引用格式:李思达,徐逸凡,刘杰,等. 基于深度迁移学习的动态频谱快速适配抗干扰方法[J]. 信息对抗技术,2024, 3(1):33-45. [LI Sida, XU Yifan, LIU Jie, et al. Rapid adaption to dynamic spectrum anti-jamming approach based on deep transfer learning[J]. Information Countermeasure Technology, 2024, 3(1):33-45. (in Chinese)]

基于深度迁移学习的动态频谱快速适配抗干扰方法

李思达[1],徐逸凡[1*],刘 杰[1],林凡迪[1],韩 昊[1],易剑波[2],徐煜华[1]

(1. 陆军工程大学通信工程学院,江苏南京,210000; 2. 海南宝通实业公司,海南海口,570100)

摘 要 机器学习逐渐发展成为一种成熟强大的技术工具,并被广泛应用于无线通信抗干扰领域。其中,较为典型的有基于深度强化学习的抗干扰方法,通过与动态、不确定通信环境的不断交互来学习最优用频策略,有效解决动态频谱接入抗干扰的问题。然而,由于外界电磁频谱空间复杂、干扰模式样式动态多变,从头开始学习复杂的抗干扰通信任务往往时效性差,导致学习效率和通信性能显著下降。针对上述问题,提出基于深度迁移学习的动态频谱快速适配抗干扰方法。首先,通过构建预训练模型对已知干扰模式进行学习;其次,使用卷积神经网络提取现实场景下的感知频谱数据,重用过往经验优先启动加速适配;最后,运用微调策略辅助强化学习实施在线抗干扰信道接入。仿真结果表明,相较于传统强化学习算法,所提方法能够有效加快算法收敛速度,提升通信设备抗干扰性能。

关键词 动态频谱抗干扰;深度迁移学习;强化学习;快速适配

中图分类号 TN 973.3+2 **文章编号** 2097-163X(2024)01-0033-13

文献标志码 A **DOI** 10.12399/j.issn.2097-163x.2024.01.004

Rapid adaption to dynamic spectrum anti-jamming approach based on deep transfer learning

LI Sida[1], XU Yifan[1*], LIU Jie[1], LIN Fandi[1], HAN Hao[1], YI Jianbo[2], XU Yuhua[1]

(1. College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China; 2. Hainan Baotong Industrial Company, Haikou 570100, China)

Abstract Machine learning has become a mature and powerful technique and has been widely used in the fields of wireless anti-jamming communication. Deep reinforcement learning (DRL), one of the typical anti-jamming approaches, that enables an agent to learn an optimal frequency-using policy by constantly interacting with dynamic and uncertain communications environments, has been proposed as effective tools to solve the problem of dynamic spectrum accessing. However, learning a complex task from scratch often results in poor timeliness due to the complexity of the state space of the external electromagnetic spectrum and the volatile variation for the jamming patterns, which may cause a significant decline of the learning efficiency as well as communication performance instead. For these problems mentioned above, this paper proposes a rapid adaption to dynamic spectrum anti-jamming(DSAL) method based on deep transfer learning(DTL). Firstly, an adequately pre-trained model is estab-

收稿日期:2023-02-17 **修回日期**:2023-08-13

通信作者:徐逸凡,E-mail: yifanxu1995@163.com

基金项目:国家自然科学基金资助项目(62071488, U22B2022);江苏省自然科学基金资助项目(BK20231027)

lished learned from known jamming patterns. Further, convolution neural network(CNN) is used to extract jamming features from sensed spectrum data in real-world scenario and reusing knowledge that comes from previous experience contributes to scale up priority-startup and fast-adaption. In addition, fine-tune strategy is adopted to assist reinforcement learning (RL) algorithm to implement the task of on-line channel accessing for anti-jamming tasks. The simulation results show that, compared with traditional RL algorithm, our improved method can increase the convergence speed and reach better anti-jamming performance.

Keywords DSAL; DTL; RL; rapid adaption

0 引言

随着人工智能技术和软件无线电技术的发展,恶意用户(如恶意干扰机)可以方便地发射低成本的噪声信号实施干扰攻击,从而使通信设备面临严重的安全威胁,因此,通信抗干扰研究已成为无线通信领域中的重要课题之一[1]。近年来,为了应对无线通信中的恶意干扰攻击,特别是针对频谱的干扰攻击,研究者提出了多种智能抗干扰决策方法[2]。例如,文献[3]提出了一种新的通信抗干扰的方法范式,即动态频谱抗干扰(dynamic spectrum anti-jamming, DSAJ)。通过感知、学习和自我决策的一体化设计,通信设备可充分利用感知到的频谱信息来学习干扰模式,找到有效应对的抗干扰策略,自适应地优化频谱接入方案,从而提高频谱利用的有效性和灵活性,解决频谱资源稀缺和资源浪费等现实问题。基于动态频谱抗干扰框架,机器学习技术已被用于无线通信抗干扰中[4-6],并取得了一些具有开创性的研究成果。其中,强化学习(reinforcement learning, RL)采用决策—反馈—调整的在线学习框架,常被应用于无线通信中的抗干扰决策问题中[7-10]。由于电磁频谱环境具有动态变化的特点,从中提取出包含原始频谱信息的状态空间观测值是非常必要的。而由于外界电磁环境的不确定性以及复杂性,恶意干扰呈现出干扰种类复合多样、类型特征各不相同等特点。基于强化学习的方法虽然可以实时感知环境并采取相应的频谱接入决策,但由于电磁频谱环境中状态空间庞大、决策维度复杂等问题,此类方法通常难以收敛至最优策略[11]。为解决此问题,利用深度神经网络拟合复杂状态空间的优势,将强化学习决策与深度神经网络相结合,可解决复杂状态空间下的智能抗干扰决策问题。与强化学习不同的是,深度 Q-网络(deep Q-network, DQN)的 Q 值不是由状态值函数计算而来的,而是基于人工神经网络通过对表征信息的感知学习得到的。

深度学习结构利用原始频谱数据提高了通信抗干扰的性能,但也带来了训练时间和计算复杂度大大增加的代价。由于深度强化学习算法对于感知到外界干扰的规律性要求较高,因此会带来算法在前期探索阶段收敛较慢的问题。每当干扰模式进行了切换,则需智能体对通信环境重新进行学习,造成现有算法在实际应用中的局限性大大增加,对抗智能干扰的效果变差,如若处于干扰模式快速切换的动态环境下,深度强化学习的方法很难甚至不能收敛。

为了解决上述问题,迁移学习(transfer learning, TL)成为一种可采取的解决方案[12-13]。迁移学习可以利用数据、任务或模型之间的相似性,将在旧领域学习过的模型和知识应用于新的领域。结合迁移学习的方法,能够帮助解决部分强化学习在初期的探索学习阶段因环境动态未知而导致迭代经验不足、收敛速度缓慢、算法性能不佳等问题。迁移学习中,使用最广泛的方法是“预训练和微调”(pre-train and fine-tune)范式——一种与神经网络相结合的深度迁移学习(deep transfer learning, DTL)方法,已被证明在获取可迁移知识方面是可行、有效的,并适用于各种下游任务[14-16],即对大量源域数据集中和目标域数据相似的参数在深度神经网络上进行训练,并导出到目标深度神经网络模型中,使用来自新场景的有限数据进行训练和微调。该方法的优势主要表现在:在相同的任务上,预训练模型与从头开始训练(train from scratch)相比,大大缩短了训练时间,加快了训练的收敛速度,当训练数据较少时,能够带来较为显著的性能提

升。文献[17-18]中相关研究证明，此方法对于深度迁移网络的学习和优化来说有着非常好的促进作用，且预训练好的模型通常都是在大数据集上进行的，无形中扩充了训练的数据量，提升了模型的鲁棒性和泛化能力；在这种设定下，使用给定的无线通信场景训练机器学习模型，然后将模型迁移部署到一个全新的无线通信环境中，可以用于预测无线通信系统的频谱状态、信道容量、信号能量等各项参数指标。

