

通用知识与专用信息共驱动的具身智能抗干扰通信策略生成方法

齐佩汉¹, 刘智豪^{1*}, 孙 雯², 梁琳琳¹

(1. 西安电子科技大学, 陕西西安 710000; 2. 中兴通讯股份有限公司, 陕西西安 710000)

摘 要: 随着新一代无线网络向通信、感知、计算一体化的复杂网络演进, 通信设备呈现爆发式增长。海量设备的并发接入引发了激烈的频谱冲突, 叠加开放信道下复杂多变的电磁干扰威胁, 导致大规模网络协同通信困难。现有基于多智能体的抗干扰算法虽能缓解上述冲突, 但其优化参数往往随智能体数量呈线性或指数级增加, 庞大的训练开销与内存消耗使其难以实现大规模扩展。针对面临的协同困难与扩展性受限问题, 提出了一种通用知识与专用信息共驱动的具身智能抗干扰通信策略生成方法。该方法通过参数共享策略构建统一的共享神经网络架构, 将优化参数的增长与智能体数量相分离, 在显著降低模型参数规模的同时提升系统的可扩展性。在此基础上, 利用共享卷积神经网络(Convolutional Neural Network, CNN)对电磁环境进行特征提取, 从原始频谱观测中学习可供所有智能体共同使用的通用知识表示。同时, 为区分个体, 方法进一步引入可学习的身份嵌入向量作为专用信息, 并与通用知识进行联合建模, 从而实现同质智能体的差异化决策, 进而生成高效的协同抗干扰通信策略。仿真结果表明, 该方法在固定干扰、扫频干扰和智能干扰场景下均具有显著优势, 有效缓解了大规模抗干扰通信中的协同冲突, 特别是在扫频干扰场景下, 该方法的自适应通信效率(Self-Adaptive Communication Efficiency, SACE)较多数智能体近端策略优化(Multi-Agent Proximal Policy Optimization, MAPPO)方法和独立DQN(Independent Deep Q-Network, IDQN)方法分别提升约22%和38%, 验证了协同机制的高效性。此外, 在不改变网络架构的前提下, 针对不同规模的多信道多用户场景进行仿真实验验证, 结果表明该方法能够稳定收敛并达到优异性能, 具有良好的可扩展性。

关键词: 具身智能; 多智能体强化学习; 协同抗干扰; 参数共享; 嵌入式身份; 分布式决策

基金项目: 国家自然科学基金(No.62171334); 国防基础科研项目(No.JCKY2023110C099)

中图分类号: TN929.5

文献标识码: A

文章编号: 0372-2112(2026)06-2727-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20260263

Embodied Intelligent Anti-Jamming Communication Strategy Generation Method Co-Driven by General Knowledge and Specialized Information

QI Peihan¹, LIU Zhihao^{1*}, SUN Wen², LIANG Linlin¹

(1. Xidian University, Xi'an, Shaanxi 710000, China; 2. ZTE Corporation, Xi'an, Shaanxi 710000, China)

Abstract: As new-generation wireless networks evolve toward complex networks integrating communication, sensing, and computing, the number of communication devices is experiencing explosive growth. The concurrent access of massive devices triggers intense spectrum conflicts, which, compounded by complex and dynamic electromagnetic interference threats in open channels, leads to difficulties in collaborative communication within large-scale networks. Although existing multi-agent-based anti-jamming algorithms can alleviate these conflicts, their optimization parameters often increase linearly or exponentially with the number of agents, and the resulting massive training overhead and memory consumption make large-scale expansion difficult. To address the challenges of collaboration and limited scalability, this paper proposes an embodied intelligent anti-jamming communication strategy generation method co-driven by general knowledge and specific information. This method constructs a unified shared neural network architecture through a parameter-sharing strategy, decoupling the growth of optimization parameters from the number of agents, thereby significantly reducing the scale of model parameters while enhancing system scalability. On this basis, a shared convolutional neural network (CNN) is utilized to extract features from the electromagnetic environment, learning general knowledge representations from raw spectrum observations that are available to all agents. Meanwhile, to distinguish individuals, the method introduces learnable identity embedding vectors as specific information and performs joint modeling with general knowledge, thereby achieving differentiated decision-making for homogeneous agents and generating efficient collaborative anti-jamming communication strategies.

Simulation results demonstrate that this method offers significant advantages in fixed jamming, sweep jamming, and intelligent jamming scenarios, effectively alleviating collaborative conflicts in large-scale anti-jamming communication; particularly in the sweep jamming scenario, the self-adaptive communication efficiency (SACE) of this method improves by approximately 22% and 38% compared to the multi-agent proximal policy optimization (MAPPO) method and the independent deep Q-network (IDQN) method, respectively, verifying the efficiency of the collaborative mechanism. Furthermore, simulation experiments on multi-channel multi-user scenarios of varying scales, conducted without altering the network architecture, show that the method achieves stable convergence and excellent performance, demonstrating good scalability.

Keywords: embodied intelligence; multi-agent reinforcement learning; collaborative anti-jamming; parameter sharing; embedded identification; distributed decision making

Foundation Item(s): National Natural Science Foundation of China (No.62171334); National Defense Basic Scientific Research Program (No.JCKY2023110C099)

0 引言

当前,新一代无线通信网络正由单纯的信息传输向通信、感知、计算一体化的复杂网络转变^[1]。在这一趋势下,无线通信设备已不再是静态终端,而是作为具身实体深度嵌入物理世界中,使得无线频谱环境逐步呈现出多主体、高动态和强对抗的发展趋势。随着此类具身设备的大规模接入,频谱接入效率和链路安全性将受到极大挑战^[2]。有限频谱资源与大量智能体并发接入需求之间的矛盾日益突出,同频干扰与链路中断问题愈发严峻^[3]。与此同时,无线信道的开放性使通信用频环境天然暴露于多源干扰与攻击之下。除了背景噪声和非故意共信道干扰外,智能化干扰机、协同压制以及战术对抗中的蓄意电磁攻击不断加剧,使通信链路出现中断、吞吐量锐减和时延突增等问题。

传统抗干扰技术,如直接序列扩频、跳频扩频、自适应调制编码和空域波束赋形等,在单系统、弱对抗或结构相对简单的环境中已得到广泛应用。然而,这些方法往往依赖固定的干扰模型和预设参数优化流程,在面对复杂时变、多主体博弈和结构未知的电磁环境时适应能力有限,在智能干扰机、协同攻击和多用户抢占频谱等场景中呈现出显著的性能退化。

近年来,具身智能理念的提出为解决上述难题提供了新范式。该范式主张智能体应通过与电磁环境的持续交互来构建认知并优化决策。深度强化学习与多智能体强化学习正是实现通信终端的智能化抗干扰决策的关键技术手段。通过构建“感知-决策-执行”的闭环,智能体能够将抗干扰决策建模为马尔可夫决策过程,利用深度Q网络(Deep Q-Network, DQN)^[4]、双重深度Q网络(Double Deep Q-Network, D-DQN)^[5]、近端策略优化(Proximal Policy Optimization, PPO)^[6]等算法在仿真环境中训练智能体,使其在未知干扰环境下学习动态策略,从而提升通信传输速率和频谱利用率。

文献[7]提出了一种基于经典Q学习的抗干扰方案,使通信节点能够通过在线决策通信频率的选择,从而自适应地规避干扰。在此基础上,文献[8]将对手建模与深度强化学习相结合,在策略学习中显式嵌入博弈推理机制,以提升对智能干扰的鲁棒性。针对无人机通信系统,文献[9]通过联合无人机功率控制与轨迹优化显著提升抗干扰性能。文献[10]利用时频谱瀑布图作为神经网络输入,从时频图中获取频谱特征提升系统抗干扰能力。面向宽带通信场景,文献[11]设计了分层式深度强化学习抗干扰框架,有效缓解了大规模频谱环境下状态-动作空间维度高、收敛缓慢的问题。文献[12]采用双动作网络和动作反馈机制进一步提高了抗干扰决策效率。文献[13]提出一种面向多域的学习型抗干扰方法,通过联合优化频率、功率等通信参数,实现对未知、动态干扰环境的高效自适应。这些研究主要集中在单用户抗干扰通信场景。随着抗干扰技术的进步,多用户抗干扰通信场景越来越受到学术界的关注。

在多用户场景中,用户不仅需要竞争有限的频谱资源,还会产生同信道干扰^[14]。面对日益复杂的多用户通信干扰环境,多智能体强化学习凭借其在协作决策、环境建模与策略优化等方面的系统性优势^[15],为多用户协同抗干扰提供了有效的理论框架与算法支撑,学术界由此提出了多种协同抗干扰策略,其中联合决策机制是早期的重要尝试。文献[16]以随机博弈建模多用户交互关系,提出联合决策式多智能体抗干扰算法,通过多智能体强化学习实现多信道阻塞干扰下的协同信道选择,既抑制外部恶意干扰,又降低用户间的内部冲突,实验表明其在数据接收率与收敛速度方面均优于跳频、感知式方法及独立强化学习策略。但联合决策方法存在一个根本性的可扩展性难题,其动作空间规模随智能体数量呈指数级增长^[17],难以在大规模网络中落地应用。

