

6G 密集网络中基于深度强化学习的资源分配策略

杨凡, 杨成, 黄杰, 张仕龙, 喻涛, 左迅, 杨川

(重庆理工大学电气与电子工程学院, 重庆 400054)

摘要: 6G 密集网络 (DN) 中通过资源分配实现小区间无交叠干扰是提升网络性能的重要技术, 但资源受限和节点密集分布使其很难通过传统的优化方法解决资源分配问题。针对此问题, 建立了基于点线图染色的交叠干扰模型, 将深度强化学习 (DRL) 和交叠干扰模型相结合, 提出一种基于竞争深度 Q 网络 (Dueling DQN) 的资源分配方法。该方法利用交叠干扰模型与资源复用率设计即时奖励, 利用 Dueling DQN 自主学习生成 6G DN 资源分配策略, 实现小区间无交叠干扰的资源分配。仿真实验表明, 所提方法可有效提高网络吞吐量和资源复用率, 提升网络性能。

关键词: 6G 密集网络; 交叠干扰; 深度 Q 网络; 资源分配

中图分类号: TN929.5

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023148

Resource allocation strategy based on deep reinforcement learning in 6G dense network

YANG Fan, YANG Cheng, HUANG Jie, ZHANG Shilong, YU Tao, ZUO Xun, YANG Chuan

School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 400054, China

Abstract: In order to realize no overlapping interference between cells, 6G dense network (DN) adopting resource allocation is the important technology of enhancing network performance. However, limited resources and dense distribution of nodes make it difficult to solve the problem of resource allocation through traditional optimization methods. To tackle the problem, a point-line graph coloring based overlapping interference model was formulated and a Dueling deep Q-network (DQN) based resource allocation method was proposed, which combined deep reinforcement learning (DRL) and the overlapping interference model. Specifically, the proposed method adopted the overlapping interference model and resource reuse rate to design the immediate reward. Then, generating 6G DN resource allocation strategies were independently learned by using Dueling DQN to achieve the goal of realizing resource allocation without overlapping interference between cells. The performance evaluation results show that the proposed method can effectively increase both network throughput and resource reuse rate, as well as enhance network performance.

Keywords: 6G dense network, overlapping interference, deep Q-network, resource allocation

收稿日期: 2023-02-20; 修回日期: 2023-07-10

通信作者: 黄杰 huangjie_cq@cqu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62301094); 重庆市自然科学基金资助项目 (No.cstc2021jcyj-msxmX0251); 重庆市教育委员会科学技术研究计划基金资助项目 (No.KJQN202101115, No.KJQN202201157, No.KJQN202301135); 重庆理工大学国家自然科学基金和社会科学基金培育计划资助项目 (No.2022PYZ017); 重庆市巴南区科技基金资助项目 (No.KY202208153976019); 重庆理工大学科研创新团队培育计划基金资助项目 (No.2023TDZ003); 重庆理工大学研究生教育高质量发展行动计划基金资助项目 (No.gzlxc20233076)

Foundation Items: The National Natural Science Foundation of China (No.62301094), The Natural Science Foundation of Chongqing (No.cstc2021jcyj-msxmX0251), The Science and Technology Research Program of Chongqing Education Commission of China (No.KJQN202101115, No.KJQN202201157, No.KJQN202301135), The Cultivation Plan of National Natural Science Foundation and Social Science Foundation of Chongqing University of Technology (No.2022PYZ017), Chongqing Banan District Scientific Research Project (No.KY202208153976019), The Cultivation Program of Scientific Research and Innovation Team of Chongqing University of Technology (No.2023TDZ003), The Funding Result of Graduate Education High-quality Development Action Plan of Chongqing University of Technology (No.gzlxc20233076)

0 引言

6G 密集网络 (DN, dense network) 是一个面向智慧医疗、全息通信、智慧城市群、应急通信抢险、智能工厂以及数字孪生等全场景的网络, 可实现通信和人工智能 (AI, artificial intelligence) 的双向互促^[1]。随着智能终端 (IT, intelligent terminal) 的空前普及和网络规模的快速扩展, 5G DN 无法完全满足不断上升的技术标准, 如自主、超大规模、高度动态和完全智能的服务。未来智能化和智能物联网网络的快速增长可能超过 5G DN 的能力。与 5G DN 相比, 6G DN 将支持 10 倍高的网络容量和十分之一的时延, 同时服务 10 倍多的终端设备数量以及提高服务质量 (QoS, quality of service)^[2-3]。6G DN 将主要由 AI 驱动, AI 技术可实现从“物联”到“智联”的跃迁, 彻底改变无线网络^[4]。对于 6G DN 而言, 优化资源分配是保证网络性能的基本机制。通过基于 AI 的智能资源分配和复用, 可以缓解海量连接对资源的巨大需求, 从而显著提高资源复用率^[5]。另一方面, 大数据促进了 AI 在资源管理中的应用, AI 可通过挖掘数据来学习各种模式 (如用户流量模式和移动模式) 或复杂的相互关系。6G DN 的主要优点是链路强度的提高和广泛的频谱复用。但 IT 节点的密集分布导致小区间无交叠干扰加剧, 网络容量下降, 同时交叠干扰造成了网络吞吐量与 IT 节点数量不成线性关系^[6], 使 6G DN 的资源利用效率和服务质量急剧降低。因此, 在密集网络下如何进行资源分配使密集网络的小区间无交叠干扰是提升 6G DN 性能亟待解决的问题^[2]。

目前, 优化方法已经用于解决密集性无线网络的资源分配问题^[7-13]。例如, 考虑密集且复杂的网络中端到端 (D2D) 用户共享相同资源产生的同频干扰, 文献[7]基于外部性匹配理论提出一种干扰感知资源分配算法, 通过最小化同频干扰来最大化网络吞吐量。考虑 D2D 通信会引入额外干扰破坏蜂窝网络连续干扰消除的解码条件, 文献[8]研究联合 D2D 模式选择和资源分配的组合优化 (CO, combinatorial optimization) 问题, 提出一种联合资源分配和模式选择方法。针对密集部署 IT 带来的严重小区干扰, 文献[9]基于回归方法对干扰建模, 提出一种面向干扰的资源分配框架。但是, 在 6G DN 资源分配过程中还需考虑功率控制对干扰的影响。为保证用户的 QoS, 文献[10]研究多目标的资源分配优化问题, 考虑每个用户的传输功率和 QoS 约束,

提出一种面向功率控制的资源分配算法。文献[11]研究联合资源分配和功率分配的优化问题, 考虑 D2D 间和小区间干扰的统计信息, 提出基于多项式时间算法的资源分配策略。文献[12]研究支持非正交多址的联合资源分配问题, 考虑功率控制、最小速率和连续干扰消除的约束, 采用 KKT (Karush-Kuhn-Tucker) 条件和对偶理论提出一种用于资源分配的梯度迭代算法。此外, 考虑到 6G DN 中存在的交叠区域会造成信道间较严重的功率干扰, 文献[13]研究密集网络的能源效率问题, 提出一种基于能量感应架构的干扰感知资源分配方法。此外, 为适应 6G DN 中多用户、多小区、多频段的无线通信场景, 资源分配过程中还需考虑用户需求差异和区分干扰类型^[14-15]。针对异构网络的同层干扰和用户需求的多样性, 文献[14]研究联合次频带和资源块分配优化问题, 考虑用户 QoS 要求和功率限制, 提出一种基于干扰管理的联合资源分配。文献[15]研究联合资源分配和用户选择问题, 考虑干扰拓扑的复杂性和资源的有限性等约束, 提出一种分布式自优化资源分配方法。然而, 使小区间无交叠干扰的资源分配是非凸且 NP-hard 的 CO 问题, 难以求解^[16]。此外, 上述方法难以适应大规模数据的实时计算, 无法充分挖掘隐藏在海量数据中有价值的信息, 不能识别网络中的通信资源复用情况, 这将造成网络吞吐量和用户 QoS 下降。因此, 基于优化方法的资源分配策略难以满足 6G DN 在不同业务对资源的需求, 对有限的网络资源利用不够充分。

