Q(s_k,i,\psi_i(s_k;\omega);\phi)$ 提升，则表明无人机悬停位置得到了优化. 因此，DPDQN 连续网络损失函数设计如下：

$$l_k^\psi(\omega)=\frac{1}{\sum_{i=1}^{n}Q(s_k,i,\psi_i(s_k;\omega);\phi)} \tag{33}$$

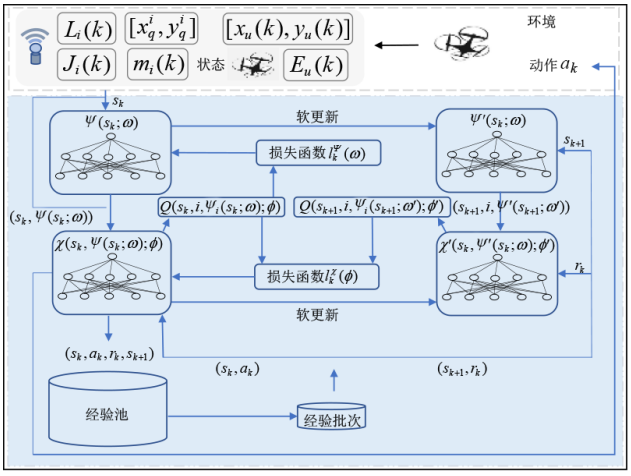


图 4 DPDQN 训练流程

DPDQN 算法使用梯度下降方法按学习率 η 对网络参数进行更新，并每隔一定步数使用软拷贝的方式更新目标网络参数，软拷贝参数为 τ.

DPDQN 算法伪代码如算法 1 所示.

4 仿真实验

4.1 仿真环境

本节通过仿真实验就算法收敛性、数据采集公平性、能量补充公平性、无人机飞行距离以及影响公平性的指标，将 DPDQN 算法与 MODDPG 算法[25]、DQN 算法和具有较强公平性的随机算法（Random）进行对比. 本

算法 1　DPDQN 算法

Input: UAV's energy E, training episodes EP, a probability distribution ξ, minibatch size U, learning rate η, and soft target update parameters τ.

Initialize network weights: ω, ω', ϕ and ϕ'.

FOR i = 0 to EP DO
  $E_u = E$.
  WHILE $E_u > 0$ DO
    Compute continuous action $\psi_i(s_k;\omega)$.
    Select action $a_k = \{i, \psi_i(s_k;\omega)\}$ according to the
    ξ-greedy policy.
    Take action a_k, observe reward r_k and the
    next state s_{k+1}.
    Store transition (s_k, a_k, r_k, s_{k+1}) into Memory pool.
    Simple U transitions $(s_u, a_u, r_u, s_{u+1})_{u\in[U]}$ randomly
    from Memory pool.
    Define the target y_u by $y_u =$

$$\begin{cases} r_u, & \text{if } s_{u+1} \text{ is the terminal state,} \\ r_u + \gamma \max\limits_{i\in\{1,2,\ldots,n\}} Q(s_{u+1}, i, \psi_i(s_{u+1};\omega');\phi'), & \text{otherwise.} \end{cases}$$

    Use data $(y_u, s_u, a_u)_{u\in[U]}$ to compute the stochastic
    gradient $\nabla_\omega l_k^\psi(\omega)$ and $\nabla_\phi l_k^\chi(\phi)$.
    Update the weights by $\omega = \omega - \eta\nabla_\omega l_k^\psi(\omega)$ and
    $\phi = \phi - \eta\nabla_\phi l_k^\chi(\phi)$.
    $E_u = E_u - (E_u^f(k) + E_u^h(k) + E_u^d(k) + E_u^c(k))$.
  END
  Update the target networks by $\omega' = \tau\omega + (1-\tau)\omega'$ and
  $\phi' = \tau\phi + (1-\tau)\phi'$.
END

文在正方形区域 $A \in \{200\times200, 300\times300, 400\times400\}$ m^2 中部署传感器节点数量为 $n \in \{50\ 100, 150\ 200\}$. 无人机从 A 的中心起飞,进行数据采集和能量补充工作. 仿真实验涉及的参数参照文献[13]和文献[15],如表 1 所示.