综上所述，本文主要考虑现实场景下的抗干扰通信，针对复杂频谱环境下干扰模式动态变化导致算法收敛较慢甚至难以收敛的问题，侧重从节约抗干扰重新适配时间成本的角度出发，在现有的抗干扰研究算法的基础上，结合深度强化学习、基于模型的迁移学习等理论和方法，以最大化用户吞吐量为优化目标，力求实现模型共享、参数复配、经验重用，降低计算算力要求，加快算法收敛速度，提升抗干扰通信性能，打破传统算法因方法重塑、模型重建、数据重训导致的时间成本大打折扣的壁垒，在战场或应急救援等时间要素致胜的场景下抢占先机，掌握主动权。

1　系统模型与问题建模

1.1　系统模型

考虑一个典型的通信场景，如图 1 所示，其中设置了一对合法用户（由一个发射端和一个接收端组成）和一个或多个干扰机。受频谱瀑布图[19]的启发，为了便于理解问题，考虑了“时间-频率”的双重维度结构：从时间的维度出发，把连续的时间（$t\in[0,\infty)$）分成一个个离散的小时隙。在持续感知的每个时隙中，在接收端所配置的智能体对频谱环境进行感知，实时生成抗干扰频谱策略，并引导发送端选择某一特定频率进行通信。当每个时隙结束时，发送端还会接收到一个指示传输是否成功的 ACK 信号；从频率的维度出发，将整个频率范围平均分配成多个离散的信道（即“信道化”），用 B_u 和 B_j（$j\in J$，假定共有 J 个干扰机）分别表示合法用户和恶意干扰的频带带宽，用于通信的共享信道数即为 $N=B_u/b_u$，其中 b_u 表示用户基带信号的带宽。由此，将可供选择的传输信道集定义为 $A_u=\{d_0,d_1,\cdots,d_{N-1}\}$。

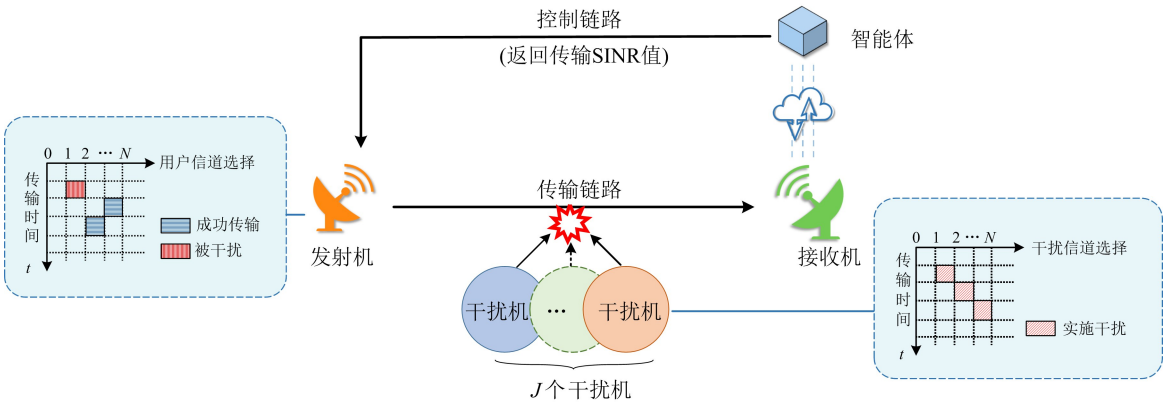


图 1　系统模型

Fig. 1　System model

图 2 所示为同一时隙下选择不同信道通信时的不同状态示意图。指引用户决策的智能体和恶意干扰机同时在时隙 t 内选择干扰和通信的信道和频率，且时隙 t 内选择的信道保持不变。在时刻 t，智能体选择信道 $a_t\in A_u$ 传输数据包并进行通信。干扰机则可以在每个时隙 t 内对一个或多个信道实施干扰攻击，力求通过功率压制覆盖信道中的通信频率，以干扰通信链路的传输。当给定时隙 t 内用户选择的通信信道没有受到干扰频率的冲击，则被认定为传输成功，否则抗干扰决策失利。

研究中，经常通过功率谱密度（power spectral density，PSD）来描述信号频率的分布，并通过图像的 RGB 值来生动表征信号强度[19]。考虑到背景噪声的影响，接收端在时刻 t 接收到的 PSD 函数可以表示为：

$$S_t(f)=g_uU(f-f_t)+\sum_{j=1}^{J}g_jJ_t^j(f-f_t^j)+N(f)\tag{1}$$

式中，$U(f)$、$J_t^j(f)$、$N(f)$ 分别表示用户、干扰机和高斯白噪声的 PSD 函数。用户在智能体的指

引下，选择中心频率为 $f_t \in [f_L, f_H]$ 的频段进行信号传输，其中 f_L 和 f_H 分别表示所选频率的上界和下界。传输功率可以看作是对用户信号在整个带宽范围上的积分，即 $\int_{-\frac{b_u}{2}}^{\frac{b_u}{2}} U(f)\mathrm{d}f$。系统中的干扰机可以任意频率实施干扰攻击，用 f_t^j 来表示所选择的频率，并用 g_u 表示从发射机到接收机的信道功率增益，用 g_j 表示从干扰机到接收机的信道功率增益。

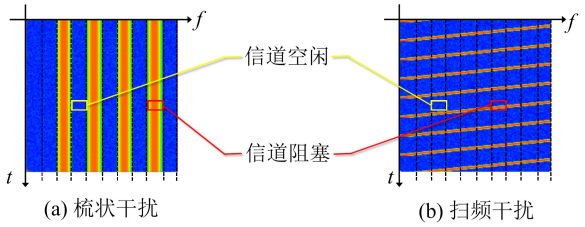


图 2 不同信道选择下的通信状态示意图

Fig. 2 A schematic diagram of the communication states under the different channel selections

使用信干噪比（signal-to-interference-plus-noise ratio，SINR）来评估通信信道的质量。因此，从接收端接收到的用户 SINR 值可以表示为：

$$\beta(f_t)=\frac{g_u p_t}{\int_{f_t-\frac{b_u}{2}}^{f_t+\frac{b_u}{2}}\left[\sum_{j=1}^{J} g_j J_t^j(f-f_t^j)+N(f)\right]\mathrm{d}f} \tag{2}$$

式中，p_t 为用户 t 时刻下接收到的信号功率。更进一步地，将在接收端持续感知到的整个通信频带的离散频谱采样向量表示为 $\boldsymbol{P}_t=(p_{t,1},p_{t,2},\cdots,p_{t,i},\cdots,p_{t,n})$，其中 t 时刻下对第 $i(i\in\{0,1,\cdots,n-1\})$ 个频段感知到的频谱能量分量表示为 $p_{t,i}=10\lg\left[\int_{f_b+i\Delta f}^{f_b+(i+1)\Delta f} S_t(f)\mathrm{d}f\right]$，其中 f_b 表示频谱感知的起始频率，样本数量 n 取决于用户的通信传输带宽 b_u 和频谱分析的分辨率 Δf。

1.2 问题建模

考虑到强化学习中用户外界环境一般受到当前状态和采取动作的影响，将这样一个动态频谱接入的序贯决策问题建模为一个确定性的马尔可夫决策过程（Markov decision process，MDP）。使用五元组 $\langle S,A,P,R,\gamma\rangle$ 来描述 MDP，其中 S 表示频谱环境状态集，A 表示可选的动作集，P 表示状态转移概率，R 表示奖励函数，γ 表示用于计算累积奖励的折扣因子。由于 MDP 的性质，使用一种基于网络（network-based）[20] 的迁移学习方法，即“预训练和微调”，将深度迁移学习应用到这样一个连续迭代的抽象过程中。