深度强化学习 (DRL, deep reinforcement learning) 作为数据、算法和算力复合驱动的机器学习技术已广泛应用于医疗、金融和交通等领域^[17]。DRL 通过挖掘数据来学习复杂的相互关系, 依靠丰富的计算能力实现实时在线决策, 可弥补以牺牲存储空间来提高系统性能的缺点, 解决没有初步离线计算结果情况下无法及时处理任务的问题。此外, DRL 可有效应对非凸和 NP-hard 挑战, 解决优化方法难以求解的 CO 问题^[18]。然而, 利用 DRL 实现小区间无交叠干扰, 提高网络吞吐量和资源复用率的资源分配方法较少且研究更多关注的是非密集网络或低密集网络的资源分配^[19-20]。考虑到 6G DN 中资源分配本质问题为 CO 问题, DRL 能有效解决在满足网络吞吐量、资源复用率以及无交叠干扰约束下的资源分配问题, 本文提出一种基于 DRL 的

资源分配方法，利用深度 Q 网络（DQN, deep Q-network）模型和竞争网络结构对 6G DN 进行通信链路的资源分配以及实现小区间无交叠干扰，以提高网络吞吐量和资源复用率。本文主要贡献如下。

1) 针对密集交叠干扰下的 6G DN 场景，构建 6G DN 覆盖模型，提出以最小化密集交叠干扰为优化目标的 CO 问题，为分析密集网络场景下网络拓扑对交叠干扰的影响提供数学依据和评估指标。

2) 提出基于点线图染色的交叠干扰模型和实现无交叠干扰的资源分配策略。首先根据欧氏距离和无交叠干扰准则，建立 6G DN 的交叠干扰模型，为计算资源分配导致的交叠干扰提供理论依据。然后，根据交叠干扰模型设计了避免交叠干扰的资源分配策略，解决 6G DN 中交叠干扰问题。

3) 为实现 6G DN 中网络无交叠干扰的资源分配，构建基于 DQN 的资源分配模型，提出基于 Dueling DQN 的资源分配算法，采用竞争网络结构感知交叠干扰特征，通过马尔可夫决策过程（MDP, Markov decision process）优化网络资源分配的过程，实现无交叠干扰的密集组网，为求解非凸和 NP-hard 的资源分配问题提供一种新的求解方法。

最后本文对所提方法进行仿真实验和实际通信测试。结果表明，基于 Dueling DQN 的资源分配算法可根据 6G DN 的网络拓扑得到资源分配方案，避免交叠干扰，有效地提高网络吞吐量和资源复用率。

1 系统模型

本文考虑如图 1 所示的复杂通信场景中的 6G DN 模型，该 6G DN 由 N 个 IT 组成，IT 之间形成 M 条通信链路。其中 IT 的覆盖范围约为 $250 \text{ m}^{[21]}$ ，主要用于较短距离间的通信。为进一步提高整个网络覆盖区域内的网络吞吐量，在 6G DN 中采用密集部署 IT 的方式。当大量 IT 密集部署在 6G DN 中时，6G DN 将产生密集交叠干扰和资源复用冲突，从而使 6G DN 的资源分配能力下降。

6G DN 的网络覆盖模型通过 IT 间的欧氏距离建立。本文通过无向图模型将资源分配问题转化为图的边染色问题，利用点线图染色理论建立交叠干扰模型。

1.1 网络覆盖模型和 6G DN 的网络拓扑

1) IT 通信覆盖与通信链路形成

图 1 中，6G DN 通信节点主要是 IT，并且 IT 间可以通过 D2D 进行通信。图 2 是图 1 中 IT_1 、 IT_2 和 IT_3 形成的自组网，任意一个 IT 都在其余 IT 的无线覆盖范围之内。 IT_1 、 IT_2 和 IT_3 间都存在数据传输的通信链路。IT 间的通信链路形成可表示为

$$d_{1,2} \leq \min(x_1, x_2) \tag{1}$$

其中， $d_{1,2}$ 表示 IT_1 和 IT_2 之间的欧氏距离； x_1 和 x_2 分别表示 IT_1 和 IT_2 的通信半径。实际情况中 IT 间存在差异，因此 $x_1 \neq x_2$ 。

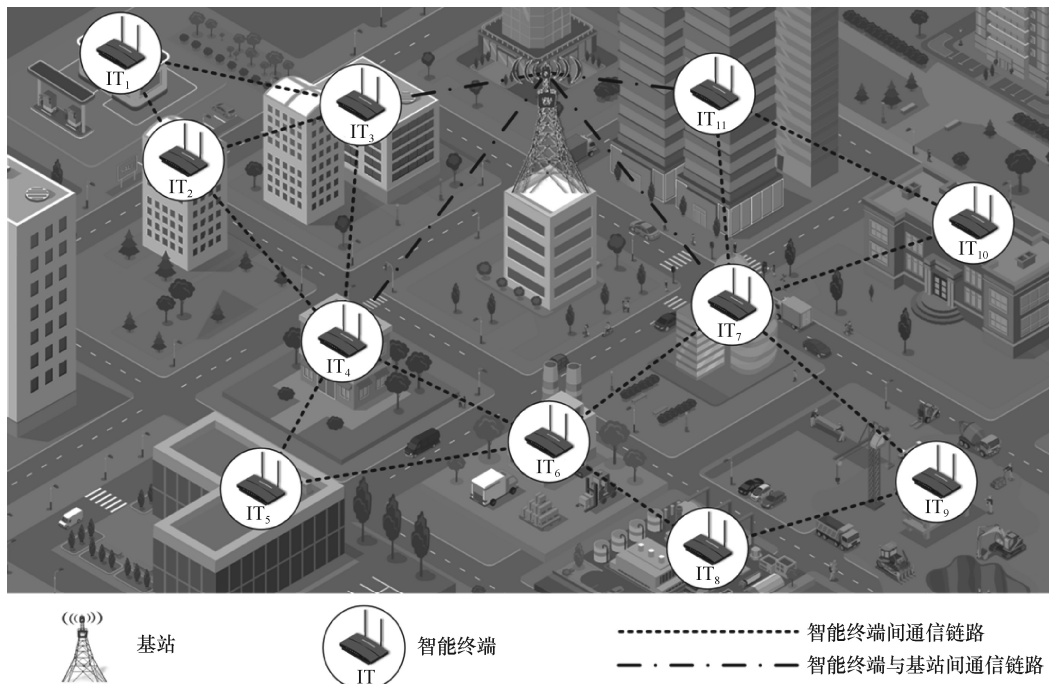


图 1 复杂通信场景中的 6G DN 模型示意

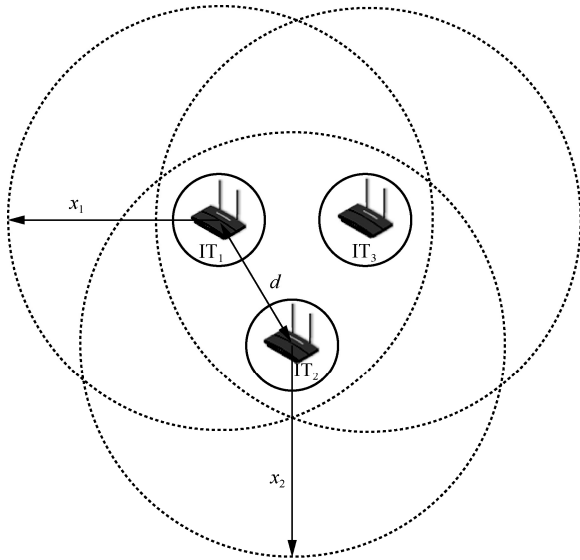


图 2 IT 通信覆盖与通信链路形成

2) 6G DN 覆盖模型

图 3 是图 1 根据 IT 覆盖范围得到的 6G DN 覆盖范围模型，其中任意 2 个 IT 可以通信。6G DN 通信覆盖范围 C_G 是所有 IT 通信覆盖范围的叠加，可表示为