DPDQN 的网络框架使用 tensorflow2.0 搭建. 连续动作网络中两层共享全连接层神经元个数分别为 256 和 128,激活函数为 ReLU. 两个分支输出层神经元个数为 n. 离散动作网络两个隐藏层中神经元个数分别为 256 和 128,输出层神经元个数为 n. DPDQN 网络涉及的参数如表 2 所示.

4.2　收敛性验证

图 5 展示了 DPDQN 算法、MODDPG 算法和 DQN 算法在 5 000 轮训练中获得的回合奖励. DPDQN 算法在 1 000 轮左右收敛,无人机同时学习运动策略和悬停位置的选择策略,获得的累计奖励最高. DQN 算法中无人机仅学习运动策略,导致获得的累计奖励低于 DPDQN 算法. 相比于以上两种算法,MODDPG 算法中无人机运动策略学习范围限于 WSN 中数据产生速率快的节点间,收敛速度明显加快. 然而,该算法中无人机学习的运动策略对 WSN 的公平性服务不足,获得的奖励明显低于其他两种算法.

4.3　数据采集公平性对比实验

表 3 展示了四种算法的数据采集公平性,实验结果显示各算法的数据采集公平性均与网络规模成正比,与区域范围成反比,且各算法的差异随问题规模的增加逐渐显现. DPDQN 算法能够指导无人机悬停在兼顾数据采集总量和数据来源广泛性的位置,数据采集公平性在各场景下均优于其他三种算法. MODDPG 算法中无人机数据采集的范围仅限于感知数据速率高的传感器节点,其他传感器节点的数据很少被采集. DQN 算法通过平衡各节点的采集次数来增加数据采集的公平性,但受限于能量补充的考虑,无人机可能选择通信半径内数据量较少的节点作为采集对象,影响了数据采集的公平性. Random 算法虽然解决了采集数据范围较小的问题,但忽略采集的数据量和能量补充效果,影响无人机对网络中各节点的采集次数.

4.4　能量补充公平性对比实验

表 4 展示了四种算法的能量补充公平性,实验结果显示各算法的能量补充公平性均与网络规模成正比,与区域范围成反比,且各算法的差异随问题规模的增加逐渐显现. DPDQN 算法在各场景下均优于其他三种算法,该算法指导无人机的飞行路线,优先为能量较少的传感器节点补充能量. 同时,DPDQN 算法优化无人

表 1　仿真参数

参数	取值	参数	取值
传感器数据缓存区大小 $L_{\max}$	10 KB	无人机初始能量 E	10^5 J
传感器初始能量 $J_{\max}$	10 J	带宽 W	1 MHz
传感器感知能耗系数 c	0.001	信道功率 P_c,P_d	40 dBm,−20 dBm
传感器传输 1 bit 最小能耗 ϵ_{el}	50 nJ/bit	LoS 和 NLoS 依赖常数 a, b	10, 0.6
随传输距离增加的额外能耗 ϵ_{amp}	0.1 nJ/(bit·m^2)	噪声功率 σ^2	−90 dBm
无人机飞行高度 H	10 m	单位信道功率增益 ζ	−30 dB
无人机飞行速度 V	15 m/s	通径损失指数 α	2.3
无人机最大连通半径 D	30 m	非视距信道额外衰减系数 μ	0.2

表2 网络参数

参数	取值	参数	取值
训练轮数EP	5 000	学习率η	10^{-4}
探索率ξ	0.9	奖励折扣因子γ	0.9
批次大小U	64	软拷贝参数τ	0.001

机的悬停位置,指导其为连通范围内较多数量的传感器节点补充能量.MODDPG算法获得的能量补充公平性低于其他三种算法,原因在于WSN中较少的节点频繁补充能量.DQN算法可优先为能量较少的节点提供服务,但无法通过调整悬停位置覆盖更多需要补充能量的节点.Random算法虽然均匀地为WSN中的传感器节点进行能量补充,但忽略了各节点剩余能量的差异,无法优先为能量较少的传感器节点补充能量.