为规避维度爆炸问题,独立决策机制受到了广泛关注。该机制允许每个智能体独立学习个体策略,无

需显式构建联合动作空间。在此机制下,文献[18]提出了一种域内知识重用辅助的多智能体强化学习,通过复用历史学习到的策略加速在动态干扰环境下的收敛。文献[19]构建了双模Q学习框架,实现多用户场景下信道与功率的联合优化分配。文献[20]设计了分布式多智能体Q学习算法,使认知无线网络中的各用户能够同时实现分布式抗干扰与无冲突频谱接入。文献[21]提出基于强化学习的诱骗式抗干扰策略,通过学习引导自适应干扰机将攻击集中到特定信道,从而提高合法用户链路的可靠性。此方法既缓解了状态-动作空间维度爆炸,也保持了对干扰环境的良好自适应能力。

在独立决策基础上,为进一步提升全局性能,引入信息共享与交互机制成为了优化抗干扰策略的新趋势。文献[22]提出了一种基于Q学习的协同抗干扰策略,通过多用户联合优化频谱接入决策以有效降低相互间干扰。文献[23]通过引入邻居无人机的局部网络状态共享,并结合迁移学习机制,提升无人机集群抗干扰通信的策略收敛速度。文献[24]提出深度交叉校验Q学习方法,允许用户观测并利用其他用户的行为信息,从而作出更优的信道选择。文献[25]利用集中训练、分布执行(Centralized Training and Decentralized Execution, CTDE)框架构建了基于多智能体离散软行为者-评论家的多用户协同抗干扰策略,将信道切换、功率控制与调制编码联合纳入建模,在多域联合抗干扰方面展现出较强的鲁棒性。文献[26]提出上下文感知的分布式协同机制,在不同干扰模式下基于博弈论建模自适应调整用户策略,以降低通信负担并提高通信速率。

上述研究促进了多智能体抗干扰技术向可扩展、通信开销可控且具备强适应性的方向持续演进,但仍面临以下局限性:第一,现有研究多集中于单链路或少量用户场景,对大规模网络下的协同通信关注不足。第二,现有分布式算法优化参数随智能体数量线性增加,这在大规模网络应用中会引发庞大的训练开销和内存需求。

为应对上述挑战,本文提出一种通用知识与专用信息共驱动的具身智能抗干扰通信策略生成方法(Embodied Intelligent Anti-Jamming communication strategy generation method co-driven by General Knowledge and Specialized Information, GKSI-EIAJ)。该方法基于参数共享策略构建了统一的共享神经网络架构,利用共享卷积网络提取电磁环境的高维特征作为通用对抗知识。并进一步引入可学习的身份嵌入向量作为专用信息,赋予同质智能体差异化的决策能力。这种设计使得通信用户能够在不增加额外通信开销的前

提下,实现自主协同抗干扰。

本文的主要贡献总结如下:

(1)构建了面向大规模对抗的可扩展分布式架构。利用参数共享策略将网络优化参数的增长与智能体数量相分离,提升了大规模多智能体系统的可扩展性。

(2)设计了环境感知与决策优化的混合驱动框架。将卷积神经网络(Convolutional Neural Network, CNN)引入具身感知模块,从高维时频谱瀑布图中精准提取干扰特征,同时结合引入自适应熵正则化的PPO算法提高搜索能力,避免局部最优,在动态对抗中平衡探索与利用。

(3)提出了基于身份嵌入的隐式协同机制。通过将身份嵌入到共享策略网络中,使智能体能够根据自身身份和对干扰环境的观测自主生成差异化的传输策略,从而实现有效的抗干扰协调。

(4)验证了方法在多样化对抗场景下的性能。在包含固定干扰、扫频干扰和智能反应干扰的复杂电磁环境,进行了广泛的仿真对比实验。结果表明,所提方法在通信效率和稳定性方面始终优于所考虑的基准方案。

1 系统模型和问题描述

1.1 多集群无人机抗干扰通信系统模型

为具体刻画上述具身智能抗干扰通信场景,本文以多集群无人机抗干扰通信系统为基础展开建模,如图1所示,系统包含总指挥所、 N 个执行不同战术任务的无人机集群以及外部恶意干扰源。在该场景中,每个无人机集群内部设置一架核心无人机作为通信节点,负责与指挥所进行信息交互,并向集群内其他无人机转发指令与协同信息,实现集群内部的统一调度与控制。由于不同集群承担的任务类型和作战功能各不相同,各集群之间需工作在相互独立的信道上,以避免跨集群同频干扰。基于上述通信结构与频谱约束,本文将每个无人机集群抽象为一个通信用户进行建模与分析,其集合表示为 $N=\{1, 2, \dots, n\}$ 。各用户需要基于对电磁环境的观测进行分布式决策,其核心目标是在动态规避外部恶意干扰的同时,与其他用户保持频域隔离,从而最大化总指挥所与各集群间通信链路的可靠性与吞吐量。文中涉及的核心符号详见表1。

N 个用户和外部恶意干扰源共用同一频带 $[f_L, f_U]$,将该频带均匀划分为 K 个窄带子信道,记可用信道集合为 $C=\{1, 2, \dots, K\}$,每个子信道带宽为 $B=(f_U-f_L)/K$ 。时间采用离散时隙结构,记时隙索引为 $t=1, 2, \dots$,在每个时隙内完成信道观测、信道选择、数据发射与接收判决等操作。第 n 个用户在 t 时隙选择的工作信道记为 $c_n(t) \in C$ 。每个用户在选定信道上的发射功率为 P_s ,背景噪声功率谱密度为 N_0 ,在带宽 B

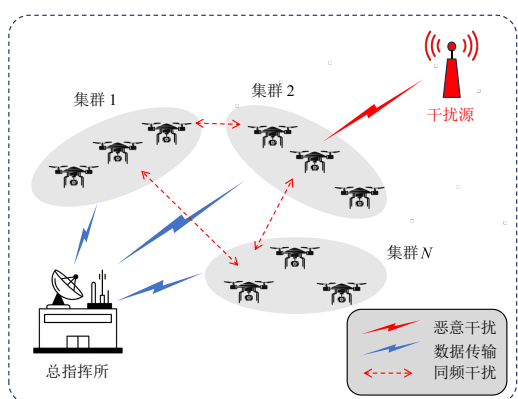


图1 系统模型

Figure 1 System model

表1 重要符号摘要

Table 1 Summary of important symbols

符号	参数描述
N	通信用户集合
J	干扰源集合
C	可用信道集合
W	系统总带宽
B	子信道带宽
t, T	时隙索引
$c_n(t)$	通信用户 n 在时隙 t 选定的工作信道
P_s	通信用户发射功率
P_j	第 j 个外部干扰源的发射功率
$g_{nm}(t)$	时隙 t 从发射节点 n 到接收节点 m 的等效增益
β	路径损耗指数
$E_n(t)$	用户 n 在时隙 t 接收到的所有外部干扰源功率之和
$I_n(t)$	多个合法用户选择相同信道时产生的内部冲突功率
$\gamma_n(t)$	用户 n 的瞬时信干噪比
γ_{th}	通信成功阈值
s_t	全局状态向量
a_t	联合动作向量
$o_n(t)$	用户 n 通过本地感知获得时频信息
H	时频图中提取的历史状态帧数
$r_n(t)$	通信用户 n 在时隙 t 的奖励函数

上的等效噪声功率记为 BN_0 。系统中存在外部干扰源,集合记为 J ,不同干扰源采用固定多音、梳状、扫频或智能反应等多种干扰策略,从而在频域上形成时变的干扰功率分布。

信道增益同时考虑大尺度路径损耗和小尺度衰落效应^[27]。第 x 个发射节点到第 y 个接收节点在时隙 t 的等效功率增益计算为式(1)所示:

$$g_{xy}(t) = d_{xy}^{-\beta} |h_{xy}(t)|^2 \quad (1)$$

其中, d_{xy} 为两节点间的几何距离, β 为路径损耗指

数^[28], $h_{xy}(t)$ 为小尺度衰落系数。对于通信用户自身链路,记发射与接收端之间的信道增益为 $g_{mn}(t)$ 。

在给定时隙 t 内,通信用户除背景噪声外还面临两类干扰:一类是来自外部干扰机的外部干扰,另一类是来自其他合法用户的同频干扰。设第 j 个干扰源的发射功率为 P_j ,其在时隙 t 作用的信道为 $k_j(t) \in C$ 。若用户 n 在该时隙选择的信道与 $k_j(t)$ 相同,则其接收端会受到该干扰源经信道衰落后的干扰功率。由此,第 n 个用户在时隙 t 所经历的外部干扰功率为式(2)所示:

$$E_n(t) = \sum_{j \in J} P_j g_{jn}(t) \delta(c_n(t) = k_j(t)) \quad (2)$$

另一方面,当多个通信用户选择了同一信道时,会在该信道上产生同频干扰。设 $m \neq n$ 为其他通信用户,其在时隙 t 的选信道为 $c_m(t)$,则用户 n 所经历的同频干扰功率为式(3)所示:

$$I_n(t) = \sum_{m \in N, m \neq n} P_s g_{mn}(t) \delta(c_m(t) = c_n(t)) \quad (3)$$

