第 46 卷 第 12 期 计 算 机 工 程 2020 年 12 月
Vol. 46 No. 12 Computer Engineering December 2020

·移动互联与通信技术· 文章编号：1000-3428(2020)12-0193-08 文献标志码：A 中图分类号：TN975

基于多域联合的无人机集群认知抗干扰算法

刘春玲[a]，刘敏捷[a,b]，丁元明[a,b]

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

摘 要： 为解决无人机集群网络在复杂通信环境中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。该算法在优势演员-评论家算法的基础上，将无人机视作智能体，并由感知到的环境频谱状态决策出干扰信道。基于 Stackelberg 博弈理论，利用功率域压制中度干扰等级的信道干扰信号，减少切换信道的时间开销。通过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而导致信道决策成功率较低的问题。仿真结果表明，相比 QL-AJ 算法与 AC-AJ 算法，该算法能够给出簇内最佳节点个数，提高接收信号信干噪比，且网络整体抗干扰性能较好。

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

开放科学（资源服务）标志码（OSID）：

中文引用格式： 刘春玲，刘敏捷，丁元明. 基于多域联合的无人机集群认知抗干扰算法[J]. 计算机工程，2020，46(12)：193-200.

英文引用格式： LIU Chunling，LIU Minti，DING Yuanming. Cognitive anti-jamming algorithm for UAV cluster based on multiple domain combination[J]. Computer Engineering，2020，46(12)：193-200.

Cognitive Anti-Jamming Algorithm for UAV Cluster Based on Multiple Domain Combination

LIU Chunling[a]，LIU Minti[a,b]，DING Yuanming[a,b]

（a. College of Information Engineering；b. Key Laboratory of Communication and Network，Dalian University，Dalian，Liaoning 116622，China）

【Abstract】 To solve the problem of weak ability of Unmanned Aerial Vehicle（UAV）cluster network to resist intelligent interference in complex communication environment，this paper proposes a multiple domain cognitive anti-jamming algorithm based on the intelligent decision-making theory. Based on the dominant actor-critic algorithm，the algorithm regards UAV as an agent and the jamming channel is determined by the perceived environmental spectrum state. Based on the Stackelberg game theory，the interference signals of the middle-level interference channels are suppressed from the power domain to reduce the time cost of channel switching. Cluster head assistance is introduced to solve the problem of low success rate of channel decision-making due to weak local spectrum sensing ability of a single agent. Simulation results show that compared with the QL-AJ algorithm and AC-AJ algorithm，the proposed algorithm can give the optimal number of nodes in the cluster，improves the Signal to Interference-plus Noise Ratio（SINR）of the received signals，and provides better overall anti-jamming performance for networks.

【Key words】 cognitive anti-jamming algorithm；dominant actor-critic algorithm；Stackelberg game；Unmanned Aerial Vehicle（UAV）cluster；distributed network

DOI： 10.19678/j.issn.1000-3428.0056784

0 概述

在未来空战中，无人机（Unmanned Aerial Vehicle，UAV）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性，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，确保无人机之间的可靠通信，将成为一项重要的研究内容[1]。

近年来，如何有效对抗智能性干扰与提高通信安全已成为研究热点[1]。在抗干扰技术研究中，认知抗干扰算法已成为研究热点方向之一[2]，该算法可归纳为如下两类：一类是基于强化学习理论[3]进

基金项目： 装备预先研究领域基金（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、博士，主研方向为智能信号处理与检测；刘敏捷，硕士研究生；丁元明，教授、博士、博士生导师。

收稿日期： 2019-12-02 **修回日期：** 2020-01-07 **E-mail：** liuchunling@dlu.edu.cn

行可用信道的选择,主动规避干扰信道,从而实现频域抗干扰。文献[4]提出基于协作 Q 学习(Q-Learning,QL)的信道选择算法,该算法可提高数据传输成功率,但当状态空间规模较大时,其面临维数灾难的问题[5-6]。针对该问题,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。当信道数量较多时,文献[8-9]利用演员-评论家(Actor-Critic,AC)算法进行信道选择,但是该算法存在方差较大以及稳定性较差的问题。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。以上算法均是仅从单个频域或者功率域角度考虑,针对智能性干扰攻击的灵活性较差[13]。

