

无线网络多用户干扰下智能发射功率控制算法

张先超^{1,2}, 赵耀³, 叶海军⁴, 樊锐⁵

(1. 东南大学移动通信国家重点实验室, 江苏 南京 210096;
2. 嘉兴学院浙江省医学电子与数字健康重点实验室, 浙江 嘉兴 314001; 3. 北京理工大学信息与电子学院, 北京 100081;
4. 北京邮电大学信息与通信工程学院, 北京 100876; 5. 中国电子科学研究院, 北京 100041)

摘 要: 针对无线网络多用户互相干扰的问题, 通过对发射功率进行智能控制, 实现干扰管理, 保证多用户通信服务质量。首先, 考虑复杂动态无线信道环境, 建立以无线通信系统加权数据速率最大化为目标的发射功率控制模型。其次, 设计以深度强化学习“行动器-评判器”为基本架构的智能发射功率控制算法, 缩短功率控制决策时间。仿真验证表明, 所提算法收敛速度快, 在 10 对收发机场景下, 计算时间缩短到传统最优算法的 $\frac{1}{4}$ 。

关键词: 无线网络; 多用户干扰管理; 发射功率控制; 深度强化学习; 功率控制时间

中图分类号: TN929.5

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2022028

Intelligent transmit power control algorithm for the multi-user interference of wireless network

ZHANG Xianchao^{1,2}, ZHAO Yao³, YE Haijun⁴, FAN Rui⁵

1. National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
2. Key Laboratory of Medical Electronics and Digital Health of Zhejiang Province, Jiaxing University, Jiaxing 314001, China
3. School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
4. School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
5. China Academic of Electronics and Information Technology, Beijing 100041, China

Abstract: To deal with the inter-user interference problem in wireless networks, an intelligent transmit power control scheme was proposed to manage the inter-user interference and guarantee multiple users' quality of service. Firstly, considering the complex dynamic wireless channel environment, a transmit power control model that aims to maximize the weighted sum-rate of the wireless communication system was established. Then, an intelligent power control algorithm based on the actor-critic framework in deep reinforcement learning was designed to shorten the power control decision time. Simulation results show that the proposed algorithm converges quickly, and when there are 10 pairs transceivers, the computation time consumed by the intelligent power control method is only a quarter of the time consumed by the traditional optimal algorithm.

Keywords: wireless network, interference management of multiple users, transmit power control, deep reinforcement learning, power control time

0 引言

近年来, 随着移动互联网和人工智能技术的快速发展, 智能手机、增强现实 (AR, augmented

reality)、虚拟现实 (VR, virtual reality) 等智能无线设备与远程医疗、工业 4.0、自动驾驶等智能应用进入爆发式增长阶段, 无线网络中出现了大量无线终端, 且这些无线终端相较于现在的智能手机而

收稿日期: 2021-08-16; 修回日期: 2021-12-17

基金项目: 国家自然科学基金资助项目 (No.61941104, No.U19B2015)

Foundation Item: The National Natural Science Foundation of China (No.61941104, No.U19B2015)

言,对通信性能的要求更加苛刻与多元^[1-2]。为了利用有限的频谱来满足未来的高性能要求,研究人员提出频谱共享^[3]、D2D (device-to-device) 技术^[4]及超密集网络^[5]等大幅提高频谱利用率的新技术,但这些技术在使用过程中不可避免地带来基站之间、基站与用户之间或多用户之间的互相干扰,而发射功率与干扰影响密切相关,功率低则相互干扰程度小,但自然通信质量差,功率高则会加剧相互干扰,因此,发射功率控制是降低相互干扰、保证多用户的通信服务质量与体验的有效途径^[5],也一直是通信研究的热点之一。

目前的发射功率控制算法研究方向主要有三类:1) 基于模型的优化算法,将发射功率控制问题转化为优化问题进行求解;2) 基于博弈论的方法,将发射功率控制问题转化为博弈问题进行求解;3) 基于机器学习的方法,将发射功率控制问题转化为机器学习相关的问题进行求解。下面分别对三类方法的研究现状进行介绍。

1) 基于模型的优化算法。文献[6-10]分别通过加权最小均方差(WMMSE, weighted minimum mean square error)算法、分数规划(FP, fractional programming)算法、连续凸逼近算法、内点法和最大化最小系统频谱效率准则的优化算法对功率控制问题进行适当转换与求解。以上算法通过采用优化理论求解系统最佳的功率分配,但由于干扰环境下的系统模型非凸,导致求解过程十分复杂,即使系统的性能提高,但是基于模型的优化算法的复杂度高,求解时间大于信道相干时间,结果难以在实际中应用^[11]。