将元组中各分量分别定义如下：

1）状态空间。对连续采样到的状态矩阵 $\boldsymbol{S}_t$（即频谱瀑布）可以看作是由每个离散感知向量 $\boldsymbol{P}_t$ 叠加的时频特征热力学图，其中包含了背景频谱的历史信息：

$$\begin{aligned}\boldsymbol{S}_t&=(\boldsymbol{P}_t,\boldsymbol{P}_{t-1},\cdots,\boldsymbol{P}_{t-m+1})^{\mathrm{T}}\\&=\begin{pmatrix}p_{t,1}&p_{t,2}&\cdots&p_{t,n}\\p_{t-1,1}&p_{t-1,2}&\cdots&p_{t-1,n}\\\vdots&\vdots&\ddots&\vdots\\p_{t-m+1,1}&p_{t-m+1,2}&\cdots&p_{t-m+1,n}\end{pmatrix}\end{aligned} \tag{3}$$

表示为一个 $m\times n$ 的二维矩阵，其中 m 表示回溯的历史状态数目。

2）动作空间。可用于通信的频段被划分为 N 个信道，智能体根据感知到的环境状态采取相应的动作 a_t。于是，智能体选择转换通道的动作空间定义为：

$$A\triangleq\{a_t:a_t\in\{d_0,d_1,\cdots,d_{N-1}\}\} \tag{4}$$

3）瞬时奖励函数。用户在智能体的引导下，采取动作 $a_t\in A_u$ 选择信道接入实施抗干扰决策，并获得即时奖励，用来评估通信效果，同时用来验证知识迁移的有效性，定义为：

$$R(\boldsymbol{S}_t,a_t)=F_t(\cdot) \tag{5}$$

式中，$F_t(\cdot)$ 为判断迁移效果好坏的评价函数，即根据效果评估是正迁移（positive transfer）还是负迁移（negative transfer）[21]。

考虑到在频域维度上，射频（radio frequency）设备需要额外的启动时间来重建传输链路，由于在不同频率上可能具有不同的传输特性，这导致在数字信号处理时可能存在差异[22]。因此，本文将用户在相邻时隙之间的信道切换成本视为一定程度上的性能损失，改进的吞吐量和频道切换成本的奖励函数定义为：

$$R(\boldsymbol{S}_t,a_t)=\varphi(a_t)-\mu(a_t) \tag{6}$$

式中，

$$\varphi(a_t)=\delta(\beta(f_t)\geqslant\beta^*)=\begin{cases}1, & \beta(f_t)\geqslant\beta^*\\0, & \beta(f_t)<\beta^*\end{cases} \tag{7}$$

代表 t 时刻的归一化评价指示函数，用来评估信号传输的成功与否，即在任意时隙内，只有当通

信 SINR 值超过门限 SINR 阈值 β^* 时，才会被认为传输成功，其中 $\delta(\cdot)$ 代表指示函数。我们规定，如果一个抗干扰策略实施成功，那么 $\varphi(a_t)=1$；相反，如果用户信号被干扰信号所覆盖，那么 $\varphi(a_t)=0$。令

$$\mu(a_t)=\begin{cases}0, & a_t=a_{t-1}\\ \lambda\cdot|a_t-a_{t-1}|, & a_t\neq a_{t-1}\end{cases} \tag{8}$$

表示测量相邻时隙间跳频切换代价的信道切换公式，其中 λ 为信道切换代价因子。切换代价值和当前动作 a_t 与前一个动作 a_{t-1} 选择的信道之间跨频幅度成正比，这样设置的目的是避免信道重新选择路由带来的不必要的损失，减少重建时间[23]。

此外，为了最小化直接影响学习算法收敛速度的探索-利用窘境(exploration and exploitation dilemma)，在训练期间采用了 ε-贪婪衰减策略(ε-greedy decay strategy)，其中 ε 值设定为随机选择动作进行探索的概率。

上述场景下的抗干扰问题，其最优的信道选择接入策略是通过智能体和环境之间的不断交互得到的，优化目标是选择抗干扰决策的方式，在考虑长期奖励的同时最大化目标函数，即

$$\max_{a_t\in A_u}\sum_{t=0}^{\infty}\gamma^t R(\boldsymbol{S}_t,a_t) \tag{9}$$

式中，$\gamma\in(0,1)$ 表示平衡累积奖励重要性的折扣因子。

2　动态频谱快速适配抗干扰方法

深度强化学习算法对于感知到的外界干扰的规律性要求较高，因此会带来算法在前期探索阶段收敛较慢的问题。每当干扰模式进行了切换，则需智能体对通信环境重新进行学习，造成现有算法在实际应用中的局限性大大增加，对抗智能干扰的效果变差(如若处于干扰模式快速切换的动态环境下，深度强化学习的方法很难甚至不能收敛)。

2.1　方法阐释

针对上述使用传统 DQN 方法中存在的适应环境突变时学习收敛速度慢的问题，本文提出基于深度迁移学习的动态频谱快速适配抗干扰方法，其整体流程框图如图 3 所示。在这样一个“预训练离线学习＋强化学习在线决策”的双重网络架构之上，所提方法的目标是：基于前期训练阶段采集的数据，构建一个通用的预训练模型，减少大量数据再训练的时间损耗和计算算力的要求，快速适配抗干扰策略的生成实施。其核心是根据实时感知的频谱状态热力图，自动识别并调取图谱相近、任务相似的经验数据，根据通信场景调整信道接入策略，以做到对不同的任务都能采取最优策略，达成一致性的成功抗干扰目标。

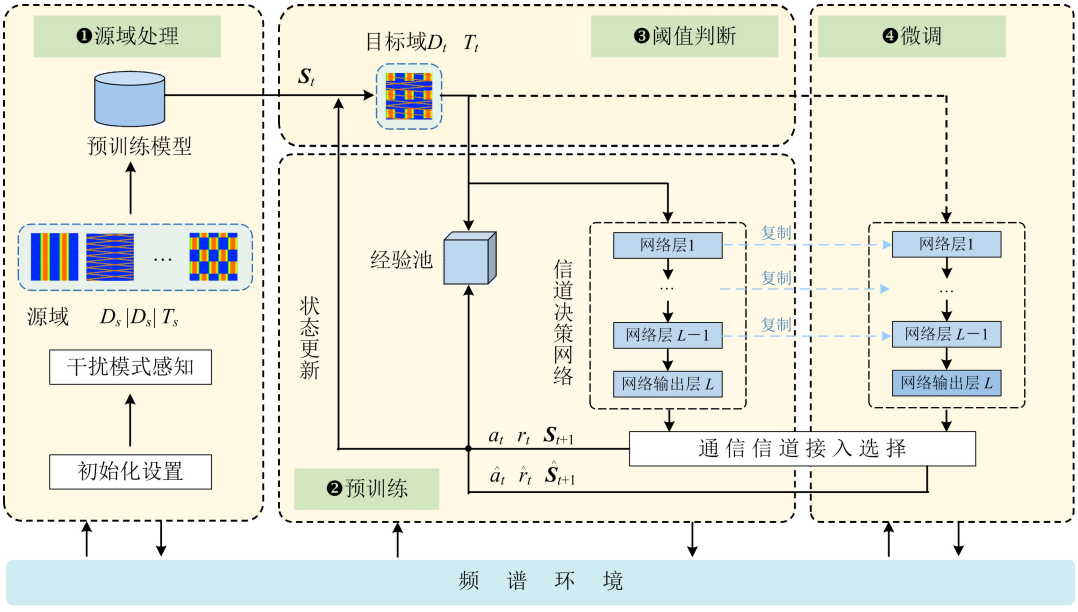


图 3　动态频谱快速适配抗干扰流程框图

Fig. 3　Flow block diagram of rapid re-adaption to dynamic spectrum anti-jamming

如图 4 所示，“预训练和微调”的方法框架作为迁移学习的一个应用实例，其目的是构建一个

通用模型，用于解决通信抗干扰决策场景中的快速模型适配任务。简单地说，通过“预训练和微调”的辅助框架，最终的目标是帮助智能体更好地感知频谱环境，加速收敛过程。“预训练”和“微调”2 个模块是所提方法性能提升的关键，将它们拆分开来，分别做如下讨论。

