doi: 10.12052/gdutxb.220090

多无人机辅助数据收集系统的智能路径规划算法

苏天赐，何梓楠，崔　苗，张广驰
（广东工业大学　信息工程学院, 广东　广州 510006）

摘要: 无人机具有高度灵活和小巧轻便等优点，已被广泛应用于无线传感器网络的数据收集。本文考虑一个用户随机分布且处于移动状态的无线传感器网络，研究如何规划多个无人机的飞行路径以有效收集网络用户的数据。通过优化多架无人机的飞行路径，使无人机在用户位置无法预测的动态环境中实现数据收集平均吞吐量最大化，同时系统受限于无人机最短飞行时间与范围约束、无人机起点与终点约束、通信距离约束、用户通信约束和无人机防碰撞约束。使用已有优化决策方法求解该问题的计算复杂度较高，同时难以求得全局最优解。针对这一情况，本文提出一种基于Dueling Double Deep Q-network(Dueling-DDQN) 的深度强化学习算法。该算法采用Dueling架构，增强算法的学习能力，提高训练过程的鲁棒性和收敛速度，同时结合了Double DQN (DDQN) 算法的优势，能有效避免因过大估计Q值而导致获取次优无人机轨迹策略。仿真结果表明，此算法可以高效优化无人机的飞行路径，与已有的基准算法相比，所提算法具有更佳的收敛性和鲁棒性。

关键词: 无人机通信；数据收集；路径规划；深度强化学习
中图分类号: TN929.5　　**文献标志码:** A　　**文章编号:** 1007–7162(2023)04–0077–08

Intelligent Path Planning Algorithm for Multi-UAV-assisted Data Collection Systems

Su Tian-ci, He Zi-nan, Cui Miao, Zhang Guang-chi
(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: With the advantages of high flexibility and lightweight, unmanned aerial vehicles (UAVs) have been widely used in data collection of wireless sensor networks. For a multi-UAV-assisted wireless sensor network with randomly distributed and moved users, how to plan the flight paths of the UAVs to effectively collect data from the users remains a challenging problem. This paper aims to maximize the average throughput of data collection in a dynamic environment where the user's location cannot be predicted by optimizing the flight path of the UAVs, which is subject to the shortest flight time and range constraints of UAVs, the constraints of UAV start and end points, the communication distance constraints, the user communication constraints, and the UAV collision avoidance constraints. The resultant problem can be solved by using existing optimization methods with high complexity, which however is difficult to obtain the globally optimal solution. To address this problem efficiently, this paper proposes a deep reinforcement learning algorithm based on Dueling Double DQN (Dueling-DDQN) . The proposed algorithm adopts the Dueling network architecture, which enhances the learning ability of the algorithm and improves the robustness and convergence speed of tracked in suboptimal solutions due to the over-estimation on the Q value. Simulation results show that the proposed algorithm can efficiently obtain the flight paths of multiple UAVs under all constraints . In particular, our proposed algorithm has encouraging convergence and stability performance in comparison with the existing benchmark algorithms.

Key words: UAV communication; data collection; path planning; deep reinforcement learning

近年来，由于无人机的高机动性、自主性和灵活性，无人机得到了广泛的关注，并已应用在生活中许多领域。例如，通过无人机观察实时路况，从而控制城市交通，以及使用无人机进行地震后的搜救工作[1-2]。

收稿日期： 2022-05-22
基金项目： 广东省科技计划项目(2022A0505050023，2022A0505020008)；广东省海洋经济发展项目(粤自然资合[2023]24号)；广东省特支计划项目(2019TQ05X409)；江西省军民融合北斗通航重点实验室开放基金项目(2022JXRH0004)
作者简介： 苏天赐(1998–)，男，硕士研究生，主要研究方向为无人机辅助无线通信和深度强化学习等
通信作者： 张广驰(1982–)，男，教授，主要研究方向为新一代无线通信技术，E-mail: gczhang@gdut.edu.cn

文献[3]研究通过无人机来实现图像采集以及高分辨率的地形探测。除此之外，由于无人机可以通过安装小型通信设备，作为中继节点为地面用户提供通信服务，或者作为移动基站采集地面无线传感网络的数据，通过优化无人机轨迹来最大化地对地面设备进行数据收集。目前无人机辅助无线通信已经成为研究热点，跨越了多个方向，如缓存网络[4]、无线传感器网络[5]、异构蜂窝网络[6]、大规模多输入多输出[7]、设备到设备通信[8]和灾难通信[9-10]。

对于单无人机辅助数据收集的问题，前人进行了诸多研究。文献[11]考虑了一种无人机对多个地面用户进行数据收集的场景，通过对无人机轨迹和数据采集时间等进行联合优化，最大化最小数据量。文献[12]在无人机支持的无线传感器网络中对节能数据收集进行了研究，通过联合优化无人机轨迹和传感器节点唤醒时间以最小化所有传感器节点的最大能耗，同时确保能可靠地收集数据。文献[13]考虑一个无人机支持的无线传感器网络，其中无人机用于从多个传感器节点收集数据，研究联合优化无人机轨迹和通信调度，最大化所有传感器节点的最小平均数据收集率。已有研究表明，用无人机辅助无线通信对地面用户或传感器节点进行数据收集具有可行性和高效性。