4.5 无人机飞行距离对比实验

表5展示了四种算法中无人机的飞行距离,实验结果显示DPDQN算法在各场景下的飞行距离均为最短.DPDQN算法通过优化节点访问次序和悬停位置,实现缩短无人机飞行距离的目的.MODDPG算法的飞行距离分别取决于数据感知速率较快节点的数量和相互间隔.DQN算法中无人机于上行信道质量最优的节点正上方悬停,悬停采集数据时间短,故飞行距离长.Random算法的飞行距离取决于节点间的平均距离.

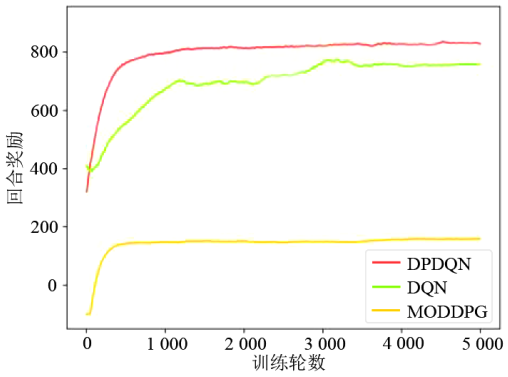


图5 DPDQN奖励收敛效果图

表3 数据采集公平性对比 单位:$\times10^3$

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		181	159	117	177	143	89	156	135	108	153	107	99
100		261	208	158	193	130	121	189	130	109	202	146	114
150		252	209	180	191	149	116	231	156	120	230	181	97
200		260	223	180	227	214	146	218	170	138	252	185	112

表4 能量补充公平性对比 单位:$\times10^3$

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		26	14	12	25	10	4	26	29	8	24	11	5
100		101	66	40	98	39	23	97	50	28	97	41	24
150		228	150	98	220	110	41	223	132	73	226	120	34
200		403	264	162	366	237	92	400	212	121	366	215	76

表5 无人机飞行距离 单位:km

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		7.625	8.621	9.142	8.612	8.846	9.514	9.408	9.878	9.941	8.865	9.121	9.698
100		7.598	8.464	8.951	8.560	8.762	9.325	9.324	9.564	9.917	8.856	9.105	9.610
150		7.445	8.448	8.901	8.487	8.635	9.245	9.322	9.504	9.863	8.625	8.986	9.458
200		7.169	8.347	8.785	8.336	8.601	9.021	9.235	9.463	9.745	8.602	8.712	9.254

4.6 影响数据采集公平性的指标对比实验

4.6.1 数据采集量

表6展示了四种算法的数据采集量,实验结果显示DPDQN算法在各场景下采集的数据量均优于其他三种算法.DPDQN算法中无人机的飞行距离短,使得更多能量可用于增加数据采集次数和均衡WSN中传感器节点的能量分布,两者均有助于增加无人机的数据采集量.MODDPG算法和Random算法均忽略飞行路线对数据采集量的影响,将更多的能量用于无人机的飞行.DQN算法中无人机为兼顾能量补充效果,会被迫选择通信半径内数据量较少的节点作为采集目标,影响数据采集量.

4.6.2 参与数据采集的传感器数量

表7展示了四种算法中参与数据传递的传感器数

量. DPDQN算法在各场景下参与数据采集的传感器节点个数均优于另外三种算法,且在网络规模为50,区域范围不大于300 m×300 m时,网络中所有节点均参与了数据采集. 在MODDPG算法中,无人机仅对产生数据量大的传感器节点进行采集,忽略了网络中其他节点. DQN算法鼓励无人机从访问次数较少的传感器节点采集数据,以增加参与数据采集的节点数量. 然而,该算法对网络中各节点的能量补充的不均衡导致部分节点生命周期较短,影响了参与数据采集的节点数量. Random算法虽然可公平地为网络中传感器节点补充能量,但随机的数据采集策略导致部分节点长时间无法获得数据传递机会.

