



计算机工程

Computer Engineering

ISSN 1000-3428, CN 31-1289/TP

《计算机工程》网络首发论文

题目: 基于多域联合的无人机集群认知抗干扰算法
作者: 刘春玲, 刘敏提, 丁元明
DOI: 10.19678/j.issn.1000-3428.0056784
网络首发日期: 2020-01-19
引用格式: 刘春玲, 刘敏提, 丁元明. 基于多域联合的无人机集群认知抗干扰算法. 计算机工程. <https://doi.org/10.19678/j.issn.1000-3428.0056784>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。



基于多域联合的无人机集群认知抗干扰算法

刘春玲¹, 刘敏提^{1,2}, 丁元明^{1,2}

(1.大连大学 信息工程学院, 辽宁 大连 116622; 2.大连大学 通信与网络重点实验室, 辽宁 大连 116600)

摘 要: 针对复杂通信环境中无人机集群网络对抗智能性干扰能力较弱的问题, 基于智能决策理论, 提出一种多域联合的认知抗干扰算法。首先, 基于优势演员-评论家方法, 将无人机视作智能体, 由感知到的环境频谱状态决策出干扰信道; 然后, 基于 Stackelberg 博弈理论, 对中度干扰等级的信道从功率域压制干扰信号, 减少切换信道的时间开销; 最后, 引入簇头协助的方法, 解决由于单个智能体局部频谱感知能力较弱导致信道决策成功率低的问题。仿真结果表明, 所提算法能够给出簇内最佳节点个数, 相比于传统算法, 能够提高接收信号信干噪比, 网络整体抗干扰性能较好。

关键词: 认知抗干扰; 优势演员-评论家; Stackelberg 博弈; 无人机集群; 分布式网络



开放科学(资源服务)标识码(OSID):

Multi-Domain Joint Based UAV Swarms Cognitive Anti-Jamming Algorithm

Liu Chunling¹ Liu Minti^{1,2} Ding Yuanming^{1,2}

(1. College of Information Engineering, Dalian University, Dalian Liaoning 116622, China;

2. Key Laboratory of Communication and Network, Dalian University, Dalian Liaoning 116600, China)

【Abstract】 Aiming at the problem that the Unmanned aerial vehicle (UAV) cluster network has a weak ability to defense intelligent jamming in complex communication environment, based on intelligent decision theory, a multi-domain joint cognitive anti-jamming algorithm is proposed. First, on account of the advantage actor-critic (A2C) method, the UAV is considered as agent, and the channel state is determined by sensing the frequency spectrum. Then, the jamming signal is suppressed by the Stackelberg game method from the power domain on channels with moderate jamming level, which reduces time overhead of channel switching. Finally, the cluster head assistance method is introduced to solve the problem of low channel decision success rate due to the weak local spectrum sensing ability of a single agent. Simulation results show that, compared to the traditional algorithms, the proposed algorithm can give the optimal inter-cluster number, improve the signal-to-noise-to-interference ratio of the received signal and the overall anti-jamming performance of the network.

【Keywords】 Cognitive anti-jamming; Advantage actor-critic; Stackelberg game; UAV swarms; Distributed network

DOI:10.19678/j.issn.1000-3428.0056784.

0 概述

在未来空战中, 无人机集群作战将是重要的作战形式之一, 针对其高动态、网络拓扑结构多变等特点, 采用分布式网络结构, 可提高无人机集群网络的抗毁性。此外, 实现信息安全、可靠传输是其完成任务的关键, 如何确保无人机之间可靠通信, 将成为一项重要研究内容^[1]。

近年来, 如何有效对抗智能性干扰, 提高通信安全, 已成为研究的热点^[1]。在众多抗干扰技术研究中, 认知抗干扰算法已成为研究热点方向之一^[2], 该方法可归纳为如下两类: 一类是基于强化学习理论^[3], 进行可用信道选择, 主动规避干扰信道, 实现频域抗干扰。Slimen F 等人^[4]提出基于协作 Q 学习(Q-learning, QL)的信道选择算法, 可提高数据传输成功率, 但在状态空间规模较大时, 面临“维数灾

基金项目: 装发部领域基金(一般项目)基金(61403110308, 61405180402); 辽宁省自然科学基金指导计划项目: 2019-ZD-0311。

作者简介: 刘春玲(1971-), 女, 教授, 硕士生导师, 研究方向: 智能信号处理与检测; 刘敏提(1993-), 女, 硕士研究生, 研究方向: 强化学习与抗干扰通信; 丁元明(1967-), 男, 教授, 博士生导师, 研究方向: 网络仿真对抗技术与太赫兹通信。

E-mail: liuchunling@dlu.edu.cn

难”的问题^{[5][6]}。针对此问题，文献[7]提出将深度 Q 网络 (DQN) 在线学习算法应用于信道选择。当信道数量较多的时，文献[8][9]利用演员-评论家 (Actor-Critic, AC) 算法进行信道选择，然而 AC 算法的方差较大，算法稳定性较差。另一类是基于博弈论的方法^[10-12]，根据敌我双方的竞争关系，建立功率域抗干扰博弈模型，通过求解博弈均衡，得到最佳传输功率，实现从功率上压制干扰信号以达到抗干扰的目的。以上算法仅从单个频域或功率域角度考虑，针对智能性干扰攻击的灵活性较差^[13]。