2) 基于博弈论的方法。文献[12-13]分别将D2D网络中的功率控制问题建模为势博弈和Stackelberg博弈,利用分布式方法最大化多用户的通信速率。基于博弈论的功率控制方法的优势在于其能够实现分布式决策,但博弈论追求稳定的纳什均衡状态,该状态并不能保证所求结果为全局最优解。

3) 基于机器学习的方法。最近兴起了基于机器学习的功率控制方法,这是由于机器学习在计算机科学领域的成功应用,使研究人员逐渐将机器学习技术应用在无线通信中^[14]。文献[15]通过收集大量全局信道状态信息(CSI, channel state information),使用WMMSE算法来产生功率分配集作为标签,使用全局CSI集合与对应的标签进行深度神

经网络(DNN, deep neural network)的训练直到收敛,训练完成后,可以将瞬时全局CSI输入训练好的DNN,直接输出对应的最优功率分配策略。该方法需要消耗大量计算资源和时间成本来产生训练集并对神经网络进行训练,且在实际环境中很难收集到准确的全局信道状态信息来产生合理的训练集。文献[16]提出一种基于多用户深度Q网络的算法,通过不断试错来优化功率分配策略,最终可以收敛到和WMMSE接近的性能,但是该方法需要为神经网络输入大量状态信息且必须对功率进行离散化处理,无法施加连续动作。类似地,文献[17]提出一种分布式深度Q网络方法来进行D2D通信中的分布式功率与频谱分配,该方法能够有效提升动态环境下的D2D通信性能,但仍然只适用于离散动作问题。文献[18]提出多种深度强化学习方法来解决频谱共享网络中的发射功率控制问题,次级用户能够获取主用户的功率分配信息,并结合传感器的接收功率强度值对自身的发射功率进行调整,最终满足通信网络的服务质量要求并实现有效的频谱共享。

本文针对多用户通信链路之间存在干扰的情况,考虑复杂的无线信道环境,提出了以深度强化学习“行动器-评判器”为基本架构的智能发射功率控制算法,对多用户发射功率进行有效控制,实现多用户通信速率最大化。该算法基于深度强化学习(DRL, deep reinforcement learning)技术,通过与环境不断交互、自我改进的学习方式来获得最优策略,不需要带标签的训练集;采用深度确定性策略梯度(DDPG, deep deterministic policy gradient)方法,使用2个神经网络分别拟合行动器和评判器,并在训练过程中加入经验回放和目标参数软替代的方法,确保算法的收敛性;训练收敛后,利用行动器网络拟合出的最优策略,根据信道状态信息实时进行最优的功率控制。仿真结果表明,所提算法能够快速收敛,且在保证性能接近理论最优算法的前提下能够有效降低功率控制所需的运算时间。此外,算法性能不会随着网络规模的增加而下降,能够很好地适用于大规模无线网络。

1 系统模型

设有 K 对收发无线终端设备的无线通信系统,每对收发终端有一条通信链路,每条通信链路中的收发设备固定,且链路之间存在干扰,如图1所示。

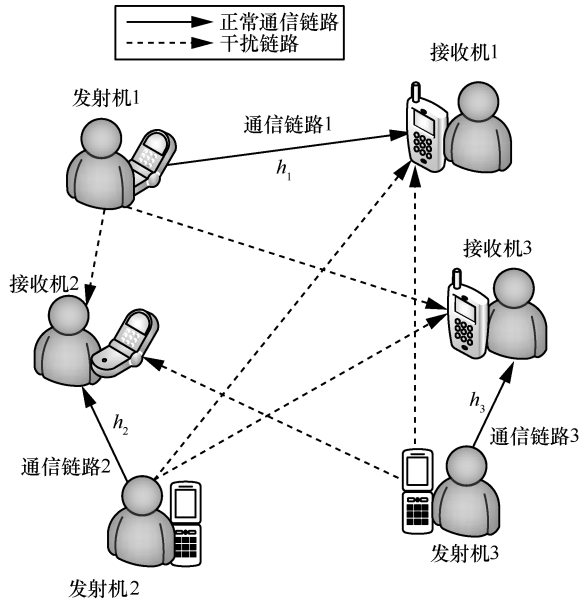


图 1 链路有干扰的无线通信系统示意

假设第 k 条链路 ($1 \leq k \leq K$) 的发射终端设备在时刻 t 的发射功率为 $P_k(t)$, 发射信号为 $x_k(t)$, 其接收端的信号为

$$y_k(t) = h_k(t) \sqrt{P_k(t)} x_k(t) + \sum_{j \neq k} h_{j,k}(t) \sqrt{P_j(t)} x_j(t) + z_k(t) \quad (1)$$

