

引用:李若宇,望运武,顾家骅,等. 基于深度强化学习的光无线融合接入网端到端网络切片映射[J]. 聊城大学学报(自然科学版),2025,38(5):669-678.

Citation:LI Ruoyu, WANG Yunwu, GU Jiahua, ZHU Min. End-to-end network slicing mapping for converged optical-wireless access networks using deep reinforcement learning[J]. Journal of Liaocheng University (Natural Science Edition), 2025, 38(5):669-678.

文章编号 1672-6634(2025)05-0669-10

DOI 10.19728/j.issn1672-6634.2025010004

基于深度强化学习的光无线融合接入网端到端网络切片映射

李若宇¹,望运武²,顾家骅²,朱敏²

(1. 东南大学 信息科学与工程学院,江苏 南京 210096; 2. 东南大学 移动通信国家重点实验室,江苏 南京 210096)

摘要 在光无线融合接入网中,光波长负责承载无线数据,其传输速率反过来影响计算资源和光带宽资源的分配。但是,光端和无线端的独立调度往往不够灵活,导致光端和无线资源的协作和利用效率低下。研究了光无线融合接入网端到端(E2E)网络切片映射问题。为了结合无线端用户需求和光端无线接入网(RAN)切片调度,建立了端到端网络切片映射模型,并提出了一种深度强化学习(DRL)方法。为了增强 DRL 代理的决策过程,提出了一种切片请求分解方案,将每个切片请求被分为两个子请求,最后通过模拟大规模的 33 节点网络来评估该算法的有效性。结果表明,提出的 DRL 优于启发式算法,在大规模网络中减少了 34.4% 的网络资源。

关键词 端到端光无线网络切片;无线接入网(RAN)切片;深度强化学习

中图分类号 TN929.1;TP18

文献标志码 A

开放科学(资源服务)标识码(OSID)



End-to-end network slicing mapping for converged optical-wireless access networks using deep reinforcement learning

LI Ruoyu¹, WANG Yunwu², GU Jiahua², ZHU Min²

(1. School of Information Science and Engineering, Southeast University, Nanjing 210096, China;

2. State Key Laboratory of Mobile Communications, Southeast University, Nanjing 210096, China)

Abstract In converged optical-wireless access networks, optical wavelengths are responsible for carrying wireless data, and their transmission rate in turn affects the allocation of computing and optical bandwidth resources. However, the independent scheduling for the optical and wireless network sides tends to be inflexible, resulting in inefficient cooperation and utilization of optical and wireless resources. In this paper, we investigate the end-to-end (E2E) optical-wireless network slicing mapping problem in converged optical-wireless access networks. To combine user requirements in the wireless side and radio access network

收稿日期:2025-01-07

基金项目:国家自然科学基金项目(62271135)资助

通信作者:朱敏,男,汉族,博士,教授,研究方向:光网络与光通信,E-mail: minzhu@seu.edu.cn.

(RAN) slicing scheduling in the optical side, we formulate an E2E network slicing mapping model and propose a deep reinforcement learning (DRL) method. To enhance the decision-making process of DRL agent, a slicing request decomposition scheme is proposed, in which each slicing request is divided into two sub-requests. We evaluate the effectiveness of the proposed algorithm through simulations on large-scale 33-node networks. The results demonstrate the superiority of our proposed DRL over the heuristic algorithm, achieving an 34.4% reduction in large-scale networks.

Keywords end-to-end optical-wireless network slicing; radio access network (RAN) slicing; deep reinforcement learning

0 引言

第三代和第四代移动网络通过无处不在的移动设备彻底改变了社会,而第五代(5th generation mobile communication technology, 5G)网络将通过万物互联来彻底改变我们的整个社会行业。例如,虚拟现实和增强现实可以带来沉浸式体验,在医疗和军事应用方面有很大的帮助,但它们具有较高的延迟要求和计算能力要求,所需带宽是高清视频的 100 倍^[1]。远程手术是一种先进的医疗技术,它允许医生通过高速互联网的连接在远离手术室的地方进行手术,但端到端延迟需要小于 300 ms,以确保安全。智慧城市将为人们提供更好的生活和服务,为企业创造更有利的商业发展环境。为了满足这些应用程序和服务的多样性和严格的需求,网络切片成为了下一代网络的开创性范例^[2]。通过软件定义的网络 (software-defined networking, SDN) 和网络功能虚拟化 (network function virtualization, NFV) 技术,能够在公共的物理基础设施上创建多个虚拟网络,也就是切片。每个切片都可以进行定制,以满足特定的性能指标,如延迟和带宽,从而确保最佳的服务交付。网络切片的灵活性和可伸缩性使网络运营商能够有效地进行管理和分配网络资源,显著改善了整体用户体验,提高了网络效率^[3]。

尽管网络切片具有许多优点,但在光无线融合接入网络中有效的资源分配仍然是一个重大的挑战。大多数现有的工作主要集中在无线接入网 (radio access network, RAN) 的切片上,无论是在无线网络^[4]还是在光网络^[5]中,通常都将这些部分视为独立的实体。这种孤立的方法导致了两个网络之间的非最优协调与资源合作。由于缺乏针对光资源和无线资源的集成调度机制,使得资源利用效率低下,网络性能下降。因此,开发一种高效的端到端 (end-to-end, E2E) 光无线网络切片映射算法至关重要,特别是对于流量不确定性和异构性显著的动态操作场景。开发这种最优策略的一个很有效的方法是使用机器学习技术,特别是深度强化学习 (deep reinforcement learning, DRL)^[6]。

近年来,基于 DRL 的算法由于优于传统的基于启发式的方法而获得了相当大的关注^[7]。DRL 在具有复杂决策过程特征的环境中特别有效,因为它通过与环境的持续交互来学习最优策略。这种对动态网络条件的适应性使 DRL 能够随着时间的推移不断提高其性能。它处理高维状态和动作空间的能力,再加上它持续学习的能力,使 DRL 成为比传统方法更优越的替代方法。通过这些能力, DRL 可以有效地解决光无线融合网络中资源分配的问题,从而提高整体网络效率和性能。

