DOI:10.19651/j.cnki.emt.2209330

基于数据辅助的无人机集群协同空域抗干扰*

姚昌华 高泽邻 韩贵真 安 蕾
(南京信息工程大学电子与信息工程学院 南京 210044)

摘 要: 研究移动中的无人机通过动态感知和学习干扰来波方向,实时调整波束成形策略来抑制干扰。针对实际场景中无人机不能获得干扰者的全部动作来进行策略训练的问题,提出使用集群内部协作收集干扰机动作数据从而补充训练数据的方法来从而提升集群抗干扰。将波束成形决策建模为马尔可夫决策过程,基于深度强化学习架构,提出了基于数据辅助的无人机集群协同空域抗干扰算法。仿真结果表明,在辅助数据分别达到 40%,60%,80%时,系统吞吐量分别得到 33%,55%,70%的提升,验证了本文提出的方法能有效提高无人机协同抗干扰能力。

关键词: 无人机集群;数据辅助;深度强化学习;抗干扰;波束成形
中图分类号: TN929.5;V279 **文献标识码:** A **国家标准学科分类代码:** 510.5015

Cooperative airspace anti-jamming of UAV cluster based on data assistance

Yao Changhua Gao Zehe Han Guizhen An Lei
(School of Electronics & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044)

Abstract: Studies the moving UAV to suppress the interference by dynamically sensing and learning the interference wave direction, and adjusting the beamforming strategy in real time. Aiming at the problem that the UAV can not obtain all the actions of the jammer for strategy training in the actual scene, a method of using the cooperation within the cluster to collect the action data of the jammer to supplement the training data is proposed to improve the anti-jamming of the cluster. The beamforming decision is modeled as a Markov decision process. Based on the deep reinforcement learning architecture, a data Aided Cooperative spatial anti-jamming algorithm for UAV cluster is proposed. The simulation results show that when the auxiliary data reaches 40%, 60% and 80%, the system throughput is improved by 33%, 55% and 70% respectively. It is verified that the method proposed in this paper can effectively improve the cooperative anti-jamming ability of UAV.

Keywords: UAV cluster;data assistance;deep reinforcement learning;anti-interference;beamforming

0 引 言

随着科技的不断发展,无人机也在飞速发展,能够在越来越多的地方发挥作用,无人机一般分为军用、民用和消费级三大类。无人机产业正在快速发展当中,就“十三五”国家战略性新兴产业发展规划[1]而言,不仅要开发市场需要的无人机,而且要可靠、低成本、适应环境,跟无人机相关的基础设施建设也要继续完善,无人机产业已经得到了国家的高度重视。无人机网络相对传统通信网络而言,其具有便携性和可移动性等优点。无人机网络也有很多不同问题值得讨论,例如宽带分配、协同合作等,一些文献对此进行了的分析[2]。而无人机发展面临的新威胁却是无人机网络能够受到恶意干扰,这极大的增加了无人机正常运行,执行任务的风险。不同于基站等终端,无人机往往在远离其控制站点的地点执行任务,这使得它容易受到欺骗、干扰和窃听等攻击,导致重要信息丢失,甚至无法满足任务需求,任务中断。

物理层抗干扰的主要思想是尽可能增加接收到的 SJNR。从物理层资源的角度来看,物理层抗干扰方法可以分为 4 个域:码域,功率域,频域和空域。通过利用诸如 BPSK 的低速率调制和编码方案,码域抗干扰可以降低所需的解调阈值,从而提高抗干扰能力[3-4]。功率域抗干扰通过增加传输功率来增加 SJNR[5-6]。频域中的抗干扰方法通过切换到无干扰的信道来规避干扰[7-8]。然而,当干扰器的发射功率较高并且干扰信号可以覆盖整个通信频带时,上

收稿日期:2022-03-19
* 基金项目:国家自然科学基金(61971439)、江苏省自然科学基金(BK20191329)、中国博士后科学基金(2019T120987)、南京信息工程大学人才启动经费(2020r100)项目资助

述3种抗干扰方法将失败。

空域抗干扰借助于MIMO技术[9]的空间分集,以增强有用信号的接收功率并消除对抗性干扰[10]。假设接收机和干扰机之间的信道状态信息为接收机所知,接收机可以通过波束形成技术有效地增加接收到的SJNR。

现有大多数工作都是在讨论接收机和干扰机之间是连续实施干扰,接收机能够获取几乎所有干扰信道的信道状态信息。而实际情况下接收机并不能知晓干扰信号所有的来波方向,因此不可能收集所有的干扰机动作数据。在数据缺少的情况下,接收机收集到的干扰机动作轨迹是不完整的,如果缺少一部分数据来进行数据分析,接收机不能每次都准确的估计干扰信道的信道状态信息,造成抗干扰决策训练的不充分,抗干扰性能必然会下降。对于干扰机动作数据的缺失,本文提出通过加入辅助无人机收集干扰信号的数据,间接补充接收无人机获得的干扰机动作数据,接收机获得的干扰机动作数据增加,接收机准确估计干扰信道的信道状态信息的概率就会变大,提高接收无人机抗干扰性能。