1）预先训练。预训练阶段，输入大量的数据样本在源域网络进行训练，对大量的已知干扰模式（包括“单一模式”“动态模式”“复合模式”3 类，其足以涵盖大部分可能出现的真实干扰类别）进行有监督的闭式学习。智能体将预训练过的经验数据存储在经验池中，并存储为模型保存下来，且每学习到一种干扰模式，模型便会保存一组权重参数到模型中。在强化学习的背景下，将每种干扰模式视为一个子域，并进一步把迁移学习源域网络和目标域网络中的领域划分为 2 种形式：① 来自于源域网络的带有先验知识的已标定的真实样本数据集，记为 D_s；② 来自目标域网络的未标定的干扰模式数据集，记为 D_t。需要注意的是，基于以上假定的前提是源域网络的数据集的数据量远远大于目标域网络的数据集的数据量，即 $|D_s| \gg |D_t|$[20]。

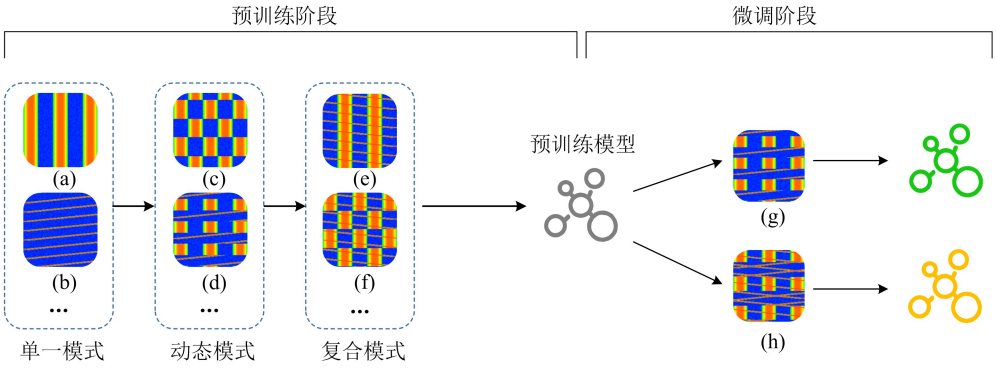


图 4 所提“预训练和微调”方法流程框架

Fig. 4 Process framework of the proposed “pre-train and fine-tune” approach

2）网络迁移。本文设定了一个阈值判定机制——应用通信场景中，每当智能体通过阈值判断感知到频谱环境状态 $\boldsymbol{S}_t$ 的迅速变化时，即强化学习算法反馈的奖励回报在设定好的时间阈值内下降到一定程度后，算法便会启用深度迁移网络，根据状态的相似程度对适配当前频谱环境下的参数配置方案进行调取，迁移模型参数，采取相应的信道接入抗干扰反制策略，将经验策略从源域 D_s（以 T_s 作为源任务，即预先学习干扰模式）迁移到目标域 D_t（以 T_t 作为目标任务，即适应未知的干扰模式）中。具体来说，将通过离线预训练的方式学习到的干扰模式可以描述为一组源域集；知识经验通过网络迁移到学习一种新干扰模式（已标定或未标定数据）的目标域上，记为 D_t，且满足 $|D_s| \gg |D_t|$ 的前提假设。当智能体感知到一个预先训练过的干扰模式（任务 T_s 和 T_t 在同一特征域内）时，就不需要在 2 个网络之间进行迁移（即重用预先训练好的抗干扰策略），模型便大致退化为经典的 DRL 算法，此时 $|D_s|=1$ 且 $D_s=D_t$。

3）微调策略。神经网络的低层通常提取一些通用特征，高层则提取对任务有强相关性的特征[24]。本文选择在合适的位置断开共享和微调深度神经网络的部分层：共享（冻结，在训练时不更新梯度）前几层通用特征（general features）层，对最后若干特定特征（specific features）层进行参数初始化[25]。具体地说，将深度神经网络中 L 个隐藏层中的前 l 层的权重向量进行迁移，即重用部分网络的权重向量 $(\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_l)$，其中 $l \in \{1,2,\cdots,L\}$。微调部分，便是对剩余层网络参数进行初始化，并更新迁移后的子网络。智能体每感知一次环境状态，即对每个任务微调一次。

一般来说，智能体在状态 $\boldsymbol{S}_t$ 下采取动作 a_t，转移到下一状态 $\boldsymbol{S}_{t+1}$，获得即时奖励 $R(\boldsymbol{S}_t,a_t)$。此外，对 MDP 元组的表达形式加以区分：在预训练阶段，使用传统 DQN 网络的源学习系统从头开始进行训练，收集到足够的交互经验并以 $(\boldsymbol{S}_t,a_t,R(\boldsymbol{S}_t,a_t),\boldsymbol{S}_{t+1})$ 的形式存储到经验池的数据集中；在知识转移阶段，则对应于 $(\hat{\boldsymbol{S}}_t,\hat{a}_t,\hat{R}(\boldsymbol{S}_t,a_t),\hat{\boldsymbol{S}}_{t+1})$ 的形式存储经验。

2.2 信道决策网络

强化学习中，智能体在特定状态下尝试行

动，从当前状态转移至下一状态，并收到环境对于采取当前动作的反馈。智能体根据其收到的即时奖励或处罚以及对其所处状态的值来评估其后果并更新 Q 值表。通过反复尝试所有状态的所有行动，它的目标是通过长期折扣奖励来判断总体上最好的行为，而不是即时奖励最大的策略。Q 值表的更新过程可以表达为：

$$Q_{\text{new}}(\boldsymbol{S}_t,a_t)\leftarrow(1-\alpha)\cdot Q_{\text{old}}(\boldsymbol{S}_t,a_t)+\alpha\cdot\left(R(\boldsymbol{S}_t,a_t)+\gamma\max_{a_t\in A_u}Q(\boldsymbol{S}_{t+1},a_t)\right)\tag{10}$$

式中，α 为学习速率。

算法在初期的探索阶段，通过对频谱状态的连续实时感知，获取并存储大量的原始信息数据。DQN 是一种将神经网络和 Q-learning 结合的方法[26]，其将从环境中感知到的原始状态作为神经网络的输入，通过网络计算出所有的动作价值，并执行其中最大值对应的动作作为下一时刻的动作。DQN 的学习样本序列由采取动作和观察值组成，由于样本序列之间的数据不尽相同，且外界环境一般受到序列样本的影响，因此一般将此类问题建模为马尔可夫决策过程，也就方便了使用强化学习来解决问题。

本文构建了如图 5 所示的深度神经网络结构。采用经典的频谱瀑布图来刻画频谱状态，并作为神经网络的输入。前 2 层隐含层分别由一组卷积层和池化层组成，最后 2 层是全连接层，网络各层参数配置见表 1 所列。神经网络直接输出为通信信道接入选择策略，强化学习根据从外界获取的状态信息对选择动作进行反馈并更新当前干扰模式下的 Q 值表。

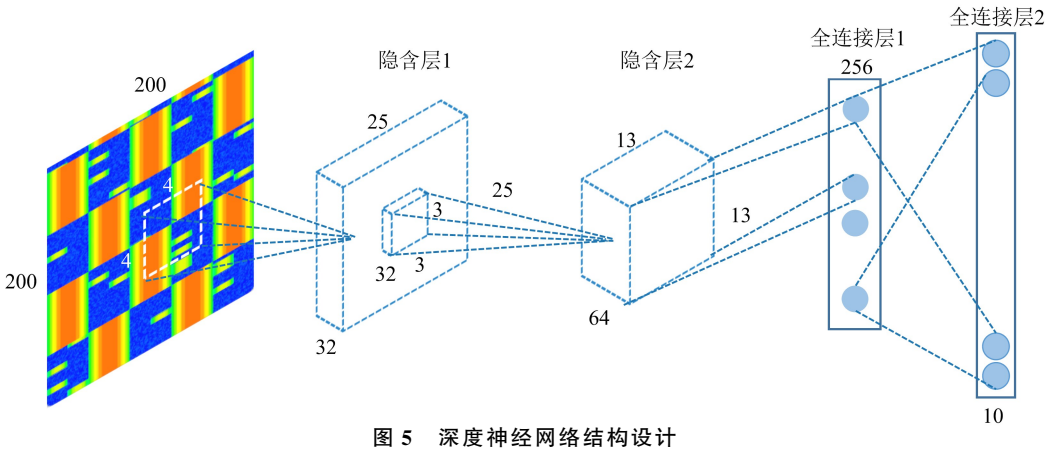


图 5　深度神经网络结构设计

Fig. 5　Design of network structure of DNN

网络训练时，通过最小化在每次迭代 i 处改变的损失函数来计算梯度并更新权值：

$$L_i(\boldsymbol{\theta}_i)=E_{\boldsymbol{S}_t,a_t\sim\rho(\cdot)}\left[(y_i-Q(\boldsymbol{S}_t,a_t;\boldsymbol{\theta}_i))^2\right]\tag{11}$$

