《工程科学学报》录用稿，https://doi.org/10.13374/j.issn2095-9389.2024.06.01.001 ©北京科技大学 2024

《工程科学学报》

基于深度强化学习的无人机时空众包资源分配[1]

刘娅汐[1)]，李旭龙[1)]，霍佳皓[1)]，皇甫伟[1)]✉

1) 北京科技大学计算机与通信工程学院，北京市融合网络与泛在业务工程技术研究中心，北京 100083

✉ 通信作者，E-mail: huangfuwei@ustb.edu.cn

摘　要　无人机时空众包资源分配是工业物联网能源管理中的重要任务之一。尽管现有方法考虑了联合反映时间敏感性和公平性的信息新鲜度指标，但忽略了无人机禁飞区和窃听者对数据新鲜度的影响。本文提出了一种基于深度强化学习的无人机时空众包资源分配方法，在考虑无人机禁飞区约束和对窃听者发送干扰信号以保障数据安全的情况下，最小化平均信息新鲜度和物联网设备能耗，从而得到最优无人机轨迹、发射干扰信号功率和物联网发射功率。然而，无人机时空众包中的资源分配复杂且存在挑战，主要表现为决策变量类型多且与考虑服务质量要求的系统性能指标关系复杂。本文将该问题建模为马尔可夫决策过程并使用先进的深度强化学习算法求解该问题，即软演员-评论家（SAC）算法。本文在多无人机场景下验证了所提出算法在解决无人机时空众包资源分配任务中的有效性和正确性。另外，SAC 算法相较于其他两种先进的深度强化学习算法，即深度确定性策略梯度算法和双延迟深度确定性策略梯度算法，具有更快的收敛速度和更优的解。最后，本文分析了最优无人机数目的选择方案。

关键词　无人机；时空众包；资源分配；物联网；深度强化学习

分类号　TN929.5

UAV spatio-temporal crowdsourcing resource allocation based on deep reinforcement learning

LIU Ya-xi[1)], *LI Xu-long*[1)], *HUO Jia-hao*[1)], *HUANGFU Wei*[1)] ✉

1) Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

✉ Corresponding author, E-mail: huangfuwei@ustb.edu.cn

ABSTRACT　Spatio-temporal crowdsourcing refers to the use of a variety of Internet of Things (IoT) devices distributed across industrial environments to collect and transmit spatio-temporal data related to industrial operations. Unmanned Aerial Vehicles (UAVs) play a crucial role in further collecting data from IoT devices in spatio-temporal crowdsourcing tasks. In the realm of industrial IoT energy management, the allocation of spatio-temporal crowdsourcing resources to UAVs represents a significant challenge. Traditional approaches to this problem have focused on optimizing the Age of Information (AoI) to ensure timely and equitable data updates. Nonetheless, these methods often overlook critical operational constraints such as UAV no-fly zones and the potential for data interception by eavesdroppers, both of which can have a detrimental effect on the freshness and integrity of the information being gathered and transmitted. To address these shortcomings, this paper introduces a novel deep reinforcement learning-based framework for UAV spatio-temporal crowdsourcing resource allocation. Our approach specifically aims to minimize the average AoI across the network while also reducing the energy consumption of IoT

收稿日期：2024-06-01

基金项目：国家自然科学基金区域基金重点项目资助项目(U22A2005)；广东省基础与应用基础研究基金资助项目（2022A1515110053）

录用稿件，非最终出版稿

devices. This is achieved by incorporating the spatial constraints imposed by UAV no-fly zones and by actively managing the transmission of jamming signals to mitigate the threat posed by eavesdroppers, thus ensuring the security of the data. However, the allocation of spatio-temporal crowdsourcing resources for UAVs is highly complex and still presents several challenges. The types of decision variables are numerous, and their numbers increase linearly with the duration of the service. Furthermore, the relationship between the performance metrics of the system and the decision variables is intricate, and there is a need to meet appropriate quality of service requirements. The problem is formalized as a Markov Decision Process (MDP), which provides a structured approach to model the decision-making scenario faced by UAVs in a dynamic environment. To solve this MDP, we employ the Soft Actor-Critic (SAC) algorithm, an advanced deep reinforcement learning method known for its sample efficiency and stability. The SAC algorithm is adept at handling the continuous action spaces typical of UAV flight paths and power control problems, making it particularly well-suited for our application. We rigorously test our proposed method in scenarios involving multiple UAVs, demonstrating not only the algorithm's ability to effectively manage the spatio-temporal allocation of resources but also its superiority in faster convergence speed and better solution over existing state-of-the-art methods such as the Twin Delayed Deep Deterministic Policy Gradient (TD3) and the Deep Deterministic Policy Gradient (DDPG) algorithms. Furthermore, the paper delves into the strategic selection of the optimal number of UAVs to balance the trade-offs between coverage, energy consumption, and operational efficiency. By analytically and empirically examining the impact of the UAV fleet size on the system's performance, we provide insights into how to configure UAV networks to achieve the best possible outcomes in terms of AoI, energy management, and security. In conclusion, our research contributes a robust and intelligent framework for UAV resource allocation. The demonstrated efficacy of the SAC algorithm in this context paves the way for its future application in other domains where secure, efficient, and intelligent resource management is paramount.

KEY WORDS Unmanned aerial vehicle; spatio-temporal crowdsourcing; resource allocation; Internet of Things; deep reinforcement learning

引言

工业物联网（IoT，Internet of Things）能源管理任务中，时空众包（STC，Spatio-Temporal Crowdsourcing）指利用分布在工业环境中的各种物联网设备，如传感器、智能仪表和摄像头，来收集和传输与工业运营相关的时空数据[1][2]。这些数据包括设备状态、能源消耗、物理环境条件等[3][4]，对于实时监控工业过程、优化运营效率、预防设备故障和支持决策制定至关重要，从而推动工业 4.0 的智能化发展[5]。无人机（UAV，Unmanned Aerial Vehicle）由于其灵活部署、大范围覆盖、实时监控和响应等优势，在时空众包任务中进一步收集物联网设备数据时发挥重要作用[6-8]。因此，无人机时空众包中的资源分配是亟待解决的关键问题之一[9]。

无人机时空众包资源分配十分复杂且尚存在一些难点。资源分配决定何时如何将何种资源分配给何种设备，该问题通常可被构建为一个优化问题，包含决策变量、目标函数和约束条件[10]。在物联网数据收集中，决策变量种类多且随着服务时长增加其数目线性增加，解空间指数级增长。衡量系统性能的指标通常可构建为目标函数，其与决策变量间关系复杂且需满足合适的服务质量需求。

无人机众包资源分配研究主要关注三个指标：能耗、时效性和公平性。为减少系统的总能耗，包括物联网设备的能耗，一些研究者把它作为优化目标，而其他研究者则通过为无人机设定能耗限制进行约束。此外，还有研究者通过减少无人机的数量来间接降低系统能耗。文献[11]不仅最小化总能耗，还在无人机电池容量的约束下，最大化系统安全性和数据收集效率。文献[12]最小化单个无人机的推进能耗和物联网设备的传输能耗。文献[13]也最小化单个无人机和物联网设备的总能耗。文献[14]和[15]最小化在多无人机场景中部署的无人机数量，间接节省了无人机的能耗。

为了确保时效性，研究者通常旨在将任务完成总时间最小化作为优化目标，或要求无人机在特定时间要求内完成数据收集任务。文献[16]旨在最小化总任务完成时间,文献[17]最小化数据收集完

录用稿件，非最终出版稿

Comment [YL1]: Ref [1] C.-X. Wang, X. You, X. et Gao, al. On the road to 6G: Visions, requirements, key technologies and testbeds. *IEEE Commun Surveys Tut*, 2023, 25(2): 905