为提高网络抗智能干扰的能力,本文将功率域和频域抗干扰方法相结合,基于优势演员-评论家(Advantage Actor-Critic,A2C)[14]与 Stackelberg 博弈(Stackelberg Game,SG),提出一种多域联合认知抗干扰(Multiple Domain Joint Cognitive Anti-Jamming,MDJC-AJ)算法。该算法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。根据设定的干扰容忍双阈值将信道干扰程度分为严重、中度与轻微 3 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。与此同时,本文采用簇头协助决策方式来协助簇内信道决策成功率较低的节点,以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,适合采用分层式结构[15]。无人机集群网络对抗智能干扰机示意图如图 1 所示。

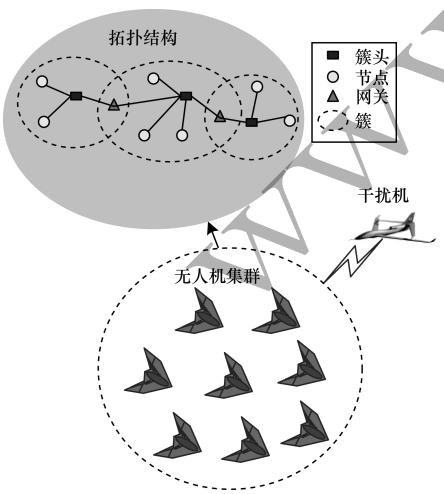

图 1 无人机集群网络对抗智能干扰机示意图

Fig. 1 Schematic diagram of UAV cluster network resist intelligent jammer

假设网络中干扰机为 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,簇间节点通过所在簇的簇头转发数据进行通信。

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可夫决策过程[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 t 时刻得到的即时奖励。马尔可夫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 $Q^*(s,a)=\max_\pi Q^\pi(s,a)=E[R_t\mid s_t=s,a_t=a]$。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为序贯决策问题,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

在 t 时刻,且无干扰机时,节点 i 接收信号的信干噪比(Signal to Interference-plus Noise Ratio,SINR)为 $\gamma_{NJ}(t)$,存在干扰机时,SINR 为 $\gamma_{YJ}(t)$,可表示为:

$$\gamma_{NJ}(t)=\frac{g_S^i P_S^i(t)}{P_S^{-i}(t)+\varepsilon} \tag{1}$$

$$\gamma_{YJ}(t)=\frac{g_S^i(t)P_S^i(t)}{g_J(t)P_J(t)+P_S^{-i}(t)+\varepsilon} \tag{2}$$

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 的其他邻居节点对节点 i 的干扰功率总和,ε 为高斯加性白噪声。

假设网络通信总带宽为 W,将其均分为 K 个子信道 b,则有 $\sum_{k=1}^{K}b_k=W$。在无干扰机的情况下,在时间步 Δt 内,节点 i 的信息传输速率如式(3)所示,存在干扰机的情况下,节点 i 的信息传输速率如式(4)所示:

$$r_S(t)=\tilde{b}_k(t)\mathrm{lb}(1+\gamma_{NJ}(t)) \tag{3}$$

$$r_J^i(t)=\tilde{b}_k(t)\mathrm{lb}(1+\gamma_{YJ}(t)) \tag{4}$$

其中,$\tilde{b}_k(t)$为 t 时刻的通信信道。

抗干扰通信的目标是确保信息安全以及可靠传输,将即时奖励 R_t 定义为通信安全容量 C_{sec}^i[17],可表示为:

$$C_{\mathrm{sec}}^i=[r_S^i(t)-r_J^i(t)]^+ \tag{5}$$

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

2.1.2　状态空间与动作空间

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

$$s_t = \boldsymbol{b}_{t-1}, s_t \in S \tag{6}$$

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 $A = \{1, 2, \cdots, K\}$,且其对应于信道索引。信道选择过程可描述为:在 t 时刻,节点由状态 s_t 选择信道 $a_t = k \in A$,并获得即时奖励 C_{sec}^i。

2.1.3　基于优势演员-评论家的频域抗干扰算法

AC 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。在实际应用中真实价值很难得到,一般采用函数近似法对价值和动作函数进行参数化,利用神经网络等机器学习算法求解,求解过程如下:

1)对于 Critic 而言,其目标是通过不断地更新参数 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:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\left\{\frac{1}{2}\left(C_{\text{sec}}^i + \lambda C_{\text{sec}}^i(s') - V_\theta(s)\right)^2\right\} \tag{7}$$

2)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:

$$w^* = \underset{w}{\operatorname{argmax}}\left\{\sum_{s\in S} d(s) \sum_{a\in A} \pi_w(s,a) C_{\text{sec}}^i(s)\right\} \tag{8}$$

其中,$d(s)$ 对应起始状态 s。

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

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,且其值越大说明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方法为:

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

其中,$H^\pi(s) = -\sum_a \pi_w(s,a)\,\text{lb}\,\pi(s,a)$ 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的权重,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:

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

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 $P_{J\text{th1}}^{\text{sec}}$ 与 $P_{J\text{th2}}^{\text{sec}}$ 将干扰功率划分为严重、中度与轻微 3 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
$0 \leqslant P_J < P_{J\text{th1}}^{\text{sec}}$	轻微	1
$P_{J\text{th1}}^{\text{sec}} < P_J < P_{J\text{th2}}^{\text{sec}}$	中度	2
$P_J > P_{J\text{th2}}^{\text{sec}}$	严重	3

2.2　基于 SG 的功率域抗干扰算法

在 2.1 节的基础上,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 Stackelberg 均衡(Stackelberg Equilibrium, 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 $G_{s,j} = \langle\{\boldsymbol{\Omega}_S, J\}, \{P_S, P_J\}, \{U_S, U_J\}\rangle$,其中,$P_S$ 和 P_J 分别为节点和干扰机的传输功率,U_S 与 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],且观测误差为 $e = |\hat{P}_S^i - P_S^i| / P_S^i$,$\hat{P}_S^i$ 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:

$$U_S^i(P_S^i, P_S^{-i}, P_J) = \gamma_{YJ} - L_S^i P_S^i \tag{11}$$

其中,$L_S^i P_S^i$ 为节点 i 的通信代价。

干扰机 J 的效用函数可表示为:

$$U_J(P_S^i, P_S^{-i}, P_J) = -\hat{\gamma}_{YJ} + L_S^i \hat{P}_S^i - L_J P_J \tag{12}$$

其中,$L_J P_J$ 为干扰代价,$\hat{\gamma}_{YJ}$ 为 γ_{YJ} 的观测值。

求解 SE 是博弈理论的重要内容。当处于 SE 状态时,节点可得到保证本身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 $(P_S^{\text{SE}}, P_J^{\text{SE}})$,且 P_S^{SE}、P_J^{SE} 分别表示为:

$$P_S^{\text{SE}} = \underset{0\leqslant P_S \leqslant P_S^{\text{Max}}}{\operatorname{argmax}} U_S(P_S^i, P_S^{-i}, P_J) \tag{13}$$

$$P_J^{\text{SE}} = \underset{0\leqslant P_J \leqslant P_J^{\text{Max}}}{\operatorname{argmax}} U_J(\hat{P}_S^i, \hat{P}_S^{-i}, P_J) \tag{14}$$

其中,P_S^{Max}、P_J^{Max} 分别为节点最大传输功率与干扰机最大干扰功率,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。

P_J^{SE} 的求解过程如下:

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 $\text{d}U_J/\text{d}P_J = g_S^i g_J \hat{P}_S^i / (\hat{P}_S^{-i} + \varepsilon + g_J P_J)^2 - L_J$,二阶导数为 $\text{d}^2U_J/\text{d}P_J^2 \leqslant 0$,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 $P_J^{\text{SE}} = \hat{P}_J = \left(\sqrt{g_S^i g_J \hat{P}_S^i / L_J} - \varepsilon - \hat{P}_S^{-i}\right) / g_J$,其中,$\hat{P}_J \in (0, P_J^{\text{Max}})$。

2)考虑极端情况,有以下 2 种情况:

(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 $P_J^{\mathrm{SE}}=P_J^{\mathrm{Max}}$。