以往的工作研究了 DRL 算法在无线网络切片映射中的应用。文献[8]使用 DRL 针对网络切片场景解决了典型的资源管理问题,包括无线电资源切片和基于优先级的核心网络切片。最后的仿真验证了 DRL 相对于几种竞争算法的优势。文献[9]利用 DRL 开发了一种在多单元 RAN 中网络切片的链路级切片方案。与最先进的方法相比,他们的方法在不平衡的切片策略下实现了高吞吐量和完美的切片隔离。文献[10]探索了 fog-RAN 切片的实现,并开发了一种基于 DRL 的方法。与基准算法相比,他们的方法取得了显著的性能改进,并显示了对时变通道的鲁棒性。文献[11]提出了一种用于 RAN 切片的分层智能资源管理方法。仿真和实验结果表明,该方法的性能均优于基准算法。文献[12]提出了一种高效的 DRL 辅助的 RAN 切片资源分配方法。仿真结果表明,他们的方法在收敛速度、计算复杂度、能源效率、可接受的用户数量和层间隔离的程度等方面都优于最先进的已发表的方法。

最近的一些研究工作提出了基于 DRL 的光网络切片映射算法。文献[13]建立了一个整数线性规划

(integer linear programming, ILP)模型,并提出了一种基于深度确定性策略梯度(deep deterministic policy gradient, DDPG)的方法来解决 RAN 函数的放置和路由问题。仿真结果表明,与 ILP 模型得到的最优结果相比,该方法获得了令人满意的网络性能。文献[14]建立了一个 ILP 模型,并提出了一种基于双深度 Q 网络(double deep Q network, DDQN)的方法来解决 DU-CU 放置和光路由中的决策问题。仿真结果表明,该方法在离线和在线情况下均能有效收敛,优于首次拟合方法。文献[15]引入了一个 ILP 模型和一个基于近似策略优化(proximal policy optimization, PPO)的方法,以解决网络中 DU-CU 的放置和路由问题。仿真结果表明,该方法在集中基带函数处理和节省光传输带宽方面取得了接近最优的性能。文献[16]在弹性光网络(elastic optical networks, EONs)中研究了细粒度的随机函数放置和路由问题。提出了一种基于 DRL 的算法来获得接近最优的解,仿真结果证明了该方法获得了更高的资源效率。以上的研究工作分别体现了 DRL 算法在无线端和光端的应用,但并没有考虑两个网络之间的协调与资源合作。因此开发了一种新的 DRL 算法,将无线网络与光网络的切片调度相结合,提供了一种整体的资源分配方法。

除了 DRL 算法的应用外,还对 RAN 的体系架构进行了拆分和优化。为了支持各种应用程序和服务的多样化和严格的需求,RAN 已经从集中式 RAN(cloud RAN, C-RAN)发展为下一代 RAN(next-generation RAN, NG-RAN)。在 NG-RAN 体系结构中,基带单元(baseband unit, BBU)功能解耦为分布式单元(distributed unit, DU)和集中式单元(centralized unit, CU),创建了一个三层网络体系结构:有源天线单元(active antenna unit, AAU)DU-CU。实时任务被分配给 DU,而非实时任务则由 CU 来处理。

近年来,5G RAN 的部署优化引起了广泛的研究。在 BBU 架构方面,文献[17]关注了具有无线前馈链路的 C-RAN,达到了低运营支出成本和不同场景下灵活部署网络的目的。对于 DU-CU 架构,文献[18]考虑到 DU-CU 池和链路的成本,使用混合整数线性规划模型优化了 DU-CU 的部署。文献[19]侧重于执行隔离感知 RAN 映射的新解决方案,提出了一种启发式算法,最终提高了带宽资源利用率。

利用了一种基于 DRL 的算法,研究了光无线网络切片映射问题。为了为现实应用提供可行的解决方案,提出了一种基于 DRL 的算法来最小化目标函数,并提出一种请求分解方案,将光无线网络切片请求分解为两个子请求,并按顺序部署。文章将无线网络中的用户数据速率要求与光网络中的 RAN 切片调度相结合,提供了一种整体的资源分配方法。仿真结果表明,与启发式方法相比,该方法显著提高了网络资源利用率,在大规模网络场景中减少了 34.4% 的网络资源。

1 系统模型与问题建立

1.1 光无线接入网模型

图 1 展示了一个从终端用户到 5G 核心网络的基底光无线融合接入网络架构。设发射端的天线数量为 M ,接收端装备的天线数量为 Q ,在不失一般性的前提下,假设发射机不知道信道状态信息(channel state information, CSI),而接收机知道 CSI。此时用户终端发射的信号矢量表示为 $E[xx^H] = \frac{P}{M}I_M$,其中 P 表示发射端的发射功率,则接受到的信号为 $y = Hx + z$,其中 $H \in \mathbb{C}^{Q \times M}$ 表示从发射端到接收端的信道矩阵, $z \sim \text{CN}(0, \sigma^2 I_Q)$ 表示接收机处的加性高斯噪声矢量。

这里的信道矩阵 H 被定义为 $H = \rho^{\frac{1}{2}} \times \tilde{H}$,其中 ρ 表示从发射机到接收机的路径损耗, $\tilde{H} \in \mathbb{C}^{Q \times M}$ 表示信道的小尺度衰落系数。这里考虑平坦非相关瑞利衰落信道,即 $\tilde{H}$ 中的变量是独立且同分布的复高斯随机变量,平均值为零,方差为 1。从文献[20]中可知,MIMO 通信系统的可实现速率为 $R = B \log_2 \det \left(I_Q + \frac{\gamma}{M} \tilde{H} \tilde{H}^* \right)$,其中 $\det(A)$ 表示矩阵 A 的行列式, I_Q 表示维度为 $Q \times Q$ 的单位矩阵, $*$ 表示转置共轭算子, $\gamma = \rho P / \sigma^2$,发射功率为 P , ρ 表示路径损耗,也就是发射器和接收器之间的阴影效应, σ^2 表示加性高斯白噪声的功率。带宽为 B ,它可以被拆分为无线电资源块 RB,它们之间的关系为 $B = 180 \text{ KHz} \times R_B$ 。从文献[20]中可知,当服务类型为 eMBB 时,该系统存在一个最小速率限制。设这个值为 r_0 ,因此有

$$B \log_2 \det \left(\mathbf{I}_Q + \frac{\gamma}{M} \tilde{\mathbf{H}} \tilde{\mathbf{H}}^* \right) \geq r_0, \quad (1)$$