即所有与用户 n 共享同一信道的其他发射节点经各自传播路径叠加后的干扰功率之和。

在上述建模下,第 n 个用户在时隙 t 选取信道 $c_n(t)$ 时的瞬时信干噪比计算为式(4)所示:

$$\gamma_n(t) = \frac{P_s g_{nn}(t)}{N_0 B + E_n(t) + I_n(t)} \quad (4)$$

其中,当 $\gamma_n(t)$ 大于预设阈值 γ_{th} 时,可表示该用户在该时隙成功通信,否则通信失败。系统在一段时间内统计所有时隙、所有用户的 $\gamma_n(t)$ 与所选信道关系,可以得到通信速率、干扰规避率、碰撞率以及切频率等多维性能指标。另一方面,系统在每个时隙构造过去 L 个时隙在所有信道上的频谱观测记录,将其排列为维度为 $L \times K$ 的时频图矩阵,作为智能体决策的主要输入特征。

上述模型中,通信用户和干扰源均通过宽带频谱感知技术获取信道增益,这在实际工程应用中具有可行性。在本文中,假设通信用户之间以及通信用户与干扰源之间的信道增益为双方已知。

1.2 多智能体交互建模与问题描述

在系统模型给定的基础上,将多用户协同抗干扰过程建模为一个多智能体部分可观测马尔可夫决策过程。用户智能体集合为 $N = \{1, 2, \dots, n\}$,每个智能体 n 在每个时隙 t 从离散动作集合 $A_n = C = \{1, 2, \dots, K\}$ 中选择一个信道索引 $a_n(t) = c_n(t)$,共同构成联合动作向量 $a_t = (a_1(t), a_2(t), \dots, a_N(t)) \in A_1 \times \dots \times A_N$ 。记时隙 t 的全局状态为 $s_t \in S$,其包含了所有用户和干扰源过去 L 个时隙在所有信道上的原始发射功率的全局时频图信息。在给定联合动作和当前状态的情况下,环境按照信道传播与干扰演化规律生成下一时隙状态

s_{t+1} , 其动态特性由转移概率 $P(s_{t+1}|s_t, a_t)$ 描述。

由于每个智能体只能通过本地频谱感知装置获得有限信息, 智能体 n 在时隙 t 的观测记为 $o_n(t) = \Omega_n(s_t)$, 其中, $\Omega_n(\cdot)$ 为观测映射。每个智能体基于自身观测 $o_n(t)$ 生成动作的随机策略可表示为 $\pi_n(a_n(t)|o_n(t))$, 所有智能体策略的组合记为联合策略 $\pi = \{\pi_n\}_{n \in N}$ 。

为了刻画多智能体协同抗干扰性能, 引入单智能体即时奖励和系统协同回报。定义以下指示量: $\chi_n^{\text{succ}}(t) = \delta(\gamma_n(t) \geq \gamma_{\text{th}})$ 表示链路是否成功; $\chi_n^{\text{coll}}(t) = \delta(\exists m \neq n, c_m(t) = c_n(t))$ 表示是否发生同频碰撞; $\chi_n^{\text{jam}}(t) = \delta(E_n(t) > 0)$ 表示是否受到外部干扰; $\chi_n^{\text{sw}}(t) = \delta(c_n(t) \neq c_n(t-1))$ 表示是否发生信道切换, 用于限制信道切换, 以避免智能体在良好信道上频繁跳频。基于上述指示量, 定义第 n 个智能体在时隙 t 的即时奖励为式(5)所示:

$$r_n(t) = r_{\text{succ}} \chi_n^{\text{succ}}(t) - \lambda_{\text{jam}} \chi_n^{\text{jam}}(t) - \lambda_{\text{coll}} \chi_n^{\text{coll}}(t) - \lambda_{\text{sw}} \chi_n^{\text{sw}}(t) \quad (5)$$

其中, $r_{\text{succ}} > 0$ 用于鼓励高 SINR 下的成功通信, $\lambda_{\text{jam}}, \lambda_{\text{coll}}, \lambda_{\text{sw}} > 0$ 为外部干扰、同频碰撞和信道切换的惩罚系数。系统的协同即时回报定义为各智能体奖励的平均值, 计算方式如式(6)所示:

$$r_{\text{coop}}(t) = \frac{1}{N} \sum_{n=1}^N r_n(t) \quad (6)$$

该回报综合反映了通信成功率、抗干扰能力、多用户冲突和跳频代价之间的权衡关系。

在上述建模框架下, 多智能体协同抗干扰问题可表述为在给定信道与干扰模型的约束下, 寻找一组联合策略 π^* , 最大化系统的折扣协同回报, 为式(7)所示:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r_{\text{coop}}(t) \right] \quad (7)$$

其中, $0 < \gamma < 1$ 为折扣因子。

2 具身智能抗干扰策略

在多用户抗干扰场景中, 分布式强化学习方法通常为每个通信用户单独构建一套强化学习网络, 分别在各自的经验轨迹上独立更新。随着智能体数量增加, 参数规模和训练开销随之线性增长, 难以支撑大规模设备的联合训练。本文引入参数共享机制, 将所有通信智能体的策略映射统一到一套模型网络之上, 由所有智能体共同驱动同一组参数的更新, 从而显著提升样本复用效率。同时, 为克服单纯参数共享可能引发的策略同质化问题, 本文在共享网络上引入可学习的智能体身份嵌入作为专用信息, 每个智能体在时频特征的基础上拼接一段低维身份向量, 使同一网络在面对不同智能体时能够产生适度差异, 从而在同一参数空间内学习到相互区分但可协同的抗干扰策略。

在训练范式上, 本文采用 CTDE 架构以兼顾决策优化与部署效率。具体而言, 该框架在训练阶段引入集中式评价网络, 利用全局环境信息评估系统协同奖励并计算优势函数, 指导共享网络的更新, 从而有效塑造多用户协同行为。在执行阶段, 系统完全去除对中心评价网络的依赖, 各智能体仅需基于本地观测, 经由训练好的策略网络独立进行分布式信道选择, 无须显式交换状态信息, 实现了轻量化的高效部署。

基于上述架构, 本文进一步引入了 CNN 特征提取网络以及熵正则化机制, 构建了 GKSI-EIAJ 算法, 算法的训练框架如图 2 所示。

在 GKSI-EIAJ 中, 网络架构主要由 CNN 特征提取

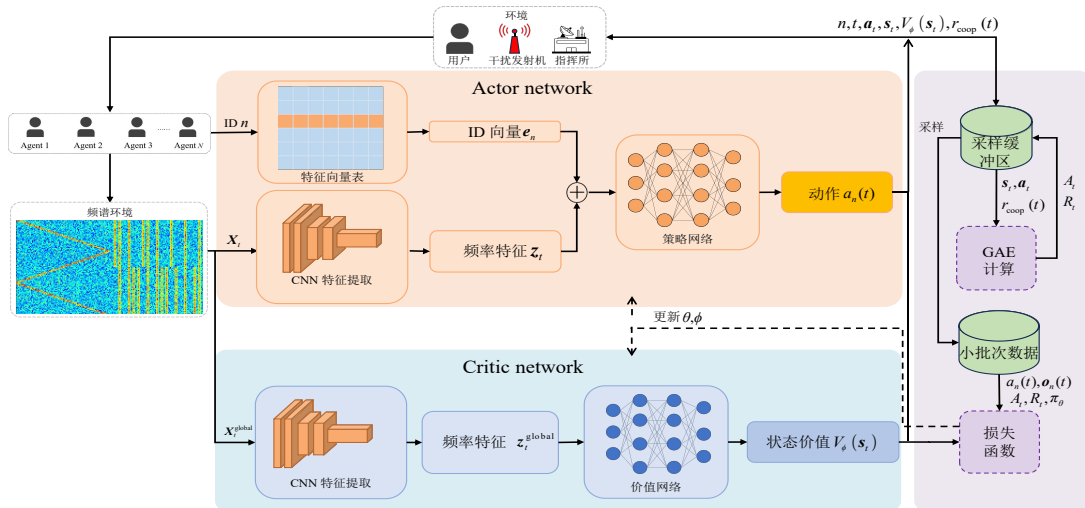


图2 GKSI-EIAJ算法框架

Figure 2 GKSI-EIAJ algorithm framework

网络、身份特征向量表、策略网络和价值网络组成。所有智能体共享一套演员 (Actor) 网络参数, 用于生成各自的信道选择策略。系统端部署一个集中式评论家 (Critic) 网络, 用于评估全局状态的价值。演员网络在共享的时频特征基础上引入智能体身份嵌入, 用于区分不同智能体。评论家网络仅依赖全局时频图, 不区分个体。