(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 $P_J^{\mathrm{SE}}=0$,则求得 P_J^{SE} 为:

$$P_J^{\mathrm{SE}}=\begin{cases}0,\hat{P}_S^i\leqslant L_J(\hat{P}_S^{-i}+\varepsilon^2)/(g_S^i g_J)\\ P_J^{\mathrm{Max}},\hat{P}_S^i\geqslant L_J(P_J^{\mathrm{Max}}g_J+\hat{P}_S^{-i}+\varepsilon)^2/(g_S^i g_J)\\ \hat{P}_J=\left(\sqrt{g_S^i g_J\hat{P}_S^i/L_J}-\varepsilon-\hat{P}_S^{-i}\right)\Big/ g_J,\text{其他}\end{cases}\tag{15}$$

P_S^{SE} 的求解过程如下:

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

$$U_S(P_S^i,P_S^{-i},P_J)=\begin{cases}\dfrac{g_S^i}{(\varepsilon+P_S^{-i})}-L_S^i P_S^i\leqslant\dfrac{\varepsilon^2+P_S^{-i}}{g_S^i g_J}\\ \left(\dfrac{g_S^i}{P_J^{\mathrm{Max}}g_J+P_S^{-i}+\varepsilon}-L_S^i\right)P_S^i,P_S^i\geqslant\dfrac{L_J(P_J^{\mathrm{Max}}g_J+P_S^{-i}+\varepsilon)^2}{(g_S^i g_J)}\\ \sqrt{\dfrac{g_S^i P_S^i L_J}{g_J}}-L_S^i P_S^i,\text{其他}\end{cases}\tag{16}$$

假设函数 $H(P_S^i)=\sqrt{g_S^i P_S^i L_J/g_J}-L_S^i P_S^i$,其一阶导数 $\mathrm{d}H/\mathrm{d}P_S^i=0.5\sqrt{g_S^i L_J/P_S^i g_J}-L_S^i$,二阶导数 $\mathrm{d}_H^2/\mathrm{d}(P_s^i)^2\leqslant 0$,因此 $H(P_S^i)$ 为凹函数,则有 $\tilde{P}_S^i=\max H=L_S^i g_S^i/(4g_J L_S^{i\,2})$。由式(16)可知,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 $P_S^{\mathrm{SE}}=0$。

2)$g_S^i/(\varepsilon+P_S^{-i}+P_J^{\mathrm{Max}})<L_S^i<g_S^i/(\varepsilon+P_S^{-i})$:如果 $P_S^i\leqslant L_J(\varepsilon^2+P_S^{-i})/(g_S^i g_J)$,$U_S(P_S^i,P_S^{-i},P_J)$ 关于 P_S^i 不减,当 $P_S^i\geqslant L_J(P_J^{\mathrm{Max}}g_J+P_S^{-i}+\varepsilon)^2/(g_S^i g_J)$ 时,$U_S(P_S^i,P_S^{-i},P_J)$ 递减,此时 $P_S^{\mathrm{SE}}=\hat{P}_S^i$。

3)$L_S^i<g_S^i/(\varepsilon+P_S^{-i}+P_J^{\mathrm{Max}})$:$U_S(P_S^i,P_S^{-i},P_J)$ 关于 P_S^i 递增,此时 $P_S^{\mathrm{SE}}=P_S^{\mathrm{Max}}$,则求得 P_S^{SE} 为:

$$P_S^{\mathrm{SE}}=\begin{cases}0,L_S^i>\dfrac{g_S^i}{\varepsilon+P_S^{-i}}\\ \dfrac{g_S^i L_J}{4g_J L_S^{i\,2}},\dfrac{g_S^i}{\varepsilon+P_S^{-i}+P_J^{\mathrm{Max}}}<L_S^i<\dfrac{g_S^i}{\varepsilon+P_J^{\mathrm{Max}}}\\ P_S^{\mathrm{Max}},L_S^i<\dfrac{g_S^i}{\varepsilon+P_S^{-i}+P_J^{\mathrm{Max}}}\end{cases}\tag{17}$$

综上所述,本文提出的 MDJC-AJ 算法实现过程描述如下:

输入 训练数据 $D=\{(s_i,b_i)\mid s_i\in S,b_i\in A\}$,经验池 E

输出 最优策略的估计 π_θ^*,接收信号的信干噪比 γ

1)初始化。$C_{\mathrm{sec}}^i(0)=0$,最大迭代次数 N_{it},网络参数 $\theta\leftarrow 0$,$w\leftarrow 0$,学习率 α 与 β,折扣因子 λ,熵正则化系数 η,干扰容忍阈值 $P_{J\mathrm{th}1}^{\mathrm{sec}}$ 与 $P_{J\mathrm{th}2}^{\mathrm{sec}}$。