式中，$y_i=E_{S_t\sim\varepsilon}\left[R(\boldsymbol{S}_t,a_t)+\gamma\max_{a_{t+1}\in A_u}Q(\boldsymbol{S}_{t+1},a_{t+1};\boldsymbol{\theta}_{i+1})\,|\,\boldsymbol{S}_t,a_t\right]$ 是迭代 i 的目标，$\boldsymbol{\theta}_i$ 表示深度学习卷积神经网络在 i 次迭代的参数。在优化损失函数 $L_i(\boldsymbol{\theta}_i)$ 时，暂时固定前一次迭代的参数 $\boldsymbol{\theta}_{i-1}$，从而得到损失函数的梯度计算公式如下：

$$\nabla_{\boldsymbol{\theta}_i}L_i(\boldsymbol{\theta}_i)=E_{\boldsymbol{S}_t,a_t\sim\rho(\cdot);\boldsymbol{S}'\sim\varepsilon}\left[\left(R(\boldsymbol{S}_t,a_t)+\gamma\max_{a_{t+1}\in A_u}Q(\boldsymbol{S}',a';\boldsymbol{\theta}_{i-1})-Q(\boldsymbol{S}_t,a_t;\boldsymbol{\theta}_i)\right)\nabla_{\boldsymbol{\theta}_i}Q(\boldsymbol{S}_t,a_t;\boldsymbol{\theta}_i)\right]\tag{12}$$

式中，$\nabla_{\boldsymbol{\theta}_i}$ 表示求梯度运算。

表 1　深度神经网络各层参数配置

Tab. 1　Parameter configuration for each layer of DNN

网络层	卷积核参数	神经元数量	激活函数
隐含层 1	尺寸：4×4×1 步长：4	32	ReLU
隐含层 2	尺寸：3×3×32 步长：4	64	ReLU
全连接层 1	—	256	ReLU
全连接层 2	—	10	ReLU

基于预训练和微调的深度强化学习信道接入抗干扰优化算法如算法 1 所示。

算法 1　深度强化学习信道接入抗干扰优化算法

步骤 1　初始化迭代次数 $i=0$，最大迭代次数 I，随机

初始化网络参数 $\boldsymbol{\theta}_0$，用户观察当前环境频谱信息，将观测值 $\boldsymbol{S}_t=O(T\times N)$ 作为初始状态输入拟合好的神经网络中，加载预训练模型，开始训练；

步骤 2 循环；

步骤 3 根据 ε-贪婪策略执行通信信道选择操作 a_t；

步骤 4 计算回报值 $R(\boldsymbol{S}_t,a_t)$，并转移至下一时刻状态 $\boldsymbol{S}_{t+1}$；

步骤 5 更新 Q 值表，计算梯度 $\nabla_{\boldsymbol{\theta}_i}L_i(\boldsymbol{\theta}_i)$，更新权值 $\boldsymbol{\theta}_i$，$i=i+1$；

步骤 6 判定干扰模式切换，共享 L 层神经网络中的前 l 层的权重向量 $(\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\cdots,\boldsymbol{\omega}_l)$（冻结，不更新梯度），重复步骤 3～5，直到用户对恢复通信并保持相对稳定；

步骤 7 直到 $i>I$；

步骤 8 输出用户信道选择接入策略，保存迭代模型。

3 仿真实验与结果分析

3.1 实验设置

参照文献[19]的数值设置仿真实验参数，采用大小为 200×200 的二维矩阵图谱作为状态 $\boldsymbol{S}_t$ 的输入卷积神经网络：时间维度上，频率持续感知的更新窗口范围为 200 ms，用户在每个时隙帧内进行一轮信息传输、频谱感知、学习反馈、策略生成并实施决策，时隙帧长 5 ms，用户每帧可以选择进行信道切换（或不切换）1 次；频率维度上，通信场景中用户和干扰机可用带宽 20 MHz，频谱感知的频率分辨率为 100 kHz，用户信号带宽 2 MHz，步进 2 MHz，可供选择的信道数量为 10 个。用户和干扰的信号波形均为升余弦波，滚降系数 $\eta=0.5$。干扰功率为 30 dBm，用户的信号功率为 0 dBm，解调门限 SINR 阈值 $\beta^*=10$ dB，信道切换代价因子 $\lambda=0.2$。仿真实验参数设置见表 2 所列。

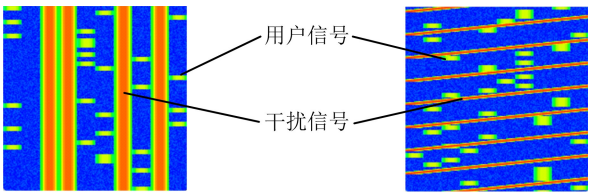
设置深度强化学习算法的学习速率为 $\alpha=0.1$，折扣因子 $\gamma=0.5$，初始贪婪度 $\varepsilon_{\text{start}}=1$，最终贪婪度 $\varepsilon_{\text{end}}=0.05$。本文系统仿真采用 Python 语言，基于 TensorFlow 深度学习框架。实验环境为 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80 GHz 型号 GPU，NVIDIA GeForce MX450 显卡。

“梳状”和“扫频”这 2 种“单一”的干扰模式，通过演变、交错、叠加等方式演变成为“动态”“复合”模式，是本文预训练学习的干扰模式的基础。其频谱瀑布图例如图 6 所示，参数设置如下：梳状干扰的梳状条数 $M=\{1\leqslant M\leqslant 10, M\in\mathbf{Z}\}$，干扰机可自主选择一条或同时选择多条梳状体实施干扰，10 个梳状体的带宽均为 2 MHz，梳状体之间无重叠，步进 2 MHz；扫频干扰可以分为左行单扫频、右行单扫频、双扫频干扰，扫频速率可以设置为 1 GHz/s 或 500 MHz/s。

表 2 仿真实验参数设置

Tab. 2 Parameters setting of simulation experiment

参数名称	参数值
窗口感知时间/ms	200
时隙帧长/ms	5
可用频率/MHz	0～20
可用信道数量	10
恶意干扰功率/dBm	30
用户信号功率/dBm	0
信干噪比解调阈值/dB	10



(a) 梳状干扰 (b) 扫频干扰

图 6 梳状干扰和扫频干扰的频谱瀑布图例

Fig. 6 Thermodynamic chart of comb and sweep jamming pattern

3.2 初步仿真论证

为了初步评估所提方法的优化效果，首先测试其在单一模式下的适配能力。选取 2 种“单一”的干扰模式：梳状干扰（选择 3 个梳状体，中心频率分别为 2、10、18 MHz）和扫频干扰（左行单扫频，扫频速率为 1 GHz/s）进行预训练，同时在场景中施加一种“动态”的干扰模式（梳状和扫频模式周期性交替干扰（如图 4(g)所示），并设置干扰切换时间为 30 ms），且通信过程中模式始终不发生变化。