Comment [YL2]: Ref [2] Q. Zhang, Y. Wang, G. Yin, et al. Two-stage bilateral online priority assignment in spatio-temporal crowdsourcing. *IEEE Trans Serv Comput*, 2022, 16(3): 2267

Comment [YL3]: Ref [3] Qi J, Wang W, Chen M, et al. Concept, architecture and key technologies of industrial internet. *Chinese J Internet Things*, 2022, 6(02): 38
(亓晋,王微,陈孟玺,等. 工业互联网的概念、体系架构及关键技术. 物联网学报, 2022, 6(02): 38)

Comment [YL4]: Ref [4] H. Zhang, M. Jiang, X. Liu, et al. PPO-based PDACB traffic control scheme for massive IoV communications. *IEEE Trans Intell Transp Syst*, 2022, 24(1): 1116

Comment [YL5]: Ref [5] Chen H, Liang J, Ma Y. Research on standardization of value creation of energy big

Comment [YL6]: Ref [6] Z. Wei, M. Zhu, N. Zhang, et al. UAV-assisted data collection for internet of things: A

Comment [YL7]: Ref [9] Guo H, Wang Y, Liu J, et al. Multi-UAV cooperative task offloading and resource

Comment [YL8]: Ref [10] Mahmood A, Vu T X, Chatzinotas S, et al. Joint optimization of 3D placement and

Comment [YL9]: Xu X, Zhao H, Yao H, et al. A blockchain-enabled energy-efficient data collection system for UAV-assisted IoT. *IEEE*

Comment [YL10]: Sun M, Xu X, Qin X, et al. AoI-energy-aware UAV-assisted data collection for IoT networks: A deep

Comment [YL11]: Khodaparast S S, Lu X, Wang P, et al. Deep reinforcement learning based energy efficient multi-UAV data

Comment [YL12]: Xu W, Xiao T, Zhang J, et al. Minimizing the deployment cost of UAVs for delay-sensitive data collection in IoT

Comment [YL13]: Zhang J, Li Z, Xu W, et al. Minimizing the number of deployed UAVs for delay-bounded data collection of IoT

Comment [YL14]: Gao Y, Liu M, Mei Z, et al. Deep reinforcement learning based UAV trajectory design for data collection scenario

Comment [YL15]: Wang Y, Gao Z, Zhang J, et al. Trajectory design for UAV-based Internet of Things data collection: A deep

成时间。文献[14]和文献[15]要求每个无人机的总巡回时间不超过最大数据收集延迟。文献[18]确保在给定时间内可靠地收集所有物联网设备的数据。文献[19]限制每个物联网设备的服务数量必须在其截止日期前达到最低水平，这反映了数据的时效性。

公平性是指每个物联网设备的数据都有平等的机会被收集。为了确保公平性，一些研究者最大化收集的数据总量或收集的物联网设备总数，而其他研究者在约束中确保在整个服务期间至少收集每个物联网设备的数据一次，或所有物联网设备的数据都在特定时间要求内被收集。此外，一些作者直接考虑在无人机飞行期间平等化每个物联网设备被收集的机会。在优化目标方面，文献[20]最大化总数据收集量，这表示无人机倾向于经过更多物联网设备以收集更多数据。文献[21]最大化服务的物联网设备数量，即尽可能地收集物联网设备数据。在优化约束方面，文献[17]要求在整个服务时长内每个物联网设备的数据至少被收集一次。文献[18]要求在指定时间内可靠地收集每个物联网设备的感测数据。类似地，文献[19]限制每个物联网设备在要求的时间内完成数据上传。文献[22]采用了基于时隙 ALOHA 和代码的组合方法，允许每个物联网设备享有平等的数据传输机会，而非部分物联网设备被频繁收集而其余设备被忽视。

现有工作独立或同时地考虑上述指标。大多数工作仅考虑单一目标优化，少数工作同时考虑时效性和公平性以进行资源分配。具体而言，部分作者结合上述目标或约束进行联合优化。文献[17]旨在最小化数据收集完成时间，同时确保在整个服务时间内至少服务每个物联网设备一次。文献[18]要求单个无人机在特定秒数内可靠地收集每个物联网设备的感测数据。文献[19]要求每个物联网设备应在规定时间内完成数据上传。部分作者引入一个新的度量指标，即物联网设备的平均信息新鲜度（AoI，Age of Information）。最小化平均 AoI 不仅意味着每个物联网设备的数据将被及时收集，同时确保了收集时的公平性。文献[12]、[23]、[24]均旨在最小化从所有地面物联网设备收集的数据的平均 AoI。

然而，现有工作仍存在一些不足。首先，在工业物联网能量管理场景下的无人机时空众包任务中，为避免可能出现的安全事故，同时确保数据收集活动的合法性和有效性，无人机禁飞区需被考虑[1]。尽管已有工作考虑无人机避障或无人机禁飞区，但它在工业物联网能量管理中的联合优化多指标时鲜少被考虑。文献[25]构建了用于无人机编队避障策略模型训练的架构。文献[13]考虑了无人机辅助地面物联网设备数据收集中的禁飞区约束，作者在实验中模拟了三个方形禁飞区。文献[16]和[17]在三维城市环境中设计了无人机的飞行轨迹，考虑了避开建筑物等障碍物。其次，工业场景下包含敏感信息的数据的安全性至关重要，它是确保能源供应稳定性和效率的关键，且可能涉及到国家安全。因此，确保在有窃听者场景下的数据安全是另一个关键问题[26]。

针对以上不足，本文提出了一种基于深度强化学习的资源分配方法在考虑无人机禁飞区域和窃听者的情况下最小化系统平均信息新鲜度和物联网设备能耗，最终得到最优的无人机飞行轨迹、发射干扰信号功率和物联网设备发射功率。本文将该时空众包资源分配问题建模为马尔可夫决策过程，基于行动者-评论家架构部署了先进的软演员-评论家（SAC，Soft Actor-Critic）算法来近似策略和价值函数以最大化奖励函数。本文在多无人机场景下验证了所提出算法的有效性和正确性。此外，本文证明了 SAC 算法在时空众包任务中表现优于另外两种先进的深度强化学习算法，即双延迟深度确定性策略梯度算法（TD3，Twin Delayed Deep Deterministic Policy Gradient）和深度确定性策略梯度算法（DDPG，Deep Deterministic Policy Gradient）算法，并且它们都优于随机算法。最后，本文讨论了最优无人机数量的选择方案。

1 系统模型与问题描述

1.1 系统模型

考虑一个实现无人机辅助物联网设备数据收集任务的能源管理场景，该场景包含K个无人机和M个物联网设备，如图 1（a）所示。常用的物联网设备包括智能电表、温度传感器、流量计、压力传感器等，主要收集能源消耗数据、电网负荷、设备运行状态、故障和异常事件记录等数据。令

Comment [YL16]: Wang Z, Liu R, Liu Q, et al. Energy-efficient data collection and device positioning in UAV-assisted IoT. *IEEE Internet Things J*, 2019, 7(2): 1122

Comment [YL17]: Samir M, Sharafeddine S, Assi C M, et al. UAV trajectory planning for data collection from time-constrained IoT devices. *IEEE Trans Wireless Commun*, 2019, 19(1): 34

Comment [YL18]: Li Y, Liang W, Xu W, et al. Data collection maximization in IoT-sensor networks via an energy-constrained UAV. *IEEE Trans Mobile Comput*, 2021, 22(1): 159

Comment [YL19]: Al-Hilo A, Samir M, Elhattab M, et al. RIS-assisted UAV for timely data collection in IoT networks. *IEEE Syst J*, 2022, 17(1): 431