尽管与无人机路径和数据收集相关的优化问题可以通过数学优化方法来解决[11-13]，但是计算复杂度较高，计算量较大。由于强化学习可以提供低复杂度的解决方案，目前被用作一种新型工具来解决以上问题。例如，文献[14]研究了无人机作为一个移动基站服务多个用户时的最优轨迹问题，利用Q-learning算法优化轨迹，最大限度地提高了传输速率。文献[15]研究了无人机数据采集场景下的轨迹优化问题，利用DQN(Deep Q-network)算法通过优化无人机轨迹采集所需数据。事实证明，利用强化学习方法可以大大减少计算量，降低计算复杂度[14-15]。

但是面对用户分布广阔且分散的情况，单无人机则显得力不从心，难以很好地完成数据收集任务，因此有必要研究使用多架无人机辅助数据收集。文献[16]研究了在分布式物联网环境下利用多无人机进行数据收集时的路径规划问题，在飞行时间和避碰约束的条件下，引入了一种基于DDQN(Double DQN)的强化学习方法解决了以上轨迹优化问题，同时最大限度地收集了物联网节点的数据。文献[17]研究通过规划多个无人机的轨迹对物联网节点的数据进行采集，为了在避免碰撞和通信干扰等约束下最小化任务时间，提出了一种三步式方法来求解这个问题，包括基于任务分配的K-means算法以及基于深度强化学习的分布式和集中式轨迹设计方法。

然而，无人机机载能量是有限的，它决定了无人机的飞行时间，而无人机的飞行时间直接影响数据采集的性能。现有的研究大多忽略了实际应用中无人机严格的飞行时间限制。同时，通常假设地面用户为静止状态且环境已知。实际上，用户往往具有移动性，且移动目的无法精确预测。传统的优化决策方法一般用来解决静态规划问题，如旅行商问题[18]。为了解决这些难题，本文使用一种新型深度强化学习算法——Dueling-DDQN来优化多无人机路径，最大化收集数据。强化学习算法相比传统优化决策方法可以更好地解决无模型动态规划问题，同时此算法既结合了DDQN将TD(Temporal-Difference)目标的动作选择和动作评估分别用不同的值函数来实现的优点，又采用了Dueling网络架构，使估算的函数值更准确，很好地解决了传统DQN算法的过估计问题[19]。

本文主要贡献：(1) 本文在吞吐量的获取、无人机轨迹和无人机成功到达终点之间取得了良好的权衡，解决了严格的飞行时间限制下无人机的路径规划问题。(2) 目前关于动态用户的研究较少，在系统模型中本文考虑用户随机分布，同时遵循随机游走模型移动，在实际应用中具有较大的适用范围。(3) 本文考虑无人机起点和终点约束、边界碰撞约束、无人机防碰撞约束、通信距离约束以及用户通信约束，为了使无人机在最短时间内到达终点，同时最大化采集地面用户数据，本文提出了基于Dueling-DDQN的多无人机辅助数据收集系统智能路径规划算法。该算法既结合了DDQN的优点，同时又采用了独特的对抗式架构，具有良好的学习能力、收敛速度以及稳定性。(4) 仿真结果表明，所提算法可以很好地完成本文研究目标，相比两个基准算法具有最优的收敛性和鲁棒性。

1 系统模型

如图1所示，本文考虑一个多无人机辅助数据收集系统，N 架无人机在目标区域内同时服务P个用户。用户能以低于v的速度自由移动，初始位置坐标从目标区域内M个半径为R_{reg}的集群中随机生成，其中每个集群存在K个用户。第m个集群中第k个用户在第t时隙的位置坐标用$X_{m,k}^{t}=(x_{m,k}^{t},y_{m,k}^{t})$表示。

由于尺寸限制，考虑无人机和用户都具有单天线，所有无人机起点和终点位置的水平坐标分别为(X_0,Y_0)，$(X_{\mathrm{f}},Y_{\mathrm{f}})$，飞行高度固定为$H$，能与其覆盖范围内的用户进行通信，其中地面覆盖范围的半径R_{uav}由无人机的飞行高度H和天线发射角度Θ决定，即$R_{\mathrm{uav}}=H\cdot\tan(\Theta)$。在第$t$时隙，无人机$i$的三维坐标用$X_i^t=$

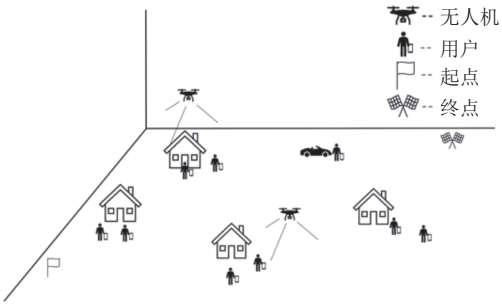

图 1　多无人机辅助数据收集系统模型

Fig.1　Model of a multi-UAV-assisted data collection system

(x_i^t, y_i^t, H)表示，$i = 1,2,\cdots,N$。X_{bound}和Y_{bound}表示目标区域长度和宽度，假设在任意时刻用户和无人机不能离开目标区域，即$0< x_i^t, x_{m,k}^t < X_{\text{bound}}$，$0< y_i^t, y_{m,k}^t < Y_{\text{bound}}$。无人机两两之间的距离为

$$d_{i,j}^t = \sqrt{(x_i^t - x_j^t)^2 + (y_i^t - y_j^t)^2} \tag{1}$$

式中：$j \in \{1,2,\cdots,N\}$，且$i \neq j$，为避免无人机之间发生碰撞，假设$d_{i,j}^t \geqslant L$。

在第t时隙，第m个集群中第k个用户到无人机i的距离为

$$d_{m,k,i}^t = \sqrt{(x_i^t - x_{m,k}^t)^2 + (y_i^t - y_{m,k}^t)^2 + H^2} \tag{2}$$