$$C_G = \bigcup_{i \in \mathcal{I}} C_i \quad (2)$$

其中， C_i 表示 IT_i 的通信覆盖范围， \mathcal{I} 表示所有 IT 的集合。

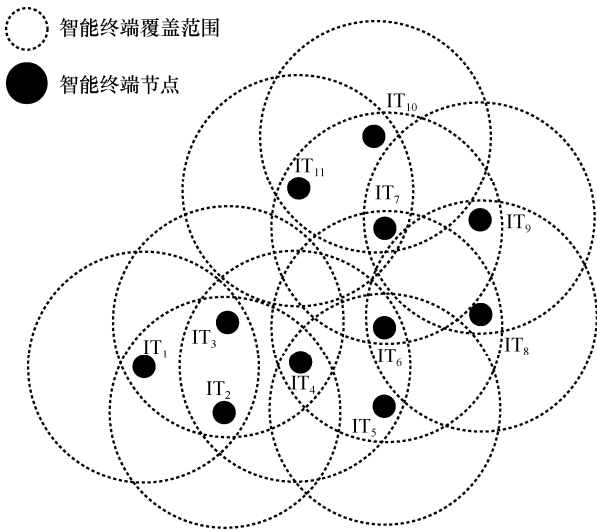


图 3 6G DN 覆盖范围模型

3) 6G DN 的网络拓扑

图 4 是图 3 根据式(1)得到的 6G DN 的网络拓扑。其中，网络拓扑中的节点表示具有通信能力的 IT，拓扑中的线表示节点间的通信链路。6G DN 点线图模型可表示为

$$G = (\mathcal{V}, \mathcal{E}) \quad (3)$$

其中， $\mathcal{V} = \{1, 2, \dots, N\}$ 表示 6G DN 中的 IT 集合， $\mathcal{E} = \{1, 2, \dots, M\}$ 表示 6G DN 中的通信链路集合。

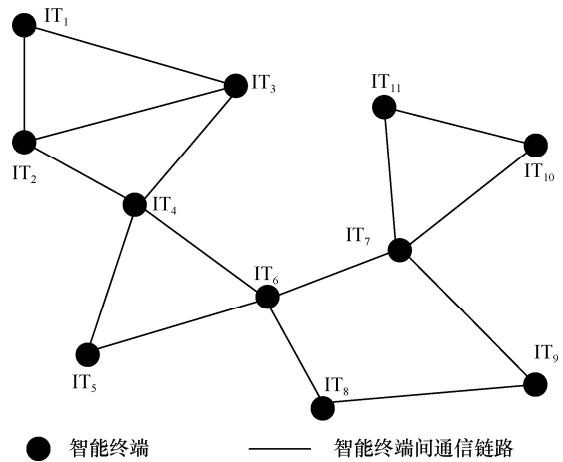


图 4 6G DN 的网络拓扑

1.2 6G DN 的资源分配问题

本节根据点线图边染色与 6G DN 的网络拓扑对应，对 6G DN 中的每一条通信链路分配通信资源。若图 4 中相同节点的通信链路具有不同的通信资源，则 6G DN 资源分配可表示为

$$\begin{aligned} \min \quad & \sum_{e \in \mathcal{E}} \sum_{e' \in \mathcal{E}} \psi(e, e') \\ \text{s.t.} \quad & l_e \cap l_{e'} = \emptyset \\ & e \neq e' \end{aligned} \quad (4)$$

其中， $\psi(e, e')$ 表示通信链路 e 和通信链路 e' 的关系，且 $\psi(e, e')$ 只有 2 个取值， $\psi(e, e') = 1$ 表示通信链路 e 和通信链路 e' 具有相同的通信资源，否则 $\psi(e, e') = 0$ ； $l_e \cap l_{e'} = \emptyset$ 表示通信链路 e 和通信链路 e' 在图 4 中不存在公共的通信节点。

2 6G DN 交叠干扰模型

本节提出一种使网络无交叠干扰的资源分配策略，建立 6G DN 的交叠干扰模型，降低在密集网络交叠干扰下的资源分配难度，实现无交叠干扰资源分配。在建立无向图网络模型后，利用关联矩阵设计资源分配策略。

2.1 基于点线图染色的交叠干扰模型

为在密集网络下进行资源分配，利用关联矩阵 H_G 表示 IT 与通信链路的关系。6G DN 交叠干扰的资源分配问题本质是一个无向图的边染色数学问题，即将无向图中的边染成不同颜色，并满足具有公共顶点的 2 条边颜色不同。6G DN 中存在 N 个 IT

其中, $\delta_G = \frac{2}{15}$ 表示 6G DN 存在 $\frac{2}{15}$ 的通信链路产生交叠干扰。

2.2 基于交叠干扰模型的资源分配策略

在所有 IT 组成 6G DN 的过程中, 根据式(10)可求得 6G DN 资源分配导致的交叠干扰程度。因此, $\delta_G = 0$ 表示 6G DN 资源分配无冲突。资源无冲突分配策略是从起始状态到最终状态的过程中, 保证 6G DN 无交叠干扰的资源分配方法。6G DN 状态 s_{t+1} 的资源分配只与状态 s_t 的资源分配结果有关, 即 6G DN 的通信链路资源分配具有马尔可夫性质。本文提出使用有限长马尔可夫链解决无向图染色问题, 建立任意时间 t 满足 $\delta_G = 0$ 条件的状态转移矩阵 \mathbf{P} 。6G DN 从初始状态 s_0 到最终状态 s_* 的状态转移矩阵 \mathbf{P} 可表示为

$$\mathbf{P} = \begin{bmatrix} p_{s_0, s_0} & p_{s_0, s_1} & \cdots & p_{s_0, s_*} \\ p_{s_1, s_0} & p_{s_1, s_1} & \cdots & p_{s_1, s_*} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s_*, s_0} & p_{s_*, s_1} & \cdots & p_{s_*, s_*} \end{bmatrix} \quad (13)$$

然而, 6G DN 无法直接从通信环境中获得状态转移矩阵 \mathbf{P} 。因此, 6G DN 需要依次对通信链路进行资源分配, 从而使状态 s_t 改变为状态 s_{t+1} 。6G DN 的资源分配动作可表示为

$$a \sim \pi(\cdot | s) \quad (14)$$

其中, s 表示 6G DN 的状态, 包含网络拓扑情况和通信链路的资源分配情况; a 表示通信链路分配通信资源; 资源分配策略 π 表示在状态 s 做出资源分配动作 a 的概率, 具体可表示为

$$\sum_i \pi(a_i | s) = 1, \forall s \quad (15)$$

根据式(14)和式(15), 6G DN 在状态 s 下进行资源分配后, 转移到状态 s' 的概率可表示为

$$p_{s, s'} = p(s' | s) = \sum_i \pi(a_i | s) p(s' | s, a_i) \quad (16)$$

当 6G DN 执行资源分配动作 a 使 $\delta_G = 0$ 时, 状态 s 转移到状态 s' 的概率为 $p(s' | s, a_i) = 1$ 。如果 $\delta_G \neq 0$, 则表示状态 s 无法转移到状态 s' , 即 $p(s' | s, a_i) = 0$ 。因此, 根据系统状态转移矩阵 \mathbf{P} 得到系统求解最优资源分配问题, 即得到最优资源分配策略 π^* 。其中, 资源分配策略 π 会产生具体的资源分配方案 τ , 并且根据不同网络状态产生的分配方案 τ 也不同。资源分配策略 π 产生资源分配方案