4.7　影响能量补充公平性的指标对比实验

4.7.1　能量补充量

表8展示了四种算法的能量补充量均与网络规模成正比,与区域范围成反比. DPDQN算法通过优化无人机的悬停位置,为更多剩余能量较低的传感器节点提供能量补充服务,能量补充效果在各场景下均优于其他三种算法. MODDPG算法的能量补充量与其他三种算法的差距随网络规模的增加逐渐变大. 该算法中能够获得能量补充的传感器节点数量有限,且能量补充频率较高导致此类节点每次补充的能量较少. DQN算法指导无人机悬停在数据采集节点正上方,导致获得补充能量的节点仅限数据节点周围,影响了网络中其他节点的能量补充. Random算法均匀地为网络中的传感器节点补充能量,但无人机悬停位置无法根据待充能传感器节点的分布和剩余能量动态调整.

4.7.2　WSN中传感器节点的能量分布

表9~12分别展示了四种算法结束时满足能量阈值$\upsilon \in \{0.2, 0.4, 0.6, 0.8\}$的传感器节点数量,实验结果显示各场景下DPDQN算法均优于其他三种算法. DPDQN算法中公平性能量补充的奖励函数结合充能量和传感器节点能量分布两个因素优化无人机的悬停位置,为更多传感器节点提供能量补充服务的同时,尽可能为能量较少的节点补充能量,使得各传感器节点的剩余能量分布更均衡. MODDPG算法中无人机频繁访问的传感器节点能够保持不低于80%能量,造成其他传感器节点逐渐因能量耗尽而无法感知和传递数据. DQN算法和Random算法的能量补充效果优于MODDPG算法,各传感器节点的剩余能量较为平均.

表6　数据采集量　　单位:KB

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		448.6	375.4	158.4	446.2	333.1	53.6	377.3	292.9	236.3	372.8	251.0	86.1
100		532.6	423.8	328.9	504.2	323.4	287.6	426.2	273.3	226.7	439.4	308.1	230.5
150		502.7	407.5	364.9	511.3	367.4	262.1	469.6	324.2	250.0	489.2	371.8	195.8
200		530.9	477.8	342.3	518.6	470.5	313.4	442.3	353.5	272.4	521.9	360.2	220.8

表7　参与数据采集的传感器数量

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		50	50	40	50	42	29	47	40	37	44	38	34
100		82	69	53	45	31	27	61	47	40	65	48	39
150		84	67	58	41	32	30	75	48	44	71	56	40
200		85	65	54	47	46	38	73	51	47	81	57	47

表8　能量补充量　　单位:J

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		526	335	300	517	268	142	534	345	221	498	285	201
100		1 019	742	521	990	501	349	990	592	374	990	555	359
150		1 527	1 105	829	1 507	891	413	1 501	1 041	653	1 527	939	358
200		2 022	1 456	1 032	1 849	1 360	701	2 017	1 298	859	1 869	1 270	585

4.8　超参数分析

4.8.1　学习率

图6展示了网络规模为100,区域范围为300 m×300 m的环境中,DPDQN算法在学习率$\eta \in \{1\times10^{-2}, 1\times10^{-3}, 1\times10^{-4}, 1\times10^{-5}\}$的奖励收敛情况. 实验结果显示,$\eta = 1\times10^{-2}$时,算法的奖励收敛值低,但收敛速

表 9 WSN 中传感器节点的能量分布 $\upsilon=0.2$

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		50	36	31	47	30	18	50	34	24	48	31	22
100		99	82	58	91	57	39	96	66	43	95	62	39
150		149	120	94	146	99	51	149	112	76	148	106	53
200		200	163	116	196	131	87	197	142	92	196	135	89

表 10 WSN 中传感器节点的能量分布 $\upsilon=0.4$

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		50	33	31	46	27	13	50	33	20	47	30	22
100		98	74	49	90	53	30	93	55	35	93	54	31
150		145	114	81	144	95	40	145	101	62	145	82	45
200		197	155	106	182	117	69	194	132	80	183	119	70