由于无人机与用户之间的通信信道由视线链路主导，在第t时隙，无人机i与第m个集群中的第k个用户之间的信道遵循自由空间路径损耗模型，可表示为

$$h_{m,k,i}^t = \beta_0 d_{m,k,i}^{t\ -2} = \frac{\beta_0}{(x_i^t - x_{m,k}^t)^2 + (y_i^t - y_{m,k}^t)^2 + H^2} \tag{3}$$

式中：β_0表示信道在参考距离d为1 m时的功率增益。

当用户处于无人机覆盖范围内，即满足通信距离约束时，第m个集群中的第k个用户在第t时隙到第i个无人机的吞吐量定义为

$$R_{m,k,i}^t = B\log_2\left(1 + \frac{p_{m,k}^t h_{m,k,i}^t}{\sum\limits_{a\neq m}^{M}\sum\limits_{b}^{K} p_{a,b}^t h_{a,b}^t + \sum\limits_{c\neq k}^{K} p_{m,c}^t h_{m,c}^t + \alpha^2}\right), \forall m,k \tag{4}$$

式中：$p_{m,k}$为传输功率，B和α^2分别是带宽和噪声功率。在第t时隙，若用户处于多个无人机覆盖范围内，则其吞吐量是与多个无人机通信产生的吞吐量之和：

$$R_{m,k}^t = \sum_{i=1}^{N} R_{m,k,i}^t \tag{5}$$

若仅被无人机i覆盖，其吞吐量为$R_{m,k}^t = R_{m,k,i}^t$。因此，从无人机起飞到第t个时隙，集群m中的第k个用户与无人机通信的总吞吐量为

$$R_{m,k} = \sum_{t=1}^{t} R_{m,k}^t, \forall m,k \tag{6}$$

本文考虑用户通信约束$P(m,k) = \{0,1\}$，$P(m,k)$初始值为0。当总吞吐量$R_{m,k}$大于吞吐量阈值$r_{\min}$时，$P(m,k) = 1$，表示此用户和无人机已完成通信，同时被标记为“已覆盖用户”且在此轮任务中不再被无人机访问。定义用户平均吞吐量为

$$T_{\text{average}} = \frac{\sum_{m}^{M}\sum_{k}^{K} P(m,k) R_{m,k}}{MK}, P(m,k) = \{0,1\} \tag{7}$$

2　问题构建

为了最大化用户平均吞吐量，保证无人机的最短飞行时间，本文在所有无人机的飞行起点和终点约束、边界碰撞约束、无人机防碰撞约束、通信距离约束以及用户通信约束下，研究了多无人机路径规划问题，优化问题可以表述为

$$\max\ T_{\text{average}} \tag{8}$$

$$\text{s.t.}\begin{cases} X_i^{\text{uavf}} = X_{\text{f}} & (9)\\ Y_i^{\text{uavf}} = Y_{\text{f}} & (10)\\ 0 \leqslant X_i(t) \leqslant X_{\text{bound}} & (11)\\ 0 \leqslant Y_i(t) \leqslant Y_{\text{bound}} & (12)\\ d_{i,j}^t \geqslant L & (13)\\ d_{m,k,i}^t \leqslant d_{\text{cons}} & (14)\\ R_{m,k} \geqslant r_{\min} & (15)\\ P(m,k) = \{0,1\} & (16) \end{cases}$$

式中：T_{average}表示本文需要优化的用户平均吞吐量；式(9) 与(10) 为多无人机终点约束，X_i^{uavf}和Y_i^{uavf}分别表示无人机终点位置的横坐标和纵坐标，X_{f}和Y_{f}为终点水平位置的横坐标和纵坐标；式(11) 与(12) 为无人机边界约束，$X_i(t)$和$Y_i(t)$分别表示无人机i在当前水平位置的横坐标和纵坐标，X_{bound}和Y_{bound}为该区域的长度和宽度；式(13) 为无人机防碰撞约束，$d_{i,j}^t$为t时隙无人机i与无人机j之间的距离，L为最小安全距离；式(14) 为用户与无人机的通信距离约束，$d_{\text{cons}} = \sqrt{R_{\text{uav}}^{\ 2} + H^2}$表示用户与无人机通信的最大有效距离；式(15) 表示用户吞吐量$R_{m,k}$需要大于最小阈值$r_{\min}$；式(16) 为用户通信约束，$P(m,k)$初始值为0，当式(15) 满足时，$P(m,k) = 1$，届时此用户将不再被无人机访问。

问题(8)～(16) 是一个非凸的优化问题，使用传统的优化决策方法需要进行复杂的数学分析和数学推导，随着无人机数量的增加，计算成本也可能会增加，且不能保证求得最优解。深度强化学习作为机器学习的一个重要分支，可为复杂系统的感知和决策问题提供低复杂度的解决方案，具有强大的数据处理能力[20-21]。

3 基于深度强化学习的多无人机路径规划

本文提出使用Dueling-DDQN算法设计多无人机轨迹，解决上述优化问题。Dueling-DDQN算法采用Dueling网络架构，同时结合了DDQN算法的优势，有效避免了因过大估计Q值而导致的过优化问题，具有很强的探索性、鲁棒性和收敛性，可应用于复杂任务中搜索可能的最优策略。

Dueling-DDQN算法的框架如图2所示，它主要由Main网络、Target网络、环境和经验缓冲区组成，在第t时隙，智能体通过与环境交互观察到状态s^t，并将状态输入到Main网络，然后根据策略选择动作a^t，得到当前奖励r^t，随后转移至下一个状态s^{t+1}，最后将经验(s^t,a^t,r^t,s^{t+1})存储在经验缓冲区中用于网络的训练。