基于以上分析，为提高网络抗智能干扰的能力，把功率域和频域抗干扰方法相结合，提出一种基于优势演员-评论家 (Advantage Actor-Critic, A2C)^[14] 与 Stackelberg 博弈 (Stackelberg Game, SG) 的多域联合认知抗干扰 (Multi-Domain Joint Cognitive Anti-Jamming, MDJC-AJ) 算法。该方法首先将可用信道探索问题转化为序贯决策问题，由感知到的环境频谱状态，进行信道选择；然后，根据设定的干扰容忍双阈值将信道干扰程度分为严重、中度、轻微三个等级，并对处于中度干扰等级的信道，建立功率域斯塔克伯格博弈模型，通过求解博弈均衡得到最佳传输功率，提高接收信号信噪比。最后，采用簇头协助决策方式，协助簇内信道决策成功率低的节点，有效地提高网络整体感知环境的准确性与干扰信道决策成功率。

1 无人机集群网络模型

无人机集群网络采用层次结构的移动 Ad Hoc 网络 (当无人机的数量大于 6 架时，宜采用分层式结构^[15])，无人机集群网络对抗智能干扰机示意图，如图 1 所示。

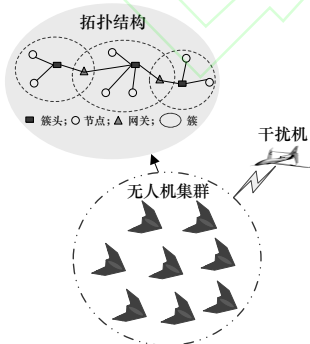


图 1. 无人机集群网络对抗智能干扰机示意图

Fig1. Schematic diagram of UAV cluster network defense intelligent jammer

设网络中干扰机为 J ，干扰机个数为 1，节点总数为 N_s ，分簇数 $M = N_s / N_c$ ， N_c 为簇内节点个数，

节点 i 的簇内邻节点个数 $C_{-i} \subset \Omega_s$ ，其中 Ω_s 为网络节点集合。假设簇头具有较高的等级，数据处理能力最强，其在簇内则充当局部控制中心的角色，簇间节点间的通信通过所在簇的簇头转发数据。

2 多域联合认知抗干扰算法

2.1 基于 A2C 的频域抗干扰算法

信道选择过程可建模为马尔科夫决策过程^[16]，由四元组 (S, A, R, λ) 来描述，其中 S 为状态空间， A 为动作空间，累计奖励函数 $R = \sum_{t=1}^T \lambda^{t-1} R_t$ ， $\lambda \in [0, 1]$ 为有损因子， R_t 为时间步 t 得到的即时奖励。马尔科夫的决策过程是序贯决策问题，其最终目标是找到最优决策序列 $\pi(s, a): S \times A \rightarrow \mathbb{R}_+$ ，以得到最大期望奖励 $Q^*(s, a)$ ，即给定状态 $s \in S$ 和动作 $a \in A$ 下，选择最优策略 $\pi(s, a)$ 获得最大期望奖励 $Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a) = E[R_t | s_t = s, a_t = a]$ 。信道选择的最终目标是根据感知环境干扰状况得到最有效的抗干扰策略。将信道选择问题转化序贯决策问题，首先需要设定信道选择过程对应的奖励函数，状态空间以及动作空间。

2.1.1 奖励函数 R

在时间步 t ，无干扰机时节点 i 的接收信号信噪比 (Signal-to-Interference-plus-Noise Ratio, SINR) 为 $\gamma_{NI}(t)$ ，有干扰机时为 $\gamma_{JI}(t)$ ，可分别表示为：

$$\gamma_{NI}(t) = \frac{g_s^i P_s^i(t)}{P_s^{-i}(t) + \varepsilon} \quad (1)$$

$$\gamma_{JI}(t) = \frac{g_s^i(t) P_s^i(t)}{g_j(t) P_j(t) + P_s^{-i}(t) + \varepsilon} \quad (2)$$

其中， $P_s^i(t)$ 为节点 i 的传输功率， $P_j(t)$ 为干扰机 J 的干扰功率； $g_s^i(t)$ 、 g_j 分别为节点 i 和干扰机 J 的信道增益， $P_s^{-i}(t) = \sum_{j \neq i} g_s^j P_s^j(t)$ 为网络中不包含节点 i 的其他邻节点对节点 i 的干扰功率总和， ε 为高斯加性白噪声。

设网络通信总带宽为 W ，将其均分为 K 个子信道 b ，则有 $\sum_{k=1}^K b_k = W$ 。无干扰机情况下，在时间步 t 内，节点 i 的信息传输速率为：

$$r_s(t) = \tilde{b}_k(t) \log_2(1 + \gamma_{NI}(t)) \quad (3)$$

存在干扰机情况下，其信息传输速率为：

$$r_j^i(t) = \tilde{b}_k(t) \log_2(1 + \gamma_{JI}(t)) \quad (4)$$

式中， $\tilde{b}_k(t)$ 为时间步 t 内的通信信道。

抗干扰通信的目标是最大化确保信息安全、可靠传输，将即时奖励 R_t 定义为通信安全容量 C_{sec}^i ^[17]，

可表示为:

$$C_{\text{sec}}^i := [r_s^i(t) - r_j^i(t)]^+ \quad (5)$$

其中, $[\cdot]^+ = \max\{0, \cdot\}$, 即时奖励 $R_t = C_{\text{sec}}^i$ 。

2.1.2 状态空间与动作空间

设环境状态空间 S 为节点 i 的前一时刻感知频谱 b_{t-1} , 则时刻 t 的状态 s_t 可表示为:

$$s_t = b_{t-1} \quad s_t \in S \quad (6)$$

节点根据观测状态 s_t 和即时奖励 R_t 进行信道选择。将动作空间定义为可选信道, 则有 $A = \{1, 2, \dots, K\}$, 对应于信道索引。信道选择过程可描述为, 在时间步 t , 节点由状态 s_t , 选择信道 $a_t = k \in A$, 获得即时奖励 C_{sec}^i 。

