

DOI:10. 20079/j. issn. 1001-893x. 230401001

“电磁信号智能感知”专题

引用格式:李博扬,刘洋,万诺天,等. 基于强化学习的无人机电磁干扰感知与抗干扰传输方法[J]. 电讯技术,2023,63(12):1855-1861. [LI B Y, LIU Y, WAN N T, et al. An electromagnetic jamming sensing and anti-jamming transmission method of UAV based on reinforcement learning [J]. Telecommunication Engineering, 2023, 63(12): 1855-1861.]

基于强化学习的无人机电磁干扰感知与抗干扰传输方法*

李博扬,刘 洋,万诺天,许 魁,夏晓晨,张月月,张 咪

(陆军工程大学 通信工程学院,南京 210007)

摘 要:无人机对于无线信道的依赖性和无线传播环境的开放性,导致其通信易受到恶意的电磁干扰。针对其中恶意的信道跟随干扰,在感知干扰信道信息的基础上,将无人机的发射功率和信道选择策略建模为马尔科夫决策过程(Markov Decision Process,MDP),利用强化学习算法对该通信系统的抗干扰方法进行智能优化,提出了基于赢或快学习策略爬山算法(Win or Learn Fast Policy Hill-climbing, WoLF-PHC)的抗干扰算法。仿真结果证明,所提算法能够将用户干信比降低至 0.1 以下,将用户可达速率在初始值基础上提升 14%,与 Q 学习算法和 PHC 算法相比具有更好的抗干扰传输性能。

关键词:无人机(UAV);抗干扰;电磁干扰感知;强化学习

开放科学(资源服务)标识码(OSID):  微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN929.5 文献标志码:A 文章编号:1001-893X(2023)12-1855-07

An Electromagnetic Jamming Sensing and Anti-jamming
Transmission Method of UAV Based on Reinforcement Learning

LI Boyang, LIU Yang, WAN Nuotian, XU Kui, XIA Xiaochen, ZHANG Yueyue, ZHANG Mi

(College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China)

Abstract: The dependence on wireless channel and the openness of wireless channel make unmanned aerial vehicle(UAV) vulnerable to malicious electromagnetic jammings. To combat channel-following jamming from jammers, a reinforcement learning-based anti-jamming strategy is proposed based on the perception of jamming spectrum information. The power control and channel access strategy of UAV is modeled as Markov Decision Process (MDP), and the anti-jamming strategy of the communication system is intelligently optimized by Reinforcement Learning Algorithm. An anti-jamming algorithm based on Win or Learn Fast Policy Hill-climbing(WoLF-PHC) algorithm is proposed. The simulation results prove that the proposed algorithm can reduce the user's Interference-to-Signal Ratio(ISR) less than 0.1, and increase the user's achievable rate by 14% on the basis of the initial value. Compared with Q-learning Algorithm and PHC Algorithm, it has better anti-interference transmission performance.

Key words: unmanned aerial vehicle (UAV); anti-jamming; electromagnetic jamming sensing; reinforcement learning

* 收稿日期:2023-04-01;修回日期:2023-05-29
基金项目:国家自然科学基金资助项目(62071485,62271503,62001513);江苏省自然科学基金项目(BK20201334,BK20201334,BK20200579,BK20231485)
通信作者:许魁

0 引 言

无人机(Unmanned Aerial Vehicle,UAV)在过去的几十年里已经得到了广泛的应用^[1]。无人机具有部署灵活、机动性高、隐蔽性强的特性,近年来被广泛应用于军事领域,其无人化的突出特点可以有效减少人员的战斗伤亡,在战场态势中拥有巨大的应用潜力。随着无人机成本的不断降低和设备体积的小型化,无人机也开始广泛应用于民用领域,包括通信中继、交通运输、目标快速检测等^[2-3]。