其中, $h_k(t)$ 为时刻 t 第 k 条链路的复信道系数; $h_{j,k}(t)$ 为时刻 t 第 j 条链路的发射机与第 k 条链路的接收机的复信道系数, 即 $\sum_{j \neq k} h_{j,k}(t) \sqrt{P_j(t)} x_j(t)$ 为时刻 t 第 k 条链路收到的来自其他链路的干扰信号; $z_k(t)$ 为独立同分布的复高斯白噪声, 噪声功率为 N_0 。

令 $\mathbf{H}(t) = \{H_{j,k}(t), \forall j, k\}$ 表示全局 CSI, 其中 $H_{j,k}(t) = |h_{j,k}(t)|^2$ 表示时刻 t 第 j 条链路的发射机与第 k 条链路的接收机之间的瞬时信道增益。这里采用无线接入网中常见的瑞利衰落信道模型, 为了表征信道的动态特性, 根据 Jakes 衰落模型^[19], 将信道系数表示为一阶复高斯马尔可夫过程

$$h_{j,k}(t) = \rho h_{j,k}(t-1) + \sqrt{1-\rho^2} e_{j,k}(t) \quad (2)$$

其中, $h_{j,k}(t)$ 和信道更新过程 $e_{j,k}(t)$ 均是独立同分布的单位方差循环对称复高斯随机变量。相关系数 $\rho = J_0(2\pi f_d T)$, 其中, $J_0(\cdot)$ 是零阶贝塞尔函数, f_d 是最大多普勒频率。

对于其中一条通信链路的信号, 其他发射机的信号将被视为噪声, 该设备的接收信号速率也将取决于信干噪比 (SINR, signal to interference plus noise ratio)。在给定信道状态信息 $\mathbf{H}(t)$ 和发射功率 $P(t) = \{P_k(t), \forall k\}$ 的情况下, 接收机 k 的接收数据速率为

$$C_k(t) = \text{lb} \left(1 + \frac{H_k(t) P_k(t)}{\sum_{j \neq k} H_{j,k}(t) P_j(t) + N_0} \right) \quad (3)$$

依据式(3), 建立多用户发射功率控制的干扰管理问题的数学模型, 如式(4)所示。

$$\begin{aligned} \max_{P(t)} \quad & \sum_{k=1}^K \alpha_k C_k(t) \\ \text{s.t.} \quad & 0 \leq P_k(t) \leq P_{\max}, k \in \{1, 2, \dots, K\} \end{aligned} \quad (4)$$

式(4)的目标是实现无线通信系统加权数据速率最大化, 其中, α_k 是设备 k 对应的正值权重, 表征链路重要性。

式(4)的优化变量同时存在于 lb 函数的分子与分母中, 该问题的优化目标函数复杂且非凸, 该问题的求解一直是无线通信领域进行干扰管理的研究重点。现有的求解算法对模型依赖度高, 且算法复杂度较高, 难以适用于未来无线网络大规模多用户接入的复杂动态场景。

2 智能无线发射功率控制算法

考虑到发射功率控制为连续动作问题, 采用深度强化学习的 DDPG 方法^[20], 构建智能无线发射功率控制算法, 整体框架如图 2 所示。

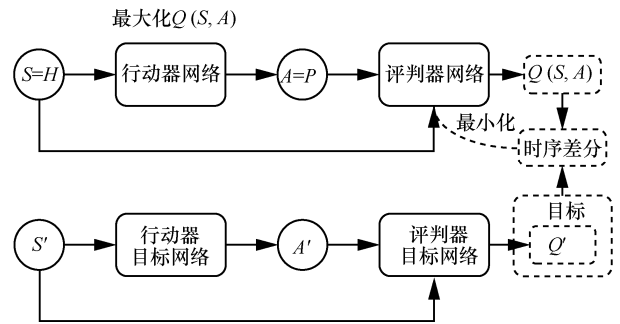


图 2 智能无线发射功率控制算法整体框架

首先, 由于无线信道环境具有马尔可夫性质, 故在此将功率控制这一动态决策的过程建模为马尔可夫决策过程。设 $S=H$ 为全局状态值, 设 $a(t) = \{P_k(t) | 0 \leq P_k(t) \leq P_{\max}, k=1, 2, \dots, K\}$ 为动作集, 在