使用表征通信传输成功与否的（归一化）奖励值函数曲线来衡量抗干扰效果，将其与基准算法进行对比。本小节中，强化学习算法在随机探索阶段的迭代步数为 100 步，学习训练阶段为 1 100 步，验证测试阶段为 300 步。

学习训练阶段中采用了 ε-贪婪策略，ε 值的选取决定了智能体在已知的全部（状态-动作）二

元组分布之外，选择其他未知动作的概率，即以一个正数 $\varepsilon\in(0,1)$ 的概率随机“试探”未知的一个动作策略，最大化长期收益；同时以 $(1-\varepsilon)$ 的概率“利用”已有经验中价值回报最大的动作 $a_{t+1}=\arg\max_{a_t\in A_u}Q(\boldsymbol{S}_t,a_t;\boldsymbol{\theta}_i)$。图 7 展示了采取不同 ε 值下的所提方法与原始算法之间的对比仿真结果，即分别取 $\varepsilon=0$（纯贪婪策略）以及 ε-贪婪衰减策略（ε 值从设定的初始值在整个训练阶段内逐步下降至最终值）。

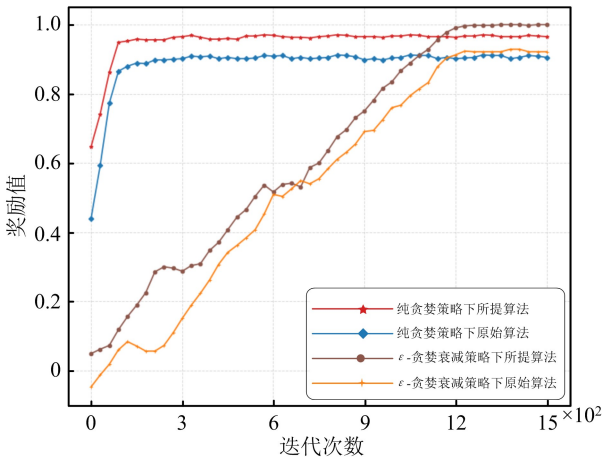


图 7　初步仿真结果

Fig. 7　Preliminary simulation results

由图 7 的奖励值曲线可知，随着迭代步数的增加，奖励值曲线均呈现出逐步提升的趋势，且无论是选择哪种贪婪策略，所提方法在绝大多数情况下取得的归一化奖励值均高于原始算法，在抗干扰的表现上更为优越。

采取纯贪婪策略时，2 种算法在测试阶段均未能收敛至最优解（归一化奖励值达到 1.0），其原因在于算法对未知环境的探索不够，导致未能及时探索并储备较为全面的经验策略；采取 ε-贪婪衰减策略之后，原始算法仍未能收敛至最优解，由此表明：在适应单一干扰模式场景时，原算法能够通过逐步学习，最终收敛于信道选择最优解，然而在相对更为复杂的干扰环境（更为有限的可供选择信道数量）下却不能找到最优策略，其中原因可能在于对策略选择的探索经验不够，导致最终的决策方式固化，并在测试阶段出现策略选择陷入局部最优的情况；相反，所提算法具备通过预先训练获得的经验数据，能够更为快速地对感知环境有针对性地进行场景适配，并最终收敛于更优解。

总之，通过使用“预训练和微调”的方法，在性能优化上带来显著的效果提升，由此证明了此方法的可行性，适用于动态频谱抗干扰通信的场景中。

3.3　仿真结果与分析

本小节中，主要分为 3 个部分将所提方法与原算法[19]（对比算法）进行仿真实验对比。首先，改变预训练阶段冻结和微调的网络层数，比对迁移效果；其次，在算法探索阶段，以相同的贪婪策略对环境进行感知学习，比对收敛速度；最后，在算法适应一种场景达到收敛之后对干扰模式进行变换，比对收敛速度和抗干扰效果。

通过尽可能多干扰模式的学习，期望构建一个能够适应多变频谱环境的普适化预训练模型，并且通过不断地感知、学习、存储新的干扰迭代模式，使得模型适应能力更稳健、更鲁棒。在模型构建上，预先对“单一”“动态”“复合”3 类 10 种干扰模式进行训练，具体做法是：将智能体设置在 10 种模式交替变换的通信场景中，每对一种模式进行探索学习就保存 1 组配置参数，每种模式迭代次数不少于 500 步且满足算法达到收敛的最少步数条件。仿真设置上，考虑了 2 种通信应用场景下的干扰模式：1）动态干扰。梳状（选择 3 个梳状体，中心频率分别为 2、10、18 MHz）和扫频模式（左行单扫频，扫频速率为 1 GHz/s）周期性交替干扰（如图 4(g)所示），并设置干扰切换时间为 30 ms；2）复合干扰。在上述动态干扰的基础上，叠加一个单一的扫频干扰（右行单扫频，扫频速率为 500 MHz/s）（如图 4(h)所示）。以上设置的 2 种复杂干扰模式，均能满足在任何时隙都存在可供选择的未被干扰的通信空闲信道。本小节中，强化学习算法在随机探索阶段的迭代步数为 100 步，学习训练阶段为 1 200 步，干扰模式在训练阶段 400 步后进行切换，验证测试阶段为 300 步。采取的 ε-贪婪衰减策略的贪婪度 ε 值在整个训练阶段内，从设定的初始值在 200 步内逐渐递减至最终值。

预训练模型的特点是通过大量数据的训练，逐步具备提取浅层基础特征和深层抽象特征的能力[14]。用于提取浅层通用特征和深层特定特征的网络之间的界限是模糊的，为了更好地区分它们，使迁移效用最大化，分别对比了冻结 1 层、2 层、3 层权重参数（即分别取 $l=1$、$l=2$、$l=3$，$L=4$）时的仿真结果，截取迭代 700～1 600 步的部分，如图 8 所示。

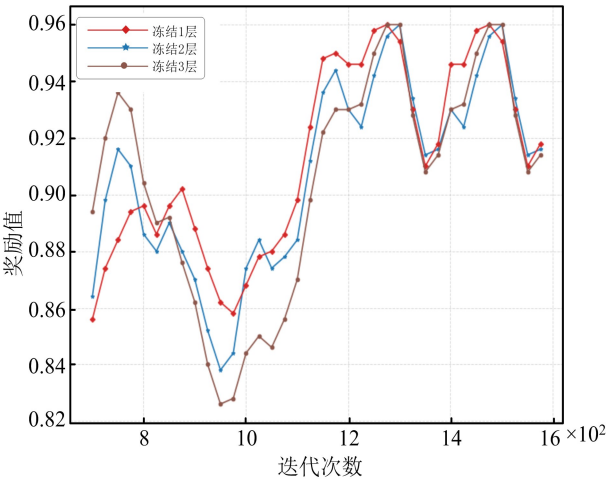


图 8 冻结不同卷积神经网络层数时的仿真结果

Fig. 8 Simulation results freezing different CNN layers

由图 8 可知，3 种微调方案的仿真结果在曲线走势上较为相似，相差不大。对于本文的建模条件来说，由于网络层数有限且受不确定性实验环境的影响等原因，致使从图中不易看出冻结不同层数的明显变化。依据选择获得的总奖励值（曲线下方与横轴围成的面积）最大方案的原则，综合选择“冻结 2 层神经网络的权重参数、对剩余 2 层进行初始化并更新梯度”作为所提算法适用于本文通信抗干扰场景中微调部分的操作方法。

上文提到，本文所提方法的性能优化提升来源于“预训练”和“微调”2 个部分。图 9 给出了对比所提“预训练和微调”方法“只预训练不微调”“只微调未预训练”以及原算法 4 种条件下的仿真结果。从中可以看出，启动速度方面：结合迁移学习方法的 3 组曲线，在仿真测试的初期阶段，都能够相较于原始方法更早更快启动，加快智能体对新场景环境的学习收敛速度。重新适配方面：在感知的干扰模式发生变化之后，原算法和“只微调未预训练”方法的奖励值曲线均出现较大范围的下跌趋势，且回升速度较为缓慢，通信中断时间较长；相反，结合“预训练和微调”方法的曲线趋势相较更为平缓且能够耗费更短的时间重新达到收敛状态。由此表明：所提算法能够基于模型历史经验数据对环境进行重新学习并迅速收敛，受到复杂频谱环境的影响冲击明显小于原算法，适应性更强，且测试阶段更能找到并选择更优策略。