表 11 WSN 中传感器节点的能量分布 $\upsilon=0.6$

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		50	32	28	45	23	12	50	32	19	47	27	20
100		98	65	45	90	51	26	93	53	34	92	53	27
150		144	106	69	142	84	35	143	92	54	144	76	37
200		194	136	87	173	110	54	193	110	70	180	109	60

表 12 WSN 中传感器节点的能量分布 $\upsilon=0.8$

		DPDQN			MODDPG			DQN			Random		
n	L	200	300	400	200	300	400	200	300	400	200	300	400
50		48	30	26	45	20	11	50	32	19	45	25	18
100		96	62	44	90	50	25	93	50	34	90	51	26
150		139	103	66	120	75	31	143	92	54	134	75	32
200		184	130	84	156	99	46	193	110	70	171	100	50

度快；1×10^{-5} 时，算法的奖励收敛值较高，但收敛速度较慢；$\eta=1\times10^{-3}$ 和 $\eta=1\times10^{-4}$ 时，算法能在较短的时间收敛到较优解.

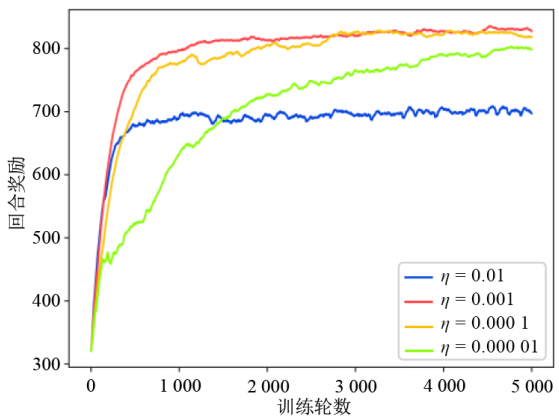


图 6 不同学习率下的奖励收敛情况

4.8.2 奖励折扣因子

图 7 展示了网络规模为 100，区域范围为 300 m×300 m 的环境中，DPDQN 算法在奖励折扣因子

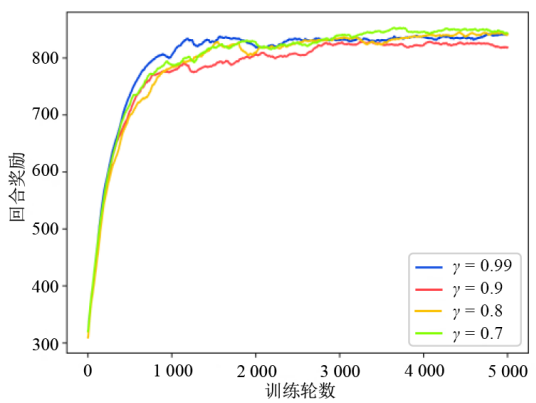


图 7 不同奖励折扣因子下的奖励收敛情况

$\gamma \in \{0.99, 0.9, 0.8, 0.7\}$的奖励收敛情况. 实验结果显示, DPDQN算法在以上几种折扣下均能以较快速度收敛.

4.8.3 软拷贝参数

图8展示了网络规模为100, 区域范围为300 m×300 m的环境中, DPDQN算法在软拷贝参数$\tau \in \{1, 1\times10^{-1}, 1\times10^{-2}, 1\times10^{-3}\}$的奖励收敛情况. 实验结果显示$\tau=1\times10^{-1}$和$\tau=1$时, 网络参数更新幅度过大, 奖励波动剧烈; $\tau=1\times10^{-2}$和$\tau=1\times10^{-3}$时, 奖励平稳收敛.