演员网络由 CNN 特征提取网络、身份特征向量转换表、策略网络组成。网络输入为当前时隙的时频观测以及智能体身份。设在时隙 t 的时频图记为 $\mathbf{X}_t \in \mathbb{R}^{L \times K}$, 其中, L 为频谱瀑布图历史长度, K 为可用子信道数。

CNN 特征提取网络参数如表 2 所示, 该模块由两个级联的卷积块和一层特征映射层构成。时频图 $\mathbf{X}_t$ 经对数压缩和归一化后, 被重塑为维度为 $1 \times H \times W$ 的张量并输入到 CNN 编码器, 其中, H 为时间维度, W 为频率维度。第一层卷积块作为特征提取的初级阶段, 该层包含 16 个尺寸为 3×3 的卷积核。采用了 (2, 1) 的非对称步长配合最大池化层, 使得特征图在时间维度上的压缩率高于频率维度, 从而在保留关键频谱特征的同时压缩历史冗余。第二层卷积块将特征通道数扩展至 $n_2 = 32$, 继续使用 3×3 卷积核进行深度特征提取与融合, 进一步捕获复杂的动态干扰模式与多智能体信道占用分布。第三层为特征映射层, 经过两层卷积与下采样后的特征图被展平为一维向量, 将高维时频信息线性映射为低维的 128 维状态特征向量 $\mathbf{z}_t \in \mathbb{R}^{d_z}$, 供后续的策略网络使用。

表 2 CNN 特征提取网络

Table 2 CNN feature extraction network

层级	参数设置	激活函数
卷积层 1	卷积核 3×3 、步长 (2, 1)、通道数 16、池化 2×2	ReLU
卷积层 2	卷积核 3×3 、步长 1、通道数 32、池化 2×2	ReLU
特征映射	输出维度 128	ReLU

为在演员网络中区分不同智能体, 引入身份特征向量表给每个智能体 n 分配一个可学习的身份嵌入向量 $\mathbf{e}_n \in \mathbb{R}^{d_e}$ 。该向量在训练过程中与演员网络其余参数一同更新, 用于编码智能体在空间位置、通信业务等方面的差异。策略网络的输入表示由时频特征向量和身份特征向量拼接而成, 如式 (8) 所示:

$$\mathbf{h}_n(t) = [\mathbf{z}_t, \mathbf{e}_n] \in \mathbb{R}^{d_z + d_e} \quad (8)$$

策略网络主体采用四层全连接结构的多层感知机 (MultiLayer Perceptron, MLP), 在前三层全连接层后采用 ReLU 非线性激活函数。 $\mathbf{h}_n(t)$ 通过策略网络进行

处理, 输出维度为 K 的分值向量 $\mathbf{l}_n(t) \in \mathbb{R}^K$ 。最终, 演员网络输出在信道的选择概率分布具体如式 (9) 所示:

$$\pi_\theta(a_n(t) = c | \mathbf{o}_n(t)) = \frac{\exp(\mathbf{l}_{n,c}(t))}{\sum_{k=1}^K \exp(\mathbf{l}_{n,k}(t))}, \quad c \in C \quad (9)$$

其中, θ 为共享演员网络参数。对于任意智能体 n , 其在时隙 t 的动作 $a_n(t)$ 通过在该分布上采样得到。可以看到, 所有智能体共享同一组网络参数 θ , 仅通过不同的身份嵌入 $\mathbf{e}_n$ 在同一特征空间内形成差异化的决策, 既避免了参数量随智能体数量增加而线性膨胀, 又在一定程度上减弱了纯参数共享所带来的策略同质化风险。

评论家网络的输入为当前时隙的全局频谱状态。全局时频图记为 $\mathbf{X}_t^{\text{global}} \in \mathbb{R}^{L \times C}$, 与单个智能体可见的时频图在结构上保持一致。评论家网络采用 CNN 特征提取网络和价值网络, 网络结构与 Actor 网络一致。 $\mathbf{X}_t^{\text{global}}$ 通过 CNN 特征提取网络映射为全局特征向量 $\mathbf{z}_t^{\text{global}}$, 随后经过价值网络输出估计价值, 计算如式 (10) 所示:

$$V_\phi(\mathbf{s}_t) = f_{\text{critic}}(\mathbf{X}_t^{\text{global}}) \quad (10)$$

其中, ϕ 为评论家网络参数, $\mathbf{s}_t$ 为时隙 t 对应的全局状态。

GKSI-EIAJ 训练阶段时, 所有智能体在同一环境中并行与频谱环境交互, 按照当前共享演员网络 π_θ 选择工作信道, 环境根据联合动作更新干扰状态和信道占用情况, 并向每个智能体反馈即时奖励。为突出系统协同性能, 将各智能体时间步 t 的即时奖励的平均值作为系统协同回报 $r_{\text{coop}}(t)$, 同时记录对应的全局时频图 $\mathbf{X}_t^{\text{global}}$, 构成经验三元组 $(\mathbf{s}_t, \mathbf{a}_t, r_{\text{coop}}(t))$ 。其中, $\mathbf{a}_t = (a_1(t), a_2(t), \dots, a_N(t))$ 为多智能体联合动作。经过固定长度的采样阶段后, 这些经验被写入采样缓冲区, 用于后续批量更新。

在采样缓冲区中, GKSI-EIAJ 采用基于集中式价值函数的广义优势估计^[29] (Generalized Advantage Estimation, GAE) 计算时序差分残差、优势函数估计和折扣回报。在折扣因子 γ 和优势平滑参数 λ 下, 其计算如式 (11)、式 (12) 和式 (13) 所示:

$$\delta_t = r_{\text{coop}}(t) + \gamma V_\phi(\mathbf{s}_{t+1}) - V_\phi(\mathbf{s}_t) \quad (11)$$

$$A_t = \sum_{i=0}^{\infty} (\gamma \lambda)^i \delta_{t+i} \quad (12)$$

$$R_t = A_t + V_\phi(\mathbf{s}_t) \quad (13)$$

GKSI-EIAJ 采用均方误差损失量化评论家网络估计的价值与计算得到的折扣回报之间的偏差, 具体计算为式 (14) 所示:

$$L_{\text{critic}}(\phi) = \mathbb{E}_t \left[\left(V_\phi(\mathbf{s}_t) - R_t \right)^2 \right] \quad (14)$$

在演员网络更新方面, GSKI-EIAJ 采用裁剪策略比率的 PPO 目标^[30], 并结合参数共享架构对多智能体进行统一处理。记 $\pi_\theta(a_n(t)|o_n(t))$ 为当前网络在观测 $o_n(t)$ 下动作 $a_n(t)$ 的概率, $\pi_{\theta_{old}}$ 为采样阶段冻结的旧策略, 则时间步 t 、智能体 n 对应的概率比率可写为式(15):

$$\rho_{n,t}(\theta) = \frac{\pi_\theta(a_n(t)|o_n(t))}{\pi_{\theta_{old}}(a_n(t)|o_n(t))} \quad (15)$$

在参数共享架构下, 将不同时刻、不同智能体的样本统一视为同一网络的独立样本, 对所有样本求和得到策略目标的经验近似。裁剪后的 PPO 目标函数具体写作式(16)所示:

$$J_{PPO}(\theta) = \mathbb{E}_t \left[\frac{1}{N} \sum_{n=1}^N \min(\rho_{n,t}(\theta) A_t, \text{clip}(\rho_{n,t}(\theta), 1-\epsilon, 1+\epsilon) A_t) \right] \quad (16)$$

其中, ϵ 为裁剪阈值。该目标在保证策略更新方向与优势估计一致的前提下, 对概率比率施加上下界, 从而在多轮迭代更新中抑制过大的策略步长, 提升训练稳定性。

为增强策略在复杂干扰环境下的探索能力, 在上述 PPO 目标基础上引入熵正则化项。对每个智能体的策略分布的熵计算方式如式(17)所示:

$$H(\pi_\theta(\cdot|o_n(t))) = - \sum_{c \in C} \pi_\theta(c|o_n(t)) \log \pi_\theta(c|o_n(t)) \quad (17)$$

同时取多智能体的平均熵作为当前时间步的策略不确定度量。带熵正则化的演员目标函数写为式(18):

$$J_{actor}(\theta) = J_{PPO}(\theta) + \alpha_t \mathbb{E}_t \left[\frac{1}{N} \sum_{n=1}^N H(\pi_\theta(\cdot|o_n(t))) \right] \quad (18)$$

其中, α_t 为熵权重系数。训练早期 α_t 取相对较高的数值, 以鼓励在信道选择上保持足够的多样性, 探索不同信道在同一干扰组合下的收益。随着训练进行, α_t 依据训练步数进行线性插值, 从初始值平滑衰减至较小的终值, 使策略逐渐由探索过渡到利用, 最终集中在少数高价值信道选择上。