2.1.3 基于优势演员-评论家的频域抗干扰算法

AC 策略算法是由行动者 (Actor) 与评论家 (Critic) 组成的强化学习算法, 其中 Actor 负责更新策略, Critic 负责更新动作值函数。与 AC 算法相比, A2C 算法通过引入基线能够降低学习过程中的方差, 以较准确的动作值指导策略更新, 可带来更好的求解效果。实际应用中真实价值很难得到, 一般采用函数近似法, 通过对价值和动作函数参数化, 利用神经网络等机器学习算法求解。求解过程如下:

对于 Critic 而言, 其目标是通过不断的更新参数 θ , 使得值函数 $V_\theta(s)$ 更加逼近真实的累计奖励值 $C_{\text{sec}}^i(s)$, 即:

$$\theta^* = \arg \min_{\theta} \left\{ \frac{1}{2} \left(C_{\text{sec}}^i + \lambda C_{\text{sec}}^i(s') - V_\theta(s) \right)^2 \right\} \quad (7)$$

对于 Actor 而言, 其目标是通过不断地更新参数 w , 使得其尽可能得到好的策略 $\pi_w(s, a)$, 即:

$$w^* = \arg \max_w \left\{ \sum_{s \in S} d(s) \sum_{a \in A} \pi_w(s, a) C_{\text{sec}}^i(s) \right\} \quad (8)$$

式中, $d(s)$ 对应起始状态 s 。

在每一步的更新中, Actor 根据当前状态 s 和策略 $\pi_w(s, a)$, 执行动作 a , 转到下一状态 s' , 得到立即奖励 C_{sec}^i 。Critic 根据真实奖励和之前标准下的评分 $C_{\text{sec}}^i + \lambda C_{\text{sec}}^i(s')$, 修正评价标准, 使得估计价值更加逼近真实的奖励值。

为增加模型探索能力, 在模型目标函数中加入策略的熵正则化项, 其可衡量概率策略分布的不确定性, 值越大意味着模型具有更好的多样性^{[18][19]}。Actor 网络的参数 w 基于策略梯度下降的计算公式为:

$$w \leftarrow w + \alpha \sum_t \nabla_w \log_2 \pi_w(s, a) \left(\underbrace{C_{\text{sec}}^i(s) - V_\theta(s)}_{\text{优势函数的估计}} \right) + \eta E_s [\nabla_w H^\pi(s)] \quad (9)$$

式中, $H^\pi(s) = -\sum_a \pi_w(s, a) \log_2 \pi(s, a)$ 为策略梯度的熵, $E[\cdot]$ 为期望, η 为策略梯度的熵在目标函数中权重, α, β 分别为 Actor 和 Critic 网络学习率, 值网络模型目标函数梯度为:

$$\theta \leftarrow \theta + \beta \sum_i \partial \left[\frac{1}{2} \left(C_{\text{sec}}^i(s) - V_\theta(s) \right)^2 \right] / \partial \theta \quad (10)$$

通过 A2C 算法决策出各信道干扰情况后, 根据给定信道干扰容忍双阈值 P_{Jth1}^{sec} 与 P_{Jth2}^{sec} 将干扰功率划分为严重、中度、轻微三个等级。信道干扰等级判定规则, 如表 1 所示。

表 1 信道干扰等级判定规则

Table 1. Decision rule of channel jamming level

干扰功率	干扰程度	等级
$0 \leq P_J < P_{Jth1}^{\text{sec}}$	轻微	1
$P_{Jth1}^{\text{sec}} < P_J < P_{Jth2}^{\text{sec}}$	中度	2
$P_J > P_{Jth2}^{\text{sec}}$	严重	3

2.2 基于 SG 的功率域抗干扰算法

在 2.1 的基础上, 当上一时刻所用信道在当前时刻被判决为等级 2 时, 则对该信道建立功率域 SG 模型, 并通过求解 SE 得到最佳传输功率, 实现功率域抗干扰。

假设网络中的节点为领导者, 干扰机为跟随者, 参与者的行动受功率约束。定义博弈模型为 $G_{s,j} = \langle \{\Omega_s, J\}, \{P_s, P_j\}, \{U_s, U_j\} \rangle$, 其中 P_s 和 P_j 分别为节点和干扰机的传输功率; U_s, U_j 分别为节点与干扰机的效用函数。以节点作为领导者, 并假定其可精确估计出干扰功率, 而干扰机对节点的功率估计存在观测误差^[10], 观测误差为: $e = |\hat{P}_s^i - P_s^i| / P_s^i$, \hat{P}_s^i 为干扰机观测的节点 i 的功率。根据式(2)得到的接收信号 SINR, 并考虑节点间的通信开销, 则节点 i 的效用函数, 可表示为:

$$U_s^i(P_s^i, P_s^{-i}, P_j) = \gamma_{ji} - L_s^i P_s^i \quad (11)$$

式中, $L_s^i P_s^i$ 为节点 i 的通信代价。

干扰机 J 的效用函数可表示为:

$$U_j(P_s^i, P_s^{-i}, P_j) = -\hat{\gamma}_{ji} + L_s^i \hat{P}_s^i - L_j P_j \quad (12)$$

式中, $L_j P_j$ 为干扰代价, $\hat{\gamma}_{ji}$ 为 γ_{ji} 的观测值。

求解 Stackelberg 均衡 (Stackelberg Equilibrium, SE) 是博弈理论的重要内容。当处于 SE 状态时, 节点可得到保证自己损失最小的传输功率, 而干扰