将 Q 设置为 4, 发射端天线数量 M 设置为 16, 由公式(1)可以得到每个请求所需的最小 RB 数量。而每个请求所需的最小 RB 数量, 影响了光网络中计算资源的占用情况。

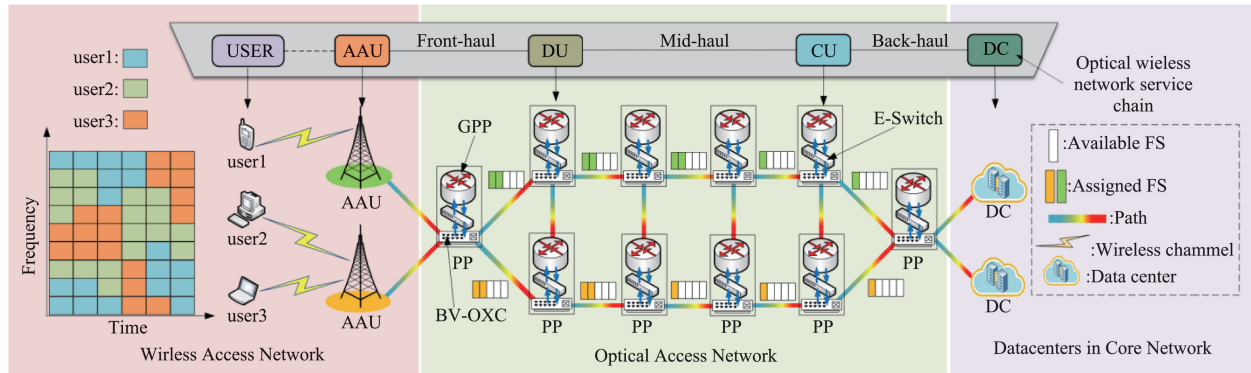


图 1 光无线融合接入网络架构

Fig. 1 The converged optical-wireless access network architecture.

1.2 RAN 功能分割模型

采用了一种具有功能分裂的高级 RAN 体系结构, 如图 1 所示。用户的信号通过天线向上传输到 AAU, 后者负责信号的模拟/数字转换和功率放大, 然后将信号传输到 BBU。当 BBU 被分为 DU 和 CU 时, DU 负责提供分组数据聚合协议层(packet data converged protocol layer, PDCP)以下的大带宽和延迟敏感的 BBU 功能, 而 CU 负责提供 PDCP 以上的功能。每个 DU 或 CU 都被放置在处理池(processing pool, PP)中。处理完成后, 信号传输到数据中心。

每个 DU 和 CU 都有其所需的计算能力, 计算能力由资源块 R_B 、天线数 Q 、MIMO 层数 L 、调制位 A 、发射端天线数量 M 和编码率 C 决定, 数字基带处理的模型是基于每秒千兆次运算(giga operations per second, GOPS)的估计复杂性, 由文献[21]可知, 每个 DU 或 CU 需要占用的计算资源为

$$C [\text{GOPS}] = \alpha_i \times \left(3A + A^2 + \frac{1}{3} M \times C \times L \right) \times \frac{R_B}{5}, \quad (2)$$

其中 α_i 是在文献[21]中定义的一个因子, DU 与 CU 的 α_i 不相同。

除了计算资源的约束之外, 延迟也是一个需要考虑的约束条件。从文献[21]可以得出, 节点之间的光纤链路的传输延迟为 $5 \mu\text{s}/\text{km}$ 。信号在每个 DU 和 CU 中的延迟为 $20 \mu\text{s}$, 如果 DU 和 CU 恰好放置在同一 PP 中, 则两者的总延迟为 $20 \mu\text{s}$ 。从 AAU 到 DU 的总链路延迟不大于 $100 \mu\text{s}$, 从 AAU 到数据中心 DC 的总延迟不大于 $1000 \mu\text{s}$ 。

除了计算容量和延迟外, 还需要考虑带宽限制。节点间链路中的带宽是固定的, 可以划分为多个相同的频槽(frequency slot, FS)。AAU 到 DU、DU 到 CU、CU 到 DC 所需的 FS 数量有所不同。从文献[21]中, 可以得到带宽需求的计算方法。一旦确定了带宽需求, 根据文献[22], 所需的 FS 的数量可以计算为 $F_i = \lceil \frac{B_i}{m \times B_{\text{FS}}} \rceil + N_g$, 其中 i 表示前传, 中传或者后传, $\lceil \cdot \rceil$ 表示向下取整取最小整数的运算, m 表示调制格式的级别, 并且对于二进制相移键控(BPSK)来说, $m = 1$ 。 B_{FS} 表示每个 FS 的带宽, N_g 表示保护间隔的频率隙。

1.3 仿真评估目标

考虑具有有限物理资源的光无线融合接入网, 包括无线电 RB、计算资源和带宽资源。在同一个时刻, 大量的光无线网络切片请求到达网络, 每个请求都有特定的数据速率要求、所需的 RB 数量要求、所需的计算和带宽资源, 以及光无线延迟约束。利用网络切片映射的特点, 所提出的 DRL 算法旨在找到一个最优的光无线网络切片映射策略, 以最小化网络中计算资源和频谱资源利用率, 同时确保满足光无线延迟约束。