Comment [YL20]: Tyrovolas D, Mekikis P V, Tegos S A, et al. Energy-aware design of UAV-mounted RIS networks for IoT data collection. *IEEE Trans Commun*, 2022, 71(2): 1168

Comment [YL21]: Yi M, Wang X, Liu J, et al. Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks. *Proceedings of IEEE Conf Comput Commun Work*. Toronto, 2020: 716

Comment [YL22]: Hu H, Xiong K, Qu G, et al. AoI-minimal trajectory planning and data collection in UAV-assisted wireless powered IoT networks. *IEEE Internet Things J*, 2020, 8(2): 1211

Comment [YL23]: Zhang Y, Duan H, Wei C. Digital twin-based obstacle avoidance method for unmanned aerial vehicle formation control using deep reinforcement learning. Chinese J Eng, 2024, 46(7): 1187
（张宇宸，段海滨，魏晨. 基于深度强化学习的无人机集群数字孪生编队避障. 工程科学学报, 2024, 46(7): 1187）

Comment [YL24]: Li M, Tao X, Li N, et al. Secrecy energy efficiency maximization in UAV-enabled wireless sensor networks without eavesdropper's CSI. *IEEE Internet Things J*, 2021, 9(5): 3346

Comment [YL25]: Figure

录用稿件，非最终出版稿

$U=\{U_1,U_2,\cdots,U_K\}$和$I=\{I_1,I_2,\cdots,I_M\}$分别表示无人机和物联网设备集合。假定无人机和物联网设备都位于三维区域$\mathcal{R}$中，满足$\mathcal{R}=\{(x,y,z)\in\mathbb{R}^3 \mid x^{\min}\le x\le x^{\max}, y^{\min}\le y\le y^{\max}, z^{\min}\le z\le z^{\max}\}$。假定总服务时长为$T$，将$T$等分为$N$个时间段，记为$N=\{t_1,t_2,\cdots,t_N\}$。假定$I_m$在$t_n$内产生数据量为$S_{m,n}$。无人机在$T$内持续飞行以收集物联网设备产生的数据，即物联网设备向无人机传输数据。同时，区域内存在J个窃听者监听物联网设备的数据，记作$E=\{E_1,E_2,\cdots,E_J\}$。在物联网设备向无人机传输数据时，由于物联网设备上装有全向发射天线，窃听者可以同时监听该数据。无人机上安装支持全双工的定向天线，即可以实现同时接收数据和发送信号，并且可以朝指定方向发送信号。因此，为保障数据安全，无人机在接收物联网设备数据的同时朝窃听者发送干扰信号。在T内无人机、物联网设备及窃听者的时间规划如图 1（b）所示。

Comment [YL26]: Figure

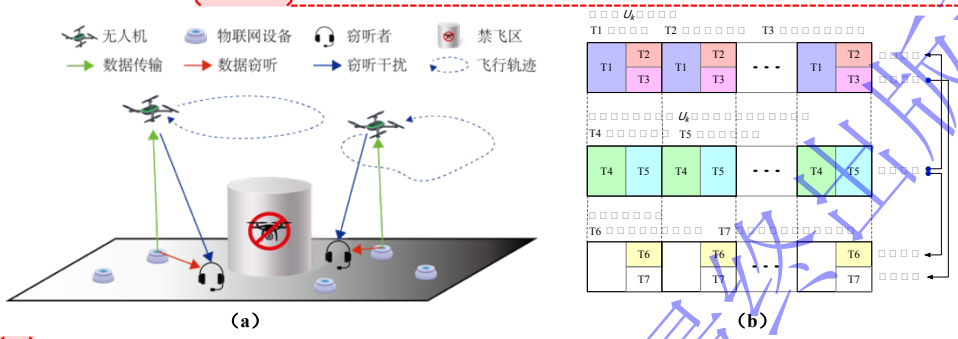


图 1 能源管理场景下的无人机辅助物联网设备数据收集任务.（a）网络架构；（b）服务时间内无人机、物联网设备及窃听者的时间规划

Comment [YL27]: Figure

Fig.1 UAV-assisted data collection from IoT devices in energy management scenario. (a) Network Framework; (b) Time schedules for UAVs, IoT devices, and eavesdroppers during the whole service time

在t_n内，U_k和I_m的位置分别为$\mathrm{u}_{k,n}^{\mathrm{U}}=\{\mathrm{x}_{k,n}^{\mathrm{U}},y_{k,n}^{\mathrm{U}},z_{k,n}^{\mathrm{U}}\}$和$\mathrm{u}_{m,n}^{\mathrm{I}}=\{\mathrm{x}_{m,n}^{\mathrm{I}},y_{m,n}^{\mathrm{I}},z_{m,n}^{\mathrm{I}}\}$。物联网设备在地面上，因此$z_{m,n}^{\mathrm{I}}=0,\forall \mathrm{m},n$。物联网设备的位置是已知且在$T$内固定不变的，因此$\mathrm{u}_{m,1}^{\mathrm{I}}=\cdots=\mathrm{u}_{m,N}^{\mathrm{I}}$。假设在$T$中，无人机能量充足因此无需返回充电站充电，无人机初始位置$\mathrm{u}_{k,0}^{\mathrm{U}}$在区域$\mathcal{R}$内。第$j$个窃听者的位置为$\mathrm{u}_{j,n}^{\mathrm{E}}=\{\mathrm{x}_{j,n}^{\mathrm{E}},y_{j,n}^{\mathrm{E}},z_{j,n}^{\mathrm{E}}\}$。窃听者也位于地面上，因此满足$z_{j,n}^{\mathrm{E}}=0,\forall \mathrm{j},n$。区域$\mathcal{R}$内存在无人机禁飞区，即在整个$T$内无人机无法在禁飞区飞行。禁飞区域可以是任意不规则三维区域，假设禁飞区为圆柱形，该禁飞区在xoy平面的圆心为$(x^{\mathrm{B}},y^{\mathrm{B}})$，半径为$r^{\mathrm{B}}$，高为$h^{\mathrm{B}}$。无人机的禁飞约束可写为$\mathrm{u}_{k,n}^{\mathrm{U}}\notin\mathcal{B}$, $\forall k,n$。其中，$\mathcal{B}$为禁飞区域。此外，无人机必须满足在$\mathcal{R}$内，因此无人机的坐标需满足边界约束：$\mathrm{u}_{k,n}^{\mathrm{U}}\in\mathcal{R},\ \forall k,n$。

1.2 数据传输速率

在t_n内，I_m是否向U_k传输信息的指示符号记作$\alpha_{m,k,n}$。若在t_n内I_m向U_k发送信号，则$\alpha_{m,k,n}=1$；否则，$\alpha_{m,k,n}=0$。每个物联网设备只能向一个无人机传输数据，但一个无人机可以同时接收多个物联网设备的数据。因此，$\sum_{\mathrm{k}=1}^{K}\alpha_{m,k,n}\le 1,\forall m,n$和$\sum_{m=1}^{M}\alpha_{m,k,n}\le M,\forall k,n$。这里，我们认为物联网设备倾向于向距离其最近且在通信距离范围$D_{k,n}^{\mathrm{Max}}$内的无人机传输数据。

录用稿件，非最终出版稿

为了衡量传输效率，在t_{n}内I_m向U_k传输信息的速率为

$$R_{m,k,n}^{\mathrm{I}\to\mathrm{U}}=B\log_2\left(1+\frac{p_{m,n}^{\mathrm{I}}h_{m,k,n}^{\mathrm{I}\to\mathrm{U}}}{\sum_{\hat{k}=1,\hat{k}\neq k}^{K}\sum_{\hat{m}=1}^{M}\alpha_{\hat{m},\hat{k},n}p_{\hat{m},n}^{\mathrm{I}}h_{\hat{m},\hat{k},n}^{\mathrm{I}\to\mathrm{U}}+N_0}\right).\#(1)$$