机可得到最佳的干扰功率。本文利用逆向归纳法求解 SE, 并定义功率约束下的 SE 为 (P_s^{SE}, P_j^{SE}) , P_s^{SE} 、 P_j^{SE} 分别表示为:

$$P_s^{SE} = \arg \max_{0 \leq P_s \leq P_s^{Max}} U_s(P_s^i, P_s^{-i}, P_j) \quad (13)$$

$$P_j^{SE} = \arg \max_{0 \leq P_j \leq P_j^{Max}} U_j(\hat{P}_s^i, \hat{P}_s^{-i}, P_j) \quad (14)$$

其中, P_s^{Max} , P_j^{Max} 分别为节点最大传输功率与干扰机的最大干扰功率, P_s^{SE} , P_j^{SE} 分别为节点最佳传输功率与干扰机最佳干扰功率。

(1) 求解 P_j^{SE}

1) 考虑一般情况。基于式(12), $U_j(\hat{P}_s^i, \hat{P}_s^{-i}, P_j)$ 关于 P_j 的一阶导数为: $\partial U_j / \partial P_j = g_s^i g_j \hat{P}_s^i / (\hat{P}_s^{-i} + \varepsilon + g_j P_j)^2 - L_j$, 二阶导数 $\partial^2 U_j / \partial P_j^2 \leq 0$, 则 $U_j(\hat{P}_s^i, \hat{P}_s^{-i}, P_j)$ 关于 P_j 为凹函数。令一阶导数等于 0, 得到 U_j^{Max} , 对应 $P_j^{SE} = \tilde{P}_j = (\sqrt{g_s^i g_j \hat{P}_s^i / L_j} - \varepsilon - \hat{P}_s^{-i}) / g_j$, 其中 $\tilde{P}_j \in (0, P_j^{Max})$; 2) 考虑极端情况。①相比阻断节点间的信息传输, 干扰机本身的损失更低, 此时干扰机的 $P_j^{SE} = P_j^{Max}$; ②相比阻断节点间的信息传输, 干扰机本身的损失更大, 此时 $P_j^{SE} = 0$ 。则求得 P_j^{SE} :

$$P_j^{SE} = \begin{cases} 0 & , \tilde{P}_j \leq L_j (\hat{P}_s^{-i} + \varepsilon^2) / (g_s^i g_j) \\ P_j^{Max} & , \tilde{P}_j \geq L_j (P_j^{Max} g_j + \hat{P}_s^{-i} + \varepsilon) / (g_s^i g_j) \\ \tilde{P}_j = (\sqrt{g_s^i g_j \hat{P}_s^i / L_j} - \varepsilon - \hat{P}_s^{-i}) / g_j & , \text{其他} \end{cases} \quad (15)$$

(2) 求解 P_s^{SE}

将式(15)代入式(11), 假设对节点 i 而言, 其认为干扰机的观测不存在误差, 即 $\hat{P}_s^i = P_s^i$, 则有:

$$U_s(P_s^i, P_s^{-i}, P_j) = \begin{cases} \frac{g_s^i}{(\varepsilon + P_s^{-i})} - L_s^i, P_s^i \leq \frac{\varepsilon^2 + P_s^{-i}}{g_s^i g_j} \\ \left(\frac{g_s^i}{P_j^{Max} g_j + P_s^{-i} + \varepsilon} - L_s^i \right) P_s^i, P_s^i \geq \frac{L_j (P_j^{Max} g_j + P_s^{-i} + \varepsilon)^2}{(g_s^i g_j)} \\ \sqrt{\frac{g_s^i P_s^i L_j}{g_j}} - L_s^i P_s^i, \text{其他} \end{cases} \quad (16)$$

设函数 $H(P_s^i) = \sqrt{g_s^i P_s^i L_j / g_j} - L_s^i P_s^i$, 其一阶导数 $dH / dP_s^i = 0.5 \sqrt{g_s^i L_j / P_s^i g_j} - L_s^i$, 二阶导数 $d^2 H / dP_s^i{}^2 \leq 0$, 故 $H(P_s^i)$ 为凹函数, 则有 $\tilde{P}_s^i = \max H = L_s^i g_s^i / (4 g_j L_s^i)$ 。由式(16), 节点 i 的 P_s^{SE} , 存在以下情况:

- 1) $L_s^i > g_s^i / (\varepsilon + P_s^{-i})$: $U_s(P_s^i, P_s^{-i}, P_j)$ 关于 P_s^i 递减, 此时 $P_s^{SE} = 0$;
- 2) $g_s^i / (\varepsilon + P_s^{-i} + P_j^{Max}) < L_s^i < g_s^i / (\varepsilon + P_s^{-i})$: 如果 $P_s^i \leq L_j (\varepsilon^2 + P_s^{-i}) / (g_s^i g_j)$, $U_s(P_s^i, P_s^{-i}, P_j)$ 关于 P_s^i 不减,

当 $P_s^i \geq L_j (P_j^{Max} g_j + P_s^{-i} + \varepsilon)^2 / (g_s^i g_j)$, $U_s(P_s^i, P_s^{-i}, P_j)$ 递减, 此时 $P_s^{SE} = \tilde{P}_s^i$;