为了评估服务提供算法的性能,定义了5个指标。(1) 目标函数:这里的目标函数是DRL算法进行不断学习优化的目标,与网络中计算资源和频谱资源的占用情况相关,定义为 $\eta_{ob} = \frac{Z}{N_{pp}} + \frac{F_{mfsi}}{F_i}$, 其中 Z 表示所有请求部署完后,网络中被占用过的PP的数量,由于PP池中提供计算资源,因此用占用的PP数量来代表计算资源。 N_{pp} 表示网络中除了数据中心节点外其它节点的数量, F_{mfsi} 表示在所有链路中使用的频率块的最大索引值,也就是被占用频率块最多的链路的最后一个被占用的频率块的索引值。 F_i 表示一个链路中的频率块的总数。(2) 光无线网络延迟:对于所有部署成功的请求,光无线网络延迟可根据1.2小节中的延迟模型来计算,最后将所有的延迟相加,获得部署成功的请求的总延迟,定义为 $D_{de} = \sum_{u \in \{U-U_{block}\}} D_u$, 其中, U_{block} 表示所有部署失败的请求的集合, U 表示所有请求的集合。(3) 无线电RB资源:由于不同的终端用户请求消耗不同数量的无线电RB,将所有部署成功的请求的RB相加,因此无线电RB占用情况被定义为 $\eta_R = \sum_{u \in \{U-U_{block}\}} N_u$ 。(4) 处理池使用率:当所有请求部署完成后,统计网络中的PP处理池占用情况,将所有被占用过的处理池相加, Z 表示网络中被占用过的PP的数量,可得 $\eta_C = Z$ 。(5) 链路中FS使用率:当所有请求部署完成后,统计网络中的所有链路中使用的频率块的最大索引值,可得 $\eta_s = F_{mfsi}$ 。

2 基于DRL的算法

本节提出了一种基于DRL的算法来解决光无线融合接入网络中的E2E光无线网络切片映射问题。首先描述了将复杂的E2E光无线网络切片请求划分为两个子请求的分解方案。然后,提出了基于DRL的算法的建模方法,包括状态表示、动作定义和奖励描述。最后,利用训练算法的伪代码详细解释了训练过程。

2.1 光无线网络切片请求分解方案

在DRL框架中,每个步骤都执行一个动作,这就需要执行多个步骤来提供一个选择性无线网络切片请求。为了方便DRL代理的决策过程,将光无线网络切片请求分解为两个子请求,并按顺序部署。具体地说,第一子请求包括前传带宽和DU的计算资源需求,而第二子请求包括CU的计算资源需求和中、回传带宽需求。这种两步分解不仅简化了部署过程,而且提高了DRL代理管理和完成网络切片请求的效率。

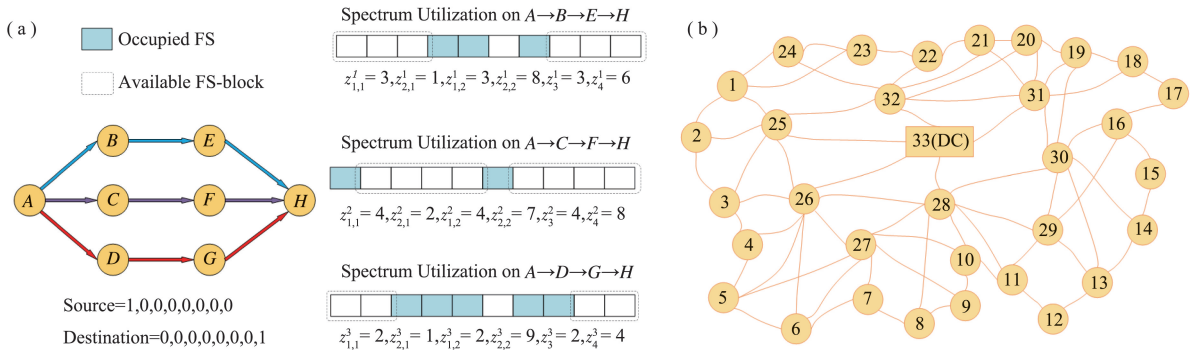


图2 (a) 基于DRL的算法中状态表示的一个示例; (b) 大规模33节点网络拓扑

Fig. 2 (a) An illustrative example of state representation in the algorithm;

(b) Simulation topology for large-scale 33-node network

2.2 DRL模型

状态表示:状态表示DRL代理从网络中收集到的环境信息,整合各种功能,提供对网络环境的全面感知,并促进最佳决策。状态 s_t 是一个 $1 \times [(2 \times |V| + (2 \times J + 2) \times K)]$ 数组。该阵列捕获在 K 个最短候选路径上的当前请求、服务持续时间和频谱利用率的信息,每个候选路径都有 J 个可用的FS频率块。其中使用的 K 的一个比较小的值(例如, $K=3$),而不是枚举所有可能的候选最短路径,这是因为使用过长的路由其路径的增益非常有限。状态数组 s_t 被定义为

$$s_t = [s_u, d_u, \{\{z_{1,j}^k, z_{2,j}^k\} \mid j \in [1, J], z_3^k, z_4^k\} \mid k \in [1, K]\}, \quad (3)$$

其中,使用 $2 \cdot |V|$ 个数来表示 s_u, d_u 。此外, $z_{1,j}^k, z_{2,j}^k, z_3^k$ 和 z_4^k 分别表示第 j 个可用 FS 块的大小、第 j 个可用 FS 块的起始 FS 索引、第 j 个可用 FS 块的平均大小以及第 k 个路径的光纤链路上可用 FS 块的总大小。如果 s_u 和 d_u 之间的候选路径的数量小于 K , 则补充空路径, 以确保总数达到 k 。对于这些空路径, 将 $\{z_{1,j}^k, z_{2,j}^k \mid j \in [1, J], z_3^k, z_4^k\}$ 的值替换为 -1 来保持状态格式的一致性。这种方法从每个候选路径中提取关键信息, 使 DRL 代理能够选择最佳动作。图 2(a) 显示了在基于 DRL 的算法中构造 s_t 的过程。为简单起见, 假设从节点 A 到节点 H 需要 2 个 FS。在每个路径上获得 2 个可用的 FS 块(用虚线框标记)。例如, 路径 $A \rightarrow B \rightarrow E \rightarrow H$ 上的第一个可用 FS 块($j=1$)的大小为 $z_1^1=3$, 起始 FS 索引为 $z_2^1=1$ 。此外, 可用的 FS 块的平均大小为 $z_3^1=3$, 而该路径上的可用的 FS 块的总数为 $z_4^1=6$ 。