1 系统模型和问题建模

1.1 系统模型

如图1所示,系统模型中主要包含了无人发送机、无人接收机、干扰机和无人辅助机各一个。每个单位均配置均匀线性天线阵列。干扰机处于静止状态,向周围某个随机方向进行干扰,如果接收无人机刚好处于干扰范围内,那么接受无人机将会被干扰,同时也收集到干扰机动作数据。发送无人机,接受无人机和辅助无人机依据集群规划的轨迹飞行。发送无人机的数据信号到接收无人机的DoA(到达方向角)为30°,辅助无人机受到干扰后将干扰信号的方向储存,经过一定时间后将储存的干扰机动作数据传输给接受无人机。

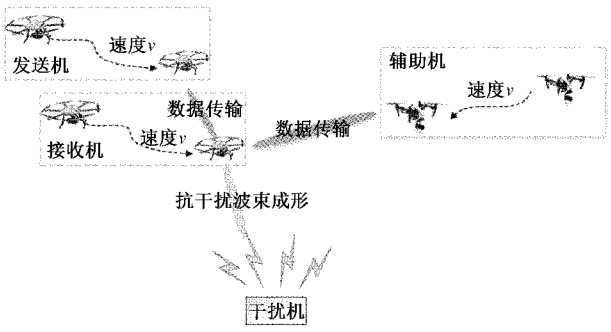

图1 空域抗干扰系统模型

将MIMO信号用物理信道模型来表示[8],接收机r到干扰机j的离开方向角(direction of departure, DoD)和到达方向角(direction of arrival,DoA)可以分别用 $\varphi_{r,j}$ 和 $_{r,j}$ 表示。接收机与干扰机之间的信道可以表示为[11]:

$$\boldsymbol{H}_{r,j}=\mu_{r,j}\mathrm{e}^{-j2\pi d_{r,j}/\lambda_c}\boldsymbol{e}_j(\theta_{r,j})\boldsymbol{e}_r^H(\varphi_{r,j}) \tag{1}$$

其中,$\mu_{r,j}$ 为接收机与干扰机之间的路径损耗,$d_{r,j}$ 为接收机与干扰机天线之间的距离,λ_c 为载波波长。发送机和接收机的天线阵列响应向量可以用 $\boldsymbol{e}_r(\varphi_{r,j})$ 与 $\boldsymbol{e}_j(\theta_{r,j})$ 表达:

$$\begin{aligned}\boldsymbol{e}_r(\varphi_{r,j})&=[1,\mathrm{e}^{j2\pi\Delta_r\cos\varphi_{r,j}},\cdots,\mathrm{e}^{j2\pi(N_r-1)\Delta_r\cos\varphi_{r,j}}]^{\mathrm{T}}\\ \boldsymbol{e}_j(\theta_{r,j})&=[1,\mathrm{e}^{-j2\pi\Delta_j\cos\theta_{r,j}},\cdots,\mathrm{e}^{-j2\pi(N_j-1)\Delta_j\cos\theta_{r,j}}]^{\mathrm{T}}\end{aligned} \tag{2}$$

其中,Δ_r 和 Δ_j 表示天线间隔距离,N_r 和 N_j 分别为接收机和干扰机的天线数。在信号传输阶段,接收机接收到的信号可以表示为:

$$\boldsymbol{y}=\sqrt{P_t}\boldsymbol{H}_{t,r}\boldsymbol{x}_t+\sqrt{P_j}\boldsymbol{H}_{j,r}\boldsymbol{x}_j+\boldsymbol{n} \tag{3}$$

其中,t 为发送机,$\sqrt{P_t}\boldsymbol{H}_{t,r}\boldsymbol{x}_t$ 表示为接收机期望接收到的有用信号,$\sqrt{P_j}\boldsymbol{H}_{j,r}\boldsymbol{x}_j$ 表示为干扰信号。$\boldsymbol{x}_t=\boldsymbol{w}_t z_t$ 表示有用信号,$\boldsymbol{w}_t$ 为发送波束成形(预编码)向量,对于发送机而言,其发送功率有限制 $\|\boldsymbol{w}_t\|=1$。$\boldsymbol{x}_j=\boldsymbol{w}_j z_j$ 表示干扰信号,$\boldsymbol{w}_j\in\mathbb{C}^{N_j\times 1}$ 为干扰方为提升干扰性能的预编码向量,干扰机干扰功率限制 $\|\boldsymbol{w}_j\|=1$。z_t 和 z_j 分别表示为发送符号和干扰符号,限制均为 $E[|z|^2]=1$。$\boldsymbol{n}$ 是加性高斯白噪声矢量。接收机通过接收波束成形向量处理接收到的信号,接收滤波向量用 $\boldsymbol{f}$ 表示,即:

$$\hat{z}_s=\boldsymbol{f}^H\boldsymbol{y}=\sqrt{P_t}\boldsymbol{f}^H\boldsymbol{H}_{t,r}\boldsymbol{w}_t z_t+\sqrt{P_j}\boldsymbol{f}^H\boldsymbol{H}_{j,r}\boldsymbol{w}_j z_j+\boldsymbol{f}^H\boldsymbol{n} \tag{4}$$