低空的无人机与地外通信或基于高海拔平台的通信相比,能够更快地部署,更灵活地重新配置,并且由于存在短距离视距链路,可能具有更好的通信通道^[4]。

然而,对电磁环境的依赖导致无人机具有易被干扰的致命弱点^[3]。近年来,许多学者针对该问题进行了一系列研究。在电磁感知研究方面,文献[5]提出了一种通过少量感知节点获取不完整采样电磁数据(Electromagnetic Data,ED)的方案,利用复合电磁图重建方法重构完整的ED;文献[6]通过设计一种生成对抗网络(Generative Adversarial Networks,GAN)用于提取时域电磁相关信息,提出了一种域映射算法,能够有效获取频域的电磁相关信息;文献[7]针对复杂电磁环境,研究了一种适应高噪声干扰强度的数字信号调制识别模型。在抗干扰策略研究方面,文献[8]在地面存在恶意电磁干扰的情况下,通过联合优化地面传感器(Ground Sensor,GS)的传输调度、无人机水平和垂直轨迹,最大化了有限飞行周期内GS之间的最小期望速率;文献[9]从博弈论的角度出发,提出了基于博弈的无人机与干扰机之间的竞争关系,根据无人机的效用函数,选择最优功率控制策略对抗干扰;文献[10]从静态博弈和动态博弈两方面分析了无人机网络的抗干扰问题,研究了无线信道衰落特性和飞行成本对静态博弈中纳什均衡(Nash Equilibrium,NE)的影响;文献[11]对无人机和用户效用函数进行了设计,利用斯坦伯格博弈(Stackelberg Game)模型模拟抗干扰场景,联合优化无人机的功率和信道选择进行抗干扰。

强化学习算法得益于可以在未知的环境中进行快速的学习,因此被广泛应用于通信抗干扰的研究中。Q学习算法是强化学习领域的经典算法,其核心思想是通过在环境中试错,不断优化智能体的决策,以提高智能体获得的奖励值^[12]。文献[13]利

用无人机的通信频率、运动轨迹和空间域的自由度提出了一种多参数规划的Q学习算法,优化了接收方的通信质量。文献[14]通过核密度估计(Kernel Density Estimation,KDE)估计有效干扰信号强度的概率密度函数(Probability Density Function,PDF),提出了一种基于深度强化学习的算法来优化系统中的功率控制策略,能够在降低功耗的同时降低通信的中断概率。文献[15]针对毫米波大规模多输入多输出(Multiple-Input Multiple-Output,MIMO)系统的抗干扰问题,提出了一种基于策略爬山(Policy Hill-climbing,PHC)算法的功率控制算法,与Q学习算法相比,能够实现更高的平均信噪比和系统可达速率。文献[16]针对无人机群抗干扰传输的场景,提出了一种智能算法,能够同时实现最大化的系统可达速率和最小化的跳频开销。文献[17]通过联合控制发射机的功率和信道接入对抗不同策略下的恶意电磁干扰,但未考虑实际通信场景的信道衰落和变化。

本文针对无人机空地抗干扰通信系统提出了一种基于强化学习的抗干扰方法。在对干扰信息进行感知的基础上,将无人机的功率和信道选择策略建模为马尔科夫决策过程(Markov Decision Process,MDP),利用强化学习算法对其进行智能优化,提出了一种基于赢或快学习策略爬山算法(Win or Learn Fast Policy Hill-climbing,WoLF-PHC)的无人机抗干扰策略,通过仿真验证了所提算法的抗干扰性能优于PHC算法和Q学习算法。

1 系统模型与问题建模

1.1 系统模型

图1所示为无人机空地通信系统模型。

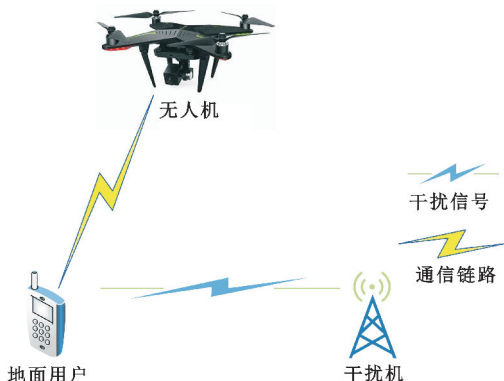


图1 系统模型

当无人机向地面用户发送信号时,在地面用户附近存在一个恶意干扰机,企图对地面用户的信道接入进行电磁干扰,阻断无人机与用户的合法通信。假设无人机-地面用户和干扰机-地面用户均为视距链路,地面用户接收到的信号为

$$y_u = \sqrt{p_t} h_{tu} s_t + \sqrt{p_j} h_{ju} s_j + n。 \tag{1}$$

式中: p_t 表示无人机的发射功率; p_j 表示干扰机的干扰发射功率,其发射功率均受最大发射功率的限制, $p_i \leq p_{imax}, i=t, j$; s_t 表示无人机的发射信号; s_j 表示干扰机的电磁干扰信号,假设 $E\{s_i\} = 0, E\{|s_i|^2\} = 1, i=t, j$; n 表示地面用户所处位置的加性高斯白噪声,服从均值为 0、方差为 σ^2 的高斯分布。