3) $L_s^i < g_s^i / (\varepsilon + P_s^{-i} + P_j^{Max})$: $U_s(P_s^i, P_s^{-i}, P_j)$ 关于 P_s^i 递增, 此时 $P_s^{SE} = P_s^{Max}$ 。则求得 P_s^{SE} :

$$P_s^{SE} = \begin{cases} 0 & , L_s^i > \frac{g_s^i}{\varepsilon + P_s^{-i}} \\ \frac{g_s^i L_j}{4 g_j L_s^i{}^2}, \frac{g_s^i}{\varepsilon + P_s^{-i} + P_j^{Max}} < L_s^i < \frac{g_s^i}{\varepsilon + P_j^{Max}} \\ P_s^{Max} & , L_s^i < \frac{g_s^i}{\varepsilon + P_s^{-i} + P_j^{Max}} \end{cases} \quad (17)$$

综上所述, 提出的多域联合认知抗干扰算法 (MDJC-AJ), 实现过程描述如下。

输入: 训练数据 $D = \{(s_i, b_i) | s_i \in S, b_i \in A\}$, 经验池 E 。
输出: 最优策略的估计 π_θ^* , 接收信号信噪比 γ 。

(1) (初始化) $C_{sec}^i(0) = 0$, 最大迭代次数 N_{it} , 网络参数 $\theta \leftarrow 0, w \leftarrow 0$, 学习率 α, β , 折扣因子 λ , 熵正则化系数 η , 干扰容忍阈值 $P_{Jth1}^{sec}, P_{Jth2}^{sec}$ 。

(2) (迭代更新) 对每个智能体(节点/簇头), 每幕执行以下操作:

- 步骤 1 (采样): 根据状态 s 和动作 b 得到采样值 C_{sec}^i 和下一个状态 s' , 数据存储 $E \leftarrow \{s, b, C_{sec}^i, s'\}$;
- 步骤 2 (执行): 利用 $\pi_\theta(\cdot | s')$ 得到动作 b' ;
- 步骤 3 (通信安全容量): $C_{sec}^i(s) \leftarrow C_{sec}^i + \lambda C_{sec}^i(s')$;
- 步骤 4 (策略更新): 基于式(9), 更新策略网络参数 w ;

步骤 5 (价值更新): 基于式(10), 更新策略网络参数 θ ;

步骤 6 (更新状态与动作): $s \leftarrow s', b \leftarrow b'$;

步骤 7 (信道选用策略): if 前一时刻被选信道在当前时刻的干扰为等级 3: 被选信道存在严重干扰, 最佳传输功率为 0, 奖励函数 $C_{sec}^i(s) = 0$ 时, 则转至步骤 1。

Else-if

等级 2: 建立功率域博弈模型, 由式(17)得到最佳传输功率

Else 等级 1: 保持使用当前信道

(3) 直到达到最大迭代次数 N_{it} , 结束。

2.3 算法复杂度分析

参考文献[12], 本文对所提 MDJC-AJ 算法的复杂度进行分析结果, 如表 2 所示。

表 2 MDJC-AJ 算法运算复杂度

Table 2 Complexity analysis of MDJC-AJ algorithm

计算	复杂度
----	-----

C_{sec}^i	$O(N_s C_1)$
更新 $C_{\text{sec}}^i(s)$	$O(N_s C_2)$
更新 w, θ, s, b	$O(N_s C_3)$
等级划分	$O(N_s C_4)$
P_J^{SE}	$O(C_5)$
P_S^{SE}	$O(N_s C_6)$

详细算法运算复杂度分析描述如下:

(1) 对于单个节点 i 或簇头, 获得观测状态和得到采样值 C_{sec}^i 的复杂度为 $O(C_1)$, C_1 为与采样过程有关的常数, 关于所有的节点运算复杂度为 $O(N_s C_1)$, 该部分对应算法迭代步骤 1;

(2) 对于单个节点, 根据策略 $\pi_\theta(\cdot|s')$, 在每个状态下执行相应动作得到奖励值的复杂度为 $O(C_2)$, C_2 为与策略类型有关的常数, 关于所有节点的运算复杂度为 $O(N_s C_2)$, 该部分对应步骤 2~3;

(3) 对于单个节点, 基于式(9)、式(10), 更新参数 w, θ 以及状态、动作, 运算复杂度为 $O(C_3)$, C_3 为与每幕的时间步长或收敛迭代次数有关的常数, 关于所有节点的运算复杂度为 $O(N_s C_3)$, 该部分对应步骤 4~6;

(4) 对于单个节点, 根据阈值进行等级划分, 运算复杂度为 $O(C_4)$, C_4 为与阈值个数有关的常数, 关于所有节点的运算复杂度为 $O(N_s C_4)$;

(5) 干扰机最佳干扰功率运算复杂度为 $O(C_5)$, C_5 为与计算式(15)有关的常数;

(6) 对单个节点, 根据式(17)计算节点最佳传输功率运算复杂度为 $O(C_6)$, C_6 为常数, 关于所有节点的运算复杂度为 $O(N_s C_6)$ 。

通过以上分析, 可得到 MDJC-AJ 算法总的运算复杂度为:

$$C_{\text{sum}} = N_{it} (O(N_s C_1) + O(N_s C_2) + O(N_s C_3) + O(C_4) + O(N_s C_5) + O(N_s C_6)) \quad (18)$$

2.4 基于簇头协助的信道选择算法

考虑到实际环境态势的多变性、信息的局部性, 存在单个节点局部频谱感知能力有限的问题, 为此引入簇头协助从节点决策的方法。基于簇头协助的无人机集群网络抗干扰示意图, 如图 2 所示。