信号与干扰噪声的比值能比较直观表现出通信质量,接收端输出的SINR可以写成:

$$SINR=\frac{E\left[\left|\sqrt{P_t}\boldsymbol{f}^H\boldsymbol{H}_{t,r}\boldsymbol{w}_t z_t\right|^2\right]}{E\left[\left|\sqrt{P_j}\boldsymbol{f}^H\boldsymbol{H}_{j,r}\boldsymbol{w}_j z_j+\boldsymbol{f}^H\boldsymbol{n}\right|^2\right]}=\frac{P_t\boldsymbol{f}^H\boldsymbol{H}_{t,r}\boldsymbol{w}_t\boldsymbol{w}_t^H\boldsymbol{H}_{t,r}^H\boldsymbol{f}}{\boldsymbol{f}^H(P_j\boldsymbol{H}_{j,r}\boldsymbol{w}_j\boldsymbol{w}_j^H\boldsymbol{H}_{j,r}^H+\sigma_n^2\boldsymbol{I})\boldsymbol{f}} \tag{5}$$

只有满足 $SINR>\lambda$,接收机才能正确解调接收到的信号,其中 λ 为解调门限。令 $\boldsymbol{R}_j=P_j\boldsymbol{H}_{j,r}\boldsymbol{w}_j\boldsymbol{w}_j^H\boldsymbol{H}_{j,r}^H+\sigma_n^2\boldsymbol{I}$,那么 $\boldsymbol{f}^H\boldsymbol{R}_j\boldsymbol{f}$ 就能表示接收机接收到的所有干扰和噪声信号。此传输速率可以表示为:

$$r=\begin{cases}\log_2\left(1+\dfrac{P_t\boldsymbol{f}^H\boldsymbol{H}_{t,r}\boldsymbol{w}_t\boldsymbol{w}_t^H\boldsymbol{H}_{t,r}^H\boldsymbol{f}}{\boldsymbol{f}^H\boldsymbol{R}_j\boldsymbol{f}}\right), & SINR<\lambda\\ 0, & SINR>\lambda\end{cases} \tag{6}$$

抗干扰波束成形方法是最小方差波束成形方法(minimum-variance beamforming scheme, MVBS)[12]。利用MIMO通信设计波束成形,主要目的还是使SINR最大化即:

$$\begin{aligned}&\max_{\boldsymbol{f},\boldsymbol{w}_t}SINR\\ &\text{s.t.}\quad\|\boldsymbol{w}_t\|=1.\end{aligned} \tag{7}$$

通过使干扰和噪声功率最小来设计接收滤波向量 $\boldsymbol{f}$,然后再计算预编码向量 $\boldsymbol{w}_t$[13],这样就可以同时生成较优的收发信机滤波向量。

对接收到的干扰信号数据进行处理,然后来估计下一

时刻干扰信道的瞬时信道状态信息。当发送无人机不工作时，接受无人机接收到的就只有干扰和噪声，接收机接收到信号为 $\boldsymbol{y}_j = \boldsymbol{H}_{j,r}\boldsymbol{w}_j z_j + \boldsymbol{n}$。对干扰和噪声信号进行处理，$\boldsymbol{R}_j$ 可以表示为：

$$\hat{\boldsymbol{R}}_j = \frac{1}{M}\sum_{n=1}^{M}\boldsymbol{y}_j(n)\boldsymbol{y}_j(n)^H \tag{8}$$

其中，M 为采样数。对该信号进行处理，可以分解为干扰子空间 $\hat{\boldsymbol{U}}_j$ 和噪声子空间 $\hat{\boldsymbol{U}}_n$[14]：

$$\hat{\boldsymbol{R}}_j = \hat{\boldsymbol{U}}_j\boldsymbol{\Lambda}_j\hat{\boldsymbol{U}}_j^H + \sigma_n^2\hat{\boldsymbol{U}}_n\hat{\boldsymbol{U}}_n^H \tag{9}$$

干扰的特征值对角矩阵为 $\boldsymbol{\Lambda}_j$。为了消除大部分干扰信号造成的影响，通过在噪声子空间中来设计滤波向量来实现。如果接受无人机接收到的干扰信号到达方向角与有用信号到达方向角接近时，这将会压制有用信号的功率。所以通过添加一个条件来计算最优接收滤波向量，以减少对有用信号的抑制影响：

$$\begin{aligned}&\min_{\boldsymbol{f}}\ \boldsymbol{f}^H\boldsymbol{R}_j\boldsymbol{f}\\&\text{s.t.}\ \ \boldsymbol{f}^H\boldsymbol{e}_r(\theta_{s,r}) = 1\end{aligned} \tag{10}$$

其中，$\boldsymbol{e}_r$ 为接收机的阵列响应向量。最优滤波向量在通过 DoA 为 $\theta_{s,r}$ 的信号时，抑制干扰和噪声的显著能量。为了解决约束优化问题，可以求助于拉格朗日乘子[15]，最优滤波向量可以表示为：

$$\boldsymbol{f}_{MVBS} = \frac{\boldsymbol{R}_j^{-1}\boldsymbol{e}_r(\theta_{t,r})}{\boldsymbol{e}_r^H(\theta_{t,r})\boldsymbol{R}_j^{-1}\boldsymbol{e}_r(\theta_{t,r})} \tag{11}$$

由此可以求得估计的最优接收滤波向量，然后通过最大比传输方法来获得预编码向量，即 $\boldsymbol{w}_t = \boldsymbol{H}_{t,r}^H / \| \boldsymbol{H}_{t,r}^H \|$。干扰机也会用此方法对接收机进行干扰，即 $\boldsymbol{w}_j = \boldsymbol{H}_{j,r}^H / \| \boldsymbol{H}_{j,r}^H \|$。