在参数更新阶段, 首先利用前述价值函数对缓冲区中的状态序列进行价值估计, 结合协同回报按广义优势估计计算每个时间步的优势 A_t 与折扣回报 R_t 。随后, 将这一批经验随机打乱并划分为若干小批量, 在每个小批量上分别更新评论家网络和演员网络。评论家网络通过最小化价值回归损失函数, 对参数 ϕ 进行梯度下降。演员网络则利用裁剪策略比率和优势构造 PPO 目标, 并叠加熵正则项, 对参数 θ 进行梯度上升。记学习率分别为 μ_ϕ 和 μ_θ , 则一轮更新可概

括为式(19)所示:

$$\phi \leftarrow \phi - \mu_\phi \nabla_\phi L_{critic}(\phi), \quad \theta \leftarrow \theta + \mu_\theta \nabla_\theta J_{actor}(\theta) \quad (19)$$

其中, 梯度均通过采样缓冲区中小批量样本的蒙特卡罗平均进行近似。

算法 1 概述了基于 GSKI-EIAJ 的多用户协同抗干扰策略训练流程。首先, 算法初始化共享演员网络、评论家网络及熵权重等超参数, 并构建固定容量的采样缓冲区。在随后的每一轮迭代中, 所有智能体利用当前策略与环境交互, 收集包含全局状态、自身观测、联合动作及协同回报的经验序列, 并以先进先出

算法 1 GSKI-EIAJ 算法

输入: 总时间步 T , 更新间隔 T_{up} , PPO 迭代次数 D , 学习率 μ_θ, μ_ϕ , 熵

参数 $\alpha_{init}, \alpha_{final}$, 智能体数量 N

输出: 优化后的共享 Actor 网络参数 θ^* 和共享 Critic 网络参数 ϕ^*

1. 初始化共享的基于 CNN 的 Actor 网络 θ 、共享 Critic 网络 ϕ 以及采样缓冲区 $B=0$;
2. 初始化熵系数 $\alpha \leftarrow \alpha_{init}$;
3. FOR $t=1$ to T
4. FOR $n=1$ to N
5. 智能体 n 观测自身时频谱 $o_n(t)$ 并获取身份嵌入 e_n ;
6. 通过共享 CNN 提取时频特征 z_t 并与 e_n 进行融合 $h_n(t)$;
7. 根据随机策略 $\pi_\theta(a_n(t)|o_n(t))$ 选择抗干扰动作 $a_n(t)$;
8. END FOR
9. 在环境中执行联合动作 $a_t = (a_1(t), a_2(t), \dots, a_N(t))$;
10. 接收全局协调奖励 r_t 和下一时刻观测 $o_n(t+1)$;
11. 将转移样本 $(o_t, s_t, a_t, r_t, s_{t+1})$ 存入采样缓冲区 B ;
12. IF $t \bmod T_{up} = 0$
13. 基于 B 计算 GAE 优势值 A_t 和目标价值 V_{target} ;
14. 基于 t/T 更新熵系数 α ;
15. FOR epoch = 1 to D
16. 利用全局状态 s_t 更新共享 Critic 网络:
 $\phi^* \leftarrow \phi - \mu_\phi \nabla_\phi L_{critic}(\phi)$;
17. 更新共享 Actor 网络:
 $\theta^* \leftarrow \theta + \mu_\theta \nabla_\theta J_{actor}(\theta)$;
18. END FOR
19. 清空采样缓冲区 $B \leftarrow \emptyset$;
20. IF (最近 M 个周期的平均协同回报趋于稳定)
21. BREAK FOR;
22. END IF
23. END IF
24. END FOR

的方式存入缓冲区。当样本累积至设定阈值时, 算法从缓冲区中采样轨迹片段构成训练批次, 利用评论家网络计算优势估计, 并在多个训练轮次内执行价值拟合与策略更新, 同时动态调整熵权重以平衡探索与利用。待迭代过程满足收敛判据后, 算法输出最终优化的共享网络参数。

3 实验结果与分析

3.1 实验设置

本次仿真实验的硬件平台为一台配备 Intel(R) Core(TM) i7 处理器和 NVIDIA RTX 5070TI GPU 的工作站。软件环境为 Python 3.9 与 PyTorch 1.13.1, 操作系统为 64 位 Linux。仿真场景中设置 3 个无人机集群, 即通信用户数量为 3, 可用频谱信道数量为 10, 各信道中心频率分别为 [101, 103, 105, 107, 109, 111, 113, 115, 117, 119] MHz, 单信道带宽 2 MHz, 干扰源数量为 4 个, 负责固定多音、梳状、扫频和智能反应四种干扰模式。无人机集群在 (1 500 m × 1 500 m) 区域内随机分布, 指挥所和干扰源则分别位于区域左下和右上位置, 与无人机集群共享频谱信道。仿真实验相关参数设置详见表 3。本文在 20 MHz 的工作频段上预设三类典型干扰场景:

(1) 固定模式干扰 (Fixed Pattern Jamming, FPJ)。在仿真开始时生成三组具有不同频谱特征的静态干扰分布场, 每隔 5 个时间步在上述三组预设分布之间进行循环切换。干扰由两部分叠加构成, 其一为多音分量, 在全频带内随机选取若干载频, 总体占用可用频带约 20%; 其二为周期梳状分量, 以 6 MHz 为固定频率间隔在频谱上周期性部署干扰, 梳状分量的初始相位随机确定。

(2) 扫频干扰 (Sweep Jamming, SJ)。干扰源发射带宽为 4 MHz 的干扰信号, 其中心频率以每个时间步 2 MHz 的恒定步长在整个频段内往返扫动, 触及频段边缘后扫频方向反转。

(3) 智能反应式干扰 (Intelligent Reactive Jamming, IRJ)。干扰源模拟具备“侦测-锁定-压制”能力的智能对抗干扰模型, 每隔 5 个时间步, 干扰机识别出当前通信用户占用的中心频率。然后在该周期内, 随机挑选 2 个目标, 在这些目标频率上发射高功率干扰信号。若未能成功锁定有效目标, 则回退为采用 FPJ 模型的背景静态干扰场。

3.2 方法对比实验

在对比实验中, 主要采用自适应通信效率 (Self-Adaptive Communication Efficiency, SACE) 作为核心性能评价指标。相较于单一的吞吐量或切频率指标, SACE 旨在量化系统在保障有效信息传输能力的同时兼顾切换代价的综合表现。其具体定义如式 (20) 所示:

$$\text{SACE} = \frac{B \log_2(1 + \text{SINR}_{\text{linear}})}{1 + \eta \cdot f_{\text{sw}}} \quad (20)$$

该指标的分子项 $B \log_2(1 + \text{SINR}_{\text{linear}})$ 量化基于多智能体平均信干噪比所推导出的系统平均理论容量。其中, $\text{SINR}_{\text{linear}}$ 为系统内所有智能体的平均线性信干噪比。分母项则引入了基于平均切频率 (Average

表 3 仿真参数设置

Table 3 Simulation parameter settings

参数名称	值
频谱瀑布图历史长度 L	32
通信用户数量 n	3
干扰源数量 j	4
折扣因子 γ	0.99
优势平滑参数 λ	0.95
裁剪阈值 ϵ	0.2
Actor 网络学习率	$3 \times 10^{-4} \rightarrow 1 \times 10^{-5}$
Critic 网络学习率	$1 \times 10^{-3} \rightarrow 5 \times 10^{-5}$
熵系数 α_t	0.01 $\rightarrow$ 0.001
通信成功奖励 r_{succ}	60
外部干扰惩罚 λ_{jam}	300
同频碰撞惩罚 λ_{coll}	250
切频惩罚 λ_{sw}	10
切换惩罚系数 η	1
背景噪声 N_0	-90 dBm/Hz
用户发射功率 P_s	2 W
单信道干扰发射功率 P_j	40 W
路径衰落因子 β	2

Switching Rate, f_{sw}) 的惩罚项, η 为切换惩罚系数。该指标通过引入切换惩罚项, 综合考虑了频率合成器锁定时间及同步开销等物理代价。SACE 值越大, 表明方法在保障有效传输的同时能有效抑制频繁切换, 实现了通信性能与工程开销之间的良好平衡。

在统一的多信道干扰仿真环境和训练步数设置下, 本文将所提出的 GKSI-EIAJ 与以下几类代表性方法进行对比:

(1) 随机跳频方法 (Random)。每个通信用户在每个时隙独立、均匀地从 10 个可用信道中随机选择一个工作信道, 不进行任何学习, 用作无智能基线。

(2) 独立 DQN 方法 (Independent Deep Q-Network, IDQN)^[31]。每个通信用户基于自身的观测独立训练一套 DQN 网络, 以其私有频谱图为输入, 输出当前时隙的信道选择, 各用户之间既不共享策略也不交换参数。