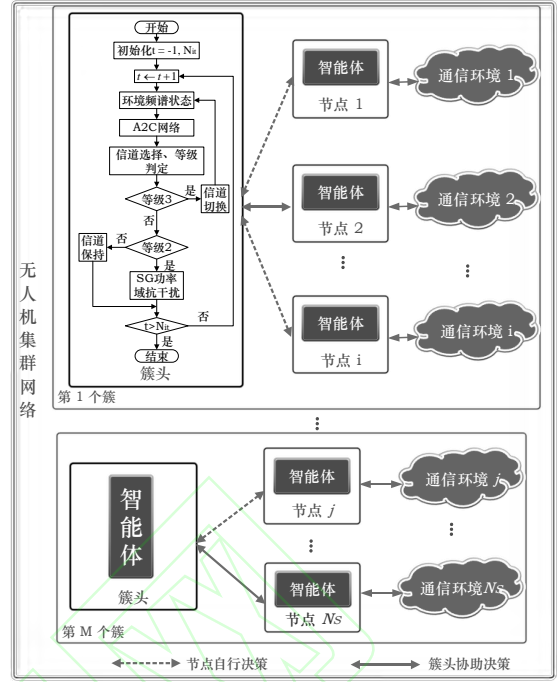


图 2 基于簇头协助的无人机集群网络抗干扰示意图

Fig.2 The schematic diagram of UAV swarms network based on cluster head assistance

簇头协助从节点决策方法可描述为: 首先, 各节点进行局部环境感知与信道决策, 若某节点所得结果无法达到期望值, 则向簇头发出 Help 信息, 簇头收到求助信息后, 则向其传输无干扰信道数据信息, 使其能够可靠通信。需要说明的是, 所有节点和簇头均采用 MDJC-AJ 算法进行抗干扰。不失一般性, 图 2 中仅对第一个簇头内部抗干扰算法实现流程作详细说明。

3 实验仿真与分析

为验证所提算法有效性, 选用卷积神经网络来拟合值函数和策略函数。仿真环境为 Intel (R) Core(TM)i7-4790 CPU@3.60GHz 四核八线程处理器, 采用 Pytorch1.2.0 深度学习框架, 与 Matlab2018a 仿真平台。

仿真条件设置如下: 每幕训练时长 $T = 5000$, 每个时间步长 $t = 0.1s$, 每 100 幕为一个训练周期, 共设有 10 个仿真周期, 经验池容量大小为 $E = 10000$ 。设定信道个数 $K = 32$, 并将状态空间重塑为 6×6 的排列作为网络的输入。为确保 Critic 有足够的时间计算奖励值, Critic 网络学习率 $\beta = 0.02$, Actor 网络学习率 $\alpha = 0.005$, 惩罚系数 $\eta = 0.3$, 干扰功率阈值 $P_{Jth1}^{\text{sec}} = 0.5$, $P_{Jth2}^{\text{sec}} = 6$ 。

Actor 网络与 Critic 网络一致, 不同的是最后的全连接层^[20]。Actor 的输出为 32 对应于 32 个信道,

是一个 vector, Critic 的输出维度为 1, 用于计算 Actor 所获奖励。网络结构参数设置, 如表 3 所示。

表 3 网络结构参数设置

Table 3 Parameters of network structure

卷积网络	卷积层 1	卷积层 2	全连接层 (Actor/Critic)
输入	1×6×6	8×4×4	64
滤波器尺寸	3×3	2×2	--
步长/填充	1/0	1/0	--
滤波器数量	8	16	32/1
激活函数	Relu	Relu	Relu
输出	8×4×4	16×2×2	32/1

仿真 1 验证所提算法信道选择性能。考虑干扰机采用智能性干扰, 即不同时间段干扰机干扰的信道和功率均不同, 为便于分析将环境状态的时变点分别设在 $t_{change} = 1500$ 和 $t_{change} = 3300$, 网络中节点个数为 4, 编队及所选簇头已最优。不同算法下信道干扰情况决策成功率, 如图 3 所示。

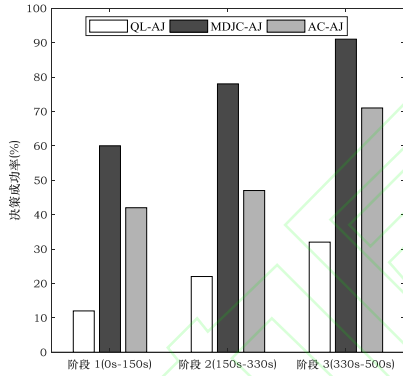


图 3 不同算法下干扰信道情况决策成功率

Fig.3 The channel decision success rate of jamming situation with different algorithm

由图 3 可知, 在各阶段内, MDJC-AJ 算法的信道干扰情况决策结果较文献[4]Q 学习抗干扰 (QL-AJ)、文献[8]演员-评论家抗干扰 (AC-AJ) 算法成功率更高。

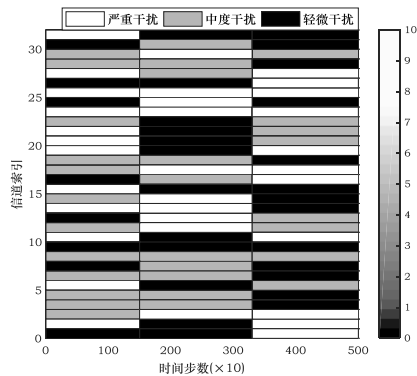


图 4 MDJC-AJ 算法对信道干扰情况判定结果

Fig.4 The decision result of channel jamming situation with MDJC-AJ

为进一步说明所提 MDJC-AJ 算法在智能性干扰情况下信道决策有效性, 由仿真所得信道干扰情况判决结果, 如图 4 所示。由图 4 可知, 所提 MDJC-AJ 算法, 在决策出可用信道索引情况下, 对信道干扰功率情况进行判决, 可为功率域抗干扰提供依据。