2)迭代更新。对每个智能体(节点/簇头),每幕执行以下操作:

(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 $E\leftarrow\{s,b,C_{\mathrm{sec}}^i,s'\}$。

(2)执行:利用 $\pi_\theta(\cdot\mid s')$ 得到动作 b'。

(3)通信安全容量:$C_{\mathrm{sec}}^i(s)\leftarrow C_{\mathrm{sec}}^i+\lambda C_{\mathrm{sec}}^i(s')$。

(4)策略更新:基于式(9),更新策略网络参数 w。

(5)价值更新:基于式(10),更新策略网络参数 θ。

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

(7)信道选用策略:判断前一时刻被选信道在当前时刻的干扰等级,当等级为 3 时,最佳传输功率为 0,奖励函数 $C_{\mathrm{sec}}^i(s)=0$ 时,则转至(1)采样操作;当等级为 2 时,建立功率域博弈模型,由式(17)得到最佳传输功率;当等级为 1 时,则保持使用当前信道。

3)直至达到最大迭代次数 N_{it},结束。

2.3 算法复杂度分析

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

表 2 MDJC-AJ 算法复杂度分析

Table 2 Complexity analysis of MDJC-AJ algorithm

计算	复杂度
C_{sec}^i	$O(N_S C_1)$
更新 $C_{\mathrm{sec}}^i(s)$	$O(N_S C_2)$
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}	$O(N_S C_6)$

本文算法的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 $O(C_i)$,C_1 为与采样过程相关的常数,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更新中的步骤 1。

2)对于单个节点,根据策略 $\pi_\theta(\cdot\mid s')$,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 为与策略类型相关的常数,所有节点的运算复杂度为 $O(N_S C_2)$,该部分对应算法迭代更新中的步骤 2、步骤 3。

3)对于单个节点,基于式(9)、式(10),更新参数 w、θ 以及状态 s、动作 b,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相关的常数,所有节点的运算复杂度为 $O(N_S C_3)$,该部分对应算法迭代更新中的步骤 4 ~ 步骤 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 为与阈值个数相关的常数,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 为与式(15)相关的常数。

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,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 算法的总运算复杂度为:

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

2.4　基于簇头协助的信道选择算法

由于实际环境态势的多变性以及信息的局部性,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。

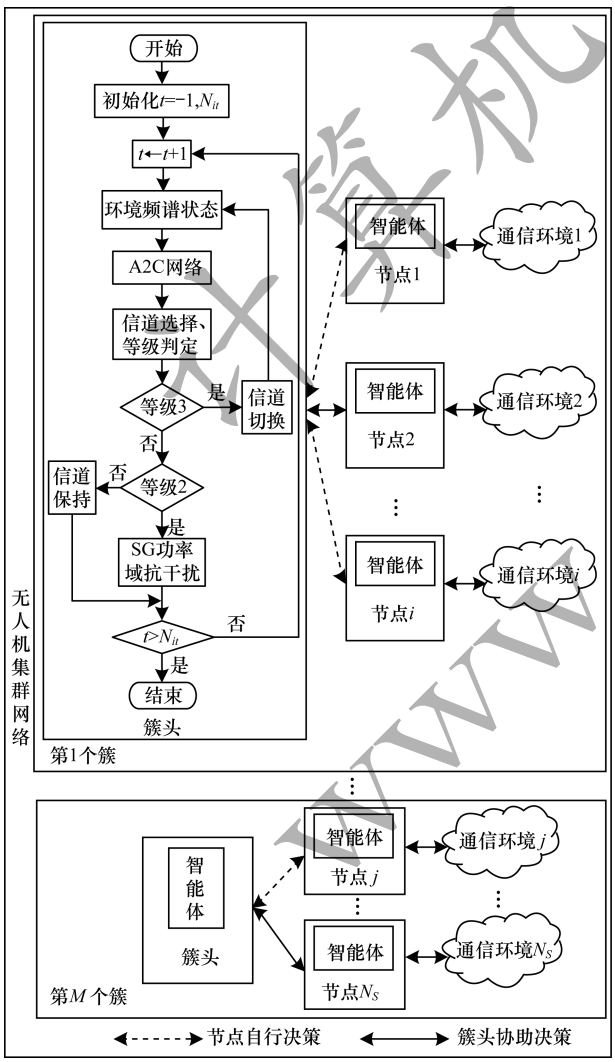

图 2　基于簇头协助的无人机集群网络抗干扰示意图

Fig. 2　Schematic diagram of UAV cluster network anti-jamming based on cluster head assistance