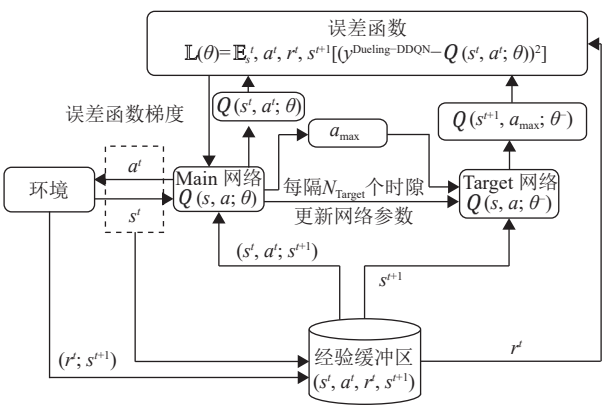

图2 Dueling-DDQN算法框架图

Fig.2 Framework of the Dueling-DDQN algorithm

Dueling-DDQN算法采用双网络结构，包括参数为θ的Main网络$Q(s,a;\theta)$和参数为θ^-的Target网络$Q(s,a;\theta^-)$。每隔N_{Target}个时隙，Target网络通过Main网络更新网络参数，使得Target网络的参数θ^-更新频率较慢，这有助于保持TD目标相对稳定，从而提高训练的收敛性。图3为Main网络(Target网络)的神经网络图，该网络是一个6层全连接层的网络，即一个输入层、4个隐藏层和一个输出层。图中输入层对应无人机位置状态s^t，其神经元个数与状态信息维数相同，输出层输出采取不同动作a^t时获得的Q值，输出层神经元个数与动作空间维数相同。Dueling层使用两个全连接序列层，而非单一序列的全连接层。其中一个序列称为状态价值函数$V_\pi(s)$，包含1个神经元，用于预测状态值；另一个序列称为依赖于状态的动作优势函数$A_\pi(s,a)$，神经元个数与动作空间维数一致，用于预测当前状态下采取不同动作的优势。与单一序列的全连接层相比，Dueling架构的优点是可以在不受动作影响的情况下学习环境的状态价值，使估算的函数值更准确。此外，还可以在不改变底层强化学习算法的情况下，跨动作进行泛化学习。

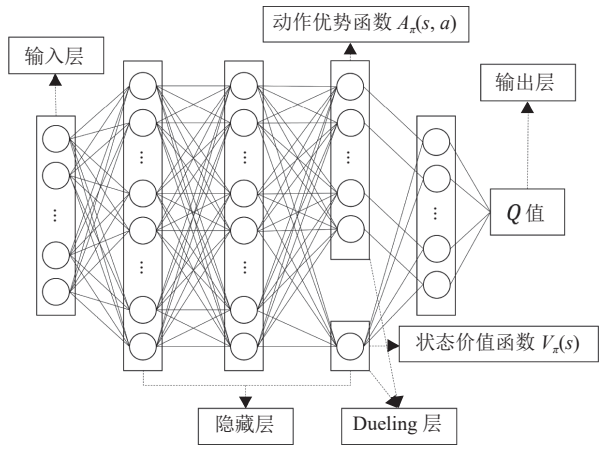

图3 Main网络(Target网络) 神经网络图

Fig.3 Neural network chart of Main network (Target network)

Dueling-DDQN是一种迭代求解贝尔曼方程的无模型强化学习算法，其状态动作价值函数定义为

$$Q_\pi(s^t,a^t)=r^t+\gamma\sum_{s^{t+1}\in S}P^{a^t}_{s^ts^{t+1}}\left[\sum_{a^{t+1}\in A}\pi(a^{t+1}|s^{t+1})Q_\pi(s^{t+1},a^{t+1})\right] \tag{17}$$

式中：$P^{a^t}_{s^ts^{t+1}}$表示智能体在状态s^t采取动作a^t后转移到状态s^{t+1}的概率，$\pi(\cdot)$表示智能体的选择策略。Dueling-DDQN算法先通过Main网络选择Q值最大的一个动作$a_{\max}$作为下一步无人机的移动方向，之后再由Target网络来评估$a_{\max}$：Target网络将$a_{\max}$作为最优动作计算$Q(s^{t+1},a_{\max};\theta^-)$的值，而非根据Target网络的参数$\theta^-$来选择最优动作计算$Q$值，这样做可以避免由于状态值过高而导致的过估计问题。

$$\begin{cases}a_{\max}=\arg\max\limits_a Q\left(s^{t+1},a;\theta\right)\\ y^{\text{Dueling-DDQN}}=r^t+\gamma Q(s^{t+1},a_{\max};\theta^-)\end{cases} \tag{18}$$

同时，此算法把动作价值函数$Q_\pi(s,a)$分为了两个部分，包括状态价值函数$V_\pi(s)$和动作优势函数$A_\pi(s,a)$，有效提高了算法学习效率和鲁棒性。

$$Q_\pi(s,a;\theta,\alpha,\beta)=V_\pi(s;\theta,\alpha)+A_\pi(s,a;\theta,\beta) \tag{19}$$

式中：θ表示$V_\pi(s)$和$A_\pi(s,a)$的互参数，α和β分别为$V_\pi(s)$和$A_\pi(s,a)$的参数。

Dueling-DDQN算法设置了经验缓冲区，将经验(s^t,a^t,r^t,s^{t+1})存储在其中，然后通过小批量随机采样对网络进行更新。这样可以避免连续更新的高度相关数据，有助于减少参数更新时的方差。