动作的定义: 采用一种启发式算法来构造动作空间。首先, 使用 KSP(最短路径算法, 即找到两点之间最短的 n 条路径, n 可以自己定义) 算法得到源节点和目标节点之间的 K 个最短候选路径, 并将这些路径上的所有 PP 节点添加到节点集 A_u 中。接下来, 根据每个 PP 剩余的可获得计算资源将 A_u 按降序排序。为了进一步加速训练过程, 使用了一个掩蔽机制^[23], 它检查每个动作是否满足所有的约束(例如, 计算资源约束和延迟约束)。这种机制消除了不合理的动作, 从而减少了候选动作的数量, 有效地缩小了动作空间。最后, 选择第一个 $|A|$ PP 来形成动作空间 A , 其中 $|A|$ 是一个决定可选动作集大小的超参数。当 DU(或 CU) 成功映射到一个 PP 节点时, 使用最短路径连接源节点和所选的 PP 节点, 而 FS 分配决策则使用第一拟合(FF)原则进行。

奖励的描述: 奖励功能在 DRL 中是必不可少的, 因为它指导了 DRL 代理的学习过程。在执行一个动作后, DRL 代理会立即收到一个奖励 r_t 。基于 DRL 的算法的目标是最小化目标函数。因此, 在一个回合结束时, r_t 设置为网络中目标函数值的相反数; 否则, r_t 设置为 0。这种奖励结构鼓励 DRL 代理优化请求部署完成后的网络中计算资源和频谱资源的利用率。奖励功能被表示为

$$r_t = \begin{cases} -\left(\frac{Z}{N_{pp}} + \frac{F_{mfsi}}{F_i}\right), & \text{当一个回合结束时,} \\ 0, & \text{其他情况,} \end{cases} \quad (4)$$

采用策略梯度(policy gradient, PG)方法进行训练, 利用累积计算奖励来更新策略。在一个回合结束时, 计算总奖励, 然后在所采取的动作序列中进行计算。这种计算奖励用于更新政策, 根据其对最终奖励的贡献为每个行动分配奖励信息。由于 DRL 代理是经过迭代训练的, 所以一个回合被定义为 DRL 代理和环境之间的一系列交互, 从一个初始状态开始, 并涉及到一系列的动作。当达到终端状态或超过最大步骤数时, 这一个回合就结束了。具体来说, 一个回合是指所有光无线网络切片请求在光无线融合接入网中部署完成的过程。这种学习结构使 DRL 代理能够从请求的整个周期中学习, 从而提高长期奖励并提高网络性能。

2.3 训练过程

算法 1 中详细介绍了基于 DRL 的算法的伪代码。在第 1, 2 行中, 每次迭代都初始化策略网络。为了确保全面的训练, 执行 EP 过程来收集不同的环境-代理交互样本, 并初始化 l_e (第 3, 4 行), 其中 l_e 跟踪训练步骤并对存储的数据进行索引化。首先, 所有到达的请求都被添加到集合 Q 中。使用不等式 $R \geq R^{\text{req}}$ 可以确定每个请求所需的最小无线电 RB 数量, 并使用 FF 原则分配这些 RB。随后, 集合 Q 中的每个请求被分解为两个子请求(第 6 行)。对于每一个子请求 $q \in Q$, 找到连接 s_q 和 d_q 的 K 条最短候选路径, 使用公式(3)来构造状态 $s_{l_e}^q$, 将该状态输入给策略网络, 进而获取动作 $a_{l_e}^q$ (第 8 行)。如果所选 PP $a_{l_e}^q$ 有资格部署 DU(或 CU), 那么就在服务期间部署 DU(或 CU)并为前传(中传和后传)提供最短路径, 而 FS 分配决策使用 FF 原则(第 9, 10 行)完成。如果 DU(或 CU)部署失败, 将撤销当前请求 q (第 11, 12 行)的无线电 RB 分配和 DU(或 CU)部署。然后根据公式(4)计算奖励 $r_{l_e}^q$, 并将元组 $(s_{l_e}^q, a_{l_e}^q, r_{l_e}^q)$ 存储在内存中(第 14 行)。在处理了所有的 E2E 光无线网络切片请求后, 使用存储的元组来更新策略网络 θ 。一旦 EP 过程完成, 从内存中的训练数据中计算返回数据和基准值(第 18, 19 行)。其中, γ 是折扣因子, 在 0 到 1 之间; r_k^e 是在时间步 k 获得的即时奖励。而基准值的计算是取为当前批次样本的回报平均值, 其中 EP 为回合的总数量, h_i^e 是第

e 回合的返回值。每一个状态-动作对的返回值是从一开始计算的奖励,而基准值,即对预期返回的估计,减少了梯度估计的方差。对于每一个回合 e ,使用梯度下降(第 20,21 行)更新策略网络参数 $\Delta\theta$,其中 α 和 γ 分别表示学习率和贴现率, $\nabla_{\theta} \log \pi_{\theta}(s_l^e, a_l^e)(h_l^e - b_l)$ 表示策略梯度估计值。该更新包括计算策略目标与网络参数相关的梯度,并调整它们以使预期返回最大化。最后,第 24 行将此更新应用于策略网络 θ 。通过迭代更新,DRL 代理通过基于在多个事件中积累的经验改进其策略来逐步提高其性能。

算法 1 基于 DRL 的网络资源配置训练算法

输入:1) 请求集 U ,其中包括无线电 RB、计算资源、带宽资源、前端传输和光无线延迟要求;2) 网络拓扑、PP 集合 V 、FS 集合 F 和无线电 RB 集合 M ;3) 每个 PP 和链路的容量。