簇头协助从节点决策方法可描述为:各节点进行局部环境感知与信道决策时,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,使其能够进行可靠通信。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。为了不失一般性,图 2 中仅详细说明第一个簇头内部抗干扰算法的实现流程。

3　实验仿真与分析

为验证本文所提算法的有效性,实验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® Core™ i7-4790 CPU@3.60 GHz 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。

仿真条件设置为:每幕训练时长 $T=5\,000$,每个时间步长 $\Delta t=0.1$ s,每 100 幕为一个训练周期,总共设有 10 个仿真周期,经验池容量大小为 $E=10\,000$。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 $P_{J\text{th}1}^{\text{sec}}=0.5$、$P_{J\text{th}2}^{\text{sec}}=6$。

Actor 网络与 Critic 网络基本一致,不同的是最后的全连接层[20]。Actor 网络输出维度为 32×1,对应 32 个待选信道,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　Parameter setting of network structure

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

仿真 1　为验证本文所提算法的信道选择性能,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 $t_{\text{change}}=1\,500$ 和 $t_{\text{change}}=3\,300$,网络中节点个数为 4,编队及所选簇头已最优。实验对文献[4]Q 学习抗干扰(QL-AJ)算法、文献[8]演员-评论家抗干扰(AC-AJ)算法与本文算法的信道干扰情况决策成功率进行比较,结果如图 3 所示。从图 3 可以看出,在各个阶段内,相比 QL-AJ 算法与 AC-AJ 算法,本文所提 MDJC-AJ 算法的信道干扰情况决策成功率更高。

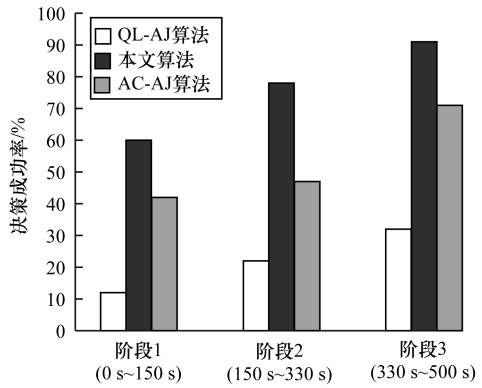

图 3 3 种算法的干扰信道情况决策成功率

Fig. 3 Channel decision success rate of jamming situation with three algorithms

为进一步说明 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。从图 4 可以看出,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。

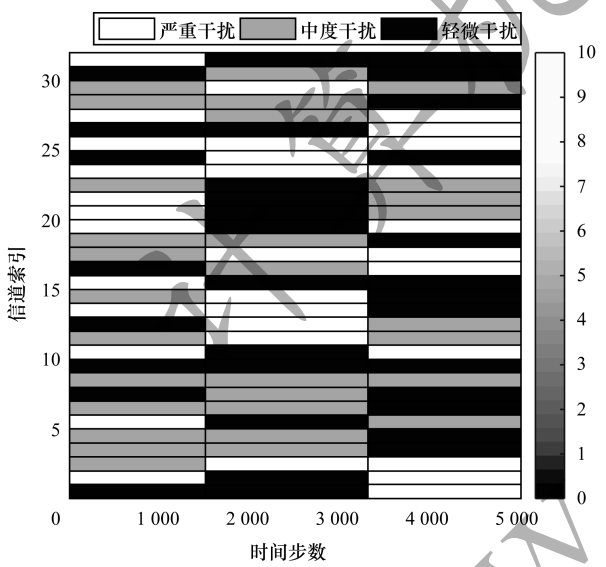

图 4 MDJC-AJ 算法对信道干扰情况判定结果

Fig. 4 Decision result of channel jamming situation by MDJC-AJ algorithm

仿真 2 考虑干扰机观测误差,实验对干扰机偏离 SE 对节点效用函数及其自身效用函数的影响进行研究,以检验网络的抗干扰性能。仿真参数设置为: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 $\varepsilon=0.5$,节点损耗系数为 $L_S^i=[0.1,0.2,0.3,0.7]$。干扰机观测误差 e 对节点效用函数之和 $U_S=\sum_i^{N_S}U_S^i$ 与干扰机效用函数 U_J 的影响如图 5 所示。