为了实现用户平均吞吐量最大化，本文依次对智能体、状态、动作和奖励进行定义。

智能体：多无人机整体充当一个智能体，通过不断与环境进行交互寻找最优策略。

状态s^t：本文把多无人机的位置定义为状态空间，即$S=\{x_1,y_1,H,x_2,y_2,H,\cdots,x_i,y_i,H\}$。在第$t$时隙，智能体的状态为$s^t=\{x_1^t,y_1^t,H,x_2^t,y_2^t,H,\cdots,x_i^t,y_i^t,H\}$。

动作a^t：每一个无人机可以根据策略选择属于动作空间A的一个动作：

$$A=\{\text{left},\text{right},\text{forward},\text{backward}\} \tag{20}$$

式中：left 和 right 分别指无人机向左和向右飞行，forward 和 backward 分别指向前和向后飞行。例如，假设目标区域内存在两个无人机，即$N=2$时有

$$a^t=\{\text{forward},\text{right}\} \tag{21}$$

表示无人机1向前移动，同时无人机2向右移动。

奖励r^t：在深度强化学习中，奖励用于评估智能体在当前状态下采取的动作的好坏。本文设计的奖励功能既依赖于用户平均吞吐量T_{average}，同时也跟无人机整条路径上所收集的其他奖励有关。当所有无人机都到达终点(不要求同时到达)时获得终点奖励R_{final}。

$$R_{\text{final}}=\begin{cases}2\,000, & X_i(t)=X_{\text{f}}\ \text{and}\ Y_i(t)=Y_{\text{f}}\\ 0, & \text{otherwise}\end{cases} \tag{22}$$

无人机飞行轨迹越长，整条路径上获得的步数惩罚R_{sp}越大，若有无人机飞出目标区域，智能体将返回上一个状态且受到边界惩罚R_{bp}。同时，若无人机之间发生碰撞将被给予碰撞惩罚R_{cp}。

因此，在第t时隙，多无人机总奖励r^t表达式为

$$r^t=T_{\text{average}}+R_{\text{sp}}+R_{\text{bp}}+R_{\text{cp}}+R_{\text{final}} \tag{23}$$

当s^{t+1}为终止状态：

$$y^{\text{Duling-DDQN}}=r^t \tag{24}$$

当s^{t+1}不是终止状态：

$$y^{\text{Duling-DDQN}}=r^t+\gamma Q(s^{t+1},\arg\max_a Q(s^{t+1},a;\theta),\theta^-) \tag{25}$$

本文所提的基于Dueling-DDQN的多无人机辅助数据收集系统智能路径规划的完整算法见算法1。

算法1　基于Dueling-DDQN的多无人机辅助数据收集系统智能路径规划算法

1. 初始化：Main网络的参数θ、Target 网络的参数θ^-和经验回放池，随机初始化用户坐标。
2. 对于每一个episode：
3. 　初始化多无人机坐标以及系统状态；
4. 　在当前系统状态s^t下，基于ε-贪婪策略，根据式(20)和式(21) 选择一个动作$a^t\in A$：

$$a^t=\begin{cases}\text{random action}, & \varepsilon\\ \arg\max\limits_a Q(s^t,a), & 1-\varepsilon\end{cases}$$

5. 　执行动作a^t，根据式(4) 计算多无人机覆盖范围内与每个用户的吞吐量，根据式(23) 估计当前奖励r^t；
6. 　获得下一个环境状态s^{t+1}；
7. 　将元组(s^t,a^t,r^t,s^{t+1})存入到经验回放池；
8. 　如果经验回放池R储存经验数量大于容量阈值C_{batch}；
9. 　随机从经验回放池中抽取N_{batch}个样本，计算样本(s^t,a^t,r^t,s^{t+1})的目标值；
10. 　用梯度下降法更新网络参数，使损失最小化：

$$\mathbb{L}(\theta)=\mathbb{E}_{s^t,a^t,r^t,s^{t+1}}[(y^{\text{Dueling-DDQN}}-Q(s^t,a^t;\theta))^2],$$

梯度更新为：

$$\nabla_\theta\mathbb{L}(\theta)=\mathbb{E}_{s^t,a^t,r^t,s^{t+1}}[(y^{\text{Dueling-DDQN}}-Q(s^t,a^t;\theta))\nabla_\theta Q(s^t,a^t;\theta)],$$

11. 　更新状态$s^t=s^{t+1}$；
12. 　每隔一定步数更新Target网络参数$\theta^-=\theta$；
13. 　如果所有无人机均到达终点，进入下一个episode，返回3；如果所有的episode都已经执行完，则训练过程结束；否则返回4。

4　仿真结果

本文设定无人机在1 000 m×1 000 m的目标区域内飞行，即$X_{\text{bound}}=Y_{\text{bound}}=1\,000$ m，同时无人机飞行高度H固定为200 m，无人机的起点和终点位置水平坐标分别为(0,0) 和(1 000,1 000) 。该目标区域内存在5个集群，每个集群中随机分布了10个用户。训练时目标区域被离散为25 m×25 m 的栅格，因此无人机每一步移动距离为40 m。训练阶段的算法迭代次数episode为3 000次，当所有无人机到达终点则当前episode结束，为便于分析，本文在仿真中每10个episode对无人机每步均覆盖率取一次平均。本文设置经验缓冲区R的容量$C_{\max}$为2×10^5，容量阈值C_{batch}为200。本文其他参数如表1所示。