所有信道均假定为准静态平坦衰落模型,地面用户接收到的信干噪比为^[13]

$$R_{SIN} = \frac{p_t |h_{tu}|^2}{x p_j |h_{ju}|^2 + \sigma^2}。 \tag{2}$$

式中: x 为干扰因子,当地面用户被干扰时 $x=1$; 未被干扰时 $x=0$, 即

$$x = \begin{cases} 0, c_t \neq c_j \\ 1, c_t = c_j \end{cases}。 \tag{3}$$

式中: c_t 表示无人机与地面用户之间的通信信道; c_j 表示干扰机的干扰信道。

1.2 问题建模

本文的优化指标为地面用户的可达速率,当干扰机对地面用户的信道接入进行恶意电磁干扰时,可以通过动态调整无人机的发射功率及信道选择规避干扰,以最大限度地提高地面用户的可达速率。优化问题表示为

$$\begin{aligned} \max \quad & \text{lb}(1+R_{SIN}) \\ \text{s. t.} \quad & C1: p_t \leq p_{imax}, \\ & C2: c_i \in \{C\}, i=t, j. \end{aligned} \tag{4}$$

式中: p_{imax} 为无人机的最大发射功率,发射功率 p_t 小于等于 p_{imax} ; $\{C\}$ 为无人机与干扰机的可用信道集,当 $c_t = c_j$ 时,地面用户受到干扰机的恶意电磁干扰,信干噪比下降。由于干扰机的干扰策略是动态未知的,因此可以采用强化学习算法,通过在动态环境中的试错探索,优化该空地通信系统的抗干扰方法,达到提升地面用户可达速率的目的。

2 MDP 建模

强化学习的核心思想是通过智能体在环境中不断交互得到的奖励值,进行自身决策优化,以达到奖

励最大化的目的。其基本要素有状态、动作、策略、奖励函数、价值函数、环境模型^[12],学习过程可以描述为如图 2 所示的马尔科夫决策过程。针对信道跟随干扰,可采取强化学习的方式,将无人机的功率和信道选择策略建模为马尔科夫决策过程^[18]。

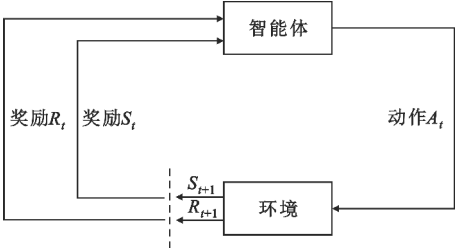


图 2 强化学习基本模型

在这一场景下,无人机空地通信系统作为环境,无人机作为智能体对干扰进行感知和学习。状态、动作、策略和奖励函数的设置分别如下^[17]:

状态: 系统状态 $s^n \in S$ 是智能体在当前时隙 n 从环境中观测到的离散化信息,包括前一时隙无人机与地面用户的通信信道 c_t^{n-1} 和前一时隙感知到的干扰信道 c_j^{n-1} :

$$s^n = \{c_t^{n-1}, c_j^{n-1}\}。 \tag{5}$$

式中:动作 $a^n \in A$ 是智能体在当前时隙 n 执行的行为,包括无人机的发射功率 p_t^n 和通信信道 c_t^n :

$$a^n = \{p_t^n, c_t^n\}。 \tag{6}$$

策略: $\pi(\cdot)$ 表示从当前系统状态映射到某个可选动作的概率分布,即 $\pi(s^n, a^n): s^n \rightarrow a^n$ 。