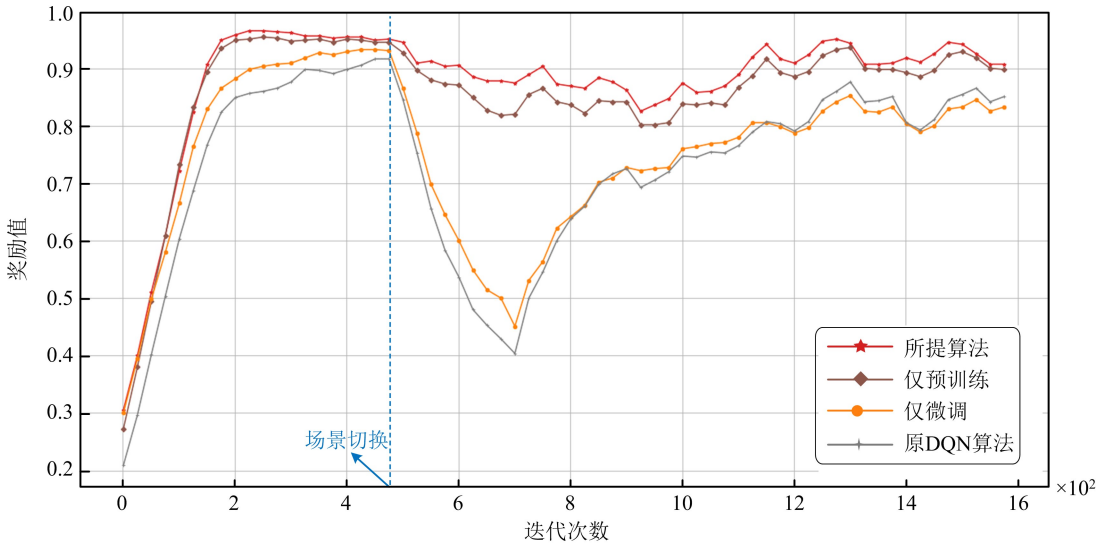


图 9 4 种实验条件下的对比仿真结果

Fig. 9 Comparative simulation results under four experimental conditions

3.4 评价指标及结果分析

为了进一步描述模型的性能表现，量化评估性能提升效果的程度，根据文献[27]总结提出的方法标准，结合本文通信场景实际，设置衡量迁移效果的多重评价指标，侧重于比较适用于应用场景的快速反应能力以及场景发生切换之后迅速恢复原有性能水平的稳健适应能力。如图 10 所示，本文主要从优先启动、性能提升和快速适配 3 个方面进行评估，评价指标包括平均优先启动比率（mean jump initiation ratio，MJIR）、平均性能提升比率（mean performance improvement ratio，MPIR）、建立通信缩短步数比率（iterations shortened ratio for establishing communication，ISREC）和恢复通信缩短步数比率（iterations shortened ratio for recovering communication，ISRRC）。

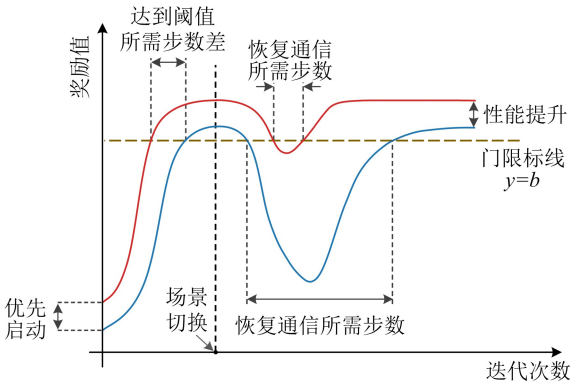


图 10　评估迁移效果的评价指标

Fig. 10　Evaluation indices estimating the effectiveness of transferring

1）优先启动。由于迁移学习在适应目标域任务时利用了源域的知识经验，因此一个良好的迁移过程在训练开始的时候就应该具有比从头学习的模型更好的表现。平均优先启动比率 MJIR 指用户开机时所提方法率先于原始算法获得的平均初始奖励值提升的比例，即

$$r_{\mathrm{MJI}}=\frac{E\left|R_{\mathrm{os}}-R_{\mathrm{is}}\right|}{E\left(R_{\mathrm{os}}\right)} \tag{13}$$

式中，R_{os} 是原始算法取得的初始奖励值，R_{is} 是所提方法取得的初始奖励值。

2）性能提升。对于复杂的任务场景来说，智能体往往无法找到最优决策策略，进而收敛至一个次优解。通过使用迁移学习的方法，能够帮助找到尽可能趋近于最优解的决策方案；同时，在整个训练过程中，迁移学习获得的总奖励理论上会高于原始方法。本文通过平均性能提升比率 MPIR 表示测试阶段所提方法最终收敛的平均奖励值相较于原始算法提升的比例，即

$$r_{\mathrm{MPI}}=\frac{E\left|R_{\mathrm{of}}-R_{\mathrm{if}}\right|}{E\left(R_{\mathrm{of}}\right)} \tag{14}$$

式中，R_{of} 是原始算法最终收敛的奖励值，R_{if} 是所提方法最终收敛的奖励值。

3）快速适配。为了更为直观地对比性能提升效果，通过人为设置能够满足达到必要通信质量门槛的阈值，使用一条平行与横轴的门限值基准线 $y=b$，其中 b 表示截距，即设定的门限奖励值，取决于通信双方能够保持全时不间断通信的实际需求。本文通过所提方法相较于原始算法迭代步数平均缩减的比例来评估快速适配效果，表示为：

$$r_{\mathrm{IS}}=\frac{E\left|C\left(R_{\mathrm{ori}}\leqslant b\right)-C\left(R_{\mathrm{imp}}\leqslant b\right)\right|}{E\left[C\left(R_{\mathrm{ori}}\leqslant b\right)\right]} \tag{15}$$

式中，R_{ori} 表示原始算法取得的奖励值，R_{imp} 表示所提方法取得的奖励值，$C(\cdot)$ 表示迭代次数的计数函数。

具体评估以下 2 项指标：

1）通信系统处于初始场景环境中，结合迁移学习的方法，在训练开始阶段达到特定性能阈值水平所需迭代步数（时间）相较于原算法压缩的比例 r_{ISEC}；

2）频谱环境发生变化（干扰机切换至新的干扰模式）后，结合迁移学习的方法，通信质量恢复至特定性能阈值水平所需迭代步数（时间）相较于原算法压缩的比例 r_{ISRC}。

表 3 给出了设定的门限奖励值 b 分别取 0.80、0.85、0.90 时，所提方法相较原始算法在各性能指标上的提升贡献程度。可以看出：在通信链路建立初期，结合所提方法能够获得接近 50% 的开机起点的提升，但最终测试集上的性能提升并不明显，提升率不足 10%；训练阶段，随着设置的归一化奖励阈值 b 的提升，r_{ISEC} 逐步增大，结合所提方法带来的达到正常通信条件所需迭代次数的压缩更为明显；干扰模式突变后，结合所提方法的表现突出，在恢复正常通信条件所需迭代次数的压缩上提升明显，r_{ISRC} 超过了 90%，b 取 0.82 以下时，所提方法取得的奖励值始终未降至阈值以下，不受通信干扰的影响，b 取 0.88 以上时，原始算法取得的奖励值曲线甚至不能收敛至门限标线以上，无法恢复通信。

表 3　性能评价指标对比

Tab. 3　Comparison of performance evaluation indices

截距 b 取值（$0<b<1$）	r_{MJI}（%）	r_{MPI}（%）	r_{ISEC}（%）	r_{ISRC}（%）
0.80	42.86	8.07	26.83	（所提算法未受干扰影响）
0.85	42.86	8.07	33.95	92.31
0.90	42.86	8.07	55.88	（原有算法无法恢复通信）