仿真考虑两个场景——“单无人机场景”与“两无人机场景”。在两个场景下将本文所提方案简称为“基于Dueling-DDQN的数据收集方案”。为进行比较，在两个场景中将DQN和DDQN算法辅助数据收集作为对比方案。这两种对比方案的状态、动作和奖励等的定义与“基于Dueling-DDQN的数据收集方案”一致。DDQN和DQN算法同样采用双网络结构，不同之处在于DDQN算法通过解耦目标Q值动作选择和计算目标Q值这两步来消除过度估计的问题，这与本文所提算法一致，但两种基准算法均未采用Dueling网络架构。本文将DDQN和DQN算法分别称为“基于DDQN的数据收集方案”和“基于DQN的数据收集方案”。此外，在“单无人机场景”中增设一个没有优化轨迹的对比方案，采用固定的无人机飞行轨迹。在固定无人机飞行轨迹的对比方案中，取所有用户

表1 仿真参数

Table 1 Simulation parameters

参数	设定值
带宽w/MHz	1
无人机传输功率/W	5
折扣因子γ	0.9
噪声功率α^2/dBm	110
信道功率增益β_0/dB	−50
路径损失指数	2
天线发射角度θ	45°
样本大小N_{batch}	128
Target 网络更新频率N_{Target}	300
学习率	0.01
贪婪率ε	0.04
最小安全距离 L/m	40
终点奖励R_{final}	1 000
边界惩罚R_{bp}	−100
碰撞惩罚R_{cp}	−200
步数惩罚R_{sp}	−1 000

的坐标中心点(X_c, Y_c)，即P个用户的初始水平坐标之和的均值，$X_c = \sum_m^M \sum_k^K x_{m,k}^0 / P$，$Y_c = \sum_m^M \sum_k^K y_{m,k}^t / P$。无人机从起点起飞，径直穿过坐标中心点大致位置，最后到达终点，同时在飞行途中计算满足通信距离约束的用户的吞吐量。将该方案称为“固定轨迹数据收集方案”，与基于强化学习的方案作对比。

由于当用户总吞吐量$R_{m,k}$大于吞吐量阈值$r_{\min}$时，$P(m,k)=1$，此用户被标记为“已覆盖用户”且在此幕中不再与无人机通信，因此，“已覆盖用户”越多则表示已收集平均吞吐量$T_{\text{average}} = \dfrac{\sum_m^M \sum_k^K P(m,k) R_{m,k}}{MK}$越大。为了使数据更直观且易于分析，本文考虑将以无人机每步均覆盖率$C = \dfrac{\text{Users}_c}{\text{Steps}}$来衡量系统性能的好坏，其中的$\text{Users}_c$表示每一幕“已覆盖用户”的数量，Steps表示在每一幕中无人机从起点飞往终点所用总步数。最大化每步均覆盖率即最大化用户平均吞吐量的同时最小化无人机到达终点总步数，用每步均覆盖率作为性能指标刚好满足本文研究目的。

图4给出了“单无人机场景”下不同方案的无人机每步均覆盖率对比图。可以看出，“固定轨迹数据收集方案”无人机每步均覆盖率基本保持不变且低于其他3种方案，这是因为在每一幕中无人机飞行的轨迹相同，所服务的地面用户基本固定不变。这说明强化学习算法具备自主学习的能力，在训练时会通过与环境交互最大化本文优化目标。在“基于DQN的数据收集方案”中，无人机每步均覆盖率在50幕之前较低，这是由于DQN算法在网络训练初期对Q值计算尚不准确时，在Target网络中遍历取最大Q值导致过估计。此方案无人机每步均覆盖率在200幕之后不再上升，训练开始收敛。同时，可以观察到“基于DDQN的数据收集方案”和本文所提方案在训练初期无人机每步均覆盖率增长速度比“基于DQN的数据收集方案”更快，这体现了目标的动作选择和动作评估分别用不同的值函数来实现的优势。但是，在整个训练过程中，“基于DDQN的数据收集方案”的每步均覆盖率波动较大，特别是在第153幕出现较大幅度下降，然后回升，在190幕左右开始收敛。在本文所提方案中，每步均覆盖率在140幕之后趋于收敛，且在训练后期具备最优的收敛值。同时，通过对各方案在训练后期(200～300幕)无人机每步均覆盖率的方差进行了计算，“基于DQN的数据收集方案”和“基于DDQN的数据收集方案”的覆盖率方差分别为1.7×10^{-4}和3.4×10^{-4}，而本文所提方案的覆盖率方差为1.5×10^{-4}，较两种基准方案更稳定，这体现了Dueling-DDQN算法采用Dueling网络架构的优势。上述结果显示，与其他3个对比方案相比，本文所提方案在“单无人机场景”下具有很好的学习能力、鲁棒性和收敛速度。

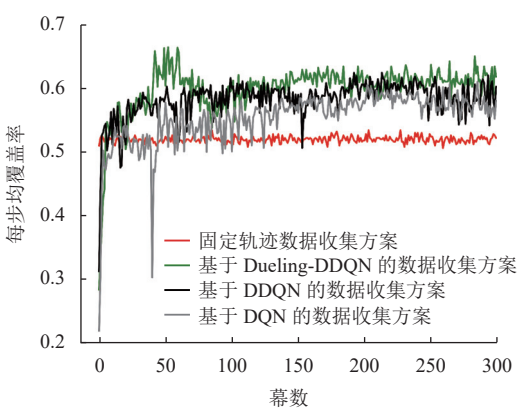

图4 “单无人机场景”下基于不同方案的无人机每步均覆盖率对比图

Fig.4 Comparison of average per-step coverage ratios of different schemes in the single-UAV scenario