仿真 2 考虑干扰机观测误差, 研究干扰机偏离 SE 对节点效用函数及其自身效用函数的影响, 检验网络抗干扰性能。仿真参数设置如下: 通信信道增益 $g_s^i = 0.7$, $g_s^{-i} = 0.2$; 干扰机损耗系数 $L_j = 0.4$; 节点 i 传输功率与干扰机干扰功率分别为 $P_s^i = 12$, $P_j = 10$; 为实现方便, 将簇内邻节点对节点 i 的互干扰功率均设为常数 $P_s^{-i} = 2$; 噪声功率 $\varepsilon = 0.5$; 节点损耗系数, $L_s = [0.1, 0.2, 0.3, 0.7]$ 。节点效用函数之和 $U_s = \sum_i^{N_s} U_s^i$ 与干扰机的效用函数 U_j 关于干扰机观测误差 e 的变化情况, 如图 5 所示。

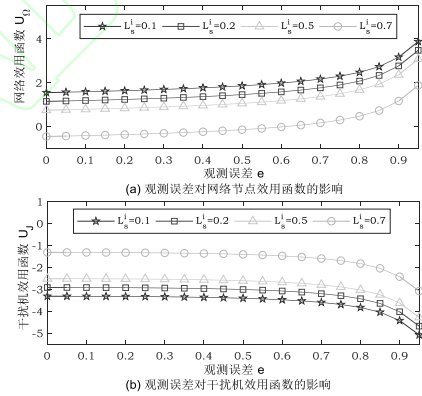


图 5 观测误差 e 对网络节点/干扰效用函数的影响

Fig.5 Utility value of network and jammer with respect to observation errors

由图 5 可知, 随着干扰机观测误差的增加, 节点效用函数之和递增, 干扰机的效用函数递减。其原因是, 观测误差的增加, 会使干扰机最佳传输功率偏离 SE, 导致其效用函数减小。干扰机观测误差等效于削弱了干扰机干扰的强度, 而这将有利于提高节点效用函数, 使其通信性能提升。

簇内节点个数对接收信号信干噪比的影响, 如图 6 所示。

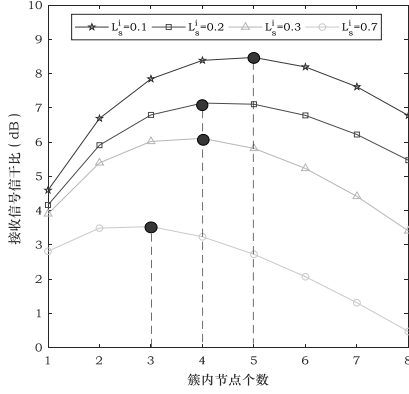


图 6 簇内节点数对整体接收信号信干噪比的影响

Fig.6 The relation of between nodes numbers inter-cluster and receive signal SINR

由图 6 可知,起初接收信号信干噪比随着节点数的增加而增大,然而,当节点个数增加到一定程度时开始出现衰减。分析原因可知,随着簇内节点数的增加,节点之间的互干扰不断增加,使得单个节点的接收信号信干噪比降低,进而导致簇内各节点接收信号信干噪比降低。此外,在不同的节点损耗系数下 $L_s^1=[0.1,0.2,0.3,0.7]$,簇内节点数存在一个最佳的个数,分别对应最佳簇内节点的个数为 $N_c=[5,4,4,3]$,一定程度上可为无人机集群网络簇的划分提供有益参考。

仿真 3 验证不同算法下抗智能干扰性能,结果如图 7 所示。

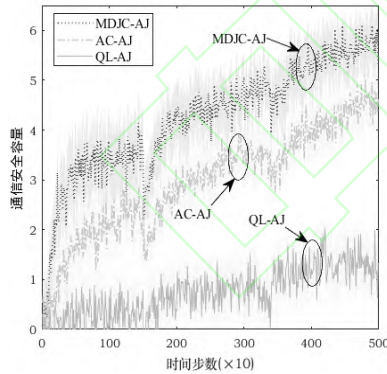


图 7 不同算法下抗智能性干扰性能对比

Fig.7 Comparison of anti-jamming performance with different algorithm

由图 7 可知,不同算法下,网络通信安全容量均随着训练时间的增加不断提高,且所提 MDJC-AJ 算法与 QL-AJ、AC-AJ 算法相比,网络通信安全容量较高。值得注意的是,在三个阶段的突变点,三种算法得到的通信安全容量均骤减,之后恢复,而 MDJC-AJ 算法较其他两种算法恢复得速度更快。其原因是,由于状态空间和动作空间较大,QL-AJ 算

法遍历 Q 表所有状态的计算量庞大,算法收敛较慢;同时注意到,AC 算法利用卷积神经网络强大的计算能力,相比 QL 的方法提高了近 4 倍的计算速度;另外,所提 MDJC-AJ 算法相比于 AC-AJ 算法能够降低学习过程的方差,减少误动作算法稳定性好、收敛更快,且通过联合功率域抗干扰减少信道切换的时间,同时提高了信号接收信干噪比,得到的通信安全容量更高。

为进一步说明 MDJC-AJ 算法的稳健性。定义单个状态 s 的均方价值误差表示为近似价值函数 $V_\theta(s)$ 与真实 $C_{\text{sec}}^i(s)$ 的差的平方,记作 \overline{VE} :

$$\overline{VE} = \sqrt{\frac{1}{|S|} \sum_{s \in S} [C_{\text{sec}}^i(s) - V_\theta(s)]^2} \quad (19)$$