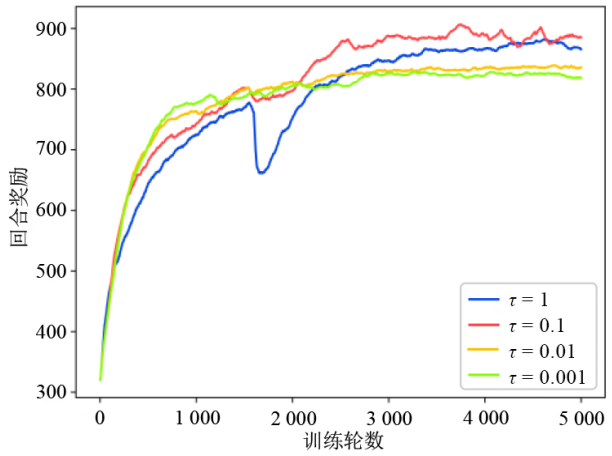


图8　不同软拷贝下的奖励收敛情况

5 结论

针对无人机辅助WSN中数据采集和能量补充公平性问题, 本文提出一种DPDQN算法规划无人机的飞行路线和悬停位置, 在能量约束下最大化数据采集和能量补充的公平性. 公平性数据采集奖励、公平性能量补充奖励和能耗代价用于优化无人机的运动策略. 仿真实验结果显示, DPDQN算法的数据采集公平性、能量补充公平性和飞行距离均优于MODDPG算法、DQN算法和Random算法. 同时, 本文从数据采集量、充能量、上传数据节点数量和节点剩余能量分布情况分析影响数据采集公平性和能量补充公平性的因素, DPDQN算法在以上4个指标的比较中均优于其他两种算法, 有效保证了无人机采集数据和提供能量补充服务的公平性. 最后, 本文通过对比不同学习率、奖励折扣因子和软拷贝参数下的DPDQN收敛效果, 进一步验证了算法的可行性和鲁棒性.

参考文献

[1] AKYILDIZ I F, SU W, SANKARASUBRAMANIAM Y, et al. Wireless sensor networks: A survey[J]. Computer Networks, 2002, 38(4): 393-422.

[2] RAWAT P, SINGH K D, CHAOUCHI H, et al. Wireless sensor networks: A survey on recent developments and potential synergies[J]. The Journal of Supercomputing, 2014, 68: 1-48.

[3] LAI X, JI X, ZHOU X, et al. Energy efficient link-delay aware routing in wireless sensor networks[J]. IEEE Sensors Journal, 2017, 18(2): 837-848.

[4] LI X, LI D, WAN J, et al. A review of industrial wireless networks in the context of Industry 4.0[J]. Wireless networks, 2017, 23: 23-41.

[5] FANG Q, PAN J, CHEN Y, et al. Construction of the supply chain of live streaming e-commerce based on blockchain and internet of things[C]//2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022). Dordrecht: Atlantis Press, 2022: 526-540.

[6] LI J, HAN Q, WANG W. Characteristics analysis and suppression strategy of energy hole in wireless sensor networks[J]. Ad Hoc Networks, 2022, 135: 102938.

[7] CICEK C T, SHEN Z J M, GULTEKIN H, et al. 3-D dynamic UAV base station location problem[J]. INFORMS Journal on Computing, 2021, 33(3): 839-860.

[8] BLISS M, MICHELUSI N. Adaptive scheduling and trajectory design for power-constrained wireless UAV relays [EB/OL]. (2023-02-05)[2023-04-02]. https://arxiv.org/pdf/2007.01228.pdf.

[9] GUO H, LIU J. UAV-enhanced intelligent offloading for Internet of Things at the edge[J]. IEEE Transactions on Industrial Informatics, 2019, 16(4): 2737-2746.

[10] YE Z, WANG K, CHEN Y, et al. Multi-UAV navigation for partially observable communication coverage by graph reinforcement learning[J]. IEEE Transactions on Mobile Computing, 2022.

[11] WANG B, ZHANG R, CHEN C, et al. Graph-based file dispatching protocol with D2D-enhanced UAV-NOMA communications in large-scale networks[J]. IEEE Internet of Things Journal, 2020, 7(9): 8615-8630.

[12] KUMAR S, RATHORE N K, PRAJAPATI M, et al. SF-GoeR: An emergency information dissemination routing in flying ad-hoc network to support healthcare monitoring[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(7): 9343-9353.