输出:目标函数。

```

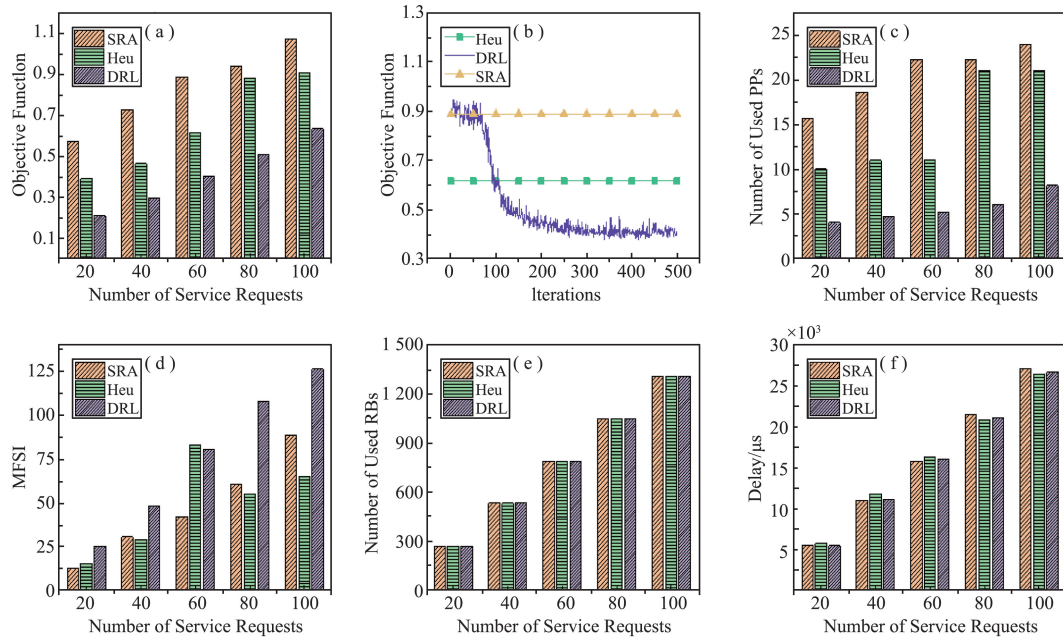
1 for 每一个循环:
2    $\Delta\theta \leftarrow 0$ ;
3   for 每一个回合:
4      $l_e \leftarrow 0, Q \leftarrow []$ ;
5     所有到达的请求都被添加到集合  $Q$  中。
6     根据等式(2),确定请求  $u$  所要求的无线电 RB 的最小数量,并使用 FF 原理分配这些无线电 RB。随后,集合
     $Q$  中的每个请求  $u$  被
      分解为两个子请求。
7     for 每个子请求:
8       找到连接  $s_q$  和  $d_q$  的  $K$  条最短候选路径,构造状态  $s_{l_e}^q$ ,将其输入策略网络,得到动作  $a_{l_e}^q$ 。
9       if 动作  $a_{l_e}^q$  可以部署 DU(或 CU) then
10        部署 DU(或 CU),并沿最短路径为前传(或中传后传)提供光链路。
11      else
12        阻止当前的子请求  $q$ ,并撤销请求  $q$  对集合  $V$ 、 $F$  和  $M$  所做的更改。
13      end if
14      获得奖励  $r_{l_e}^q$ ,存储元组  $(s_{l_e}^q, a_{l_e}^q, r_{l_e}^q)$ ,然后设置  $l_e \leftarrow l_e + 1$ 。
15    end for
16  end for
17  for 每一个  $l = 1$  到  $\{l_1, l_2, \dots, l_E\}$  的最大值:
18    计算返回值:  $h_l^e = \sum_{k=l}^{l_e} \gamma^{k-l} r_k^e$ ;
19    计算基准值:  $b_l = \frac{1}{EP} \sum_{e=1}^E h_l^e$ ;
20    for 每一个回合:
21       $\Delta\theta = \Delta\theta + \alpha \cdot \nabla_{\theta} \log \pi_{\theta}(s_l^e, a_l^e)(h_l^e - b_l)$ ;
22    end for
23  end for
24   $\theta \leftarrow \theta + \Delta\theta$ ;
25 end for

```

3 仿真性能评估与分析

在本节,给出仿真结果来评估所提出的基于 DRL 的算法的网络性能。并将其性能与两种算法进行了比较:启发式算法(Heu)和最短路径随机算法(SRA)。对于每一个请求,根据不等式 $R \geq r_0$,确定请求所要求的无线电 RB 的最小数量,并使用 FF 原则进行分配。启发式算法首先选择最短路径,然后将 FF 原理应用于 DU(或 CU)映射。如果在最短路径上没有找到可用的 PP 节点,则启发式算法将进入下一个最短路

径,以此类推。此外,SRA 算法是沿着从 AAU 到 DC 的最短路径随机选择满足约束条件的节点为 DU 和 CU。



注:(a) 目标函数;(b) 目标函数迭代;(c) 使用的 PP 数量;(d) MFSI;(e) 使用的 RB 数量;(f) 光网络延迟。

图 3 大规模 33 节点网络仿真结果

Fig. 3 Simulation results in the 30-node network

3.1 仿真环境建立

为了评估所提出的基于 DRL 的算法的网络性能,使用了一种网络场景:大规模 33 节点网络拓扑,如图 2(b)所示。在这些拓扑结构中,PP 节点之间的距离为 10~30 km,而基站与其直接连接的接入 PP 节点之间的距离为 0~3 km。对于无线部分,每个基站有 4 个天线和一个 100 MHz 的无线频谱(即 500RB,其中每个 RB 的频率范围为 180 kHz,有 12 个子载波)。对于大规模的网络,每个 PP 节点的计算容量都设置为 40 000(GOPS),每个光纤链路的容量为 358 FSS^[24],即频率块的数量为 358。所有 E2E 光无线网络切片请求的源节点都在给定的网络拓扑中随机生成,而其目标节点是相同的数据中心节点。因此,每个 E2E 光无线网络切片请求都链接到网络拓扑。对于 DRL 训练,学习率 α 和贴现率 γ 分别设置为 0.001 和 0.99。候选 PP 集合的大小($|A|$)、K-最短路径数(K)和事件数(EP)分别设置为 5、3 和 5。