τ 的概率可表示为

$$p(\tau | \pi) = \rho_0(S_0) \prod_{t=0}^{T-1} p(S_{t+1} | S_t, A_t) \pi(A_t | S_t) \quad (17)$$

其中, $\rho_0(S_0)$ 表示 6G DN 起始网络状态分布。式(17)可有效地表示每个分配方案 τ 与策略 π 之间的关系。考虑动态情况, 网络拓扑或者结构发生变化将会直接影响马尔可夫链的状态值。因此, 在 6G DN 的动态情况下, 6G DN 的资源分配过程可表示为

$$\underbrace{s_0, a_0, s_1, a_1, s_2, \dots, s_i, a_i, s_{i+1}, a_{i+1}, s_{i+2}, \dots}_{\text{网络拓扑 } 1} \quad (18)$$

其中, 状态 $\{s_0, s_1, s_2\}$ 属于 6G DN 的第 1 种网络拓扑。此外, 网络拓扑动态变化产生新的网络拓扑结构, 状态 $\{s_i, s_{i+1}, s_{i+2}\}$ 属于 6G DN 的第 i 种网络拓扑。

3 基于深度强化学习的资源分配策略

为求解 6G DN 无干扰资源分配策略问题, 本文提出基于 Dueling DQN 的资源无冲突分配 DRL 算法, 构建 6G DN 交叠干扰下的资源分配 MDP 模型, 结合竞争网络模型求解最优资源分配策略 π^* 。在 6G DN 中, MDP 通常由动作 A 、状态 S 、即时奖励 R 和动作价值函数 $Q(s, a)$ 组成, 接下来对其进行详细的介绍。

3.1 基于深度 Q 网络的资源分配模型

本文算法引入 DQN 方法来确定 6G DN 场景下资源无冲突分配的决策序列, 并进行网络拓扑信息和资源分配情况提取, 利用即时奖励函数 R_t 设计交叠干扰判断机制, 通过即时奖励函数 R_t 判断 6G DN 交叠干扰的状态和资源分配对交叠干扰的影响, 提高无线通信网络系统应对密集交叠干扰场景的资源分配能力。

1) 状态: 6G DN 所有可能的通信链路资源分配状态的集合, 即状态集。因此, 6G DN 状态可以表示为

$$S_t = \{H_G^T, K_t\} \quad (19)$$

其中, K_t 表示通信链路的资源分配状态。

2) 动作: 6G DN 观察当前状态做出相应通信链路资源分配的集合, 即动作集。动作集的大小为通信链路的数量 M 与最大通信资源数量 M_s 的乘积。因此, 6G DN 资源分配动作集可以表示为

$$A_t = \{0, 1, 2, \dots, MM_s - 1\} \quad (20)$$

3) 奖励: 6G DN 在状态 S 执行动作 A 获得的对应回报。每个奖励 R 和每个状态-动作对对应,

因此, 6G DN 奖励可表示为

$$R_t = R(S_t, A_t) = \begin{cases} M - \eta_t, & \delta_G = 0 \\ 0, & \text{其他} \end{cases} \quad (21)$$

其中, η_t 表示 6G DN 已经使用的通信资源数量。

4) 动作价值函数: 6G DN 状态 S 和动作 A 的价值估计, 即动作价值函数是基于状态 S 和动作 A 的期望回报。由于 6G DN 是根据资源分配策略 π 进行通信链路资源分配的, 动作价值函数 $Q_\pi(s, a)$ 可以表示为

$$Q_\pi(s, a) = E_{A_t \sim \pi(\cdot|S_t)} \left[\sum_{t=0}^{\infty} \gamma^t R_t \mid S_0 = s, A_0 = a \right] \quad (22)$$

其中, $R(\tau) = \sum_{t=0}^{\infty} \gamma^t R_t$ 表示资源分配方案 τ 的累积折扣回报, γ 表示折扣系数。求解最优资源分配 π^* , 需要找到一个最优的价值函数 $Q^*(s, a)$, 可以表示为

$$Q^*(s, a) = Q_{\pi^*}(s, a) = \max_{\pi} Q_{\pi}(s, a) \quad (23)$$

由于每个状态都需要穷举出所有可能的方案, 本文提出利用贝尔曼方程来化简, 具体推导过程如式(24)所示。

$$\begin{aligned} Q_\pi(s, a) &= E_{a \sim \pi(\cdot|s), s' \sim p(\cdot|s, a)} [R(\tau_{t,T}) \mid S_t = s, A_t = a] = \\ &E_{a \sim \pi(\cdot|s), s' \sim p(\cdot|s, a)} [R_t + \gamma(R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{T-1} R_T) \mid S_t = s, \\ &A_t = a] = E_{s' \sim p(\cdot|s, a)} [R(s, a) + \gamma E_{a' \sim \pi(\cdot|s')} [Q_\pi(s', a')]] \end{aligned} \quad (24)$$

动作价值函数可以表示为

$$Q^*(s, a) = E_{s' \sim p(\cdot|s, a)} \left[R(s, a) + \gamma \max_{a'} Q^*(s', a') \right] \quad (25)$$

贝尔曼方程利用时间差法中的 Q-Learning 算法进行求解。Q-Learning 值的迭代可以表示为

$$Q(s_t, a_t) \leftarrow (1 - \alpha) Q(s_t, a_t) + \alpha \left(R_t + \gamma \max_{a'} Q(s_{t+1}, a') \right) \quad (26)$$

其中, α 表示学习率。当系统的状态和动作数量庞大且连续时, 利用 Q-Learning 解决时很难实现资源分配策略的收敛。因此, DQN 用一个神经网络 $Q(s, a; \theta)$ 代替动作价值函数。DQN 的神经网络参数 θ 采用梯度下降方法, 可以表示为

$$\begin{aligned} \theta_{t+1} &= \theta_t + \alpha \nabla_{\theta} (Q(s_t, a_t; \theta_t)) \cdot \\ &\left(\underbrace{R_t + \gamma \max_{a'} Q(s_{t+1}, a; \theta_t)}_{\text{目标值}} - \underbrace{Q(s_t, a_t; \theta_t)}_{\text{估计值}} \right) \end{aligned} \quad (27)$$

其中, $\nabla_{\theta}(\cdot)$ 表示对权值 θ 进行梯度运算。此外, 在利用神经网络逼近动作价值函数时, 目标值与估计值使用相同权值 θ 的神经网络, 容易导致训练网络不稳定和估值过高。因此, 使用一个权值 $\hat{\theta}$ 的神经网络专门用于产生目标值。权值 θ 的神经网络专门用于评估策略更新参数。具有双重权值的神经网络更新可以表示为

$$\begin{aligned} \theta_{t+1} &= \theta_t + \alpha \nabla_{\theta} (Q(s_t, a_t; \theta_t)) \cdot \\ &\left(\underbrace{R_t + \gamma \max_{a'} Q(s_{t+1}, a; \hat{\theta}_t)}_{\text{目标值}} - \underbrace{Q(s_t, a_t; \theta_t)}_{\text{估计值}} \right) \end{aligned} \quad (28)$$

3.2 基于 Dueling DQN 的资源分配算法

本文所提的基于 Dueling DQN 的资源分配算法采用竞争网络模型结合 MDP 来解决交叠干扰问题。竞争网络可以有效提取 6G DN 密集交叠干扰的特征。算法的时间复杂度只与竞争网络模型的规模有关, 受实际的无线通信环境影响较小。其次, 利用 MDP 来解决资源分配问题, 将原本的 CO 问题转成序列决策问题求解, 大大降低直接求解优化问题的难度。