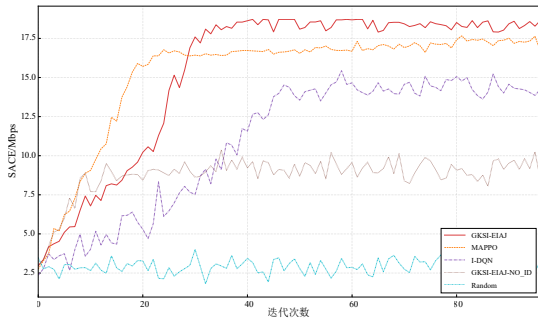
(3) 多智能体近端优化方法 (Multi-Agent Proximal Policy Optimization, MAPPO)^[32]。采用 CTDE 架构, 各用户在训练阶段利用全局信息辅助拟合价值函数。在执行阶段, 各用户仅依据自身私有观测, 通过 PPO 优化的策略网络输出信道选择决策。每个通信用户有各自的网络。

(4) 无身份嵌入的参数共享方法 (GKSI-EIAJ-NO_ID)。在共享网络输入端不引入身份嵌入, 各智能体仅依据时频观测做决策。

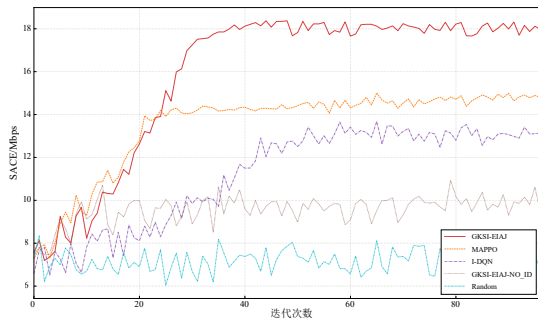
IDQN 与 MAPPO 均由全连接网络构建, 其具体的层级深度与神经元配置与 GKSI-EIAJ 保持一致。实

验结果均为 5 个不同随机种子独立重复运行后的统计平均值。

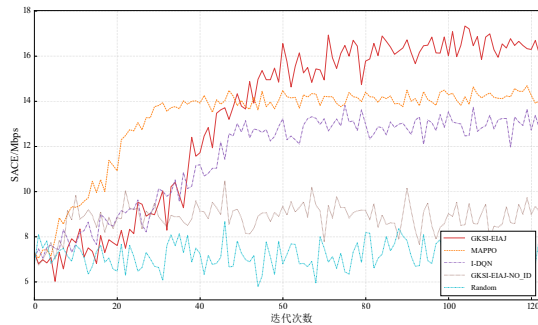
图 3 显示了本文提出的方法与四种基准方法在三种干扰环境下的 SACE 曲线。图 3(a)~(c) 分别展示了对抗 FPJ、SJ、IRJ 的结果。



(a) FPJ



(b) SJ



(c) IRJ

图 3 不同干扰模式下的 SACE

Figure 3 SACE under different interference modes

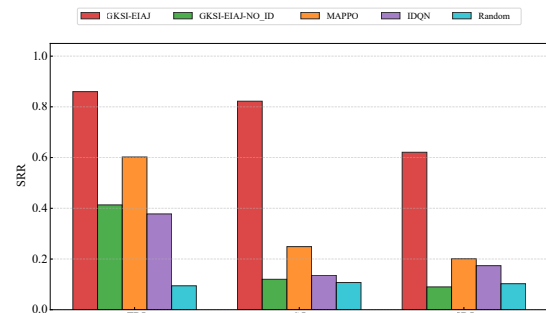
结果显示所提方法在三类干扰下均呈现出较快的收敛速度与更高的 SACE, 说明所提方法能够在较少迭代内完成对干扰的识别与对可用信道的稳定分配, 从而获得更高的有效速率并抑制不必要的切频代价。

在图 3(a) 中, FPJ 场景下 GKSI-EIAJ 收敛后 SACE 相比随机方法约提升 500%, 相比 GKSI-EIAJ-NO_ID 约提升 100%, 相比 IDQN 约提升 29%, 相比 MAPPO 约提升 7%。在图 3(b) 中, MAPPO、IDQN 方法 SACE 收

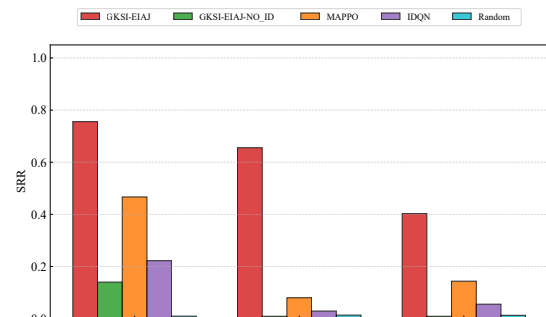
敛值相较于 FPJ 环境下分别降低了 14% 和 9%, 而 GKSI-EIAJ 只下降了 3%, 最终性能分别超过 MAPPO、IDQN、GKSI-EIAJ-NO_ID 和随机方法约 22%、38%、78% 和 146%。在图 3(c) 中, GKSI-EIAJ 收敛速度略缓于 MAPPO、IDQN, 但其性能依旧保持最优, 比 MAPPO、IDQN、GKSI-EIAJ-NO_ID 和随机方法分别提高了约 16%、26%、79% 和 124%。

本文进一步构建持续稳定通信占比 (Sustained Stable Communication Ratio, SRR) 以表征通信链路的稳定性。该指标统计抗干扰过程中用户所选信道满足同一信道连续通信成功阈值 (τ 时隙) 约束的时隙总数与总时隙数之比, 刻画模型对长时间连续可用信道的识别、选择与保持能力。本文设置阈值 $\tau=2$ (SRR-2) 和 $\tau=3$ (SRR-3), 对比分析五种方法在不同程度下的持续稳定通信能力。

图 4 为五种方法在 SRR-2 和 SRR-3 下的对比结果。在 SRR-2 指标下, GKSI-EIAJ 在固定与扫频干扰模式下分别达到 86% 与 82%。在智能对抗模式, 所提方法 SRR-2 达到 62%, 相较于表现最好的基线方法 MAPPO 的 19%, 约提升 226%。而在更为严苛的 SRR-3 指标下, 所有方法的性能均出现下滑。在扫频干扰环境下, MAPPO 和 IDQN 等基线方法分别降至 8% 和 3%, 但所提方法保持在 65%, 相对于 MAPPO 提高了约 713%。



(a) SRR-2



(b) SRR-3

图 4 不同干扰模式下的 SSR

Figure 4 SSR under different interference modes

SRR-2 与 SRR-3 指标的综合对比揭示了各方法生成策略在通信稳定性方面的差异。基线方法从 SRR-2 到 SRR-3 性能衰减严重, MAPPO 在扫频模式下的性能从 SRR-2 的 25% 跌至 SRR-3 的 8%, 跌幅超过 67%, 说明基线方法缺乏对干扰演化规律的预测, 无法识别长时间可用信道。相比之下, GKSI-EIAJ 的性能衰减平缓, 在不同干扰模式和不同间隙约束下均优于基线方法, 表明其在动态复杂环境中表现出更强的通信稳定性。

表 4 为 GKSI-EIAJ 在保持网络架构不变的情况下, 于多用户多信道场景下 FPJ、SJ、IRJ 三种干扰的 SACE 仿真结果。由表 4 可知, 所提 GKSI-EIAJ 在网络架构固定的条件下, 能够同时适配多用户多信道场景, 并均取得了优异的通信性能。

同时 IDQN 中每个智能体均需维护一套独立的 Q 网络, MAPPO 中每个智能体均维护一套独立的 Actor 网络和一个共享 Critic 网络, 二者可训练参数量可表示为 $P_{IDQN}(N) = NP_Q$ 和 $P_{MAPPO}(N) = NP_A + P_C$ 。而 GKSI-EIAJ 采用共享网络结构, 仅维护一套共享 Actor 和 Critic 网络, 其总参数量为 P_{AC} , 因此参数量可表示为 $P_{GKSI-EIAJ}(N) = P_{AC}$ 。

表 5 超参数敏感性分析

Table 5 Hyperparameter sensitivity analysis

参数名称	测试范围	SACE 影响/%
折扣因子 γ	[0.95, 0.995]	<1.71
优势平滑参数 λ	[0.90, 0.98]	<0.48
裁剪阈值 ϵ	[0.10, 0.30]	<1.82
Actor 网络学习率	$[1 \times 10^{-4}, 5 \times 10^{-4}]$	<2.87
Critic 网络学习率	$[1 \times 10^{-3}, 5 \times 10^{-3}]$	<1.44

3.3 消融对比实验

为验证 GKSI-EIAJ 网络架构的贡献, 本文设计了时频特征提取网络消融实验、参数共享身份建模消融实验。

3.3.1 时频特征提取网络消融实验

在本实验中将以下两种变体方法进行消融对比实验。变体一采用基于门控循环单元 (Gated Recurrent Unit, GRU) 的循环神经网络代替原有的 CNN 时频特征提取网络, 为 GKSI-EIAJ (GRU)。变体二为移除 CNN 时频特征提取网络的 GKSI-EIAJ (MLP)。