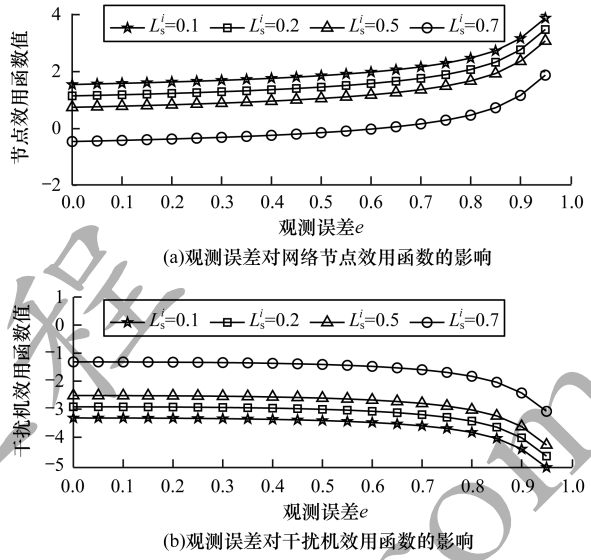

图 5 观测误差 e 对网络节点与干扰机效用函数的影响

Fig. 5 Influence of observation error e on utility function of network node and jammer

从图 5 可以看出,随着干扰机观测误差 e 的增加,节点效用函数之和呈现递增趋势,然而干扰机的效用函数呈现递减趋势。这是因为随着观测误差的增加,使得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的 SINR 的影响如图 6 所示。

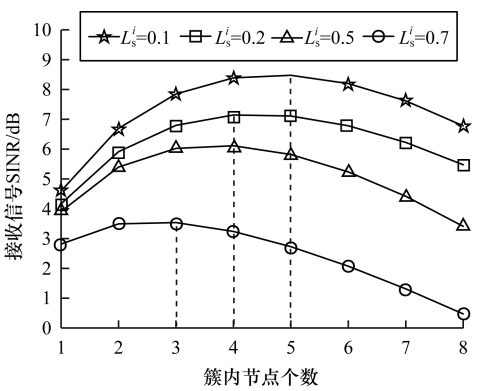

图 6 簇内节点数对整体接收信号 SINR 的影响

Fig. 6 Influence of the number of nodes in a cluster on the overall received signal SINR

从图 6 可以看出,随着节点数的增加,接收信号的 SINR 呈现先增大后减小的趋势。这是由于随着簇内节点数的增加,节点之间的互干扰不断增加,使得单个节点接收信号的 SINR 降低,进而导致簇内各节点接收信号的 SINR 降低。此外,在不同的节点损耗系数 $L_S^i=[0.1,0.2,0.3,0.7]$下,簇内节点数存在一个最佳个数,分别对应最佳簇内节点的个数为 $N_C=[5,4,4,3]$,在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　实验比较了 QL-AJ 算法、AC-AJ 算法与本文算法的抗智能干扰性能,如图 7 所示。从图 7 可以看出,在 3 种不同算法下,网络通信安全容量均随着训练时间的增加而不断提高,且与 QL-AJ 算法、AC-AJ 算法相比,本文算法的网络通信安全容量更高。值得注意的是,在 3 个阶段的突变点,上述 3 种算法得到的通信安全容量均骤减,之后恢复,然而本文算法较其他 2 种算法恢复的更快,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,AC 算法利用卷积神经网络强大的计算能力,相比 QL 算法提高了近 4 倍的计算速度;另外,相比于 AC-AJ 算法,本文算法能够降低学习过程的方差,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,同时提高了接收信号 SINR,从而得到的通信安全容量更高。