奖励: 智能体通过环境反馈的奖励值优化下一步的动作,因此奖励函数的设计与系统性能的期望目标相关联。在本文中,优化目标为最大限度地提高地面用户的可达速率,但同时也应考虑尽可能地降低无人机的功耗,包括发射功率代价以及信道切换代价。因此,奖励函数定义为

$$r^n = \begin{cases} R_{SIN}^n - J_p p_t^n - J_c z^n, R_{SIN}^n \geq R_{SIN0} \\ R_{SIN}^n - J_p p_t^n - J_c z^n - E, R_{SIN}^n < R_{SIN0} \end{cases}。 \tag{7}$$

式中: J_p 表示无人机的发射功率代价系数; J_c 表示无人机的信道切换代价系数; z^n 表示当前时隙的信道切换状态,定义为^[17]

$$z^n = \begin{cases} 0, c_t^n = c_t^{n-1} \\ 1, c_t^n \neq c_t^{n-1} \end{cases}。$$

若相邻时隙间信道发生变化则产生信道切换代价 J_c 。为了激励智能体的动作正向探索,设信干噪比门限值为 R_{SIN0} ,当用户接收信干噪比大于门限

值,即 $R_{\text{SIN}}^n \geq R_{\text{SIN}0}$ 时,奖励值表示为信干噪比减去功率和信道切换代价;若 $R_{\text{SIN}}^n < R_{\text{SIN}0}$,则在奖励值基础上减去一个惩罚值 E ,表示在当前状态下对智能体做出当前行为的惩罚。

3 算法实现

3.1 电磁干扰感知算法

电磁干扰感知是指通过对电磁信号的监测和分析,及时发现和定位电磁干扰信号的时域、频域特征的技术,是对抗恶意的电磁干扰的前提手段^[19]。

常用的电磁干扰感知方法有功率谱估计法^[20]、小波频谱感知^[21]、学习算法类频谱感知^[22]等,其中,小波感知算法不需要检测信号的先验信息,属于半盲检测,由于小波变换具有多分辨率特性,因此能够对频带的高频部分进行更为细致的频谱分析^[23];学习算法类频谱感知在干扰信息动态未知的抗干扰传输场景中更具有优势,可以分析识别获得干扰机的攻击信道。

本文基于长短期记忆(Long Short-Term Memory,LSTM)神经网络算法^[24]实现电磁干扰感知来获取对方的信道信息。干扰机采用信道跟随干扰的策略,以某一恒定功率值对地面用户进行干扰,即干扰机当前时刻的干扰信道始终与其感知到的上一时刻的通信信道保持一致。如图3所示,绿色部分代表无人机与地面用户合法通信信道,橙色部分代表干扰机攻击的信道。

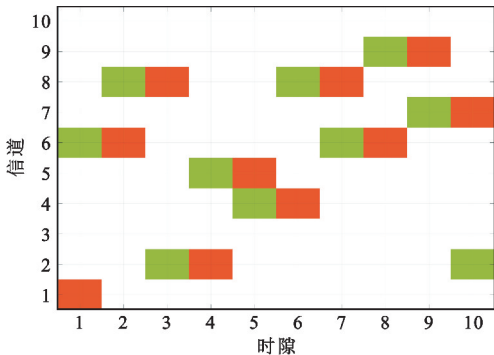


图3 通信合法信道与被干扰信道信息

无人机通过电磁干扰感知获得上一时刻的干扰信息后,依据链路反馈的通信质量对抗干扰决策进行优化,通过强化学习算法对决策过程进行训练,以提升用户的可达速率,降低干信比。

3.2 基于 WoLF-PHC 的抗干扰算法实现

Q 学习算法是强化学习中的典型算法,其中智

能体的目标是最大化当前时刻的长期累积折现奖励,即

$$R^n = \sum_{i=0}^{\infty} \gamma^i r^{(n+i+1)} \tag{9}$$

式中: $\gamma \in (0, 1]$ 为折现因子,用于权衡潜在的未来奖励对于当下的影响。Q 学习是将状态和动作张成一个二维 Q 表来存储 Q 值,Q 值函数 $Q(s^n, a^n)$ 表示在状态 s^n 下执行动作 a^n 的价值,可以表示为

$$Q(s^n, a^n) = E \left[\sum_{i=0}^{\infty} \gamma^i r^{(n+i+1)} \mid s^n = s, a^n = a \right] \tag{10}$$