其中，N_0表示高斯白噪声的功率；B表示总可用带宽；$p_{m,n}^{\mathrm{I}}$表示I_m在t_{n}内的发射功率。干扰项包括I_m向除了U_k以外的其他无人机发送的信号。由于在向同一架无人机传输数据的物联网设备之间采用正交频分多址接入，因此向同一无人机发送数据的相邻物联网设备间的信号干扰可以被避免。$h_{m,k,n}^{\mathrm{I}\to\mathrm{U}}$表示$I_m$到$U_k$传输信息的信道，可以计算为$h_{m,k,n}^{\mathrm{I}\to\mathrm{U}}=\frac{g_0G_0}{\|u_{k,n}^{\mathrm{U}}-u_{m,n}^{\mathrm{I}}\|_2^2}$。其中，$g_0$表示参考距离为 1 米时的信道功率增益，$G_0$表示天线增益。在$t_{\mathrm{n}}$内，每一个窃听者都监听每一个物联网设备传输的数据，E_j监听I_m的数据速率为

$$R_{\mathrm{m},j,n}^{\mathrm{I}\to\mathrm{E}}=B\log_2\left(1+\frac{p_{m,n}^{\mathrm{I}}h_{m,j,n}^{\mathrm{I}\to\mathrm{E}}}{\sum_{\hat{k}=1,\hat{k}\neq k}^{K}\sum_{\hat{m}=1}^{M}\alpha_{\hat{m},\hat{k},n}p_{\hat{m},n}^{\mathrm{I}}h_{\hat{m},\mathrm{j},n}^{\mathrm{I}\to\mathrm{E}}+\sum_{\hat{k}=1}^{K}p_{\hat{k},n}^{\mathrm{U}}h_{\hat{k},\mathrm{j},n}^{\mathrm{U}\to\mathrm{E}}+N_0}\right).\#(2)$$

这里，k表示I_m向U_k传输数据的无人机索引；$p_{k,n}^{\mathrm{U}}$表示U_k发射干扰信号的功率。窃听者监听数据的干扰不仅包括其他物联网设备向其他无人机发送的传输信号，还包括全部无人机对窃听者发送的干扰信号。$h_{m,j,n}^{\mathrm{I}\to\mathrm{E}}$和$h_{\mathrm{k},j,\mathrm{n}}^{\mathrm{U}\to\mathrm{E}}$分别表示$I_m$和$U_k$分别向$E_j$发送信号的信道，分别计算为

$$h_{m,j,n}^{\mathrm{I}\to\mathrm{E}}=\frac{g_0G_0}{\left\|u_{m,n}^{\mathrm{I}}-u_{j,n}^{\mathrm{E}}\right\|_2^{\mu}},\ h_{k,j,\mathrm{n}}^{\mathrm{U}\to\mathrm{E}}=\frac{g_0G_0}{\left\|u_{k,n}^{\mathrm{U}}-u_{j,n}^{\mathrm{E}}\right\|_2^2}.\#(3)$$

由于地面物联网设备到空中无人机通信链路和空中无人机到地面窃听者通信链路为空/地通信链路，因此仅考虑视距路径。而地面物联网设备到地面窃听者通信链路为地到地链路，还需考虑非视距路径，因此μ表征由非视距路径和多径引起的小尺度衰落包络系数。

1.3 问题描述

本文旨在及时地收集物联网设备的信息并且尽可能地减少窃听者收集到的信息。窃听者收集的信息会影响信息的时效性，一旦信息被窃听者获得，则认为信息不再新鲜。因此，定义I_m在t_{n}内的信息新鲜度为

$$A_{m,n}=\begin{cases}\frac{T_{m,n}}{\delta}, & T_{m,n}\leq\delta,\\ \min\left\{A_{m,n-1}+1,A^{\mathrm{Max}}\right\}, & \text{Otherwise.}\end{cases}\#(4)$$

其中，其中参数δ表示数据传输的最大容忍延迟。信息传输时间可定义为

$$T_{m,n}=\frac{S_{m,n}}{R_{m,k,n}^{\mathrm{I}\to\mathrm{U}}-\max\limits_{\mathrm{j}}R_{\mathrm{m},j,n}^{\mathrm{I}\to\mathrm{E}}}.\#(5)$$

如果信息在δ内传输完毕，则信息传输成功。但考虑到窃听者会监听数据，信息传输速率描述为物联网设备向无人机传输速率减去所有窃听者监听该物联网设备的最大速率。算法倾向于最小化窃听者获得信息的最大速率以避免信息被窃听者获得。信息新鲜度表征收集的物联网设备数据的新鲜程度。如果在某时间段内某物联网设备的信息被收集，则该物联网设备的信息新鲜度降为一个很小的

录用稿件，非最终出版稿

值；若未被收集，则信息新鲜度在历史新鲜度数值基础上累加 1。可以看出，信息新鲜度越低，信息越新鲜。另外，本文旨在最小化物联网设备的能耗，I_m在t_n内的能耗计算为

$$E_{m,n}=p_{m,n}^{I}\min\{T_{m,n},\delta\}.\ \#(6)$$

本文旨在最小化全部物联网设备在总服务时长的信息新鲜度和物联网设备的能耗的加权和，优化变量为无人机位置集合、物联网设备发射功率集合和无人机发射干扰功率集合，分别记作$\mathrm{V}=\{u_{1,1}^{U},\cdots,u_{K,1}^{U},\cdots,u_{K,N}^{U}\}$，$\boldsymbol{P}^{I}=\{p_{1,1}^{I},\cdots,p_{M,1}^{I},\cdots,p_{M,N}^{I}\}$和$\boldsymbol{P}^{U}=\{p_{1,1}^{U},\cdots,p_{K,1}^{U},\cdots,p_{K,N}^{U}\}$。因此，本文的优化问题可以描述为

$$P1:\ \min_{\mathrm{V},\boldsymbol{P}^{I},\boldsymbol{P}^{U}}\frac{1}{N}\sum_{n=1}^{N}O_n(\mathrm{V},\boldsymbol{P}^{I},\boldsymbol{P}^{U})=\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}A_{m,n}+\frac{\kappa}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}E_{m,n}\#(7a)$$

$$\text{s.t. } C1:\ u_{k,n}^{U}\in\mathcal{R},\ \forall k,n,\ C2:\ u_{k,n}^{U}\notin\mathcal{B},\ \forall k,n,\#(7b)$$

$$C3:\ p_{\min}^{I}\le p_{m,n}^{I}\le p_{\max}^{I},\ \forall m,n,\quad C4:\ p_{\min}^{U}\le p_{k,n}^{U}\le p_{\max}^{U},\ \forall k,n,\ \#(7c)$$

$$C5:\ \left|x_{k,n}^{U}-x_{k,n-1}^{U}\right|\le x_{\max}^{U},\forall k,n,\ C6:\ \left|y_{k,n}^{U}-y_{k,n-1}^{U}\right|\le y_{\max}^{U},\forall k,n,\#(7d)$$

$$C7:\ \left|z_{k,n}^{U}-z_{k,n-1}^{U}\right|\le z_{\max}^{U},\forall k,n.\ \#(7e)$$

其中，κ表示物联网设备能耗相对系统信息新鲜度的优化权重；$p_{\min}^{I}$和$p_{\max}^{I}$分别表示物联网设备最小和最大发射功率；$p_{\min}^{U}$和$p_{\max}^{U}$分别表示无人机最小和最大干扰功率；$x_{\max}^{U}$，$y_{\max}^{U}$和$z_{\max}^{U}$分别表示无人机在x轴、y轴和z轴方向上单个时间段内的最大移动距离。