1.2　时隙结构

时隙结构如图 2 所示。在感知阶段 T_S，接收机感知干扰的 DoA，此时发送机不向接收机发送有用信号。当进入数据传输阶段 T_{DT} 时，发送机开始将数据传输到接收机。每隔一段时间增加一个辅助数据传输 T_{DTA}，辅助机将接收到的干扰机动作数据发送到接收机。辅助数据传输和数据传输的时间之和与以往的数据传输的时间一致。在学习阶段 T_L，接收机运行抗干扰算法并生成下一个时隙的接收滤波向量 f。在发送机向接收机传输数据后，ACK 传输阶段 T_{ACK} 将确认接收机是否接收到数据。此阶段还将决定下一个时隙的滤波向量。

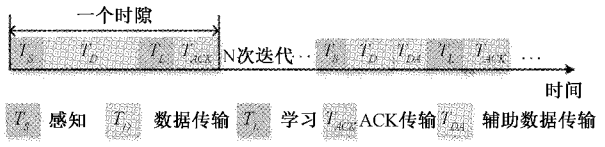

图 2　时隙结构

1.3　问题建模

对于收发信机而言，传输速率通常是越大越好，为了能使传输速率最大，接收机就需要找到较优的滤波向量。策略 π 是空间谱 m_t 到动作的映射函数，可以表示为：

$$\pi^* = \underset{\pi}{\text{argmax}} r[\pi(m_t)] \tag{12}$$

因为无人机不是固定一个位置，相对于实际传输过程中所感知到的干扰信道而言，感知阶段所感知的干扰信道具有一定的延迟，通过动态空间算法来减少时延带来的影响。

2　基于深度强化学习的动态空间谱抗干扰方法

随着收发无人机位置的变化，接收无人机需要实时调整其波束成形方向。对于动态环境下的序列决策问题，可以采用 MDP 来建模[16-17]。

2.1　马尔可夫决策过程框架

一个马尔可夫决策过程可以由状态，动作，奖励函数和状态转移概率组成 (S, A, r, P_r)。如果一组 (s_t, a_t) 转移到了下个状态 s_{t+1}，那么奖励函数可记为 $r(s_{t-1} \mid s_t, a_t)$。P_r 表示的是在状态 s_t 下，执行 a_t 后，会转移到的其他状态的概率。一个从环境状态到动作的映射（即行为策略），记为策略 $\pi: S \to A$。

而强化学习奖励往往不是立即产生的，需要一定的时间反馈，就如在第 n 步取得奖励，但是之前所以奖励都归为 0。所以奖励就不能说明状态和动作选择的好坏。因此需要状态价值函数 $V(s_t) = \max_{a_t \in A} Q(s_t, a_t)$ 来表示该策略能获得的总的期望奖励。MDP 的优化目标为找到最大化期望累积奖励的最优决策策略，即：

$$\max_{\pi} E\left[\sum_{i=1}^{\infty}\gamma^{i-1}r_{t+i}\right] = \max_{\pi} V^{\pi}(s_t) \tag{13}$$

γ 为折扣因子，范围在 0～1。

1）动作：在此模型中，将接收机生成的滤波向量做为动作，由于干扰信号的到达方向角是时刻变化的，可以根据其变化来设定动作，然后根据 MVBS 抗干扰波束成形公式来生成接收机的滤波向量。动作集 A 的大小根据干扰信号到达方向角的范围而变化。

2）奖励：将通信的传输速率设为奖励，通过判断奖励值的大小来评估选择的动作是否合理。根据 SINR 和传输速率公式，如果接收机选择了相对较差的滤波向量，抗干扰效果较差，导致低的奖励值回报。

3）状态：在马尔可夫决策过程中，往往解决的是动态问题。在系统模型中可以看出，生成滤波向量需要知道干扰信号的到达方向角，但是干扰信号的到达方向角是时刻变化的，所以将其设为状态。MUSIC 算法是一种基于矩阵特征空间分解的方法[14]，MUSIC 空间谱由 P_M 表示。在对干扰信号进行分解后，其到达方向角估计公式为：

$$\underset{\theta_{j,r}^t}{\text{argmax}} P_M(\theta_{j,r}^t) = \underset{\theta_{j,r}^t}{\text{argmax}} \frac{\boldsymbol{e}_r^H(\theta_{j,r}^t)\boldsymbol{e}_r(\theta_{j,r}^t)}{\boldsymbol{e}_r^H(\theta_{j,r}^t)\hat{\boldsymbol{U}}_n\hat{\boldsymbol{U}}_n^H\boldsymbol{e}_r(\theta_{j,r}^t)} \tag{14}$$

2.2　深度强化学习框架

DQN 算法流程如图 3 所示。将输入 s_t 进行预处理，送

入神经网络,输出每个动作的 Q 值。根据 $\in-greedy$ 算法计算应该采取的动作 $a_t = arg\max_a Q(s_{t-1},a';\theta^-)$。执行动作 a_t,转移到下一状态 s_{t+1},收到环境给出的回报,将经验 (s_t,a_t,r_t,s_{t+1}) 存入经验池。从经验池进行一个 batch 的采样,训练神经网络。神经网络(当前网络)训练若干次之后,目标网络将复制其中的参数[18]。