[13] BAEK J, HAN S I, HAN Y. Optimal UAV route in wireless charging sensor networks[J]. IEEE Internet of Things Journal, 2019, 7(2): 1327-1335.

[14] QIAN L P, ZHANG H, WANG Q, et al. Joint multi-domain resource allocation and trajectory optimization in UAV-assisted maritime IoT networks[J]. IEEE Internet of

Things Journal, 2022, 10(1): 539-552.

[15] HU H, 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, 2020, 8(2): 1211-1223.

[16] BENMAD I, DRIOUCH E, KARDOUCHI M. Data collection in UAV-assisted wireless sensor networks powered by harvested energy[C]//2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Piscataway: IEEE, 2021: 1351-1356.

[17] LIU Y, XIONG K, LU Y, et al. UAV-aided wireless power transfer and data collection in Rician fading[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(10): 3097-3113.

[18] 黄晓舸, 何勇, 陈前斌, 等. 无人机群辅助的数据采集能耗优化方法[J]. 电子与信息学报, 2023, 45(6): 2054-2062.
HUANG X G, HE Y, CHEN Q B, et al. Optimization method for energy consumption in data acquisition assisted by UAV swarms[J]. Journal of Electronics & Information Technology, 2023, 45(6): 2054-2062. (in Chinese)

[19] FU S, TANG Y, WU Y, et al. Energy-efficient UAV-enabled data collection via wireless charging: A reinforcement learning approach[J]. IEEE Internet of Things Journal, 2021, 8(12): 10209-10219.

[20] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2018, 41(1): 1-27.
LIU Q, ZHAI J W, ZHANG Z Z, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27. (in Chinese)

[21] LI K, NI W, TOVAR E, et al. On-board deep Q-network for UAV-assisted online power transfer and data collection[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 12215-12226.

[22] ZHANG J, YU Y, WANG Z, et al. Trajectory planning of UAV in wireless powered IoT system based on deep reinforcement learning[C]//2020 IEEE/CIC International Conference on Communications in China (ICCC). Piscataway: IEEE, 2020: 645-650.

[23] SUN M, XU X, QIN X, 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.

[24] ZHANG Z, XU C, LI Z, et al. Deep reinforcement learning for aerial data collection in hybrid-powered noma-iot networks[J]. IEEE Internet of Things Journal, 2022, 10(2): 1761-1774.

[25] YU Y, TANG J, HUANG J, 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.

[26] XIONG J, WANG Q, YANG Z, et al. Parametrized deep q-networks learning: Reinforcement learning with discrete-continuous hybrid action space[EB/OL]. (2018-10-10)[2023-04-02]. https://arxiv.org/pdf/1810.06394.pdf.

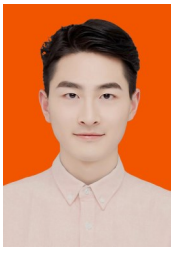
[27] ZENG Y, XU J, ZHANG R. Energy minimization for wireless communication with rotary-wing UAV[J]. IEEE Transactions on Wireless Communications, 2019, 18(4): 2329-2345.

作者简介

高思华 男,1987年出生,河北承德人. 现为中国民航大学讲师、硕士生导师. 主要研究方向为强化学习理论、最优化理论、无线传感器网络和无人机系统.
E-mail: shgao@cauc.edu.cn

李军辉 男,1996年出生,安徽阜阳人. 现为中国民航大学计算机科学与技术学院硕士研究生. 主要研究方向为强化学习理论、无人机路径规划.
E-mail: li2531512787@163.com

李建伏 女,1979年出生,河北沧州人. 现为中国民航大学计算机科学与技术学院副教授、硕士生导师. 主要研究方向为深度学习、机器学习、推荐技术及应用.
E-mail: jfli@cauc.edu.cn



刘宝煜 男,1999年出生,河南漯河人. 现为中国民航大学计算机科学与技术学院硕士研究生. 主要研究方向为强化学习理论、无人机路径规划.
E-mail: by_liu529@163.com