2　基于深度强化学习的无人机时空众包资源分配算法设计

2.1 强化学习问题建模及基本元素

深度强化学习是机器学习的一个领域，它涉及智能体在环境中学习如何采取行动以最大化某种累积奖励。深度强化学习在决策无人机飞行轨迹方面有良好的表现[27]。为使用强化学习进行资源分配算法设计，优化问题P1需重新描述为马尔科夫决策过程问题以决定最优无人机轨迹，物联网设备和无人机的发射功率变化。该问题可由四元变量组成，分别为状态空间、动作空间、状态转移概率和奖励，记作$(\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R})$。下面将详细介绍这四种基本元素：

（1）状态空间集合$\mathcal{S}$：状态空间包含对环境的完整描述。在时间段t_n中，状态$\mathcal{S}_n$包括无人机的位置和发射功率，物联网设备的位置、发射功率、信息新鲜度和新产生数据量，某物联网设备是否向某无人机传输数据指示符号，窃听者位置，记作

$$\mathcal{S}_n=\left\{\begin{matrix}u_{1,n}^{U},\cdots,u_{K,n}^{U},p_{1,n}^{U},\cdots,p_{K,n}^{U},u_{1,n}^{I},\cdots,u_{M,n}^{I},p_{1,n}^{I},\cdots,p_{M,n}^{I},A_{1,n},\cdots,A_{M,n},\\ S_{1,n},\cdots,S_{M,n},\alpha_{1,1,n},\cdots,\alpha_{1,K,n},\cdots\alpha_{M,1,n},\cdots,\alpha_{M,K,n},u_{1,n}^{E},\cdots,u_{J,n}^{E}\end{matrix}\right\}.\#(8)$$

（2）动作空间集合$\mathcal{A}$：在给定状态$\mathcal{S}_n$下，行动是智能体可以采取的决策，智能体为无人机和物联网设备。在时间段t_n中，动作$\mathcal{A}_n$包括无人机的移动距离，无人机和物联网设备的发射功率变化，记作

$$\mathcal{A}_n=\{\Delta x_{1,n}^{U},\Delta y_{1,n}^{U},\Delta z_{1,n}^{U},\cdots,\Delta x_{K,n}^{U},\Delta y_{K,n}^{U},\Delta z_{K,n}^{U},\Delta p_{1,n}^{I},\cdots,\Delta p_{M,n}^{I},\Delta p_{1,n}^{U},\cdots,\Delta p_{K,n}^{U}\}.\#(9)$$

Comment [YL28]: Ou Y, Guo Z, Luo D, et al. Collaborative air combat maneuvering decision-making method based on graph convolutional deep reinforcement learning. Chinese J Eng, 2024, 46(7): 1227
（欧洋, 郭正玉, 罗德林, 等. 基于图卷积深度强化学习的协同空战机动决策方法. 工程科学学报, 2024, 46(7): 1227）

Comment [YL29]: Equation

录用稿件，非最终出版稿

（3）状态转移函数$\mathcal{P}$：状态转移函数指在状态$\mathcal{S}_{\mathrm{n}}$下采取行动$\mathcal{A}_{\mathrm{n}}$后转移到下一个状态$\mathcal{S}_{\mathrm{n}+1}$的概率，它为智能体与环境交互提供了一个基本框架，表示为$\mathcal{P}=P\left(\mathcal{S}_{\mathrm{n}+1}\middle|\mathcal{S}_{\mathrm{n}},\mathcal{A}_{\mathrm{n}}\right)$，其中$P(A\middle|B)$表示条件概率函数。

（4）奖励$\mathcal{R}$：奖励是对智能体在给定状态$\mathcal{S}_{\mathrm{n}}$下采取特定行动$\mathcal{A}_{\mathrm{n}}$时，从环境中反馈的有效性的评价。智能体通过最大化其奖励来调整其策略。在时间段t_{n}中，奖励被定义为

$$\mathcal{R}_{\mathrm{n}}=-\omega O_{\mathrm{n}}\left(\mathrm{V},\boldsymbol{P}^{\mathrm{I}},\boldsymbol{P}^{\mathrm{U}}\right)+Q_{n}\left(\mathrm{V},\boldsymbol{P}^{\mathrm{I}},\boldsymbol{P}^{\mathrm{U}}\right)+\mathrm{c}.\#(10)$$

其中，$O_{\mathrm{n}}\left(\mathrm{V},\boldsymbol{P}^{\mathrm{I}},\boldsymbol{P}^{\mathrm{U}}\right)$为优化问题$P1$的目标函数；$Q_{n}\left(\mathrm{V},\boldsymbol{P}^{\mathrm{I}},\boldsymbol{P}^{\mathrm{U}}\right)$为罚函数以满足优化问题的边界约束和禁飞区约束，定义为

$$Q_{n}\left(\mathrm{V},\boldsymbol{P}^{\mathrm{I}},\boldsymbol{P}^{\mathrm{U}}\right)=\begin{cases}-q, & \text{在}t_{\mathrm{n}}\text{内}C1\text{或}C2\text{不满足,}\\ 0, & \text{Otherwise.}\end{cases}\#(11)$$

上述四种基本要素、智能体、环境、策略函数$\pi(\mathcal{A}\middle|\mathcal{S})$的关系如图 2 所示。

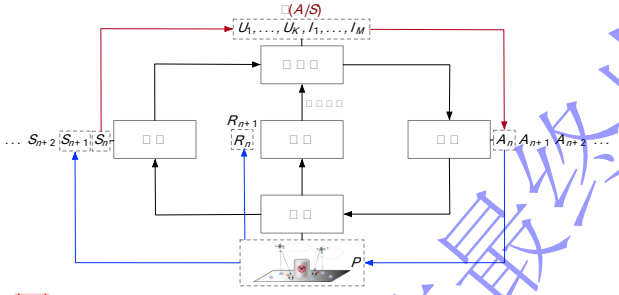


图 2 四种基本要素、智能体、环境与策略函数$\pi(\mathcal{A}\middle|\mathcal{S})$关系图.

Fig.2 Relationship among four fundamental elements, agent, environment, and policy function.

2.2 软演员-评论家算法

软演员-评论家算法是先进的深度强化学习算法之一，它基于演员-评论家架构，在策略优化中引入了熵作为额外的奖励，以鼓励探索并避免过早收敛到次优策略。在 SAC 算法中考虑六个网络，包括一个策略网络$\pi_{\theta}(\mathcal{A}\middle|\mathcal{S})$，两个 Q 值网络$Q_{\phi_1}(\mathcal{S},\mathcal{A})$和$Q_{\phi_2}(\mathcal{S},\mathcal{A})$，目两个目标 Q 值网络$Q_{\phi_1'}(\mathcal{S},\mathcal{A})$和$Q_{\phi_2'}(\mathcal{S},\mathcal{A})$和一个软价值网络$V_{\psi}(\mathcal{S})$。软价值网络估算给定状态下的预期回报，即状态价值函数。SAC 算法首先进行环境交互，将状态转移$(\mathcal{S},\mathcal{A},\mathcal{R},\mathcal{S}')$存储在经验回放缓冲区$D$中。然后在$D$中随机抽取一批状态转移进行经验回放，并更新网络的参数，更新过程如下：

（1）更新 Q 值网络$Q_{\phi_i}(\mathcal{S},\mathcal{A})$：首先，目标值计算为

$$\mathrm{y}=\mathcal{R}+\gamma\left(\min_{\mathrm{i}=1,2}Q_{\phi_i'}(\mathcal{S}',\tilde{\mathcal{A}}')-\zeta\log\pi_{\theta}(\tilde{\mathcal{A}}'\middle|\mathcal{S}')\right).\#(12)$$