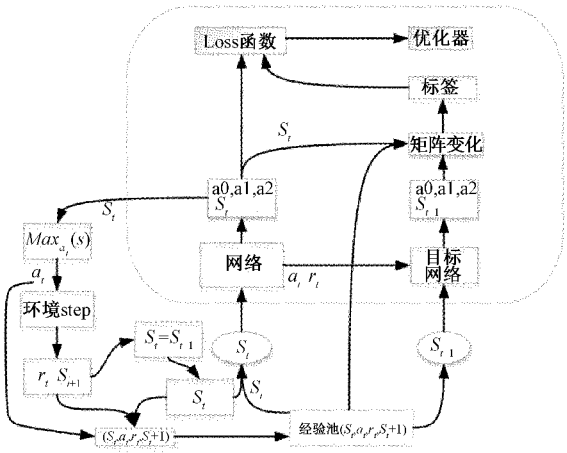

图 3 DQN 算法流程

在神经网络当中,将带有权值 θ 的神经网络称为 Q 网络[19]。在训练过程中,通过调整权重 θ 来减少贝尔曼方程中的均分误差,最佳目标值将会被近似目标值 $y_i = r + \gamma\max_{a'}Q(s',a';\theta_{i-1})$ 替代,θ_{i-1} 为之前过程中一些网络权重参数。最佳目标值与近似目标值产生了差值,因此产生了损失函数:

$$L_i(\theta_i) = \mathbb{E}_{s,a,r,s'}[y_i - Q(s,a;\theta_i)]^2 \tag{15}$$

值得注意的是,目标值取决于网络权重参数。在优化损失函数的同时,权重 θ_i 也会与上一次的权重参数 θ_{i-1} 相同,这导致了一系列的优化问题。通过对损失函数相对于权重的微分,得到了下面的梯度:

$$\nabla_{\theta_i}L_i(\theta_i) = \mathbb{E}_{s,a,r,s'}[(y_i - Q(s,a;\theta_i))\nabla_{\theta_i}Q(s,a;\theta_i)] \tag{16}$$

经验回放:在深度强化学习中经验一般表示为 (s_t,a_t,r_t,s_{t+1})。一开始,算法初始化一个经验池 D,容量为 $|D|$,并通过 $\in-$贪婪策略将部分经验填充至其中。经验池存储了带标签的一个个数据样本,训练神经网络是需要带标签的样本。然后,算法从 D 中随机抽取经验(即批次)进行 Q 网络训练。这种机制的优点是在训练过程中,神经网络通过随机抽样解决了训练数据之间的相关性和非静态发布问题。

根据数据辅助的需要和实际过程,本文算法初始化同时生成一个存放辅助经验的辅助机 D1,容量也为 $|D|$,但是与经验进入经验池(接收机)D 中不同,已经进入经验池(接收机)D 中的经验不会再次进入辅助机 D1,未进入经验池(接收机)D 中的部分经验才会进入辅助机 D1,以控制不会有重复的经验。每经过 n 次迭代后,将辅助机 D1 的经验补充至经验池(接收机)D。此过程即为辅助机将接收到的空间谱数据传输给接收机。与此同时,辅助机的经验清空。如果经验池(接收机)D 中经验数超过容量,那么会删除较早的一部分经验,这样以保持经验池(接收机)D 中的经验都是最新的。

目标 2 网络:在训练过程中使用单独的神经网络来生成目标 Q 值,该网络被称为目标 Q 网络 2,且拥有与 2$(s_t,a_t;\theta)$ 相同的网络结构。算法根据损失函数的更新公式来更新网络中的参数,目标网络中的参数在训练过程中不变,每训练 N 次,复制 Q 网络中的参数。使用目标 Q 网络,能够让目标 Q 值保持一段时间不发生变化,使其减少与估计 Q 值的关联,从而使得训练时损失值震荡发散的可能性降低,算法的稳定性提高。

2.3 基于深度强化学习的动态空间谱抗干扰方法

1) 动态空间谱构建

由于马尔可夫决策过程一般是用来解决动态问题的,将 t 时刻的空间谱表示为 $m_t = [o_1,o_2,\cdots,o_s]^T$,其中 o_s 为空间谱在角度为 $180\times s/S$ 度时的观测值,S 为空间角度分辨率。构建动态空间谱为 $s_t = S_t = [m_t,m_{t-1},\cdots,m_{t-h+1}]^T$,$s_t$ 是此算法的状态。动态空间谱的矩阵大小为 $h\times S$,h 表示拥有 h 个时隙的空间谱数据。在 $t+1$ 时刻,$t+1$ 的空间谱 m_t 进入动态空间谱,与此同时,删除较早的空间谱,达到更新空间谱的目的。