智能功率控制过程中, 假设存在一个集中控制器能够收集全部信道状态信息, 向智能体输入状态值; 智能体将基于状态值与自身随机策略做出决策, 输出具体动作 (信道状态信息与功率控制信息一般通过控制链路传递, 不占用数据链路带宽^[21]), 即 $a(t) \sim \pi(a(t) | S(t))$, 之后根据状态转移函数进入下一个状态 $S(t+1) \sim \Pr(S(t+1) | S(t), a(t))$ 。此外, 智能体将得到一个对应的奖赏函数 $r(t) = r(S(t), a(t))$, 并得到自身对新状态的观测量 $S(t+1)$ 。训练目标是追求最大化长期回报 $R = \sum_{t=1}^T \gamma^{t-1} r(t)$, 其中, γ 为折扣系数, T 为时间范围。

根据优化问题式(4), 可以将奖赏函数定义为

$$r(t) = \sum_{k=1}^K \alpha_k C_k(t)$$

为了得到最佳的功率控制策略, 强化学习需要不断试错, 并迭代进行策略评估与策略改进^[22]。深度强化学习则使用深度神经网络来进行策略评估与策略改进, 分别对应评判器网络和行动器网络, 但由于强化学习训练过程中前后序列的强相关性, 传统的“行动器-评判器”算法难以收敛。为此, 这里采用收敛性更优的深度确定性策略梯度方法。具体地, 设行动器深度神经网络为 $\mu(s | \theta^\mu)$, 其中, θ^μ 为行动器深度神经网络的权重系数, 行动器目标神经网络选用不同的权重系数 $\theta^{\mu'}$, 目标行动器深度神经网络为 $\mu(s | \theta^{\mu'})$ 。类似地, 将评判器深度神经网络表示为 $Q(s, a | \theta^Q)$, 其目标网络表示为 $Q(s, a | \theta^{Q'})$, θ^Q 和 $\theta^{Q'}$ 分别对应各自神经网络的权重系数。后续对神经网络的训练即对 θ^Q 和 $\theta^{\mu'}$ 这 2 个权重系数的更新, 更新的目的是使评判器网络能够对功率控制策略做出更精准的评估, 使行动器网络输出价值更大, 即系统传输速率更大的发射功率。

为了进一步提升训练效果, 利用经验回放方法, 增加搜索广泛性。每次在状态 $s(t)$ 下根据策略施加动作 $a(t) = \mu(s(t) | \theta^\mu) + \zeta$, 其中, ζ 为一个随机变量, 作为动作噪声来增加探索性; 之后达到新的状态 $s(t+1)$, 并获得相应奖赏 $r(t)$ 。将此时的经验 $g(t) = \{s(t), a(t), r(t), s(t+1)\}$ 存入回放缓存, 形成经验集, 在训练神经网络时从回放缓存中随机选取批量经验进行学习, 以此打破强化学习训练步骤前后的相关性, 保证训练的稳定性和收敛性。

为了训练评判器神经网络, 从回放缓存中随机采样 N 组经验, 选用合适的优化器来最小化该批经验的期望预测误差 (即损失函数), 如式(5)所示。

$$L(\theta^Q) = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2 \quad (5)$$

其中, y_i 为

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} | \theta^{\mu'}) | \theta^{Q'}) \quad (6)$$

据此更新评判器神经网络的参数 θ^Q 。目标评判器神经网络参数 $\theta^{Q'}$ 的更新则采取软更新方法, 即

$$\theta^{Q'} \leftarrow \tau \theta^{Q'} + (1 - \tau) \theta^Q \quad (7)$$

其中, $\tau \in [0, 1]$ 是目标评判器网络的学习率。

行动器神经网络的训练目标是最大化价值函数 $Q(s, a | \theta^Q)$ 的期望, 即

$$J(\theta^\mu) = \mathbb{E}[Q(s, a | \theta^Q)] \quad (8)$$

其中, $\mathbb{E}[\cdot]$ 为期望函数。使用采样策略的梯度下降法更新参数, 计算式(8)的梯度为

$$\nabla_{\theta^\mu} J(\theta^\mu) \approx \frac{1}{N} \sum_i \nabla_a Q(s, a | \theta^Q) \Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s | \theta^\mu) \Big|_{s_i} \quad (9)$$