式中, $|S|$ 为系统状态个数。

为验证所提方法算法收敛性能,对比了 QL-AJ、AC-AJ 算法两种算法收敛情况。10 个仿真周期的平均价值均方误差,如图 8 所示。

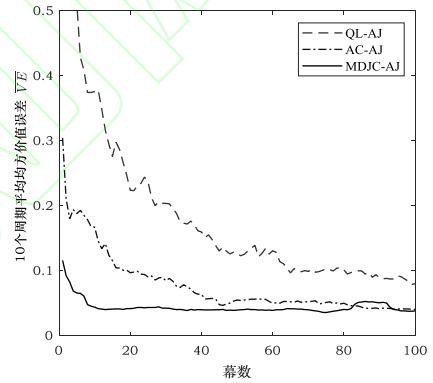


图 8 10 个周期平均价值均方误差变化曲线

Fig.8 Curve of mean value square error over 10 periods

由图 8 仿真结果表明,经过 10 幕左右所提算法将收敛,较其他两种算法收敛性能较好,且最后得到的价值误差更小。

4 结束语

为解决无人机集群网络对抗智能性干扰能力较弱的问题,本文提出了一种多域联合的认知抗干扰算法。该算法基于 A2C 频域算法,利用感知到的频谱状态信息进行信道选择,能够提高算法的收敛速度与信道决策成功率。在此基础上,根据得到的功率干扰等级,从功率域进行抗干扰,减少信道切换时间,提高接收信号信干噪比。并通过仿真对比 QL-AJ 与 AC-AJ 两种算法,证明了所提算法整体抗干扰性能较好。另外,提出簇头协助的方法,进一步改善网络抗干扰性能。下一步将考虑干扰机为多

个的情况,同时研究不完全观测信息条件下的抗干扰问题。

参考文献

- [1] Gupta L, Jain R, Vaszkun G. Survey of Important Issues in UAV Communication Networks [J]. IEEE Communications Surveys & Tutorials, 2016, 18(2):1123-1152.
- [2] Li H, Luo J, Liu C. Selfish Bandit-Based Cognitive Anti-Jamming Strategy for Aeronautic Swarm Network in Presence of Multiple Jammers [J]. IEEE Access, 2019, 7(3):30234-30243.
- [3] Lin Y, Wang T, Wang S. UAV-Assisted Emergency Communications: An Extended Multi-Armed Bandit Perspective [J]. IEEE Communications Letters. 2019, 23(5): 938-941.
- [4] Slimeni F, Chtourou Z, Scheers B, et al. Cooperative Q-learning based channel selection for cognitive radio networks [J]. Wireless Networks, 2018(4):1-11.
- [5] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540):529-533.
- [6] Silver D, Huang A, Maddison C J, et al. Mastering the game of go with deep neural networks and tree search [J]. Nature, 2016, 529(7587):484-489.
- [7] Wang S, Liu H, Gomes P H, et al. Deep Reinforcement Learning for Dynamic Multichannel Access in Wireless Networks [J]. IEEE Transactions on Cognitive Communications and Networking, 2018, 4(2): 257-265.
- [8] Bhowmik M, Malathi P. Spectrum Sensing in Cognitive Radio Using Actor-Critic Neural Network with Krill Herd-Whale Optimization Algorithm[J]. Wireless Personal Communications, 2019, 105(1):335-354..
- [9] Wei Y, Yu F, Han E, et al. User Scheduling and Resource Allocation in HetNets with Hybrid Energy Supply: An Actor-Critic Reinforcement Learning Approach [J]. IEEE Transactions on Wireless Communications, 2018, 17(1):680-692.
- [10] 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.
- [11] Ahmed I K, Fapojuwo A O. Stackelberg Equilibria of an Anti-Jamming Game in Cooperative Cognitive Radio Networks[J]. IEEE Transactions on Cognitive Communications and Networking, 2018,4(1):121-134.
- [12] Xu Y, Ren G, Chen J, et al. A One-Leader Multi-Follower Bayesian Stackelberg Game for Anti-Jamming Transmission in UAV Communication Networks [J].IEEE Access, 2018, 6():21697-21709.
- [13] Jia L, Xu Y, Sun Y, et al. A Multi-Domain Anti-Jamming Defense Scheme in Heterogeneous Wireless Networks[J].IEEE Access, 2018,6(8):40177-40188.
- [14] Simone P, Tangkaratt V, Peters J, et al. TD-Regularized Actor-Critic Methods [J]. Machine Learning, 2019, 108(9): 1467-1501.
- [15] Tinnirelloi, Bianci G, Xiao Y. Refinements on IEEE 802.11 Distributed Coordination Function Modeling Approaches [J]. IEEE Transactions on Vehicular Technology, 2010, 59(3):1055-1067.
- [16] Naparstek O, Cohen K. Deep Multi-User Reinforcement Learning for Distributed Dynamic Spectrum Access [J]. IEEE Transactions on Wireless Communications, 2019, 18(1):310-323.
- [17] Fang X, Zhang N, Zhang S, et al. On Physical Layer Security: Weighted Fractional Fourier Transform Based User Cooperation [J]. IEEE Transactions on Wireless Communications, 2017, 16(8): 5498-5510.
- [18] Haarnoja T, Zhou A, Pieter A, et al. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with A Stochastic Actor [C] // Proceeding of the International Conference on Machine Learning (ICML), 2018, 2(1).
- [19] O'donoghue B, Munos R, Kavukcuoglu K, et al. PQG: Combining policy gradient and Q-learning [J]. Statistics, 2016.
- [20] Mao H, Netravali R, Alizadeh M. Neural adaptive video streaming with pensieve [C] // Proceedings of the Conference of the ACM Special Interest Group on Data Communication. ACM, 2017: 197-210.