2) 神经网络结构

卷积神经网络在图像处理方面发挥着重要的作用[20]。采用卷积层提取空间谱特征,采用全连接层来提高 Q 值函数的逼近能力。Q 网络的网络结构如图 4 所示。

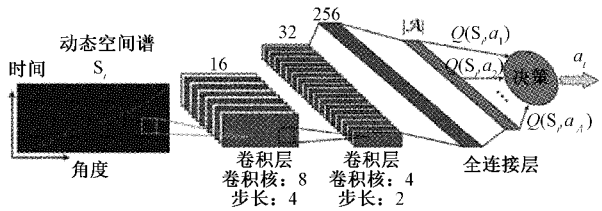

图 4 深度 Q 网络结构

第一层卷积层对动态空间谱进行卷积输出,卷积核为 8 * 8,步长为 4,共 16 个。第二层卷积层对第一层卷积层进行卷积,卷积核为 4 * 4,步长为 2,共 32 个。然后进入全连接层,全连接层对卷积层进行分类,最终将获得每个动作的奖励组合在一起。另外,网络每一层后面都有一个激活函数,本文采用修正线性函数。具体算法如算法 2.1 所示。

算法 2.1：基于数据辅助的无人机集群协同空域抗干扰算法

1. 初始化：生成 Q 网络，权重 θ 随机赋值，目标网络 $\hat{Q}$，权重为 $\theta^{-}=\theta$。根据公式 f_{MVBS}，以不同的干扰 DoA $\theta_{j,r}$，生成动作集 A。
2. 循环：$t=1,2,\cdots,T$
3. 基于以下策略选择动作：
4. 以概率 ε 随机选择接收滤波向量 $\boldsymbol{a}_t$。
5. 以概率 $1-\varepsilon$ 贪婪选择接收滤波向量，即 $\boldsymbol{a}_t=arg\max_a Q(s_{t|1},a';\theta)$。
6. 感知当前空间谱 o_{t-1}。用接收滤波向量 $\boldsymbol{a}_t$ 接收信号，获得通信速率奖励 r_t。
7. 更新 $s_{t|1}=S_{t-1}$，将一部分经验 (s_t,a_t,r_t,s_{t-1}). 存入经验池 D 中。
8. 辅助机把接收到的干扰来波方向数据存储到自己的辅助数据集中；
9. 计时到辅助机传输辅助数据时刻，辅助机将其辅助数据集中的干扰来波方向数据传送给接收机。
10. 接收机接收到辅助机传来的辅助数据，用来更新经验池 D。
11. 接收机从经验池 D 中对经验进行随机批次采样。令 $y_i=r_j+\gamma\max_{a'}Q(s_{j-1},a';\theta)$，计算 $\nabla_\theta L_k(\theta)$ 并更新 θ。
12. 每训练 C 次，令 $\theta^{-}=\theta$。
13. 结束循环

3　仿真分析

本节通过仿真结果证明该方案的可行性。假设无人机是按照设定好的轨迹飞行。通过模拟随机游走来生成 15 条无人机轨迹，其中一条轨迹如图 5 所示。在每次训练过程中，无人机都会从生成的 15 条不同的轨迹中随机选择一条轨迹。干扰机的功率 P_j 为 20 dB，发送机的功率 P_t 为 10 dB。解调阈值 $\lambda=5$ dB。设定一个时隙通信需要 0.1 s，其中 T_S 为 0.03 s、T_D 为 0.05 s、T_L 为 0.01 s 和 T_{ACK} 为 0.01 s。当有辅助数据传输时，辅助数据传输与数据传输时间总和为 0.05 s，整个时隙总体时间不变。

在此算法中，将折扣率设为 0.8。假设干扰信号到达方向角的范围是 1°～180°，接收机每隔 3°生成一个滤波向量，所以此动作集大小为 60。动态空间谱每次拥有 10 个时隙的空间谱数据，空间谱矩阵大小为 10×180。经验池 D 和 D1 容量为 $M=5\,000$。本文采用 ADAM 优化器[21]来训练网络，每次迭代从经验池中采样的样本数为 32。总迭代次数各不相同，均用来训练 Q 网络。随机探索概率 ε 随着迭代次数的增加从 1 线性递减至 10^{-3}。目标网络 $\hat{Q}$ 的权重每 100 次更新 1 次。

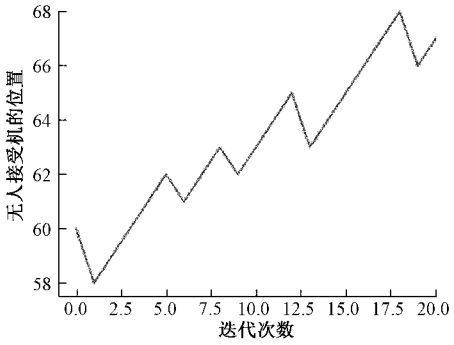

图 5　随机游走轨迹

3.1　算法收敛性分析

在迭代过程开始时，接收机随机地从动作集 A 选取滤波向量。随机选择动作会随着迭代次数的增加慢慢减少，对于每个状态，接收机就会根据之前获得的经验来选择能否获得高收益的动作，从而提高吞吐量性能。

图 6 显示了训练过程中平均 Q 值的变化，该变化是通过取 Q 网络输出的平均值获得的，其中每个点是每 1 000 次迭代的滑动平均值。通过图 6 可以看出，曲线逐渐趋向收敛，验证了该算法是收敛的。