其中，ζ表示熵正则化系数，它表明了策略熵相对于奖励的重要性；$\tilde{\mathcal{A}}'\sim\pi_{\theta}(\cdot\middle|\mathcal{S}')$。Q 值网络可以通过最小化损失函数$J_{Q}(\phi_{\mathrm{i}})$更新，计算为

$$J_{Q}(\phi_{\mathrm{i}})=\mathbb{E}\left[\left(Q_{\phi_i}(\mathcal{S},\mathcal{A})-y\right)^2\right],i=\{1,2\}.\#(13)$$

其中，$\mathbb{E}[\cdot]$表示期望函数。

（2）更新软价值网络$V_{\psi}(\mathcal{S})$：该网络可以通过最小化损失函数$J_{V}(\psi)$得到，计算为

$$J_{V}(\psi)=\mathbb{E}\left[\left(V_{\psi}(\mathcal{S})-Q_{\mathrm{m}}(\mathcal{S},\tilde{\mathcal{A}})+\zeta\log\pi_{\theta}(\tilde{\mathcal{A}}\middle|\mathcal{S})\right)^2\right].\#(14)$$

Comment [YL30]: Equation

Comment [YL31]: Figure

Comment [YL32]: Figure

录用稿件，非最终出版稿

其中，$\tilde{\mathcal{A}}\sim\pi_\theta(\cdot\,|\,\mathcal{S})$且$Q_{\min}(\mathcal{S},\tilde{\mathcal{A}})=\min_{i=1,2}Q_{\phi_i}(\mathcal{S},\tilde{\mathcal{A}})$。

（3）更新策略函数$\pi_\theta(\mathcal{A}\,|\,\mathcal{S})$：该网络可以通过最大化奖励加熵的期望值来更新，定义为

$$J_\pi(\theta)=\mathbb{E}\left[\zeta\log\pi_\theta(\tilde{\mathcal{A}}\,|\,\mathcal{S})-Q_{\min}(\mathcal{S},\tilde{\mathcal{A}})\right].\#(15)$$

（4）更新目标 Q 值网络$Q_{\phi_1'}(\mathcal{S},\mathcal{A})$和$Q_{\phi_2'}(\mathcal{S},\mathcal{A})$：网络可以更新为

$$Q_{\phi_i'}\leftarrow\xi Q_{\phi_i}+(1-\xi)Q_{\phi_i'},i\epsilon\{1,2\}.$$

其中，ξ表示软更新参数，它控制目标网络参数向当前对应网络参数靠近的速率。

（5）为了平衡探索和利用，熵正则化系数通过最小化损失函数$J_\zeta=\mathbb{E}\left[-\zeta\log\pi_\theta(\mathcal{A}\,|\,\mathcal{S})-\zeta H\right]$来自适应地调整，其中为H目标策略熵。

算法 1 总结了基于 SAC 的无人机时空众包资源分配算法。

Comment [YL33]: Algorithm

Comment [YL34]: Alg

算法 1　基于 SAC 的无人机时空众包资源分配算法

1. 初始化π_θ, Q_{ϕ_1},Q_{ϕ_2},$Q_{\phi_1'}$,$Q_{\phi_2'}$,V_ψ,D.
2: **for** 迭代次数 **do**
3:　　观察初始状态$\mathcal{S}_0$
4:　　**for** n = 1,⋯,N **do**
5:　　根据$\mathcal{S}_n$选择动作$\mathcal{A}_n\sim\pi_\theta(\cdot\,|\,\mathcal{S}_n)$，执行该动作并观察奖励$\mathcal{R}_n$和下一个状态$\mathcal{S}_{n+1}$
6:　　将转移状态$(\mathcal{S},\mathcal{A},\mathcal{R},\mathcal{S}')$存储到经验回放缓冲区$D$中，且令$\mathcal{S}_n\leftarrow\mathcal{S}_{n+1}$
7:　　**end for**
8:　　**for** 更新次数 **do**
9:　　D中随机采样一批$(\mathcal{S},\mathcal{A},\mathcal{R},\mathcal{S}')$，计算目标值 y，并通过最小化$J_Q(\phi_i)$更新 Q 值网络
10:　　通过最小化$J_V(\psi)$更新软价值网络，通过最大化$J_\pi(\theta)$更新策略网络
11:　　更新$Q_{\phi_i'}\leftarrow\xi Q_{\phi_i}+(1-\xi)Q_{\phi_i'},i\epsilon\{1,2\}$，通过最小化$J_\zeta$更新熵正则化系数
12:　　**end for**
13: **end for**

3　仿真实验结果与分析

本文在多无人场景中基于所提出算法进行仿真实验以验证算法的有效性、收敛效率，并探讨场景中最佳无人机个数选择方案。考虑一个部署在$\mathcal{R}\epsilon[-500\,\mathrm{m},500\,\mathrm{m}]\times[-500\,\mathrm{m},500\,\mathrm{m}]\times[0\,\mathrm{m},200\,\mathrm{m}]$的多无人机辅助物联网设备能量管理数据收集网络，该网络中有$K=2$架无人机，$M=2$个物联网设备，J = 2个监听者，如图 3（a）所示。网络中存在圆柱形无人机禁飞区，参数为$(x^{\mathrm{B}},y^{\mathrm{B}})=(0\,\mathrm{m},0\,\mathrm{m})$，$r^{\mathrm{B}}=200\,\mathrm{m}$，$h^{\mathrm{B}}=150\,\mathrm{m}$。在总服务时长内共有$N=100$个时间段。初始无人机、物联网设备和窃听者的位置都在非禁飞区域内随机生成。假设在整个服务时长内无人机高度需保持在$[100\,\mathrm{m},200\,\mathrm{m}]$内。其他系统参数设置为$g_0G_0=50$，$A^{\mathrm{Max}}=100$，$\mu=3.5$，$B=1\,\mathrm{GHz}$，$\delta=0.03\,\mathrm{s}$，$p_{\min}^{\mathrm{I}}=0.01\,\mathrm{W}$，$p_{\max}^{\mathrm{I}}=0.1\,\mathrm{W}$，$p_{\min}^{\mathrm{U}}=0.01\,\mathrm{W}$，$p_{\max}^{\mathrm{U}}=0.1\,\mathrm{W}$，$x_{\max}^{\mathrm{U}}=50\,\mathrm{m}$，$y_{\max}^{\mathrm{U}}=50\,\mathrm{m}$，$z_{\max}^{\mathrm{U}}=5\,\mathrm{m}$，$\xi=0.01$，$q=20$，$\kappa=10^4$，$\omega=1/30$，$c=1.3$，$N_0$在$10^{-7}$到$10^{-6}$间随机生成。SAC 算法在 Python 3.7, Pytorch 1.9.0+cpu 下执行。设置算法训练参数折扣因子$\gamma=0.98$，经验池大小$|D|=10^5$，训练轮次为10^4，批次大小为 256，步长为 0.01，优化器为 Adam。其他算法训练参数如表 1 所示，表中 IL、HL 和 OL 分别代表输入层、隐藏层和输出层。