智能体根据 Q 值来选取能够获得最大奖励的动作。传统的 Q 学习的 Q 值函数按照下式更新^[12]:

$$Q(s^n, a^n) \leftarrow (1 - \alpha) Q(s^n, a^n) + \alpha (r^n + \gamma \max_a Q(s^{n+1}, a^n)) \tag{11}$$

式中: $\alpha \in (0, 1]$ 表示学习率。为了避免学习结果陷入局部最优,Q 学习算法采用贪婪策略来平衡挖掘和探索之间的关系^[25]。在贪婪策略中,智能体以概率 ε 选择该状态下 Q 值最大的动作,以概率 $1 - \varepsilon$ 随机选择动作。智能体的动作选择概率表示为

$$\pi(s^n, a^n) = \begin{cases} \arg\max_a Q(s^n, a^n), & p_r < \varepsilon \\ a_{\text{random}}, & p_r \geq \varepsilon \end{cases} \tag{12}$$

Q 学习需要准确地估计每个状态-动作对应的 Q 值,每一步对 Q 表中的 Q 值进行更新,学习智能体在每一步做出动作,与环境进行交互,从而不断优化决策动作的能力,达到提升奖励值的目的。

WoLF-PHC 算法是在 Q 学习的核心思想基础上进行的拓展^[26],该算法结合了 WoLF (Win or Learn Fast) 和 PHC (Policy Hill-climbing) 算法,采用平均策略来近似均衡策略,通过可变的学习率增强了智能体的探索能力和效率。

在智能体执行一次动作到达下一状态且更新 Q 值后,平均策略的更新方式如下:

$$\bar{\pi}(s^n, a^n) \leftarrow \bar{\pi}(s^n, a^n) + \frac{1}{C(s)} (\pi(s^n, a^n) - \bar{\pi}(s^n, a^n)) \tag{13}$$

式中: $C(s)$ 表示状态 s^n 被访问过的次数。当智能体的表现比期望值好的时候,降低学习速度;反之,加快学习速度,寻找更优策略:

$$\pi(s^n, a^n) \leftarrow \pi(s^n, a^n) + \begin{cases} \delta, & a^n = \arg\max_a Q(s^n, a) \\ -\delta, & \text{其他} \end{cases} \frac{1}{|A| - 1} \tag{14}$$

$$\delta=\begin{cases} \delta_{\text{win}}, \sum_{a^n \in A} \pi(s^n, a^n) Q(s^n, a^n) > \sum_{a^n \in A} \bar{\pi}(s^n, a^n) Q(s^n, a^n) \\ \delta_{\text{lose}}, \text{其他} \end{cases} \quad (15)$$

- 基于 WoLF-PHC 的抗干扰算法步骤如下:
- 1 初始化: $Q(s, a) = 0, \pi(s, a) = \frac{1}{|A|}, C(s) = 0, \alpha, \gamma, \delta$
 - 2 开始迭代: $n = 1, 2, \dots, N$
 - 3 获取初始状态 s^0 ;
 - 4 在当前状态 s^n 下根据策略 $\pi(s, a)$ 选择动作 a^n , 获得奖励 r^n 并转移到下一状态 s^{n+1} ;
 - 5 更新 Q 表 $Q(s^n, a^n) \leftarrow (1 - \alpha) Q(s^n, a^n) + \alpha (r^n + \gamma \max_a Q(s^{n+1}, a^n))$
 - 6 更新平均策略: $C(s) = C(s) + 1; \bar{\pi}(s^n, a^n) \leftarrow \bar{\pi}(s^n, a^n) + \frac{1}{C(s)} (\pi(s^n, a^n) - \bar{\pi}(s^n, a^n))$
 - 7 根据式 (14)、(15) 更新策略;
 - 8 更新状态。