其中, ∇ 表示求解函数梯度。同样地, 目标行动器神经网络参数 $\theta^{\mu'}$ 的更新也采取软更新方法, 即

$$\theta^{\mu'} \leftarrow \tau \theta^{\mu'} + (1 - \tau) \theta^\mu \quad (10)$$

算法流程如算法 1 所示。

算法 1 智能无线发射功率控制算法

随机初始化功率策略的评判器网络 $Q(s, a | \theta^Q)$ 与行动器网络 $\mu(s | \theta^\mu)$ 的权重系数 θ^Q 与 θ^μ 。初始化目标网络 Q' 和 μ' 的权重系数 $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$; 初始化回放缓存 R 。

- 1) 循环
- 2) 初始化随机过程 $N(t)$ 进行行动探索, 获得初始全局状态观测值 $s(0) = H(0)$;
- 3) 循环
- 4) 根据当前策略和探索噪声选择动作 (即每个发射器的发射功率) $a(t) = \mu(s(t) | \theta^\mu) + N(t)$;
- 5) 基于动作 $a(t)$ 分离出每个发射器的发射功率 p_k 并计算得到对应奖励值 $r(t) = \sum_{k=1}^K \alpha_k C_k(t)$ 和

新的状态值 $s(t+1) = H(t+1)$;

6) 将组合 $(s(t), a(t), r(t), s(t+1))$ 存入 R 中;

7) 如果经验缓存 R 已满, 则

8) 从 R 中随机抽取最小批量数目 N 个组合 $(s(i), a(i), r(i), s(i+1))$, 令 $y(i) = r(i) + \gamma Q'(s(i+1), \mu'(s(i+1)|\theta^\mu)|\theta^\mu)$;

9) 根据前面抽取的批数据更新评判器网络: 最小化损失函数 $L(\theta^Q) = \frac{1}{N} \sum_i (y(i) - Q(s(i), a(i)|\theta^Q))^2$;

10) 使用采样梯度下降方法更新评行动器网络并改进策略

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta^Q) \Big|_{s=s(i), a=\mu(s(i))} \nabla_{\theta^\mu} \mu(s|\theta^\mu) \Big|_{s(i)}$$

11) 更新目标网络: $\theta^Q \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$, $\theta^\mu \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$;

12) until t 大于或等于最大时刻 T ;

13) until 回合数大于或等于最大回合数。

3 仿真分析

3.1 仿真设计

设有 10 对无线终端组成的无线通信系统, 随机分布在直径 1 km 的区域内, 通信总可用带宽为 2 MHz, 通信信道路径损耗为 $120.9 + 37.6\text{lb}d$ (单位为 dB, d 为发射端和接收端之间的距离), 多普勒频率为 10 Hz, 噪声功率 $N_0 = -174 \text{ dBm/Hz}$ 。文献[21]分别选取 5 对和 10 对收发机进行仿真验证, 为更好地进行验证, 本文也分别对 5 对和 10 对收发机进行仿真验证。发射机最大发射功率为 1 W, 总时间步长为 0.1 s, 分为 100 个时间块, 设置所有的链路权重 α_k 均为 1。使用 Python 开源第三方库 Tensorflow 2.4.0 和 Keras 对神经网络进行构建与训练, 以下所有仿真均在同一块 10 代 i5-CPU 上进行, 训练深度神经网络的超参数设置如表 1 所示。

3.2 算法收敛性分析

图 3 展示了基于 DDPG 的智能无线发射功率控制算法的收敛情况。从图 3 可以看到, 算法在一开始需要收集一定数量的经验, 此时进行随机探索, 强化学习的回报值基本没有提升。当经验缓存达到训练要求数量, 即算法开始训练后, 回报值将随着回合数的增加逐渐升高, 证明神经网络得到了有效的训练, 并在较短的时间内就能够收敛。