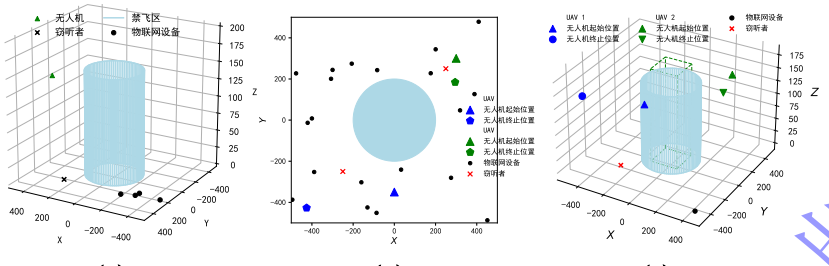
Comment [YL35]: Figure

表 1　SAC 算法训练参数

录用稿件，非最终出版稿

Table 1 Training parameters in SAC algorithm

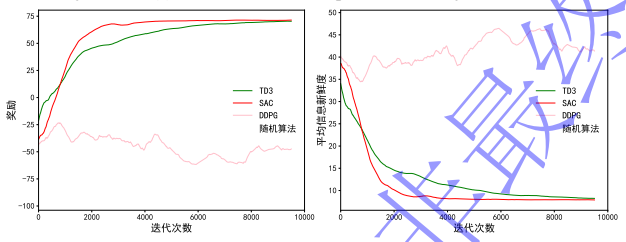
Network	IL number/size/activation function	HL number/size/activation function	OL number/size/activation function
Policy network	$1/4N+4M$/NA	1/(128, 128)/(Relu, Relu)	$2/(4N+M, 4N+M)$/(Tanh, Tanh)
Q-value network 1	$1/8N+5M$/NA	2/(128, 128)/(Relu, Relu)	1/1/NA
Q-value network 2	$1/8N+5M$/NA	2/(128, 128)/(Relu, Relu)	1/1/NA



（a） （b） （c）

图 3 仿真实验场景与无人机轨迹.（a）初始场景；（b）二维无人机最优轨迹；（c）三维无人机最优轨迹

Fig.3 Simulation experiment scenario and UAV trajectories. (a) Initial scenario; (b) Two-dimensional optimal UAV trajectories; (c) Three-dimensional optimal UAV trajectories



（a） （b）

图 4 SAC、TD3、DDPG 及随机算法效果图.（a）奖励随迭代次数变化；（b）平均信息新鲜度随迭代次数变化

Fig.4 Algorithm performance of SAC, TD3, DDPG, and random algorithms. (a) Reward versus episode; (b) Average age of information versus episode

为验证算法有效性和正确性，SAC 算法训练10^4代后得到的最优网络参数下的二维和三维最优无人机轨迹分别如图 3（b）和图 3（c）所示。首先，两个无人机在整个服务时长内均未进入禁飞区域，这证明禁飞约束是有效的。其次，两个无人机都在物联网设备的沿途飞行，这是由于无人机倾向于收集尽可能多的物联网设备数据，以保证信息新鲜度。此外，无人机在物联网设备更密集的区域重复飞行多次以最大化统计上意义上的信息时效性。最后，无人机倾向于在窃听者附近尽可能多地飞行，这是由于当无人机靠近窃听者时，干扰信号将大大增强，监听数据速率将随之降低，从而提升系统整体信息时效性。因此，本文所提出的 SAC 算法是有效的。

为进一步证明算法的有效性、收敛性和效率，SAC 算法的奖励函数和平均信息新鲜度与迭代次数关系图分别如图 4（a）和图 4（b）所示。由图可知，随着迭代次数增加，SAC 算法的奖励逐渐增加后趋于定值，信息新鲜度逐渐减小后趋于定值，这证明了算法的有效性和收敛性。另外，作为 SAC 算法的对比算法，TD3、DDPG 和随机算法的奖励和平均信息新鲜度与迭代次数关系也在图中给出。在本实验中，DDPG 算法并未收敛，效果不理想。TD3 算法通过引入双重 Q 学习和延迟策略更新等技术来改进 DDPG，从而提升算法稳定性。在本实验中，TD3 算法随着迭代次数增加逐渐收敛，但其收敛速度仍要慢于 SAC 算法。这是由于 TD3 通常依赖于动作噪声进行探索，而 SAC 使用熵正则化鼓励更有效的探索。此外，SAC 中的熵正则化策略提供了额外的稳健性。最后，随机算法并未收敛，效果相比以 SAC、TD3 为代表的可收敛的深度强化学习而言更差。

Comment [YL36]: Figure

Comment [YL37]: Figure

Comment [YL38]: Figure

Comment [YL39]: Figure

Comment [YL40]: Figure

Comment [YL41]: Figure

录用稿件，非最终出版稿

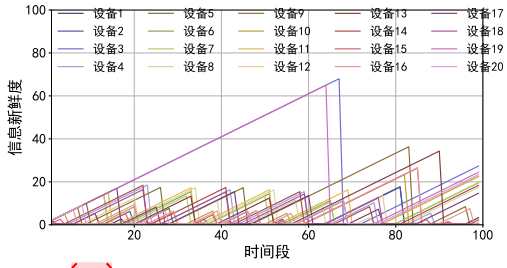
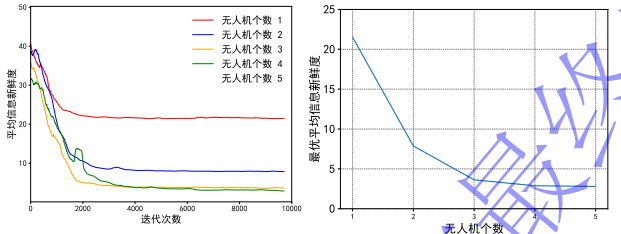


图 5 SAC 算法优化后物联网设备信息新鲜度.

Fig.5 Age of information for IoT devices after the optimization of SAC algorithm.

图 5 展示了 SAC 算法优化后物联网设备信息新鲜度。图中信息新鲜度函数的分布符合预期，即在某时间段内物联网设备数据被无人机收集时，信息新鲜度降到低值；但若未被收集，信息新鲜度将在上一个时间段基础上累积增加。由图 5 可知，仅有 2 个物联网设备信息新鲜度超过 40，即在 40 个时间段内未被收集信息。其余物联网设备在 40 个时间段内都至少被收集信息一次。多数物联网设备信息新鲜度都始终保持在 20 以下。



（a） （b）

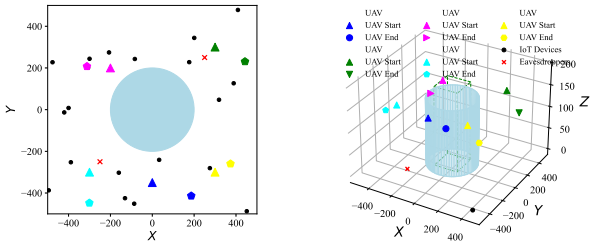
图 6 最优无人机个数选择.（a）不同无人机个数下平均信息新鲜度与迭代次数关系；（b）10000 代后最优平均信息新鲜度与无人机个数关系

Fig.6 Optimal UAV number decision. (a) Average age of information versus episode under different UAV numbers; (b) Optimal average age of information versus UAV number after 10000 episodes

为了决定最优无人机个数，不同无人机个数下的平均信息新鲜度随迭代次数关系如图 6（a）所示。由图可以看出在不同无人机个数下的平均信息新鲜度均随迭代次数增加而减小并最终收敛，这证明了不同无人机参数下的实验的正确性。其次，随着无人机个数增加，最终收敛得到的最优平均信息新鲜度降低，这与预期相符合。为清楚起见，图 6（b）展示了 10000 代后最优平均信息新鲜度与无人机个数关系图，结果与图 6（a）相吻合。当无人机数目增加，无人机可收集物联网设备数据范围更广，因此物联网设备的数据在更短时间段内即可被迅速收集，因此平均信息新鲜度会随着新引入无人机而降低。但当无人机数目增大到一定值时，该指标降幅将不再明显。这是由于当无人机数量足够时，无人机仅在初期飞行到最优位置，中后期在最优位置处小幅度盘旋飞行为附近的物联网设备在整个飞行期间提供持续的数据收集服务，SAC 算法训练10^4代后得到的最优网络参数下五架无人机最优飞行轨迹如图 7 所示。当大多数物联网设备的数据被频繁收集时，信息新鲜度的降幅将不再明显，但无人机耗能和算法复杂度却仍然线性或指数增加。由图 6 可知，在本实验中当无人机数目达到 3 时，增加无人机个数将不再显著降低平均信息新鲜度，因此在该实验场景中最优无人机数目为 3。