4 仿真与分析

本节对无人机空地通信系统基于强化学习的抗干扰进行仿真实验,对比在不同学习算法训练下的抗干扰性能指标。仿真参数如表 1 所示。

表 1 仿真参数	
参数	值
蒙特卡洛仿真次数	10 000
时隙数	100
学习率 α	0.8
折扣因子 γ	0.8
贪婪系数 ε	0.8
信道切换代价 J_c	0.2
功率发射代价 J_p	0.1
门限值 R_{SIN0}	10
惩罚值 E	5
信道数	10
功率集/W	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
噪声功率 σ^2 /W	3×10^{-10}
快学习速率 δ_{lose}	0.05
慢学习速率 δ_{win}	0.03

如图 4 所示,以地面用户为坐标轴原点建立直角坐标系,地面用户所处位置为 (0,0);无人机在以 (100 m,100 m) 为圆心、10 m 为半径的圆内悬停,并

且会发生 10 m 内的小幅度的位移抖动;干扰机位于 (150 m,0) 位置处。

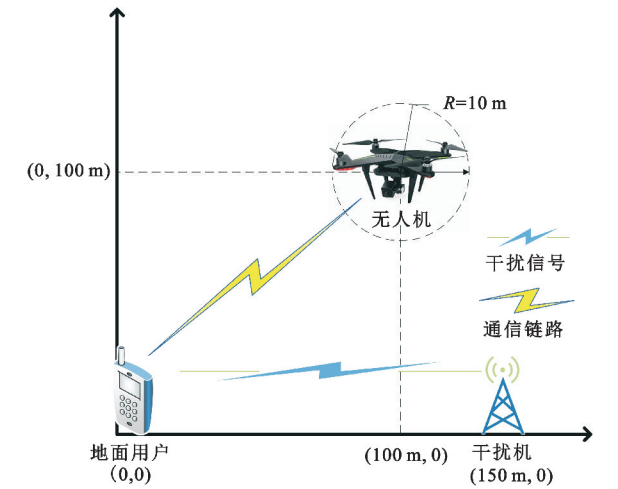


图 4 系统模型位置示意

无线信道中的路径损耗表示为 $L_p = L_{p0} - 10\beta \lg(d/d_0)$ 。式中: $L_{p0} = 30$ dB;参考距离 d_0 为 1 m; β 为信道中的路径损耗指数,无人机-地面用户链路与干扰机-地面用户链路的路径损耗指数均为 2.2。

针对电磁干扰,进行了干扰强度感知的仿真,利用干信比表示干扰信号的强度,定义为瞬时干扰信号强度与用户接收到的有用信号强度之比。图 5 所示为随机策略、Q 学习、PHC 算法和 WoLF-PHC 算法的干扰感知结果,可见随着学习进程的加深干信比随之降低, WoLF-PHC 算法最终收敛到的干信比值低于其他算法。

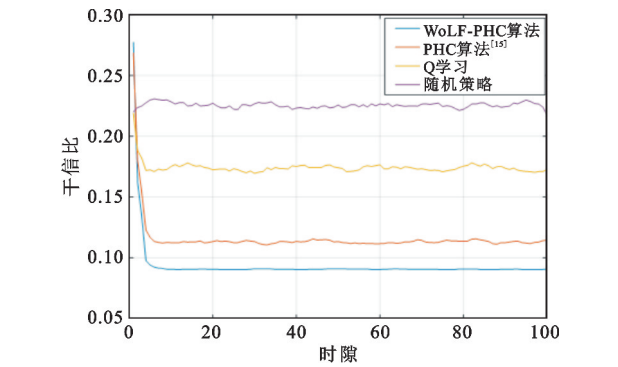


图 5 不同算法下的瞬时干信比

图 6 所示, WoLF-PHC 算法能够挖掘出智能体更大的潜力,最终能够收敛到更高的奖励值。智能体使用 WoLF-PHC 算法经过一段时间的学习后瞬时奖励由 12.3 提升到 16.2,提升了 31%。