图 5 展示了本文提出的 GKSI-EIAJ 与上述变体在三种不同的干扰模式下的 SACE 性能对比曲线。图 5(a)~(c) 分别展示了对抗 FPJ、SJ、IRJ 的结果。

由图 5 可知, 所提方法凭借卷积网络对时频特征的深度提取能力在三类干扰场景下均取得了最优的收敛值。在 FPJ 下各模型性能差异较小, 所提方法较对比方法仅提升约 4%。当干扰模式向动态性更强的 SJ 演变, 所提方法的优势逐渐显著并实现了约 11%

表 4 多场景配置下算法性能

单位: Mbps

Table 4 Algorithm performance under multiple scenario configurations

unit: Mbps

场景配置	FPJ	SJ	IRJ
3 用户-10 信道	18.71	18.30	16.88
5 用户-12 信道	18.69	17.61	16.77
8 用户-20 信道	18.44	17.63	16.74
10 用户-30 信道	18.65	18.14	16.86
20 用户-30 信道	18.08	17.60	16.71

由上述关系可以看出, IDQN 与 MAPPO 的参数规模都会随智能体数量近似线性增长, 且增长项分别对应完整的 Q 网络或 Actor 网络, 而本方法凭借卷积核的平移不变性、参数共享和身份嵌入机制, 将模型参数规模的增长与智能体数量解耦, 大幅降低了存储开销与部署复杂度, 展现了其在动态可扩展的大规模多智能体系统中的巨大潜力。

为了分析学习率和折扣因子等超参数对收敛性能的影响, 进行了敏感性分析实验, 如表 5 所示。表 5 显示, 所有配置的 SACE 收敛性能变化不超过 2.87%, 证明了 GKSI-EIAJ 的超参数稳健性。

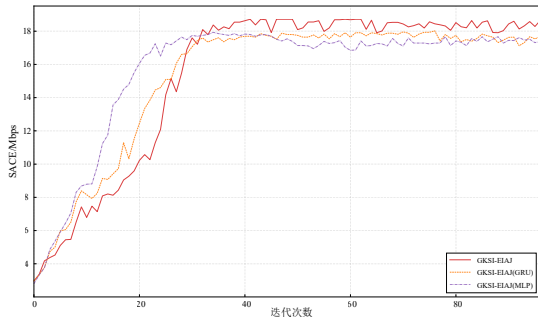
的性能增益, 而在应对更困难的 IRJ 时, 传统 MLP 与 GRU 模型因难以捕捉复杂的非平稳干扰特征过早陷入瓶颈, 而所提方法依然维持了较高的通信效率, 其 SACE 较对比方法分别提升约 14% 和 18%。随环境复杂度增加而扩大的性能差异结果充分证明了引入 CNN 特征提取网络对于提升多智能体系统在复杂电磁环境中稳定通信具有关键作用。

3.3.2 参数共享身份建模消融实验

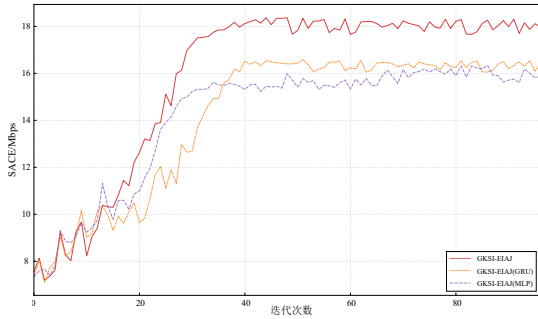
在本节中, 将与以下两种变体方法进行消融对比实验。变体一为不含身份嵌入的 GKSI-EIAJ-NO_ID。变体二为采用随机噪声向量代替身份嵌入的 GKSI-EIAJ-NOISE。

图 6 为三种方法在 FPJ、SJ、IRJ 干扰下的 SACE 对比结果。

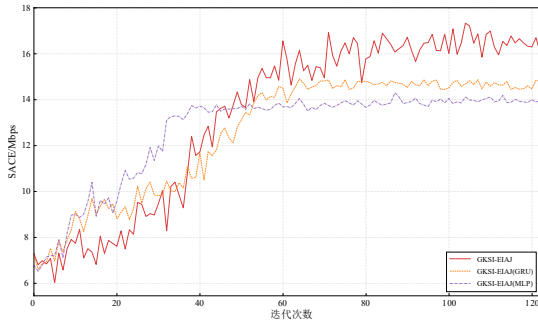
如图 6 所示, 两个消融变体均未能获得有意义的抗干扰策略, 其 SACE 在收敛后一直保持在约 10 Mbps 的较低水平。在共享频谱环境中, 纯参数共享使多智能体策略更新方向高度趋同。在缺乏身份区分信息



(a) FPJ



(b) SJ

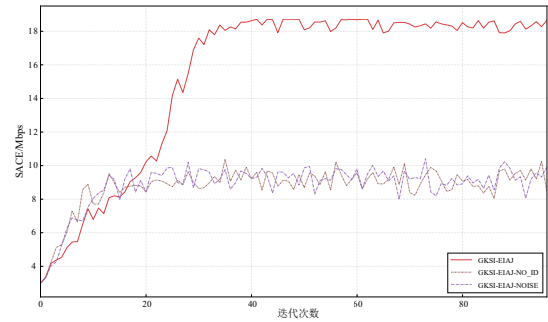


(c) IRJ

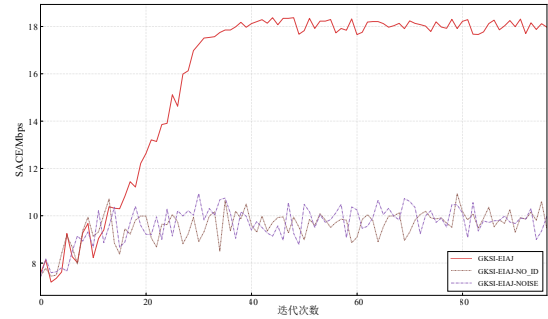
图5 时频主干网络SACE消融结果

Figure 5 Ablation results of the time-frequency backbone network for SACE

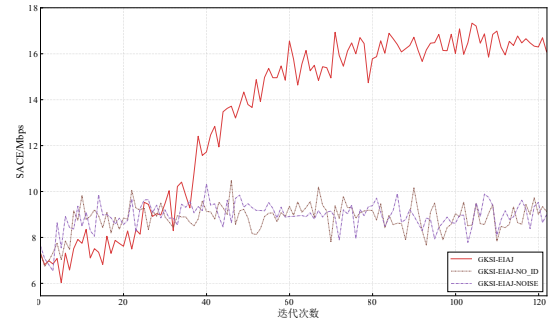
的情况下,不同智能体很容易学习到相似甚至相同的动作偏好,尤其是在某些信道于当前时刻表现出较高即时收益时,多个智能体会反复收敛到同一看似最优的信道选择。由于各智能体无法在策略层面形成有效的分化和协同行为,这种同质化决策会导致频繁的同频接入与信道碰撞,使系统无法充分利用有限频谱资源。同时噪声变体方法的失败表明,仅仅引入随机扰动不足以诱导一致的行为分化。相比之下,完整的GKSI-EIAJ能够稳定地收敛到接近最优的性能。实验表明,一个稳定的、可学习的身份表示对于实现参数共享下多智能体间的稳健协同至关重要。



(a) FPJ



(b) SJ



(c) IRJ

图6 身份嵌入SACE消融结果

Figure 6 Ablation results of identity embedding for SACE

4 结束语

本文提出了一种通用知识与专用信息共驱动的具身智能抗干扰通信策略生成方法,旨在解决大规模多智能体抗干扰通信中协同困难与扩展性受限的挑战。该方法在CTDE和参数共享架构下融合CNN时频特征提取网络,提取电磁环境特征作为通用知识,并通过引入可学习的身份向量作为专用信息,将通用知识和专用信息联合建模,有效解决了传统参数共享机制引发的策略同质化问题,在显著压缩模型参数规模的同时,成功实现了多智能体间的差异化决策与高效协同,从而最大限度地保障了用户通信链路的连续性与可靠性。实验表明,所提方法生成的策略具备优

异的通信稳定性与场景可扩展性,能够以极低的存储代价适配不同规模的通信系统。未来的研究将在此基础上进一步拓展,重点关注具备不同通信需求与硬件能力的异构用户频谱接入问题,提高模型在动态不确定环境下的鲁棒性与实用性。