表 1 深度神经网络超参数设置

超参数	数值
训练回合数 EP	500
训练步长 T	100
最小批次大小	64
折扣系数 γ	0.9
评判器网络学习率 LR_C	0.001
行动器网络学习率 LR_A	0.001
回放缓存大小	10 000
优化器	Adam
激活函数	ELU
神经网络参数初始化	he_normal
评判器网络隐藏层层数	2
评判器网络隐藏层神经元个数	(128, 64)
行动器网络隐藏层层数	2
行动器网络隐藏层神经元个数	(128, 64)
目标网络更新系数 τ	0.01

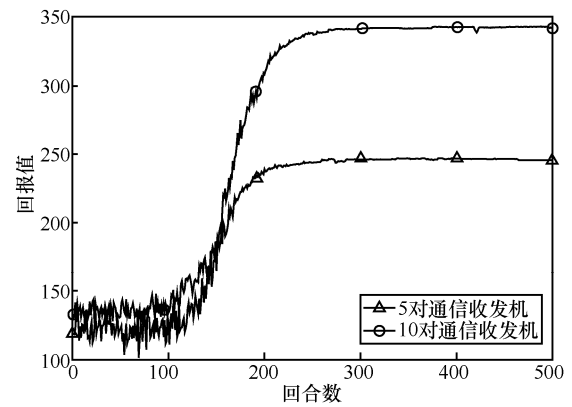


图 3 基于 DDPG 的智能无线发射功率控制算法收敛情况

由于超参数的选择对于深度学习的训练至关重要, 图 4 和图 5 给出了神经网络中典型超参数学习率和隐藏层数对算法收敛情况的影响。从图 4 可以看出, 学习率过高或者过低的情况下, 基于 DDPG 的智能无线发射功率控制算法均容易收敛至局部最优解, 选取适当的学习率对训练效果有很大影响。

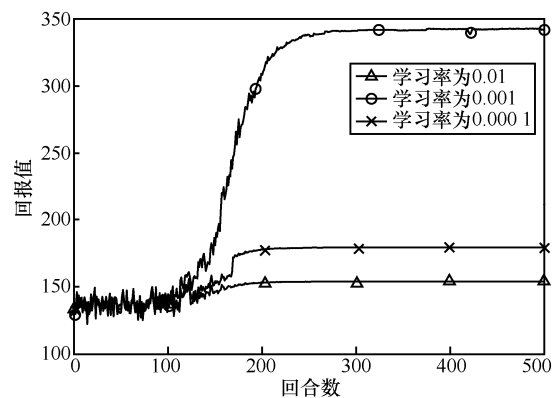


图 4 学习率对算法收敛情况的影响

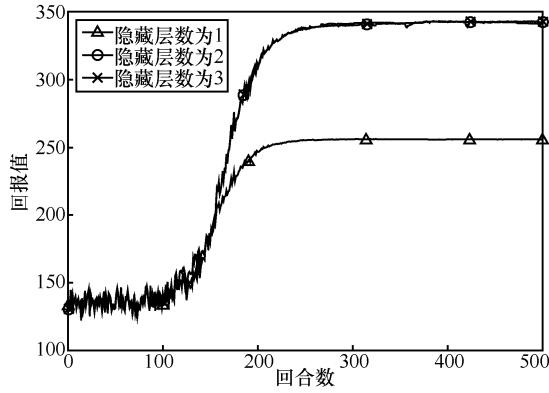


图 5 隐藏层数对算法收敛情况的影响

从图 5 可以看出，只使用一层隐藏层时神经网络不能很好地在强化学习中进行策略评估与策略改进，最终收敛至局部最优值，而使用两层及以上隐藏层时算法可以较好地收敛到全局最优值。但考虑到过多的隐藏层会增加计算与存储负担，且过多的隐藏层存在过拟合的风险，故本文最终选择使用两层隐藏层的深度神经网络。

3.3 算法性能分析

本节采用仿真手段，将本文所提智能无线发射功率控制算法训练得到的计算模型与传统优化算法 WMMSE^[6]和随机分配发射功率的方法进行比较。WMMSE 算法使用 MMSE-SINR 等式^[23]，即 $\text{mmse}_k = \frac{1}{1 + \text{sinr}_k}$ ，将非凸的通信速率最大化问题式(4)转换为更高维度的可解的信号检测问题，MMSE 指用户的最小均方误差，如式(11)所示。

$$\begin{aligned} \min_{\{w_k, u_k, v_k\}_{k=1}^K} & \sum_{k=1}^K \alpha_k (w_k e_k - \log(w_k)) \\ \text{s.t.} & 0 \leq v_k \leq \sqrt{P_k}, k=1, 2, \dots, K \end{aligned} \quad (11)$$