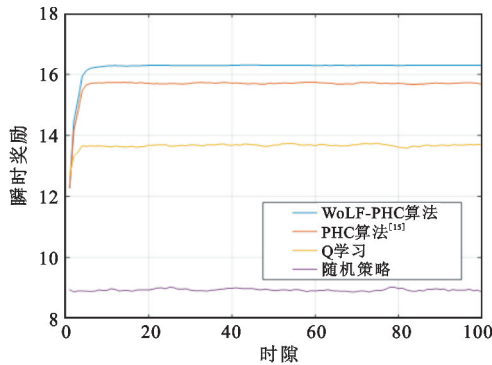


图 6 不同算法下的瞬时奖励

图 7 给出了不同算法下地面用户的中断概率,当地面用户的瞬时信干噪比小于 5 dB 时,判断无人机与地面用户间的通信发生中断。仿真结果显示,随机策略下中断概率维持在 0.22 不发生变化;使用 Q 学习算法,地面用户的中断概率由 0.12 降低到 0.08;使用 PHC 算法,中断概率从 0.125 降低到 0.025;使用 WoLF-PHC 算法,中断概率从 0.10 降低到接近于 0。

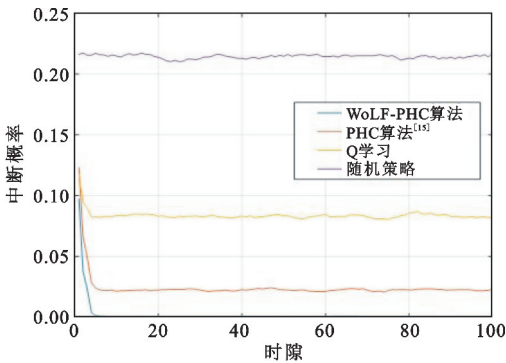


图 7 不同算法下的中断概率

图 8 所示为智能体使用不同算法学习后地面用户可达速率的变化,地面用户可达速率为本文优化的目标量。仿真结果显示,WoLF-PHC 算法能够达到较好的可达率提升效果,在初始值基础上提升了 14%;PHC 算法提升了约 11%;Q 学习算法提升了约 2%。

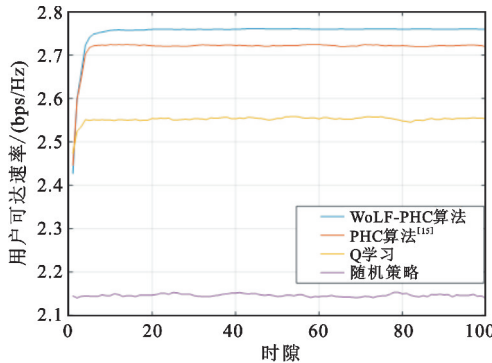


图 8 不同算法下的用户可达速率

5 结束语

本文针对无人机空地通信系统中的抗干扰问题,在对干扰进行电磁频谱感知的基础上利用强化学习算法对无人机进行训练学习,使其具有智能抗干扰传输的能力。通过联合控制无人机发射功率和信道选择对抗干扰机的恶意干扰,有效提升了用户可达速率,降低了干信比。通过仿真实验对比了 4 种不同算法的抗干扰性能,结果验证了所提的 WoLF-PHC 算法的抗干扰传输性能优于 PHC 算法、Q 学习算法以及随机抗干扰策略。

在下一步的研究中,将对强化学习的算法进行优化,以降低算法复杂度,缩短学习收敛的时间。

参考文献:

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

[2] CHENG Q,WANG H,ZHU B,et al. A real-time UAV target detection algorithm based on edge computing[J]. Drones,2023,95:1-17.

[3] CHENG Q,LI X,ZHU B,et al. Drone detection method based on MobileViT and CA-PANet[J]. Electronics,2023,12(1):223:1-16.

[4] GUPTA L,JAIN R,VASZKUN G. Survey of important issues in UAV communication networks[J]. IEEE Communications Surveys and Tutorials,2015,18(2):1123-1152.

[5] LIN D M,WANG H J,LIN Z A,et al. An effective scheme of building electromagnetic map for spectrum sensing[EB/OL]. [2023-03-15]. https://kar.kent.ac.uk/96014/1/Wang_An%20effective%20scheme.pdf.

[6] WANG H,LIN D,SHEN Z,et al. Two highly accurate electromagnetic map reconstruction methods[J]. IEEE Transactions on Vehicular Technology,2022,71(11):12419-12424.

[7] ZHA Y,WANG H,SHEN Z,et al. Intelligent identification technology for high-order digital modulation signals under low signal-to-noise ratio conditions[J]. IET Signal Processing,2023,17(2):1-11.