为防止 DQN 估值过高, 利用竞争网络进一步提高算法网络学习的鲁棒性。竞争网络将动作价值函数 $Q(s, a)$ 分成状态价值函数 $V(s)$ 和优势函数 $A(s, a)$ 。 $A(s, a)$ 表示在状态 s 下选取资源分配动作 a 的合理性。竞争网络的动作价值函数可表示为

$$Q(s, a; \theta, w_1, w_2) = V(s; \theta, w_1) + A(s, a; \theta, w_2) \quad (29)$$

其中, θ 是 DQN 公共部分的神经网络权值, w_1 是状态价值函数部分的神经网络权值, w_2 是优势函数部分的神经网络权值。此外, 为提高竞争网络的稳定性, 本文增加平均值基准 $\sum_{a'} \frac{A(s, a'; \theta, w_2)}{N_A}$ 。具有平均值基准的动作价值函数可表示为

$$\begin{aligned} Q(s, a; \theta, w_1, w_2) &= V(s; \theta, w_1) + \\ &\left(A(s, a; \theta, w_2) - \frac{\sum_{a'} A(s, a'; \theta, w_2)}{N_A} \right) \end{aligned} \quad (30)$$

其中, $N_A = MM_s$ 表示所有资源分配动作的数量。

本文提出基于 Dueling DQN 的资源分配算法来解决 6G DN 无交叠干扰分配问题。算法结构如图 7 所示, 算法包含 6G DN 环境、 ε -greedy 策略、深度 Q 网络、目标 Q 网络、损失函数、经验池和抽样小批量样本。其中, 6G DN 环境提供环境状态 S 和根据动作更新通信链路的状态; ε -greedy 策略为

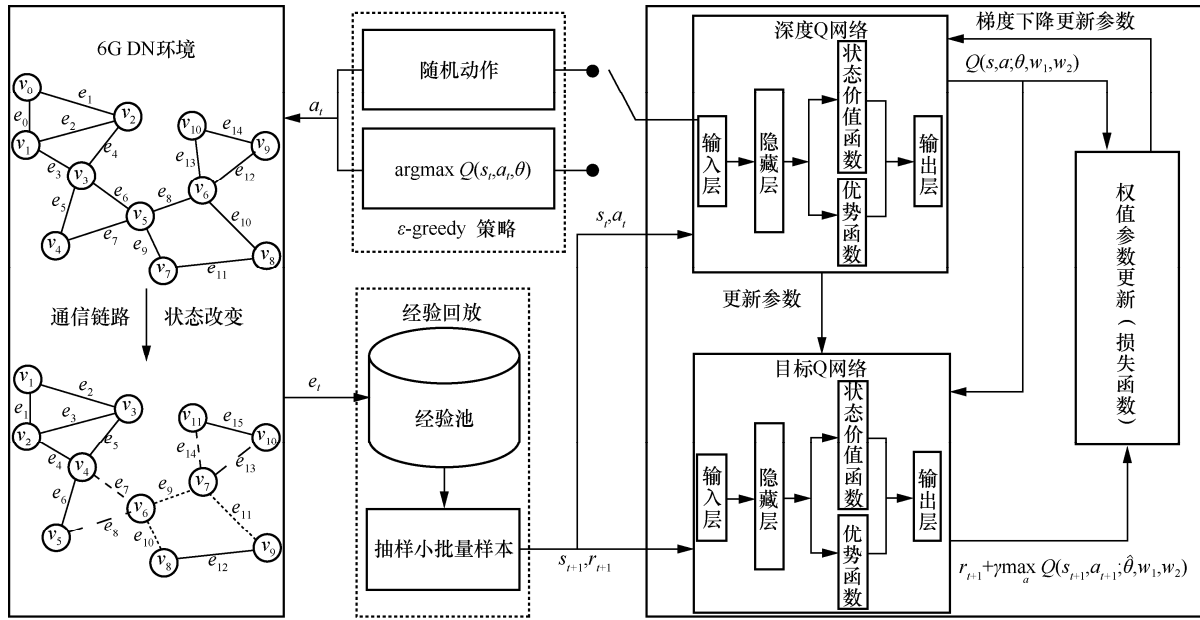


图 7 基于 Dueling QN 的资源分配算法结构

深度 Q 网络提供“探索-利用”的能力；深度 Q 网络评估动作的价值；目标 Q 网络则选择当前价值最大的动作；经验池和抽样小批量样本则增加样本之间的随机性；损失函数根据式(28)对深度 Q 网络的参数进行更新。

基于 Dueling DQN 的资源分配算法如算法 1 所示。

算法 1 基于 Dueling DQN 的资源分配算法

输入 网络拓扑 $G(V, E)$ ，奖励折扣因子 γ ，经验回放大小 N_D ，目标网络更新频率 L ，学习率 α ，样本批采样大小 Γ ，训练次数 N_{train} ，每次训练执行时间长度 T

输出 神经网络权值 θ, w_1, w_2

- 1) for episode = 1, 2, ..., N_{train} do
- 2) 初始化 6G DN 状态，得到初始状态 s_0 ；
- 3) for $t = 1, 2, \dots, T$ do
- 4) 对于每个资源分配方案 τ 中的每步，采用 ϵ -greedy 策略选择动作，随机选择一个动作或者 $a_t = \arg \max_a Q(s_t, a; \theta, w_1, w_2)$ ；
- 5) 6G DN 执行资源分配动作 a_t 后对通信链路产生即时奖励 r_t 和状态 s_{t+1} ；
- 6) 6G DN 将得到的经验 $\langle s_t, a_t, r_t, s_{t+1} \rangle$ 存入经验池中；
- 7) 6G DN 从经验池中抽取一批样本；
- 8) 6G DN 使用式(28)更新训练深度 Q 网络

的权值；

- 9) 每隔 L 步，6G DN 更新目标网络的权值；
- 10) end for；
- 11) end for

4 仿真结果与性能分析

实验采用的硬件平台为个人计算机，CPU 为 Intel(R) Xeon(R) Gold 6242R CPU @ 3.10 GHz，GPU 为 NVIDIA RTX 3080Ti，内存为 64 GB。随着通信链路数量增加，网络复杂度和密集程度增加。

根据表 1 参数进行仿真实验，分别得到本文算法与其他 3 种算法网络吞吐量和资源复用率性能的对比如表 1 所示。其中，对比算法 1 为无竞争网络的深度 Q 网络算法，对比算法 2 为基于随机匹配的资源分配算法，对比算法 3 为基于贪婪匹配的资源分配算法。

表 1 仿真参数设置

参数	数值
学习率	0.000 1
目标网络更新频率	20
经验回放大小	500
经验池容量	10 000
奖励折扣因子	0.95
贪婪因子	0.05
神经网络每层神经元数量	[256, 256, 256, 128]
频谱带宽/MHz	20
最大可容忍误码率	70%

4.1 时间复杂度分析与算法对比

本文算法和对比算法 1 的时间复杂度主要由神经网络的权值迭代决定, 时间复杂度分别为 $O(G_1 L_1)$ 和 $O(G_2 L_2)$ 。其中, G 表示神经网络的隐藏层数; L 表示每层神经元的数量。对比算法 2 的时间复杂度由通信链路数量和通信资源决定, 但每次分配资源优先考虑资源无冲突, 因此时间复杂度为 $O(M M_s)$ 。对比算法 3 的时间复杂度由通信链路数量和通信资源决定, 但每次分配资源优先考虑复用资源, 时间复杂度为 $O(M \log(M_s))$ 。

本文算法和对比算法 1 在神经网络的隐藏层数和每层神经元的数量相近时, 两者时间复杂度近似。但本文算法采用竞争网络架构, 将动作价值函数分成状态价值函数和优势函数, 当通信链路或通信资源的数量改变时, 只改变对应的神经网络权值即可。而无竞争网络的对比算法 1 则需要进行整体改变。因此随着 6G DN 的网络规模的增加, 本文算法的实际时间复杂度会低于对比算法 1。对比算法 2 和对比算法 3 直接受到通信链路和通信资源的数量影响, 两者的时间复杂度随着 6G DN 的网络规模的增加而线性增加。