3.2 大规模网络场景的结果分析

目标函数:在图 3(a)中,可以看出提出的基于 DRL 的算法优于现有的 SRA 和启发式算法。与严格遵循固定策略来调度动态切片请求的启发式方法不同,基于 DRL 的方法可以灵活地适应网络状态的变化。启发式算法优于 SRA 算法,因为它在选择节点时,总是选择剩余计算容量最大的节点,而 SRA 算法是随机选择路径上的节点,相对来说资源利用率不高。

图 3(b)展示了基于 DRL 的算法在 33 节点网络中的训练过程,体现了在请求数量为 60 的情况下的目标函数值的变化。最初,由于随机探索,资源利用率不高,但随着算法的学习,目标函数值迅速降低,最终收敛到 0.4。这证明了基于 DRL 的算法具有较强的收敛性能。在迭代 100 次时,DRL 方法在目标函数方面超过了启发式算法,在 300 次迭代后训练曲线稳定。与启发式算法相比,提出的基于 DRL 的算法可以将目标函数值降低 34.4%。

图 3(c)显示了 33 节点网络中请求部署完成后占用的 PP 的数量。与启发式和 SRA 算法相比,基于 DRL 的算法在各种请求数量下实现了更低的 PP 池占用数量。这是因为在选择节点时,使用的节点集是按照剩余计算容量排序的。启发式算法与 SRA 算法相比,在所有请求数量下占用的 PP 池数量都更少,这是因为 SRA 算法是随机选择路径上的节点,相对来说 PP 池的利用效率不高。

图 3(d)为在一个 33 节点网络计算的 MFSI 结果。随着请求数量的增加,由于对频谱资源的需求不断

增加,几乎所有三种算法的 MFSI 都在增加。而 DRL 算法在几乎所有请求数量下的 MFSI 都是最大,但由于使用的 PP 数量最少,因此总的目标函数值达到最小,DRL 的训练有效实现了 PP 池和 MFSI 资源之间的权衡。当请求数量为 60 时,启发式算法的 MFSI 高于另外两种算法,这是因为此时启发式算法占用的 PP 数量较少,相对来说目标函数的另一个指标 MFSI 就会增大。

图 3(e)显示了 33 节点网络的使用的 R_B 数量的结果。随着请求数量的增加,所有三种算法使用的 R_B 数量都在增加,这是因为每个请求都需要占用一定量的 R_B ,请求数量越多,占用 R_B 数量就越多。而在请求数量一定的情况下,三种算法占用的 R_B 数量一样多,这是因为无论请求数量为多少,所有的请求最终都被处理成功,因此无论使用哪种算法,最终占用的 R_B 数量都是一样的。

在图 3(f)显示了在一个 33 节点网络延迟的结果。它出现了一个这样的趋势:随着请求数量的增加,三种算法的延迟都在递增,这是因为延迟与占用节点和使用的链路长度有关。而请求数量相同时,三种算法延迟量相差不大,因为迭代训练的奖励与延迟无关,而且延迟主要与路径长度有关,使用最短路径算法后,选择的路径长度差别不大。

4 结论

提出了一种基于 DRL 的算法来解决光无线融合接入网络中的光无线网络切片映射问题。引入了一个计算 E2E 光无线网络切片的计算资源、带宽资源和延迟的数学模型。详细阐述了部署规则和约束条件。在基于 DRL 的方法中,提出了一种切片请求分解方案来增强 DRL 代理的决策过程。具体来说,每个切片请求被分为两个子请求,它们是按顺序被提供的。大量的仿真证明了所提出的 DRL 算法的优越性,与启发式算法相比,在大规模 33 节点网络中当请求数量为 60 时,目标函数值降低了 34.4%。

参 考 文 献

- [1] ALMEIDA G M, PINTO L D L, CARDOSO K V, et al. Optimal joint functional split and network function placement in virtualized ran with splittable flows[J]. IEEE Wireless Communications Letters, 2022, 11: 1684-1688.
- [2] ZHOU X, LI R, CHEN T, et al. Network slicing as a service: enabling enterprises, own software-defined cellular networks[J]. IEEE Communications Magazine, 2016, 54(7): 146-153.
- [3] HUA Y, LI R, ZHAO Z, et al. GAN-powered deep distributional reinforcement learning for resource management in network slicing[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(2): 334-349.
- [4] FOUKAS X, PATOUNAS G, ELMOKASHFI A, et al. Network slicing in 5G: survey and challenges[J]. IEEE Communications Magazine, 2017, 55(5): 94-100.
- [5] WANG Y, KONG L, ZHU M, et al. Dynamic RAN slicing mapping under the sliding scheduled traffic model based on deep reinforcement learning in elastic optical networks[J]. Journal of Lightwave Technology, 2024, 42(8): 2636-2647.
- [6] WANG Z, WEI Y, HAN Z, et al. Utility optimization for resource allocation in multi-access edge network slicing: a twin-actor deep deterministic policy gradient approach[J]. IEEE Transactions on Wireless Communications, 2022, 21(8): 5842-5856.
- [7] ZHU M, WANG Y, GU J, et al. DRL-assisted joint allocation of antenna, radio, and front-haul resources in TWDM-PON-based NG-RANs with mMIMO-enabled beamforming[J]. Journal of Optical Communications and Networking, 2023, 15(5): 241-254.
- [8] LI R, ZHAO Z, SUN Q, et al. Deep reinforcement learning for resource management in network slicing[J]. IEEE Access, 2018, 6: 74429-74441.
- [9] WANG T, CHEN S, ZHU Y, et al. Linkslice: fine-grained network slice enforcement based on deep reinforcement learning[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(8): 2378-2394.
- [10] XIANG H, YAN S, PENG M. A realization of fog-RAN slicing via deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2020, 19(4): 2515-2527.
- [11] LI J, WU J, ZHANG C, et al. Hierarchical intelligent radio access network slicing for differential service level agreement guaranteeing[J]. IEEE Transactions on Industrial Informatics, 2024, 20(3): 4124-4136.
- [12] AZIMI Y, YOUSEFI S, KALBKHANI H, et al. Energy-efficient deep reinforcement learning assisted resource allocation for 5G-RAN slicing[J]. IEEE Transactions on Vehicular Technology, 2022, 71(1): 856-871.
- [13] ZHU P, ZHANG J, XIAO Y, et al. Deep reinforcement learning-based radio function deployment for secure and resource-efficient NG-RAN slicing[J]. Engineering Applications of Artificial Intelligence, 2021, 106(10): 4490.