其中

$$e_k = (u_k |h_{kk}| v_k - 1)^2 + \sum_{j \neq k} (u_k |h_{kj}| v_j)^2 + \sigma_k^2 u_k^2 \quad (12)$$

运用块坐标下降法^[24]求解式(11)得出原问题的最优解^[6]。

算法效果方面，图 6 给出了不同功率控制算法的平均传输速率的比较，本文算法在不同通信收发机数量的场景下均实现了超过随机分配算法的性能，且本文的智能功率控制算法能够实现平均传输速率逼近理论上最优的 WMMSE 算法。

算法效率方面，表 2 和表 3 给出在进行 100 步功率控制情况下，不同算法在不同数量收发机场景

中进行最优功率控制所需的运算时间。可以看出，WMMSE 算法所需运算时间随收发机数量的增长而快速增加，而本文算法只有小幅度改变。具体地，在 5 对收发机的情况下，本文算法运算时间略少于 WMMSE，但在 10 对收发机的情况下，本文算法进行最优功率控制所需时间仅为 WMMSE 算法的 $\frac{1}{4}$ 。

这是因为针对不同数量收发机的神经网络规模相同，故而本文所提算法的运算时间并不会发生显著变化。本文所提算法具备良好的可扩展性，能够有效适用于大规模用户的管理。

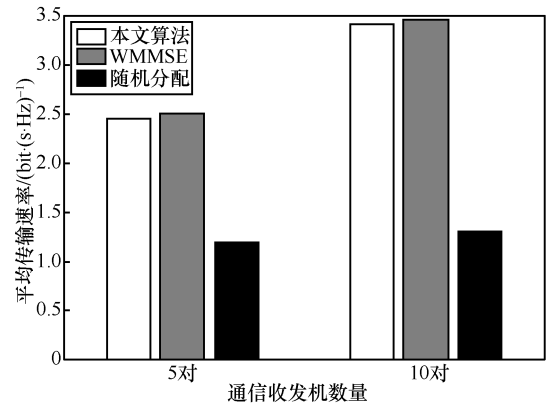


图 6 不同功率控制算法的平均传输速率

表 2 5 对收发机的最优功率控制运算时间

算法	运算时间/s
WMMSE 算法	0.044 9
本文所提算法	0.030 9

表 3 10 对收发机的最优功率控制运算时间

算法	运算时间/s
WMMSE 算法	0.131 6
本文所提算法	0.034 9

4 结束语

本文对多用户干扰情景下的智能无线发射功率控制算法进行了研究，提出了深度强化框架下的智能控制算法，以最大化通信系统的传输速率为目标优化发射功率控制策略。该算法借鉴深度强化学习中的深度确定性梯度下降技术，对行动器与评判器的 2 个深度神经网络进行训练，进而获得对策略的精准评估与合理改进，采用经验回放和目标网络参数软更新的方法，确保算法具有良好的收敛性。仿真结果表明，该算法具有良好的收敛性，计算结果接近理论最优。

参考文献:

- [1] ZHANG L, LIANG Y C, NIYATO D. 6G Visions: mobile ultra-broadband, super Internet-of-things, and artificial intelligence[J]. *China Communications*, 2019, 16(8): 1-14.
- [2] SAAD W, BENNIS M, CHEN M Z. A vision of 6G wireless systems: applications, trends, technologies, and open research problems[J]. *IEEE Network*, 2020, 34(3): 134-142.
- [3] ZHANG L, XIAO M, WU G, et al. A survey of advanced techniques for spectrum sharing in 5G networks[J]. *IEEE Wireless Communications*, 2017, 24(5): 44-51.
- [4] ASADI A, WANG Q, MANCUSO V. A survey on device-to-device communication in cellular networks[J]. *IEEE Communications Surveys & Tutorials*, 2014, 16(4): 1801-1819.
- [5] LIU J Y, SHENG M, LIU L, et al. Interference management in ultra-dense networks: challenges and approaches[J]. *IEEE Network*, 2017, 31(6): 70-77.
- [6] SHI Q J, RAZAVIYAYN M, LUO Z Q, et al. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel[J]. *IEEE Transactions on Signal Processing*, 2011, 59(9): 4331-4340.
- [7] SHEN K M, YU W. Fractional programming for communication systems—part I: power control and beamforming[J]. *IEEE Transactions on Signal Processing*, 2018, 66(10): 2616-2630.
- [8] LIU G, DENG H G, QIAN X W, et al. Joint pilot allocation and power control to enhance max-min spectral efficiency in TDD massive MIMO systems[J]. *IEEE Access*, 2019, 7: 149191-149201.
- [9] YANG R Z, LIN B, YU F R, et al. A dynamic pilot and data power allocation for TDD massive MIMO systems[C]//*Proceedings of 2018 IEEE Global Communications Conference*. Piscataway: IEEE Press, 2018: 1-6.
- [10] CHENG H V, BJÖRNSSON E, LARSSON E G. Uplink pilot and data power control for single cell massive MIMO systems with MRC[C]//*Proceedings of 2015 International Symposium on Wireless Communication Systems (ISWCS)*. Piscataway: IEEE Press, 2015: 396-400.
- [11] ZAPPONE A, DI-RENZO M, DEBBAH M. Wireless networks design in the era of deep learning: model-based, AI-based, or both? [J]. *IEEE Transactions on Communications*, 2019, 67(10): 7331-7376.
- [12] ABRARDO A, MORETTI M. Distributed power allocation for D2D communications underlying/overlying OFDMA cellular networks[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(3): 1466-1479.
- [13] LYU J B, CHEW Y H, WONG W C. A Stackelberg game model for overlay D2D transmission with heterogeneous rate requirements[J]. *IEEE Transactions on Vehicular Technology*, 2016, 65(10): 8461-8475.
- [14] 梁应敞, 谭俊杰, Dusit Niyato. 智能无线通信技术研究概况[J]. *通信学报*, 2020, 41(7): 1-17.
- LIANG Y C, TAN J J, NIYATO D. Overview on intelligent wireless communication technology[J]. *Journal on Communications*, 2020, 41(7): 1-17.
- [15] SUN H R, CHEN X Y, SHI Q J, et al. Learning to optimize: training deep neural networks for interference management[J]. *IEEE Transactions on Signal Processing*, 2018, 66(20): 5438-5453.
- [16] NASIR Y S, GUO D N. Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks[J]. *IEEE Journal on Selected Areas in Communications*, 2019, 37(10): 2239-2250.
- [17] ZOU Z Q, YIN R, CHEN X F, et al. Deep reinforcement learning for D2D transmission in unlicensed bands[C]//*Proceedings of 2019 IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops)*. Piscataway: IEEE Press, 2019: 42-47.
- [18] ZHANG H J, YANG N, HUANGFU W, et al. Power control based on deep reinforcement learning for spectrum sharing[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(6): 4209-4219.
- [19] KIM T, LOVE D J, CLERCKX B. Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems?[J]. *IEEE Transactions on Signal Processing*, 2011, 59(4): 1654-1669.
- [20] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. *arXiv Preprint, arXiv: 1509.02971*, 2015.
- [21] WANG W, ZHANG F, LAU V K N. Dynamic power control for delay-aware device-to-device communications[J]. *IEEE Journal on Selected Areas in Communications*, 2015, 33(1): 14-27.
- [22] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[J]. *IEEE Transactions on Neural Networks*, 1998, 9(5): 1054.
- [23] SCHMIDT D A, SHI C X, BERRY R A, et al. Minimum mean squared error interference alignment[C]//*Proceedings of 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. Piscataway: IEEE Press, 2009: 1106-1110.
- [24] HONG M Y, RAZAVIYAYN M, LUO Z Q, et al. A unified algorithmic framework for block-structured optimization involving big data: with applications in machine learning and signal processing[J]. *IEEE Signal Processing Magazine*, 2016, 33(1): 57-77.

[作者简介]



张先超(1984—), 男, 安徽合肥人, 博士, 东南大学在站博士后, 嘉兴学院教授, 主要研究方向为无线网络资源管理、人工智能等。

赵耀(1996—), 男, 河南南乐人, 北京理工大学博士生, 主要研究方向为智能无线通信、无线资源管理等。

叶海军(1979—), 男, 安徽池州人, 中国电子科学研究院研究员, 主要研究方向为空基网络化信息系统体系作战、总体设计与综合集成。

樊锐(1989—), 男, 安徽滁州人, 博士, 中国电子科学研究院高级工程师, 主要研究方向为网络信息体系、人工智能。