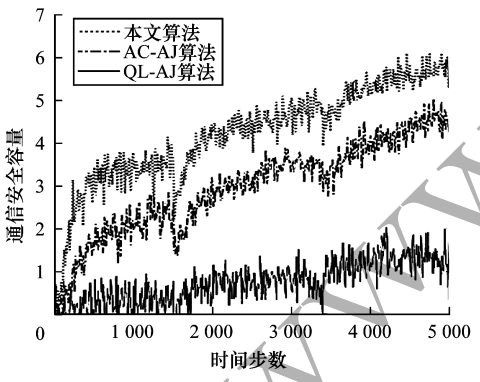

图 7　3 种算法的抗智能性干扰性能对比

Fig. 7　Comparison of anti-intelligence-jamming performance of three algorithms

为进一步说明本文算法的稳健性,定义单个状态 s 的均方值误差表示为近似值函数 $V_\theta(s)$ 与真实 $C_{\text{sec}}^i(s)$ 差的平方,并记作$\overline{\text{VE}}$:

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

其中,$|S|$为系统状态个数。

为验证所提方法算法收敛性能,实验对比了 QL-AJ 算法、AC-AJ 算法与本文算法的收敛情况。10 个仿真周期的平均均方值误差如图 8 所示。从图 8 可以看出,本文算法在经过 10 幕左右后已经收敛,比其他 2 种算法的收敛性能好,且得到的平均均方值误差更小。

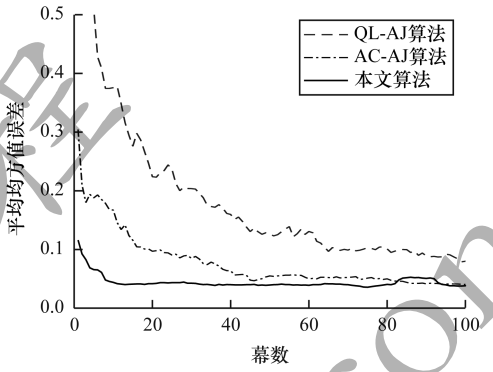

图 8　3 种算法的平均均方值误差变化曲线

Fig. 8　Average mean square error change curves of three algorithms

4　结束语

针对无人机集群网络对抗智能性干扰能力较弱的问题,本文提出一种 MDJC-AJ 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,以提高算法的收敛速度与信道决策成功率,并在此基础上,根据得到的功率干扰等级,利用功率域进行抗干扰,以减少信道切换时间、提高接收信号 SINR。通过仿真对比 QL-AJ 算法与 AC-AJ 算法,说明本文所提 MDJC-AJ 算法的整体抗干扰性能较好。同时,本文采用簇头协助的方法进一步改善网络的抗干扰性能。后续将考虑实际物理场景中存在不完全观测信息的情况,开展基于贝叶斯博弈理论的抗干扰方法研究,以满足实际工程需要。

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 Haitao, LUO Jiawei, LIU Changjun. Selfish bandit-based cognitive anti-jamming strategy for aeronautic swarm network in presence of multiple jammer [J]. IEEE Access, 2019, 7: 30234-30243.

[3] LIN Yu, WANG Tianyu. 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, 2019, 25(7): 4161-4171.
[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 X, LIU H P, 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 F, YU F R, SONG M, 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 Liang, CHEN Tianhua, LIU Jinliang, 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 Yifan, REN Guochun, CHEN Jin, 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 Luliang, XU Yuhua, SUN Youming, et al. A multi-domain anti-jamming defense scheme in heterogeneous wireless networks[J]. IEEE Access, 2018, 6: 40177-40188.
[14] PARISI S, TANGKARATT V, PETERS J, et al. TD-regularized actor-critic methods[J]. Machine Learning, 2019, 108(8/9): 1467-1501.
[15] TINNIRELLO I, BIANCHI 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 Xiaojie, ZHANG Ning, ZHANG Shan, 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, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[EB/OL]. [2019-11-01]. https://arxiv.org/abs/1801.01290.
[19] O'DONOGHUE B, MUNOS R, KAVUKCUOGLU K, et al. PQG: combining policy gradient and Q-learning[EB/OL]. [2019-11-01]. https://deepmind.com/research/publications/pgq-combining-policy-gradient-and-q-learning.
[20] MAO H Z, NETRAVALI R, ALIZADEH M. Neural adaptive video streaming with pensieve[C]// Proceedings of the Conference of the ACM Special Interest Group on Data Communication. New York, USA: ACM Press, 2017: 197-210.

编辑 刘继娟