综上所述, 相对于传统资源分配算法, 深度强化学习算法在时间复杂度方面有着稳定的优势。由于神经网络可以更好地适应 6G DN 的网络规模的动态变化, 时间复杂度只与神经网络模型的大小有关, 受实际的无线通信环境影响较小。因此使用基于 Dueling DQN 的资源分配算法可以更好地解决 6G DN 下的密集交叠干扰问题。

4.2 竞争网络结构的效能分析

为验证本文所提出的基于 Dueling DQN 的资源分配算法在通信链路数量较大的情况下所产生的优化效果, 本文主要从每次方案中的累积回报 U 进行实验验证。累积回报 U 可表示为

$$U_{t=0:T} = R(\tau) = \sum_{t=0}^T R_t \quad (31)$$

如图 8 所示, 基于 Dueling DQN 的资源分配算法的累积回报随着迭代次数的增加而变化。其中, 本文算法相比无竞争网络结构的算法探索资源分配策略的速度更快, 累积回报曲线上升的趋势更加迅速。图 8 中有竞争网络结构的算法得到的累积回报值高, 表明有竞争网络结构可有效处

理交叠干扰和优化资源分配, 即网络吞吐量和资源复用率更高。

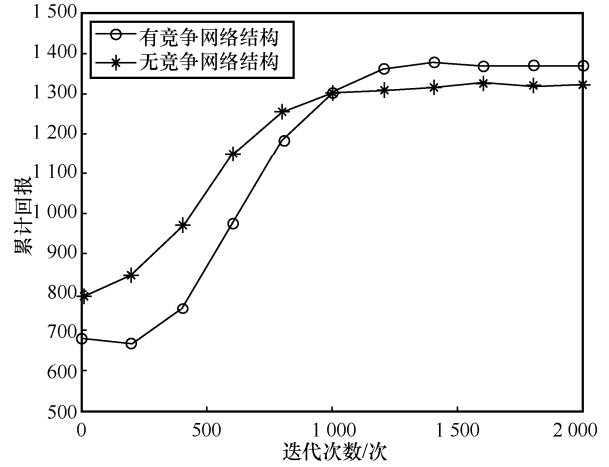


图 8 竞争网络效果

4.3 6G DN 的性能评价指标

1) 网络吞吐量。该性能指标可评价资源分配算法分配完所有通信链路资源后 6G DN 的网络吞吐量^[22], 表示为

$$A = B \log \left(1 + \frac{\bar{g}}{-\frac{2}{3} \ln \frac{P_b}{2}} \right) \quad (32)$$

其中, B 为通信链路的频谱带宽, \bar{g} 为平均信干噪比, P_b 为最大可容忍误码率。

2) 资源复用率。该性能指标可评价资源分配算法使 6G DN 所有通信链路交叠无干扰后 6G DN 的通信资源复用率^[23], 表示为

$$\gamma = \frac{M - \eta_*}{M} \quad (33)$$

其中, η_* 为最终使用的通信资源数量。

4.4 网络吞吐量性能对比

图 9 为本文算法和 3 种对比算法在不同通信链路数量下的最大网络吞吐量。随着通信链路数量的增加, 通信系统中的 IT 交叠区变多, 4 种算法得到的最大网络吞吐量整体都呈现上升趋势。对比算法 1、对比算法 2 和对比算法 3 得到的最大网络吞吐量明显低于本文算法。当通信链路数量为 20、35、40 时, 有竞争网络结构的本文算法相对于对比算法 1 提高 20% 及以上的网络性能。这表明本文算法可有效提高网络吞吐量和处理干扰能力的上限。

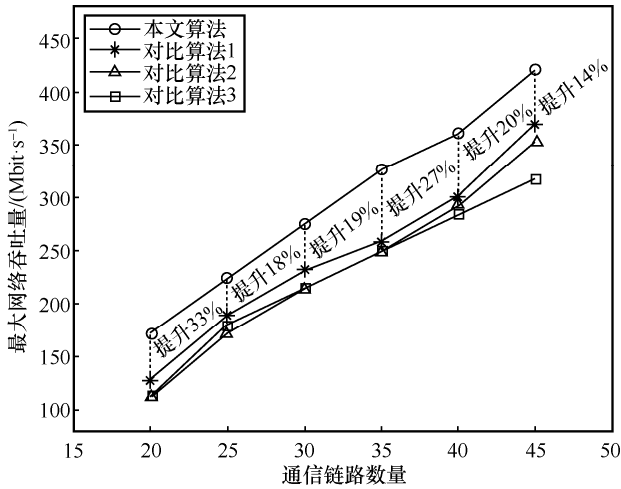


图 9 6G DN 最大网络吞吐量

图 10 为本文算法和 3 种对比算法在不同通信链路数量下的最小网络吞吐量。随着通信链路数量的增加，本文算法和 3 种对比算法得到的网络吞吐量呈现上升趋势。本文算法和对比算法 1 相对于对比算法 2、对比算法 3 有效提高了最小网络吞吐量。其中，有竞争网络结构的本文算法相对于对比算法 1 提高 5% 及以上的网络吞吐量。这表明竞争网络结构可有效地提高系统处理交叠干扰能力的下限。

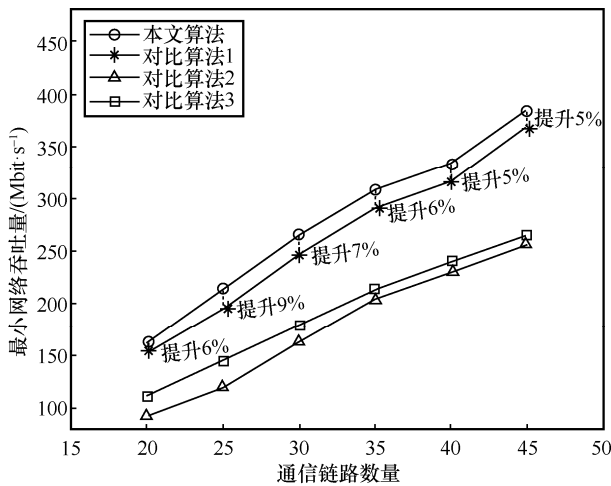


图 10 6G DN 最小网络吞吐量

图 11 为本文算法和 3 种对比算法在不同通信链路数量的平均网络吞吐量。随着通信链路数量的增加，本文算法得到的网络吞吐量呈现稳定上升趋势。本文算法得到的平均网络吞吐量明显高于对比算法 2 和对比算法 3，且相对于对比算法 1 提高 6% 及以上的网络吞吐量。

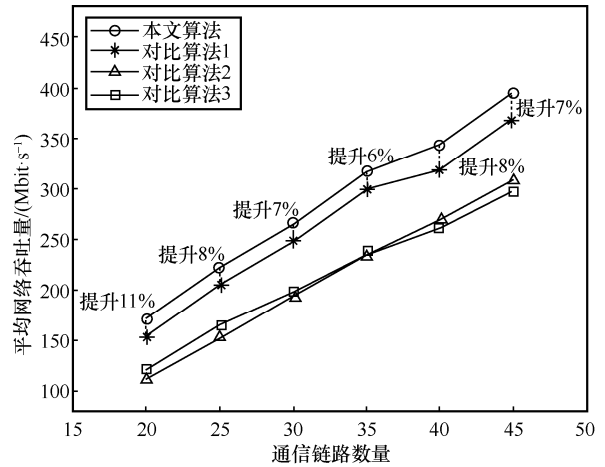


图 11 6G DN 平均网络吞吐量

针对网络吞吐量，本文所提基于 Dueling DQN 的资源分配算法利用 DRL 的在线决策能力，降低计算所需的内存消耗（将网络状态作为竞争网络的输入得到资源分配策略的输出），其中，竞争网络提取小区间无交叠干扰状态，通过资源分配实现小区间无交叠干扰，从而大幅度提高网络吞吐量，保证网络的稳定性。