图5给出了“两无人机场景”下不同方案的无人机每步均覆盖率对比图。从图中可以观察到，“基于DQN的数据收集方案”在训练初期依旧存在过估计问题，同时训练过程波动较大，在200幕左右开始收敛。“基于DDQN的数据收集方案”虽然解决了DQN方案训练初期的过估计问题，但是由于存在网络训练波动较大的问题，在150～230幕之间无人机每步均覆盖率下滑严重，最终在230幕之后开始收敛。本文所提方案在训练初期每步均覆盖率增长迅速，在第123幕之后开始收敛，且相比上述两种基准

方案具有更佳的收敛值。经计算，上述3种方案在训练后期(200～300幕)的覆盖率方差分别为1.7×10^{-3}、1.4×10^{-3}、8×10^{-4}，可以发现本文所提方案方差最小，具有更加稳定的性能。综合图4和图5分析得出，本文所提算法优于两种基准算法，既能很好地解决DQN算法的过估计问题，又能克服两种对比算法波动较大的问题，具有更好的学习能力、收敛速度和鲁棒性，更适合用于解决无人机路径规划和数据收集问题。

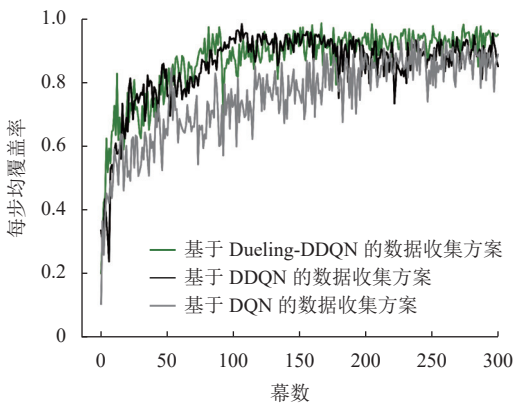

图 5　“两无人机场景”下基于不同方案的无人机每步均覆盖率对比图

Fig.5　Comparison of average per-step coverage ratios of different schemes in the dual-UAV scenario

图6为本文所提算法在“单无人机场景”和“两无人机场景”下的无人机二维飞行轨迹对比图。可以观察到，在这两种场景下无人机都能在最短时间内到达终点。在“单无人机场景”下无人机轨迹分布居中呈“之”字型，这样可以充分利用单个无人机的覆盖范围最大化数据收集。“两无人机场景”下无人机路径未出现重合现象，能够很好地避免无人机之间产生碰撞。同时轨迹分散在两侧，这样能扩大无人机对用户的覆盖范围，充分利用无人机的数量优势最大化收集数据。

图7给出了本文所提算法在不同场景下的无人机每步均覆盖率对比图。可以发现，在“两无人机场景”无人机每步均覆盖率的值接近1.0，而“单无人机场景”下每步均覆盖率在0.6左右，远低于前者。结合图6无人机轨迹可以发现，本文所提算法在两种场景下都能使无人机在最短时间内到达终点。经简单计算可以得出，在“两无人机场景”下无人机可以覆盖约50个用户，接近饱和，而“单无人机场景”下在训练收敛以后无人机覆盖用户数约为30。分析得出，本文所提算法可以很好地完成研究目标，能够使无人机在最短时间内从起点飞往终点，并在此过程中最大化收集用户平均吞吐量。同时，在本文所考虑的系统模型中两无人机辅助数据收集相比单无人机性能提升明显，基本可以实现用户全覆盖，解决了单无人机在用户分布分散且具备移动能力时覆盖不全的问题，体现了研究多无人机辅助数据收集的有效性和必要性。

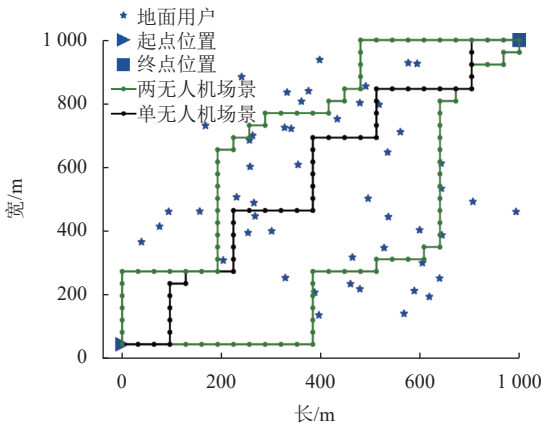

图 6　基于Dueling-DDQN的算法在不同场景下的无人机轨迹对比图

Fig.6　Comparison of UAV trajectories in different scenarios based on the Dueling-DDQN scheme

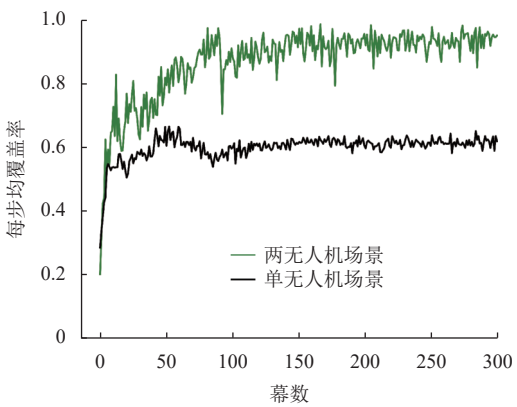

图 7　基于Dueling-DDQN的算法在不同场景下的无人机每步均覆盖率对比图

Fig.7　Comparison of average per-step coverage ratios in different scenarios based on the Dueling-DDQN scheme