- [14] GAO Z, YAN S, ZHANG J, et al. Deep reinforcement learning-based policy for baseband function placement and routing of RAN in 5G and beyond[J]. *Journal of Lightwave Technology*, 2022, 40(2): 470-480.
- [15] WANG R, ZHANG J, GU Z, et al. Edge-enhanced graph neural network for DU-CU placement and lightpath provision in X-Haul networks[J]. *Journal of Optical Communications and Networking*, 2022, 14(10): 828-839.
- [16] WANG Y, ZHU M, CAI X, et al. DRL-assisted fine-grained function placement and routing of 5G RAN slice with reuse scheme in elastic optical networks[C]// 2023 IEEE International Conference on Communications. Rome, Italy: IEEE, 2023: 1958-1963.
- [17] WANG D, WANG Y, MENG S, et al. Game based wireless fronthaul C-RAN baseband function splitting and placement[C]// 12th International Conference on Mobile Ad-Hoc and Sensor Networks. Hefei, China: IEEE, 2016: 400-404.
- [18] AROUK O, TURLETTI T, NIKAEIN N, et al. Cost optimization of cloud-RAN planning and provisioning for 5G networks[C]// IEEE International Conference on Communications. Kansas City, MO, USA: IEEE, 2018: 1-6.
- [19] YU H, MUSUMECI F, ZHANG J, et al. Isolation-aware 5G RAN slice mapping over WDM metro-aggregation networks[J]. *Journal of Lightwave Technology*, 2020, 38(6): 1125-1137.
- [20] YAN D, NG B. K, KE W, et al. Deep reinforcement learning based resource allocation for network slicing with massive MIMO[J]. *IEEE Access*, 2023, 11: 75899-75911.
- [21] XIAO Y, ZHANG J, JI Y. Can fine-grained functional split benefit to the converged optical-wireless access networks in 5G and beyond? [J]. *IEEE Transactions on Network and Service Management*, 2020, 99: 1-1.
- [22] CHEN X, LI B, PROIETTI R. DeepRSA: a deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks[J]. *Journal of Lightwave Technology*, 2019, 37(16): 4155-4163.
- [23] SHIMODA M, TANAKA T. Mask RSA: end-to-end reinforcement learning-based routing and spectrum assignment in elastic optical networks[C]// European Conference on Optical Communication. Bordeaux, France: IEEE, 2021.
- [24] FENG W, ZENG M, LIU X, et al. Joint spectrum and IT resource allocation for efficient VNF service chaining in inter-datacenter elastic optical networks[J]. *IEEE Communications Letters*, 2016, 20(8): 1539-1542.

(责任编辑:赵海军)

(上接第 662 页)

- [10] MEHRDOUST F, NOORANI I, XU W. Uncertain energy model for electricity and gas futures with application in spark-spread option price[J]. *Fuzzy Optimization and Decision Making*, 2023, 22(1): 123-148.
- [11] LI H X, YANG X F, NI Y D. Pricing of shout option in uncertain financial market[J]. *Fuzzy Optimization and Decision Making*, 2024, 23(3): 449-467.
- [12] LIO W J, LIU B D. Residual and confidence interval for uncertain regression model with imprecise observations[J]. *Journal of Intelligent & Fuzzy Systems*, 2018, 35(2): 2573-2583.
- [13] YE T Q, LIU B D. Uncertain hypothesis test for uncertain differential equations[J]. *Fuzzy Optimization and Decision Making*, 2023, 22(2): 195-211.
- [14] 朱佳璇, 刘利英. 基于不确定时间序列的外汇储备智能预测[J]. *聊城大学学报(自然科学版)*, 2024, 155(5): 41-47.
- [15] KE H, YAO K. Block replacement policy with uncertain lifetimes[J]. *Reliability Engineering & System Safety*, 2016, 148: 119-124.
- [16] SAMUEL C, CHOOKAH M, MOHAMMAD M. Development of a probabilistic mechanistic model for reliability assessment of gas cylinders in compressed natural gas vehicles[J]. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2009, 223(4): 289-299.
- [17] TERJE A, KROHN B S. A new perspective on how to understand, assess and manage risk and the unforeseen[J]. *Reliability Engineering & System Safety*, 2014, 121: 1-10.
- [18] DER A, KIUREGHIAN A, DITLEVSEN O. Aleatory or epistemic? Does it matter? [J]. *Structural safety*, 2009, 31(2): 105-112.
- [19] ZENG Z G, KANG R, WEN M L, et al. A model-based reliability metric considering aleatory and epistemic uncertainty[J]. *IEEE Access*, 2017, 5: 15505-15515.
- [20] ZENG Z G, KANG R, WEN M L, et al. Uncertainty theory as a basis for belief reliability[J]. *Information Sciences*, 2018, 429: 26-36.
- [21] ZHANG Q Y, KANG R, WEN M L. Belief reliability for uncertain random systems[J]. *IEEE Transactions on Fuzzy Systems*, 2018, 26(6): 3605-3614.
- [22] LIU B D. Extreme value theorems of uncertain process with application to insurance risk model[J]. *Soft Computing*, 2013, 17(4): 549-556.
- [23] TIAN C L, JIN T, YANG X F, et al. Reliability analysis of the uncertain heat conduction model[J]. *Computers and Mathematics with Applications*, 2022, 119: 131-140.

(责任编辑:赵海军)