4.5 资源复用率性能对比

图 12 为本文算法和 3 种对比算法在不同通信链路数量下的最大资源复用率。随着通信链路的增加，本文所提的基于 Dueling DQN 的资源分配算法得到的整体资源复用率明显高于其他算法。其中，本文算法比对比算法 1 提高 19% 的最大资源复用率；相较于对比算法 2 和对比算法 3 分别提高 25% 和 26% 的最大资源复用率。这表明本文采用 MDP 解决资源分配问题，可有效增加整个系统资源分配能力的上限。图 12~图 14 中，不同横线表示不同算法在 5 次不同通信链路数量的资源复用率均值。

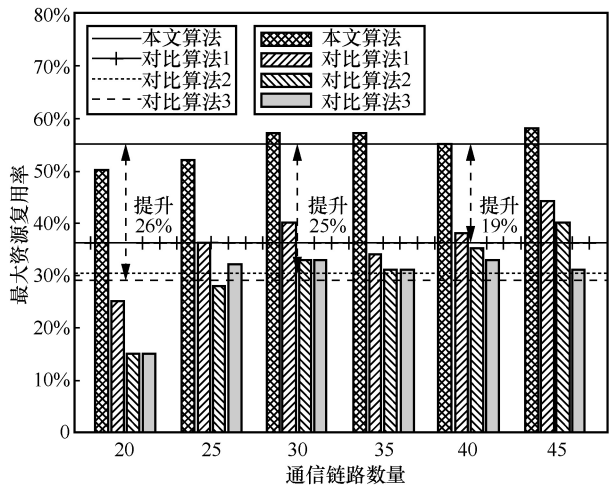


图 12 最大资源复用率

图 13 为本文算法和 3 种对比算法在不同通信链路数量的最小资源复用率。随着通信链路的增加，本文算法和对比算法 1 得到的最小资源复用率明显高于对比算法 2、对比算法 3。其中，本文算法对比算法 1 提高 6% 的最小资源利用；相较于对比算法 2 和对比算法 3 分别提高 37% 和 31% 的最小资源复用率。这表明本文算法可提高资源分配能力的下限。

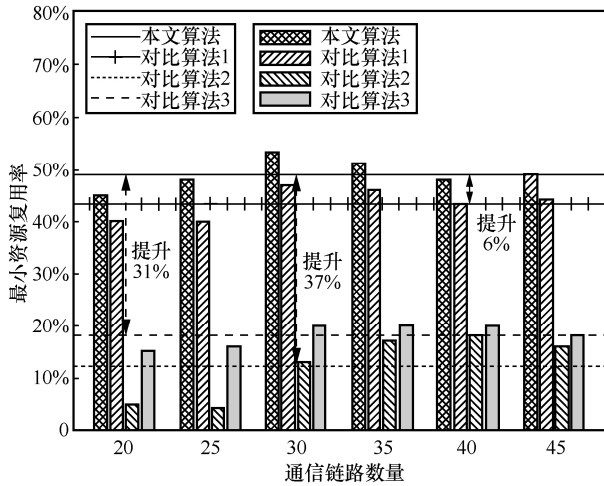


图 13 最小资源复用率

图 14 为本文算法和 3 种对比算法在不同通信链路数量的平均资源复用率。随着通信链路数量的增加，4 种算法得到的平均资源复用率处于平稳波动状态。其中，本文所提的基于 Dueling DQN 的资源分配算法对比算法 1 提高 7% 的平均资源复用率；相较于对比算法 2 和对比算法 3 分别提高 27% 和 26% 的平均资源复用率。这表明本文算法可有效提高系统资源分配能力。

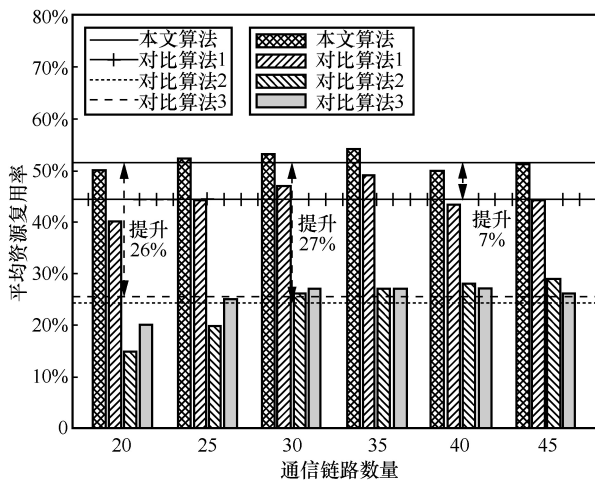


图 14 平均资源复用率

针对资源复用率，本文所提基于 Dueling DQN 的资源分配算法利用 MDP 能够实时掌握网络的运

作状态（为每一时刻的状态提供一个对应资源分配动作），找到一种满足实时性和资源效率要求的资源分配策略，增强资源分配能力，提高通信资源复用率。

4.6 仿真测试与实际测试对比

实际测试采用戴尔 (DELL) X86 服务器、美国国家仪器公司 (NI, national instruments) 通用软件无线电外设 N310 硬件平台和欧洲电信学院的开源空口 (OAI, open air interface)。根据 6G DN 相关文章中的仿真场景，搭建密集网络^[24-25]，并按照所提算法中的参数来配置实际密集网络系统的参数，如表 2 所示。

参数	数值
系统带宽/MHz	10
频点/GHz	3.5
发射功率/dBm	24
天线增益/dBi	10
信道模型	ITU 信道
IT 分布	室内随机分布
IT 连接密度/km ²	450 000
IT 速度/(km·h ⁻¹)	0

如图 15 所示，随着通信链路数量的增加，网络吞吐量整体都呈上升趋势，其中，本文算法仿真得到的网络吞吐量始终大于实际通信环境的测试结果。基准算法的实际测试结果明显低于本文算法的实际测试结果。当通信链路数量为 45 时，本文算法比基准算法最大提高 37% 的网络吞吐量；当通信链路数量为 30 时，本文算法比基准算法最小提高 17% 的网络吞吐量。通过部署基于 Dueling DQN 的资源分配算法可以降低交叠干扰，提高网络吞吐量。实际测试结果验证了本文算法的有效性。

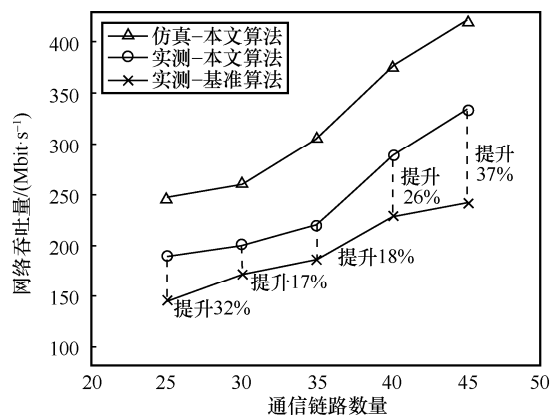


图 15 实际测试结果

5 结束语

为通过资源分配实现小区间无交叠干扰, 提高 6G DN 的网络性能, 本文讨论了 6G DN 中的资源分配问题, 但资源受限和节点密集分布导致难以通过传统的优化方法解决资源的有效分配问题。针对此问题, 本文提出了基于 Dueling DQN 的资源分配算法。该算法利用在线决策的能力弥补以牺牲存储空间来提高网络吞吐量的资源分配技术缺点, 比优化方法有更少的内存消耗。仿真实验结果表明, 所提算法可以有效提高网络吞吐量和资源复用率, 并且随着网络规模和密集程度的增加对方法的实时性和内存消耗的影响较小, 资源复用率处于稳定状态。在实际测试中, 本文算法比基准算法最大提高了 33% 的网络吞吐量和 37% 的资源复用率, 有效提升了网络性能。