参考文献

- [1] 王先梅, 刘向博, 任语铮, 等. 人工智能驱动的雷达通信一体化研究综述[J]. 雷达学报(中英文), 2025, 14(4): 1071-1091.
Wang Xianmei, Liu Xiangbo, Ren Yuzheng, et al. Review of research on artificial intelligence-driven joint radar communication[J]. Journal of Radars, 2025, 14(4): 1071-1091. (in Chinese)
- [2] Zhang Hao, Zhou Fuhui, Du Hongyang, et al. Revolution of wireless signal recognition for 6G: Recent advances, challenges and future directions[J]. IEEE Communications Surveys & Tutorials, 2026, 28: 3521-3563.
- [3] 许辰人, 马翔天, 徐昊天, 等. 5G 抗干扰技术综述[J]. 电子学报, 2023, 51(3): 765-778.
Xu Chenren, Ma Xiangtian, Xu Haotian, et al. A survey of 5G anti-interference technology[J]. Acta Electronica Sinica, 2023, 51(3): 765-778. (in Chinese)
- [4] 宋佰霖, 许华, 蒋磊, 等. 一种基于深度强化学习的通信抗干扰智能决策方法[J]. 西北工业大学学报, 2021, 39(3): 641-649.
Song Bailin, Xu Hua, Jiang Lei, et al. An intelligent decision-making method for anti-jamming communication based on deep reinforcement learning[J]. Journal of Northwestern Polytechnical University, 2021, 39(3): 641-649. (in Chinese)
- [5] 凌耀, 谢世珺, 梁豪, 等. 基于多奖励值 DDQN 智能通信抗干扰决策方法[J]. 电讯技术, 2025, 65(11): 1820-1827.
Ling Yao, Xie Shijun, Liang Hao, et al. An intelligent communication anti-interference decision algorithm based on multiple reward value DDQN[J]. Telecommunication Engineering, 2025, 65(11): 1820-1827. (in Chinese)
- [6] 潘筱茜, 张皎, 刘琰, 等. 基于深度强化学习的多域联合干扰规避[J]. 信号处理, 2022, 38(12): 2572-2581.
Pan Xiaolian, Zhang Jiao, Liu Yan, et al. Multi-domain joint interference avoidance based on deep reinforcement learning[J]. Journal of Signal Processing, 2022, 38(12): 2572-2581. (in Chinese)
- [7] Zhang Zixuan, Wu Qin hao, Zhang Bo, et al. Intelligent anti-jamming relay communication system based on reinforcement learning[C]//2019 2nd International Conference on Communication Engineering and Technology (ICCET). Piscataway: IEEE, 2019: 52-56.
- [8] Li Wen, Xu Yuhua, Chen Jin, et al. Know thy enemy: An opponent modeling-based anti-intelligent jamming strategy beyond equilibrium solutions[J]. IEEE Wireless Communications Letters, 2023, 12(2): 217-221.
- [9] Wang Haichao, Ding Guoru, Chen Jin, et al. UAV anti-jamming communications with power and mobility control[J]. IEEE Transactions on Wireless Communications, 2023, 22(7): 4729-4744.
- [10] Liu Xin, Xu Yuhua, Jia Luliang, et al. Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach[J]. IEEE Communications Letters, 2018, 22(5): 998-1001.
- [11] Li Yangyang, Xu Yuhua, Xu Yitao, et al. Dynamic spectrum anti-jamming in broadband communications: A hierarchical deep reinforcement learning approach[J]. IEEE Wireless Communications Letters, 2020, 9(10): 1616-1619.
- [12] Li Xiangchen, Chen Jienan, Ling Xiang, et al. Deep reinforcement learning-based anti-jamming algorithm using dual action network[J]. IEEE Transactions on Wireless Communications, 2023, 22(7): 4625-4637.
- [13] Li Yongcheng, Wang Jinchi, Gao Zhenzhen. Learning-based multi-domain anti-jamming communication with unknown information[J]. Electronics, 2023, 12(18): 3901.
- [14] 王为念, 苏健, 陈勇, 等. 基于多智能体深度强化学习的车联网频谱共享[J]. 电子学报, 2024, 52(5): 1690-1699.
Wang Weinian, Su Jian, Chen Yong, et al. Multi-agent reinforcement learning enabled spectrum sharing for vehicular networks[J]. Acta Electronica Sinica, 2024, 52(5): 1690-1699. (in Chinese)
- [15] 韩光洁, 朱胜超, 林川, 等. 多智能体强化学习: 从基础理论到前沿算法[J]. 电子学报, 2025, 53(12): 4756-4786.
Han Guangjie, Zhu Shengchao, Lin Chuan, et al. Multi-agent reinforcement learning: From foundational theory to cutting-edge algorithms[J]. Acta Electronica Sinica, 2025, 53(12): 4756-4786. (in Chinese)
- [16] Zhou Quan, Li Yonggui, Niu Yingtao. Intelligent anti-jamming communication for wireless sensor networks: A multi-agent reinforcement learning approach[J]. IEEE Open Journal of the Communications Society, 2021, 2: 775-784.
- [17] 陈平平, 张旭, 谢肇鹏, 等. 基于多智能体近端策略优化的多信道动态频谱接入[J]. 电子学报, 2024, 52(6): 1824-1831.
Chen Pingping, Zhang Xu, Xie Zhaopeng, et al. Multi-channel dynamic spectrum access based on multi-agent proximal policy optimization[J]. Acta Electronica Sinica, 2024, 52(6): 1824-1831. (in Chinese)
- [18] Zhou Quan, Niu Yingtao, Xiang Peng, et al. Intra-domain knowledge reuse assisted reinforcement learning for fast

- anti-jamming communication[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 4707-4720.
- [19] Zhang Xiaobo, Wang Hai, Ruan Lang, et al. Joint channel and power optimisation for multi-user anti-jamming communications: A dual mode Q-learning approach[J]. IET Communications, 2022, 16(6): 619-633.
- [20] Shan Qiucheng, Xiong Jun, Ma Dongtang, et al. Distributed multi-agent Q-learning for anti-dynamic jamming and collision-avoidance spectrum access in cognitive radio system[C]//2018 24th Asia-Pacific Conference on Communications (APCC). Piscataway: IEEE, 2018: 428-432.
- [21] Pourranjbar A, Kaddoum G, Ferdowsi A, et al. Reinforcement learning for deceiving reactive jammers in wireless networks[J]. IEEE Transactions on Communications, 2021, 69(6): 3682-3697.
- [22] Yao Fuqiang, Jia Luliang. A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks[J]. IEEE Wireless Communications Letters, 2019, 8(4): 1024-1027.
- [23] Lv Zefang, Xiao Liang, Du Yousong, et al. Multi-agent reinforcement learning based UAV swarm communications against jamming[J]. IEEE Transactions on Wireless Communications, 2023, 22(12): 9063-9075.
- [24] Elleuch I, Pourranjbar A, Kaddoum G. Deep cross-check Q-learning for jamming mitigation in wireless networks[J]. IEEE Wireless Communications Letters, 2024, 13(5): 1448-1452.
- [25] Jing Xiaorong, Wang Rui, Lei Hongjiang, et al. Multi-agent discrete soft actor-critic algorithm-based multi-user collaborative anti-jamming strategy[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 5025-5038.
- [26] Xu Yifan, Chen Jin, Feng Zhibin, et al. Hierarchical coordinated anti-jamming channel access in clustering networks: A multi-leader multi-follower Stackelberg game approach[J]. EURASIP Journal on Advances in Signal Processing, 2021, 2021(1): 42.
- [27] Goldsmith A. Wireless communications[M]. Cambridge: Cambridge University Press, 2005.
- [28] Rappaport T S. Wireless communications: Principles and practice[M]. 2nd ed. Upper Saddle River: Prentice Hall, 2002.
- [29] Schulman J, Moritz P, Levine S, et al. High-dimensional continuous control using generalized advantage estimation[PP/OL]. V6. arXiv (2018-10-20) [2026-01-03]. <https://arxiv.org/abs/1506.02438>.
- [30] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[PP/OL]. V2. arXiv (2017-08-28) [2026-01-03]. <https://arxiv.org/abs/1707.06347>.
- [31] Yang Ning, Zhang Haijun, Berry R. Partially observable multi-agent deep reinforcement learning for cognitive resource management[C]//GLOBECOM 2020 - 2020 IEEE Global Communications Conference. Piscataway: IEEE, 2020: 1-6.
- [32] Zhang Feng, Niu Yingtao, Zhou Wenhao. Intelligent anti-jamming decision algorithm for wireless communication based on MAPPO[J]. Electronics, 2025, 14(3): 462.

作者简介



齐佩汉 男, 1986年9月出生于河南省永城市。现任西安电子科技大学通信工程学院教授。主要研究方向为压缩感知、认知无线网络频谱感知及智能信号处理。中国电子学会会员编号: E190027763M。

E-mail: phqi@xidian.edu.cn



刘智豪 男, 2002年1月出生于河南省三门峡市。现为西安电子科技大学博士研究生。主要研究方向为无线通信、具身智能和强化学习。

E-mail: lzh7532022@163.com



孙雯 女, 1987年12月出生于江西省南昌市。现为中兴通讯软件测试无线资深专家。主要研究方向为无线通信领域软件测试技术与质量保障。

E-mail: sun.wen3@zte.com.cn



梁琳琳 男, 1986年10月出生于河南省驻马店市。现任西安电子科技大学网络与信息安全学院副教授。主要研究方向为隐蔽通信、物联网应用与安全。

E-mail: llliang@xidian.edu.cn