5　结论

本文研究了一种多无人机辅助数据收集系统的智能路径规划算法，无人机作为移动基站为地面用户提供通信服务。本文的研究目标是通过规划多架无人机的轨迹，使无人机在最短时间内到达终点的同时最大化收集数据。考虑到飞行时间限制、用户分布分散等问题，本文提出了基于Dueling-DDQN的多无人机辅助数据收集系统智能路径规划算法。仿真结果表明，所提算法可以实现本文研究目标，在本文所考虑的系统模型中解决了单无人机辅助数据收集

时用户覆盖不全的问题。同时，所提算法与两种基准算法相比具有更好的收敛性、鲁棒性和学习能力。

参考文献：

[1] ZHAO N, LU W D, SHENG M, *et al*. UAV-assisted emergency networks in disasters [J]. IEEE Wireless Communications, 2019, 26(1): 45-51.

[2] ZENG Y, ZHANG R, LIM T J. Wireless communications with unmanned aerial vehicles: opportunities and challenges [J]. IEEE Communications Magazine, 2016, 54(5): 36-42.

[3] GAO M, XU X, KLINGER Y, *et al*. High-resolution mapping based on an unmanned aerial vehicle (UAV) to capture paleoseismic offsets along the Altyn-Tagh fault, China [J]. Sci Rep, 2017, 7(1): 1-11.

[4] ZHONG C, GURSOY M C, VELIPASALAR S. Deep reinforcement learning-based edge caching in wireless networks [J]. IEEE Transactions on Cognitive Communications and Networking, 2020, 6(1): 48-61.

[5] GONG J, CHANG T, SHEN C, *et al*. Flight time minimization of UAV for data collection over wireless sensor networks [J]. IEEE Journal on Selected Areas in Communications, 2018, 36(9): 1942-1954.

[6] WU H, WEI Z, HOU Y, *et al*. Cell-edge user offloading via flying UAV in non-uniform heterogeneous cellular networks [J]. IEEE Transactions on Wireless Communications, 2020, 19(4): 2411-2426.

[7] HUANG H, YANG Y, WANG H, *et al*. Deep reinforcement learning for UAV navigation through massive MIMO technique [J]. IEEE Transactions on Vehicular Technology, 2020, 69(1): 1117-1121.

[8] MOZAFFARI F, SAAD W, BENNIS M, *et al*. Unmanned aerial vehicle with underlaid device-to-device communications: performance and tradeoffs [J]. IEEE Transactions on Wireless Communications, 2016, 15(6): 3949-3963.

[9] DUONG T Q, NGUYEN L D, TUAN H D, *et al*. Learning-aided realtime performance optimisation of cognitive UAV-assisted disaster communication[C]//2019 IEEE Global Communications Conference (GLOBECOM) . Waikoloa: IEEE, 2019: 1-6.

[10] DUONG T Q, NGUYEN L D, NGUYEN L K, *et al*. Practical optimization of path planning and completion time of data collection for UAV-enabled disaster communications[C]// 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC) . Tangier: IEEE, 2019: 372-377.

[11] WANG K, TANG, LIU P, *et al*. UAV-based and energy-constrained data collection system with trajectory, time, and collection scheduling optimization[C]// International Conference on Communications in China (ICCC) . Xiamen: IEEE, 2021: 893-898.

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

[13] YOU C, ZHANG R. 3D trajectory optimization in Rician fading for UAV-enabled data harvesting [J]. IEEE Transactions on Wireless Communications, 2019, 18(6): 3192-3207.

[14] BAYERLEIN H, DE KERRET P, GESBERT D. Trajectory optimization for autonomous flying base station via reinforcement learning[C]// 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) . Kalamata: IEEE, 2018: 1-5.

[15] ZHANG B, LIU C H, TANG J, *et al*. Learning-based energy-efficient data collection by unmanned vehicles in smart cities [J]. IEEE Transactions on Industrial Informatics, 2018, 14(4): 1666-1676.

[16] BAYERLEIN H, THEILE M, CACCAMO, *et al*. Multi-UAV path planning for wireless data harvesting with deep reinforcement learning [J]. IEEE Open Journal of the Communications Society, 2021, 2: 1171-1187.

[17] XU S, ZHANG X, LI C, *et al*. Deep reinforcement learning approach for joint trajectory design in multi-UAV IoT networks [J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 3389-3394.

[18] MA J, ZHANG Y, ZHANG J, *et al*. Solution to traveling agent problem based on improved ant colony algorithm[C]// 2008 ISECS International Colloquium on Computing, Communication, Control, and Management. Guangzhou: IEEE, 2008: 57-60.

[19] HUANG Z, LIN H, ZHANG G. The USV path planning based on an improved DQN algorithm[C]// 2021 International Conference on Networking, Communications and Information Technology (NetCIT) . Manchester: IEEE, 2021: 162-166.

[20] XU W, CHEN L, YANG H. A comprehensive discussion on deep reinforcement earning[C]// 2021 International Conference on Communications, Information System and Computer Engineering (CISCE) . Beijing: IEEE, 2021: 697-702.

[21] TEJA K V S S R, LEE M. Efficient practice for deep reinforcement learning[C]// 2019 IEEE Symposium Series on Computational Intelligence (SSCI) . Xiamen: IEEE, 2019: 77-84.

（责任编辑：赵少飞）