参考文献:

- [1] 张平, 田辉. 由智生简、以简促智, 创构网络发展新生态——“面向 6G 的智能至简网络关键技术”专题导读[J]. 通信学报, 2022, 43(6): 1-2.
ZHANG P, TIAN H. From wisdom to simplicity, introduction to the special topic of “key technologies for 6G-oriented smart and simplified networks”: simplicity to wisdom, creating a new ecosystem for network development[J]. Journal on Communications, 2022, 43(6): 1-2.
- [2] SHAHJALAL M, KIM W, KHALID W, et al. Enabling technologies for AI empowered 6G massive radio access networks[J]. ICT Express, 2023, 9(3): 341-355.
- [3] CHO H, MUKHERJEE S, KIM D, et al. Facing to wireless network densification in 6G: challenges and opportunities[J]. ICT Express, 2023, 9(3): 517-524.
- [4] ALWARAFY A, ALBASEER A, CIFTLER B S, et al. AI-based radio resource allocation in support of the massive heterogeneity of 6G networks[C]//Proceedings of 2021 IEEE 4th 5G World Forum (5GWF). Piscataway: IEEE Press, 2021: 464-469.
- [5] LIN M T, ZHAO Y P. Artificial intelligence-empowered resource management for future wireless communications: a survey[J]. China Communications, 2020, 17(3): 58-77.
- [6] QIN M, YANG Q H, CHENG N, et al. Machine learning aided context-aware self-healing management for ultra dense networks with QoS provisions[J]. IEEE Transactions on Vehicular Technology, 2018, 67(12): 12339-12351.
- [7] SHAMAEI S, BAYAT S, HEMMATYAR A M A. Interference-aware resource allocation algorithm for D2D-enabled cellular networks using matching theory[J]. IEEE Transactions on Network and Service Management, 2023: doi.10.1109/TNSM.2023.3283993.
- [8] DAI Y P, SHENG M, LIU J Y, et al. Joint mode selection and resource allocation for D2D-enabled NOMA cellular networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(7): 6721-6733.
- [9] PENG T, GUO Y C, WANG Y C, et al. An interference-oriented 5G radio resource allocation framework for ultradense networks[J]. IEEE Internet of Things Journal, 2022, 9(22): 22618-22630.
- [10] TANG X, REN P Y, GAO F F, et al. Interference-aware resource competition toward power-efficient ultra-dense networks[J]. IEEE Transactions on Communications, 2017, 65(12): 5415-5428.
- [11] GORANTLA B V R, MEHTA N B. Interplay between interference-aware resource allocation algorithm design, CSI, and feedback in underlay D2D networks[J]. IEEE Transactions on Wireless Communications, 2022, 21(5): 3452-3463.
- [12] ALI Z, KHAN W U, IHSAN A, et al. Optimizing resource allocation for 6G NOMA-enabled cooperative vehicular networks[J]. IEEE Open Journal of Intelligent Transportation Systems, 2021, 2: 269-281.
- [13] LI Y, GAO Z, HUANG L, et al. Energy-aware interference management for ultra-dense multi-tier HetNets: architecture and technologies[J]. Computer Communications, 2018, 127: 30-35.
- [14] CELIK A, RADAYDEH R M, AL-QAHTANI F S, et al. Joint interference management and resource allocation for device-to-device (D2D) communications underlying downlink/uplink decoupled (DUDe) heterogeneous networks[C]//Proceedings of 2017 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2017: 1-6.
- [15] LIU Y M, YU F R, LI X, et al. Distributed self-optimizing interference management in ultra-dense networks with non-orthogonal multiple access[J]. Wireless Networks, 2020, 26(4): 2809-2823.
- [16] DU X, WANG T, FENG Q, et al. Multi-agent reinforcement learning for dynamic resource management in 6G in-X subnetworks[J]. IEEE Transactions on Wireless Communications, 2023, 22(3): 1900-1914.
- [17] 廖晓闽, 严少虎, 石嘉, 等. 基于深度强化学习的蜂窝网资源分配算法[J]. 通信学报, 2019, 40(2): 11-18.
LIAO X M, YAN S H, SHI J, et al. Deep reinforcement learning based resource allocation algorithm in cellular networks[J]. Journal on Communications, 2019, 40(2): 11-18.
- [18] FAN C J, ZENG L, SUN Y Z, et al. Finding key players in complex networks through deep reinforcement learning[J]. Nature Machine Intelligence, 2020, 2(6): 317-324.
- [19] CHEN J X, CAO X B, YANG P, et al. Deep reinforcement learning based resource allocation in multi-UAV-aided MEC networks[J]. IEEE Transactions on Communications, 2023, 71(1): 296-309.
- [20] CHUA T J, YU W H, ZHAO J. Resource allocation for mobile meta-verse with the Internet of vehicles over 6G wireless communications: a deep reinforcement learning approach[C]//Proceedings of 2022 IEEE 8th World Forum on Internet of Things (WF-IoT). Piscataway: IEEE Press, 2023: 1-7.
- [21] 汪东洋. 无线多跳网络中路由及相关技术研究[D]. 北京: 北京邮电大学, 2015.
WANG D Y. Research on routing and related technologies in wireless multi-hop networks[D]. Beijing: Beijing University of Posts and Telecommunications, 2015.
- [22] HUANG J, YANG F, CHAKRABORTY C, et al. Opportunistic capacity based resource allocation for 6G wireless systems with network slicing[J]. Future Generation Computer Systems, 2023, 140: 390-401.
- [23] 滑思忠, 丁爱玲, 郭达伟, 等. 基于二分图的 D2D 通信资源复用分配算法[J]. 计算机应用研究, 2017, 34(7): 2096-2098, 2103.
HUA S Z, DING A L, GUO D W, et al. Resource allocation algorithm in D2D cellular network based on bipartite graph[J]. Application Research of Computers, 2017, 34(7): 2096-2098, 2103.

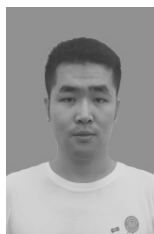
[24] PANEK M, POMYKAŁA A, JABŁOŃSKI I, et al. 5G/5G+ network management employing AI-based continuous deployment[J]. Applied Soft Computing, 2023, 134: 109984.

[25] BENLLOCH-CABALLERO P, WANG Q, CALERO J M A. Distributed dual-layer autonomous closed loops for self-protection of 5G/6G IoT networks from distributed denial of service attacks[J]. Computer Networks, 2023, 222: 109526.

[作者简介]



杨凡（1983- ），男，湖北广水人，博士，重庆理工大学副教授，主要研究方向为无线宽带自适应传输、无线通信网络、下一代移动通信技术、无线通信中的编码技术等。



张仕龙（1996- ），男，河北保定人，重庆理工大学硕士生，主要研究方向为无线通信资源分配、下一代移动通信等。



喻涛（1999- ），男，重庆人，重庆理工大学硕士生，主要研究方向为认知无线电和无线通信资源分配。



杨成（1999- ），男，重庆人，重庆理工大学硕士生，主要研究方向为无线宽带物理层、智能物理层技术等。



左迅（1998- ），男，重庆人，重庆理工大学硕士生，主要研究方向为下一代无线通信技术、无线自组网技术等。



黄杰（1988- ），男，重庆人，博士，重庆理工大学讲师，主要研究方向为无线通信理论、通信网络、下一代移动通信技术、认知无线电、无线通信资源分配等。



杨川（1999- ），男，重庆人，重庆理工大学硕士生，主要研究方向为无线通信网络、无线通信接入技术等。