Comment [YL42]: Figure
Comment [YL43]: Figure
Comment [YL44]: Figure
Comment [YL45]: Figure
Comment [YL46]: Figure
Comment [YL47]: Figure
Comment [YL48]: Figure
Comment [YL50]: Figure
Comment [YL49]: Figure

录用稿件，非最终出版稿



（a） （b）

图 7 五架无人机最优飞行轨迹.（a）二维无人机最优轨迹；（b）三维无人机最优轨迹

Fig.7 UAV trajectories with five UAVs. (a) Two-dimensional optimal UAV trajectories; (b) Three-dimensional optimal UAV trajectories

Comment [YL51]: Figure

4 结论

本文提出了一种基于深度强化学习的无人机时空众包资源分配方法。实验结果表明本文所提出的 SAC 算法在解决时空众包资源分配任务上是有效的。另外，SAC 算法在收敛性上优于另外两种先进的强化学习算法，且均优于随机算法。最后，本文分析了最优无人机数目选择方法并给出了在仿真实验场景下的最优无人机数目。在未来工作中，定量地确保窃听者无法获得信息将被研究。

录用稿件，非最终出版稿

参 考 文 献

[1] Wang C X , You X, Gao X, et al. On the road to 6G: Visions, requirements, key technologies and testbeds. *IEEE Commun Surveys Tut*, 2023, 25(2): 905

[2] Zhang Q, Wang Y, Yin G, et al. Two-stage bilateral online priority assignment in spatio-temporal crowdsourcing. *IEEE Trans Serv Comput*, 2022, 16(3): 2267

[3] Qi J, Wang W, Chen M, et al. Concept, architecture and key technologies of industrial internet. *Chinese J Internet Things*, 2022, 6(02): 38
(亓晋,王微,陈孟玺,等. 工业互联网的概念、体系架构及关键技术. 物联网学报, 2022, 6(02): 38)

[4] Zhang H, Jiang M, Liu X, et al. PPO-based PDACB traffic control scheme for massive IoV communications. *IEEE Trans Intell Transp Syst*, 2022, 24(1): 1116

[5] Chen H, Liang J, Ma Y. Research on standardization of value creation of energy big data. China Standardization, 2023, (17): 35
(陈浩敏,梁锦照,马赟. 能源大数据技术发展趋势及标准化动向研究. 中国标准化, 2023, (17): 35)

[6] Wei Z, Zhu M, Zhang N, et al. UAV-assisted data collection for internet of things: A survey. *IEEE Internet Things J*, 2022, 9(17): 15460

[7] Zhang H, Huang M, Zhou H, et al. Capacity maximization in RIS-UAV networks: A DDQN-based trajectory and phase shift optimization approach. *IEEE Trans Wireless Commun*, 2022, 22(4): 2583

[8] Zhang H, Song W, Liu X, et al. Intelligent channel prediction and power adaptation in LEO constellation for 6G. *IEEE Netw*, 2023, 37(2): 110

[9] Guo H, Wang Y, Liu J, et al. Multi-UAV cooperative task offloading and resource allocation in 5G advanced and beyond. *IEEE Trans Wireless Commun*, 2023, 23(1): 347

[10] Mahmood A, Vu T X, Chatzinotas S, et al. Joint optimization of 3D placement and radio resource allocation for per-UAV sum rate maximization. *IEEE Trans Veh Technol*, 2023, 72(10): 13094.

[11] Xu X, Zhao H, Yao H, et al. A blockchain-enabled energy-efficient data collection system for UAV-assisted IoT. *IEEE Internet Things J*, 2020, 8(4): 2431

[12] Sun M, Xu X, Qin X, et al. AoI-energy-aware UAV-assisted data collection for IoT networks: A deep reinforcement

learning method. *IEEE Internet Things J*, 2021, 8(24): 17275

[13] Khodaparast S S, Lu X, Wang P, et al. Deep reinforcement learning based energy efficient multi-UAV data collection for IoT networks. *IEEE Open J Veh Technol*, 2021, 2: 249

[14] Xu W, Xiao T, Zhang J, et al. Minimizing the deployment cost of UAVs for delay-sensitive data collection in IoT networks. *IEEE/ACM Trans Network*, 2021, 30(2): 812

[15] Zhang J, Li Z, Xu W, et al. Minimizing the number of deployed UAVs for delay-bounded data collection of IoT devices. *Proceedings of IEEE Conf Comput Commun*. Vancouver, 2021: 1

[16] Gao Y, Liu M, Mei Z, et al. Deep reinforcement learning based UAV trajectory design for data collection scenario with no-fly zones // *Proceedings of IEEE 8th Int Conf Comput Commun*. Chengdu, 2022: 765

[17] Wang Y, Gao Z, Zhang J, et al. Trajectory design for UAV-based Internet of Things data collection: A deep reinforcement learning approach. *IEEE Internet Things J*, 2021, 9(5): 3899

[18] Wang Z, Liu R, Liu Q, et al. Energy-efficient data collection and device positioning in UAV-assisted IoT. *IEEE Internet Things J*, 2019, 7(2): 1122

[19] Samir M, Sharafeddine S, Assi C M, et al. UAV trajectory planning for data collection from time-constrained IoT devices. *IEEE Trans Wireless Commun*, 2019, 19(1): 34

[20] Li Y, Liang W, Xu W, et al. Data collection maximization in IoT-sensor networks via an energy-constrained UAV. *IEEE Trans Mobile Comput*, 2021, 22(1): 159

[21] Al-Hilo A, Samir M, Elhattab M, et al. RIS-assisted UAV for timely data collection in IoT networks. *IEEE Syst J*, 2022, 17(1): 431

[22] Tyrovolas D, Mekikis P V, Tegos S A, et al. Energy-aware design of UAV-mounted RIS networks for IoT data collection. *IEEE Trans Commun*, 2022, 71(2): 1168

[23] Yi M, Wang X, Liu J, et al. Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks. *Proceedings of IEEE Conf Comput Commun Work*. Toronto, 2020: 716

[24] Hu H, Xiong K, Qu G, et al. AoI-minimal trajectory planning and data collection in UAV-assisted wireless powered IoT networks. *IEEE Internet Things J*, 2020, 8(2): 1211

[25] Zhang Y, Duan H, Wei C. Digital twin-based obstacle avoidance method for unmanned aerial vehicle formation control using deep reinforcement learning. *Chinese J Eng*, 2024, 46(7): 1187
（张宇宸, 段海滨, 魏晨. 基于深度强化学习的无人机集群数字孪生编队避障. 工程科学学报, 2024, 46(7): 1187）

[26] Li M, Tao X, Li N, et al. Secrecy energy efficiency maximization in UAV-enabled wireless sensor networks without eavesdropper's CSI. *IEEE Internet Things J*, 2021, 9(5): 3346

[27] Ou Y, Guo Z, Luo D, et al. Collaborative air combat maneuvering decision-making method based on graph convolutional deep reinforcement learning. *Chinese J Eng*, 2024, 46(7): 1227
（欧洋, 郭正玉, 罗德林, 等. 基于图卷积深度强化学习的协同空战机动决策方法. 工程科学学报, 2024, 46(7): 1227）

录用稿件，非最终出版稿