[8] DUO B,WU Q,YUAN X,et al. Anti-jamming 3D trajectory design for UAV-enabled wireless sensor networks under probabilistic LoS channel[J]. IEEE Transactions on Vehicular Technology,2020,69(12):16288-16293.

[9] 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.

[10] LV S,XIAO L,HU Q,et al. Anti-jamming power control game in unmanned aerial vehicle networks [C]// Proceedings of 2017 IEEE Global Communications Conference. Singapore;IEEE,2017;1-6.

[11] 姚昌华,程康,张建照,等. 利用部分重叠信道的多无人机自主协同优化抗干扰[J]. 电讯技术,2023,63(2):179-186.

[12] SUTTON R S,BARTO A G. Reinforcement learning: an introduction[M]. Cambridge:MIT Press,2018.

[13] XIAO L,LU X,XU D,et al. UAV relay in VANETs against smart jamming with reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2018, 67(5):4087-4097.

[14] MA N,XU K,XIA X,et al. Reinforcement learning-based dynamic anti-jamming power control in UAV networks; an effective jamming signal strength based approach [J]. IEEE Communications Letters, 2022, 26(10):2355-2359.

[15] XIAO Z,GAO B,LIU S,et al. Learning based power control for mmWave massive MIMO against jamming [C]// Proceedings of 2018 IEEE Global Communications Conference. Abu Dhabi;IEEE,2018;1-6.

[16] 康雅洁,林艳,张一晋. 基于贝叶斯 Q 学习的无人机集群抗干扰智能快跳频算法[J]. 航天控制,2022,40(2):73-78.

[17] 李芳,熊俊,赵肖迪,等. 基于快速强化学习的无线通信干扰规避策略[J]. 电子与信息学报,2022,44(11):3842-3849.

[18] 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.

[19] YUCEK T,ARSLAN H. A survey of spectrum sensing algorithms for cognitive radio applications [J]. IEEE Communications Surveys and Tutorials, 2009, 11(1):116-130.

[20] OHTA M,TAKYU O,FUJII T,et al. Novel two-stage spectrum sensing for energy detection with FFT [C]// Proceedings of 2016 IEEE Wireless Communications and

Networking Conference. Doha;IEEE,2016;346-350.

[21] 聂生鹏,刘金虎. 基于小波去噪的联合频谱感知优化算法[J]. 计算机应用与软件,2020,37(2):73-77.

[22] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1989,11(7):674-693.

[23] VERMA P,SINGH B. On the decision fusion for cooperative spectrum sensing in cognitive radio networks [J]. Wireless Networks, 2017, 23:2253-2262.

[24] 卢光跃,施聪,吕少卿,等. 基于 LSTM 神经网络的频谱感知算法[J]. 信号处理,2019,35(12):2070-2076.

[25] PEI X F,WANG X M,YAO J N,et al. Joint time-frequency anti-jamming communications: a reinforcement learning approach [C]//Proceedings of the 11th International Conference on Wireless Communications and Signal Processing. Xi'an;IEEE,2019;1-6.

[26] SCHWARTZ H M. Multi-agent machine learning: a reinforcement approach [M]. New Jersey: Wiley Publishing,2014.

作者简介:

李博扬 男,2001 年生于陕西韩城,2022 年获学士学位,现为硕士研究生,主要研究方向为智能抗干扰。

刘洋 女,1996 年生于吉林延边,2017 年获学士学位,现为硕士研究生,主要研究方向为通信抗干扰、强化学习。

万诺天 男,2001 年生于江西抚州,2022 年获学士学位,现为硕博连读研究生,主要研究方向为无线通信、移动通信。

许魁 男,1982 年生于安徽蚌埠,2009 年获博士学位,现为教授、博士生导师,主要研究方向为无线通信、移动通信。

夏晓晨 男,1987 年生于河北涿鹿,2017 年获博士学位,现为副教授、硕士生导师,主要研究方向为无线通信、移动通信。

张月月 女,1998 年生于江苏无锡,2023 年获硕士学位,现为博士研究生,主要研究方向为通信感知。

张咪 女,1999 年生于江苏镇江,2021 年获学士学位,现为硕士研究生,主要研究方向为信号处理。