综上所述，结合“预训练和微调”的方法，能够辅助传统 DRL 算法在优先启动、性能提升、快速适配等方面带来不同程度的性能提升。并且，对保持通信的质量要求越高、通信场景越复杂，越能体现出所提方法的优越性。需要注意的是，所提方法的优势主要体现在能够快速对通信场

景做出反应并对环境变化做出适应，相较于传统算法有更快的学习收敛速度，但在抗干扰性能上的提升并不明显。

4 结束语

本文研究了恶意干扰环境动态未知情况下的频谱接入问题。考虑到目前智能算法在面对复杂干扰环境时很难甚至不能收敛至最优解的现实实际，提出了一种基于深度迁移学习的快速适配抗干扰方法。该方法旨在构建一个通用适用性模型，通过对已知的干扰模式进行预先充分的学习训练，帮助强化学习智能体利用过往经验快速适应外界干扰环境；根据现实通信场景中提取的频谱特征采取微调策略，探索并学习最佳的通信信道选择策略，实现通信抗干扰。仿真结果验证了所提出的抗干扰通信方法的有效性以及实用性，并通过计算优化比率量化评估所提方法带来的性能提升程度，充分说明该算法适应通信场景的收敛速度以及性能效用优于传统的强化学习抗干扰算法，为“迁移学习”的思想应用于动态频谱接入抗干扰的研究课题提供了现实依据。

参 考 文 献

[1] PIRAYESH H, ZENG H C. Jamming attacks and anti-jamming strategies in wireless networks: a comprehensive survey[J]. Communications Surveys & Tutorials, 2022,24(2): 767-809.

[2] JIA L L, XU Y H, SUN Y M, et al. A multi-domain anti-jamming defense scheme in heterogeneous wireless networks[J]. IEEE Access, 2018, 6: 40177-40188.

[3] WANG X M, WANG J L, XU Y H, et al. Dynamic spectrum anti-jamming communications: challenges and opportunities[J]. IEEE Communications Magazine, 2020, 58(2): 79-85.

[4] ABUZAINAB N, ISLER V, YENER A, et al. QoS and jamming-aware wireless networking using deep reinforcement learning[C]//Proceedings of 2019 IEEE Military Communications Conference. [S. l.]: IEEE, 2019: 610-615.

[5] XU Y F, XU Y H, DONG X, et al. Convert harm into benefit: a coordination-learning based dynamic spectrum anti-jamming approach[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 13018-13032.

[6] XU Y F, XU Y H, REN G C, et al. Play it by ear: context-aware distributed coordinated anti-jamming channel access[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 5279-5293.

[7] LI Y Y, WANG X M, LIU D X, et al. On the performance of deep reinforcement learning-based anti-jamming method confronting intelligent jammer[J]. Applied Sciences, 2019, 9(7): 1361.

[8] TAN X, ZHOU L, WANG H J, et al. Cooperative multi-agent reinforcement-learning-based distributed dynamic spectrum access in cognitive radio networks [J]. IEEE Internet of Things Journal, 2022, 9(19): 19477-19488.

[9] ZHONG C, WANG F, GURSOY M C, et al. Adversarial jamming attacks on deep reinforcement learning based dynamic multichannel access[C]//Proceedings of 2020 IEEE Wireless Communications and Networking Conference. [S. l.]: IEEE, 2020: 1-6.

[10] HAN H, XU Y F, JIN Z, et al. Primary-user-friendly dynamic spectrum anti-jamming access: a GAN-enhanced deep reinforcement learning approach [J]. IEEE Wireless Communications Letters, 2021, 11 (2): 258-262.

[11] LIU S Y, XU Y F, CHEN X Q, et al. Pattern-aware intelligent anti-jamming communication: a sequential deep reinforcement learning approach[J]. IEEE Access, 2019, 7: 169204-169216.

[12] TONG X H, PAN H Y, LIU S C, et al. A novel approach for hyperspectral change detection based on uncertain area analysis and improved transfer learning [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 2056-2069.

[13] CYRIAC M, SHEEJA M K. Design of an optical transfer function classifier based on machine learning and deep learning for optical scanning holography [C]//Proceedings of 2021 Photonics North Conference. [S. l.]: IEEE, 2021: 1.

[14] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [J]. Advances in Neural Information Processing Systems, 2014, 27:3320-3328.

[15] WANG G F, FANG Z, LI P, et al. Transferring knowledge from human-demonstration trajectories to reinforcement learning[J]. Transactions of the Institute of Measurement and Control, 2018, 40 (1): 94-101.

[16] PHAN T V, SULTANA S, NGUYEN T G, et al. Q-TRANSFER: a novel framework for efficient deep

transfer learning in networking[C]//Proceedings of 2020 International Conference on Artificial Intelligence in Information and Communication. [S. l.]: IEEE, 2020: 146-151.

[17] LIN F, CHEN J, SUN J C, et al. Cross-band spectrum prediction based on deep transfer learning[J]. China Communications, 2020, 17(2): 66-80.

[18] RAWAT D B. Deep transfer learning for physical layer security in wireless communication systems[C]//Proceedings of the 3rd IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications. [S. l.]: IEEE, 2021: 289-296.

[19] LIU X, XU Y H, JIA L L, et al. Anti-jamming communications using spectrum waterfall: a deep reinforcement learning approach[J]. IEEE Communications Letters, 2018, 22(5): 998-1001.

[20] TAN B, ZHANG Y, PAN S J, et al. Distant domain transfer learning[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. [S. l.: s. n.], 2017: 2604-2610.

[21] NIU S T, LIU Y X, WANG J, et al. A decade survey of transfer learning(2010-2020)[J]. IEEE Transactions on Artificial Intelligence, 2020, 1(2): 151-166.

[22] YAO F Q, JIA L L. A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks[J]. IEEE Wireless Communications Letters, 2019, 8(4): 1024-1027.

[23] JIA L L, XU Y H, SUN Y M, et al. Stackelberg game approaches for anti-jamming defence in wireless networks[J]. IEEE Wireless Communications, 2018, 25(6): 120-128.

[24] NEYSHABUR B, SEDGHI H, ZHANG C. What is being transferred in transfer learning? [C]//Proceedings of 2020 Conference on Advances in Neural Information Processing Systems. [S. l.: s. n.], 2020: 512-523.

[25] TAJBAKHSH N, SHIN J Y, GURUDU S R, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? [J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1299-1312.

[26] LYU L, SHEN Y, ZHANG S C. The advance of reinforcement learning and deep reinforcement learning [C]//Proceedings of 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms. [S. l.]: IEEE, 2022: 644-648.

[27] TAYLOR M E, STONE P, LIU Y X. Transfer learning via inter-task mappings for temporal difference learning[J]. Journal of Machine Learning Research, 2007, 8(9): 2125-2167.

作 者 简 介

李思达

男，1996 年生，硕士研究生，研究方向为动态频谱抗干扰

E-mail：960656891@qq.com

徐逸凡

男，1995 年生，博士，讲师，研究方向为动态频谱抗干扰、智能博弈决策

E-mail：yifanxu1995@163.com

刘　杰

男，1983 年生，讲师，研究方向为无线通信技术

E-mail：594563512@qq.com

林凡迪

男，1992 年生，博士研究生，研究方向为无线通信、认知无线电、深度学习

E-mail：jedilv1_st@163.com

韩　昊

男，1996 年生，博士研究生，研究方向为动态频谱对抗、智能通信抗干扰、博弈论、机器学习

E-mail：haohanself@foxmail.com

易剑波

男，1976 年生，工程师，研究方向为短波通信技术

E-mail：yiwangcom2009@sina.com

徐煜华

男，1983 年生，博士，教授，博士研究生导师，研究方向为认知无线电、智能通信抗干扰、博弈论

E-mail：xuyuhua365@163.com

责任编辑　安　蓓