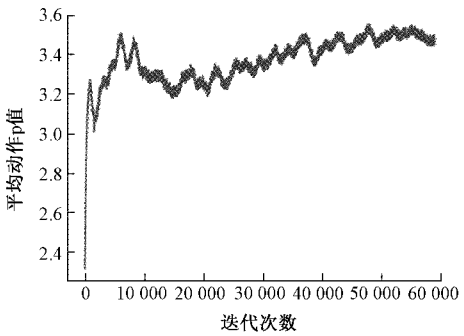

图 6　平均 Q 值收敛性能

3.2　传输速率分析

图 7 给出了接收无人机在不同程度数据辅助的干扰机动作数据量下平均传输速率性能对比，其中每个点是每 1 000 次迭代的滑动平均值。如图 7 所示，假设接收机一开始只有 20%的干扰机动作数据时，抗干扰性能是较差的。辅助机通过数据传输将接收到 20%的干扰机动作数据发送给接收机，即接收机拥有 40%的干扰机动作数据，抗干扰性能提高约 33%。辅助机再通过数据传输将接收到 20%的干扰机动作数据发送给接收机，即接收机拥有 60%的干扰机动作数据，抗干扰性能总体比例提高约 55%。辅助机再通过数据将发送接收到 20%的干扰机动作数据传输给接收机，即接收机拥有 80%的干扰机动作数据，由于接收机已经拥有大部分干扰机动作数据，抗干扰性能总体比例提高约 70%，但不如干扰机动作数据少有数据辅助时抗干扰性能提升明显。当辅助机通过数据传输将接收到的干扰机动作数据发送给接收机，使得接收机拥有 100%干

扰机动作数据时,抗干扰性能也能有所提升。总体而言,辅助机辅助的干扰机动作数据逐渐增多,接收机抗干扰性能也逐渐提高。

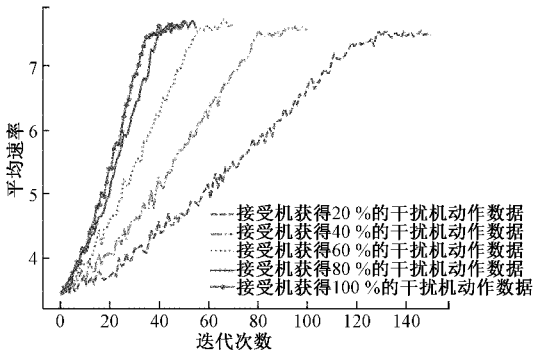

图 7 不同程度数据辅助下平均传输速率对比

图 8 给出了接收机不能得到所有的干扰机动作数据,在辅助机在不同时延情况下补充一部分数据后,平均传输速率性能对比,其中每个点是每 1 000 次迭代的滑动平均值。如图 8 所示,接收机在有辅助机的辅助获得干扰机动作数据的情况下,时延比较低的时候,接收机只需花费较少的时间就能较好地进行抗干扰,就如辅助机每 2 000,5 000 次通过数据传输将接收到 20%的干扰机动作数据发送给接收机,辅助机的作用较大。而时延比较高的时候。再如辅助机每 10 000 次通过数据传输将接收到 20%的干扰机动作数据发送给接收机,这时接收机的抗干扰性能较差,辅助机几乎不起作用。由此可知,辅助机提供干扰机动作数据越快,接收机抗干扰性能提升越快,辅助机提供干扰机动作数据较慢,接收机抗干扰性能提升也相对较慢。

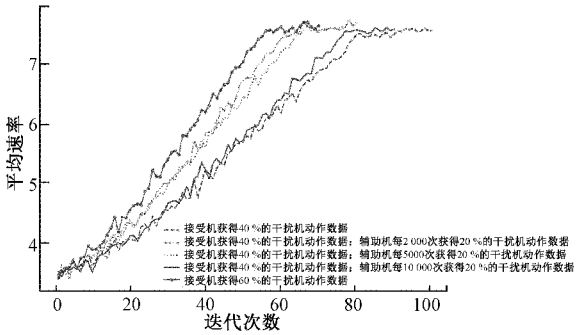

图 8 不同时延辅助下平均传输速率对比

而图 9 给出了接收机只能获得少部分的干扰机动作数据,在辅助机在不同时延情况下补充一部分数据后,平均传输速率性能对比,其中每个点是每 1 000 次迭代的滑动平均值。如图 9 所示,接收机在有辅助机的辅助获得干扰机动作数据的情况下,时延比较低的时候,接收机只需花费较少的时间就能较好地进行抗干扰,就如辅助机每 2 000,5 000 次通过数据传输将接收到 20%的干扰机动作数据发送给接收机,辅助机的作用较大。而时延比较高的时候,再如辅助机每 10 000 次通过数据传输将接收到 20%的干扰机动作数据发送给接收机,相对能获得大部分的干扰机动作数据而言,这时接收机的抗干扰性能也能大幅提升,辅助机也能起到作用。

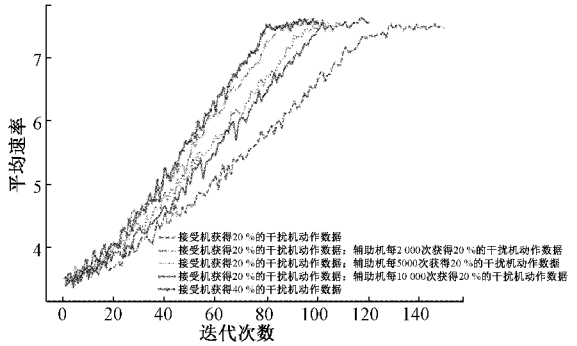

图 9 不同时延辅助下平均传输速率对比

4 结 论

本文研究了无人机集群基于数据辅助的波束成形抗干扰问题。考虑到估计干扰信道的信道状态信息由于移动性而延迟,将抗干扰波束形成决策建模为 MDP。为了在动态场景中寻找最优决策策略,提出了基于深度强化学习的波束成形抗干扰决策框架。进一步,针对深度强化学习在实际场景中缺乏对干扰机的行为数据收集的问题,提出了一种基于数据辅助的无人机集群协同空域抗干扰算法。最后,进行了不同数据量下的传输速率和不同时延辅助下的传输速率仿真实验。仿真验证了本文所提深度强化学习波束成形抗干扰算法的收敛性,以及数据辅助的有效性:辅助机提供的干扰机动作数据越多,接收机抗干扰性能越高。辅助机提供干扰机动作数据越快,接收机抗干扰性能提升越快。

参考文献

[1] “十三五”国家战略性新兴产业发展规划[J]. 中国产经,2016(12):95-97.

[2] GUPTA L, JAIN R, VASZKUN G. Survey of important issues in UAV communication networks [J]. IEEE Communications Surveys & Tutorials, 2016, 18(2): 1123-1152.

[3] FIROUZBAKHT K, NOUBIR G, SALEHI M. On the performance of adaptive packetized wireless communication links under jamming [J]. IEEE Transactions on Wireless Communications, 2014, 13(7):3481-3495.

[4] NOUBIR G, RAJARAMAN R, BO S, et al. On the robustness of IEEE 802.11 rate adaptation algorithms against smart jamming [C]. Acm Conference on Wireless Network Security, ACM, 2011.

[5] YANG D, XUE G, ZHANG J, et al. Coping with a smart jammer in wireless networks: A stackelberg

game approach[J]. IEEE Transactions on Wireless Communications, 2013, 12(8):4038-4047.
[6] XIAO L, CHEN T, LIU J, et al. Anti-jamming transmission stackelberg game with observation errors[J]. IEEE Communications Letters, 2015, 19(6):949-952.
[7] YAO F, JIA L, SUN Y, et al. A hierarchical learning approach to anti-jamming channel selection strategies[J]. Wireless Networks, 2017(9):1-13.
[8] JIA L, F YAO, SUN Y, et al. Bayesian stackelberg game for antijamming transmission with incomplete information[J]. IEEE Communications Letters, 2016, 20(10):1991-1994.
[9] 彭青. 认知无线电 MIMO 中基于博弈论的功率控制算法[J]. 电子测量技术, 2012, 35(11):129-133, DOI: 10.19651/j.cnki.emt.2012.11.033.
[10] 王霖郁, 杨旭, 项建弘. 基于多域联合处理的 MIMO 抗干扰技术[J]. 应用科技, 2019, 46(2):42-46.
[11] CHEN T, LIU J, LIANG X, et al. Anti-jamming transmissions with learning in heterogenous cognitive radio networks[C]. Wireless Communications & Networking Conference Workshops, IEEE, 2015.
[12] YAN Q, ZENG H, JIANG T, et al. MIMO-based jamming resilient communication in wireless networks[C]. Infocom, IEEE, 2014.
[13] ZHANG S, HUANG K, LI X. Adaptive transmit and receive beamforming based on subspace projection for anti-jamming [C]. in 2014 IEEE Military Communications Conference, 388-393, IEEE, 2014.
[14] 任晓航, 单宝堂, 吴昊. 新型快速 DOA 估计算法[J]. 国外电子测量技术, 2016, 35(8):22-25, DOI: 10.19652/j.cnki.femt.2016.08.006.
[15] 郭志军. 拉格朗日乘子法在有约束条件的最优化问题研究[J]. 邢台学院学报, 2013, 28(4):170-171.
[16] 卫星, 陆阳, 朱峰, 等. 基于马尔科夫决策过程的井下无线基站切换策略[J]. 电子测量与仪器学报, 2018, 32(7):108-114, DOI: 10.13382/j.jemi.2018.07.016.
[17] 肖铮. 基于马尔可夫决策过程的算法研究[J]. 河北软件职业技术学院学报, 2021, 23(1): 8-11, DOI: 10.13314/j.cnki.jhbsi.2021.01.003.
[18] 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.
[19] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015: 529-533, DOI: 10.1038/nature14236.
[20] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J]. 计算机应用, 2016, 36(9):2508-2515, 2565.
[21] 杨观赐, 杨静, 李少波, 等. 基于 Dopout 与 ADAM 优化器的改进 CNN 算法[J]. 华中科技大学学报(自然科学版), 2018, 46(7): 122-127, DOI: 10.13245/j.hust.180723.

作者简介

姚昌华, 博士, 教授, 主要研究方向为无人集群优化、无线网络、网络安全、数据分析和人工智能。
E-mail: ych2347@163.com

高泽郜(通信作者), 硕士研究生, 主要研究方向为无人机通信抗干扰、深度强化学习。
E-mail: 787933469@qq.com

韩贵真, 硕士研究生, 主要研究方向为智能无人集群、博弈论。
E-mail: han263840@163.com

安蕾, 硕士研究生, 主要研究方向为智能无人集群、博弈论。
E-mail: 1178535838@qq.com