

面向 B5G 多业务场景基于 D3QN 的双时间尺度网络切片算法

陈赓¹, 齐书虎¹, 沈斐², 曾庆田¹

(1. 山东科技大学电子信息工程学院, 山东 青岛 266590; 2. 中国科学院上海微系统与信息技术研究所, 上海 200050)

摘要: 为了有效满足不同切片的差异化服务质量需求, 面向 B5G 多业务场景提出了一种基于竞争双深度 Q 网络 (D3QN) 的双时间尺度网络切片算法。研究了联合资源切片和调度问题, 以归一化处理后的频谱效率和不同切片用户服务质量指数的加权和作为优化目标。在大时间尺度内, SDN 控制器根据每种业务的资源需求利用 D3QN 算法预先分配给不同的切片, 然后根据基站负载状态执行基站级资源更新。在小时间尺度内, 基站通过轮询调度算法将资源调度到终端用户。仿真结果表明, 所提算法在保证切片用户服务质量需求、频谱效率和系统效用方面具有优异的性能。与其他 4 种基准算法相比, 所提算法的系统效用分别提升了 3.22%、3.81%、7.48% 和 21.14%。

关键词: 双时间尺度; 资源分配; 网络切片; 竞争双深度 Q 网络; 系统效用

中图分类号: TN929.5

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2022207

Dual time scale network slicing algorithm based on D3QN for B5G multi-service scenarios

CHEN Geng¹, QI Shuhu¹, SHEN Fei², ZENG Qingtian¹

1. College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China

2. Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China

Abstract: To effectively meet the differentiated quality of service (QoS) requirements of different slices, a dual time scale network slicing resource allocation algorithm based on dueling double DQN (D3QN) was proposed for B5G multi-service scenarios. The joint resource slicing and scheduling problem was formulated, with the weighted sum of the normalized spectral efficiency (SE) and the QoS of users indices of different slices as the optimization objective. On large time scale, the SDN controller used the D3QN algorithm to pre-allocate resources to different slices based on the resource requirements of each service, and then performed BS-level resource updating based on the load status of BS. On small time scale, the BS scheduled resources to end-users by using the round-robin scheduling algorithm. The simulation results show that the proposed algorithm has excellent performance in ensuring the QoS requirements of slice users, SE and system utility. Compared with the other 4 baseline algorithms, the system utility is improved by 3.22%, 3.81%, 7.48% and 21.14%, respectively.

Keywords: dual time scale, resource allocation, network slicing, dueling double DQN, system utility

收稿日期: 2022-08-17; 修回日期: 2022-10-18

通信作者: 曾庆田, qtzeng@sdust.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61701284, No.61871370); 国家重点研发计划基金资助项目 (No.2019YFE0120700, No.2019YFB1803101); 山东省自然科学基金资助项目 (No.ZR2022MF226); 青岛市应用基础研究计划基金资助项目 (No.19-6-2-1-cg); 山东科技大学菁英计划基金资助项目 (No.skr21-3-B-048); 中国科学院百人计划基金资助项目 (No.E06BRA1001); 山东省泰山学者计划基金资助项目 (No.ts20190936); 山东省高校青年创新团队科技支撑计划基金资助项目 (No.2019KJN024)

Foundation Items: The National Natural Science Foundation of China (No.61701284, No.61871370), The National Key Research and Development Program of China (No.2019YFE0120700, No.2019YFB1803101), The Natural Science Foundation of Shandong Province (No.ZR2022MF226), The Innovative Research Foundation of Qingdao (No.19-6-2-1-cg), The Elite Plan Project of Shandong University of Science and Technology (No.skr21-3-B-048), The Hundred Talent Program of Chinese Academy of Sciences (No.E06BRA1001), The Taishan Scholar Program of Shandong Province (No.ts20190936), The Science and Technology Support Plan of Youth Innovation Team of Shandong Higher School (No.2019KJN024)

0 引言

5G 的关键技术产生了 3 种典型的应用场景：增强型移动宽带（eMBB, enhanced mobile broadband）、超可靠低时延通信（URLLC, ultra-reliable and low-latency communication）和大连接物联网（mMTC&eMTC, massive machine-type communication&enhanced machine-type communication）^[1]。eMBB 面向高数据速率的应用，如高清视频或虚拟现实；URLLC 支持具有超高可靠性和极低时延的服务，如自动驾驶或远程手术；mMTC&eMTC 涵盖需要连接各种设备的服务，并简化操作流程以提供较长的电池寿命，如物联网用例^[2]。在同一网络上服务不同的业务类型会使资源分配的复杂性大大增加，因此必须进行资源管理以确保得到较高的系统频谱效率（SE, spectral efficiency）以及不同切片用户服务质量（QoS, quality of service）。为了支持这 3 种典型的应用场景，研究者引入了无线电接入网（RAN, radio access network）切片作为新一代蜂窝网络的关键支持技术^[3]。网络切片提供了构建多个独立逻辑网络的能力，每个网络切片都适应特定服务的需求。因此，每个 RAN 切片都可以定制并专用于支持具有独特特征和要求的特定服务^[4]，利用软件定义网络（SDN, software defined network）提供的网络可编程性，动态地管理 RAN 切片资源^[5]。

近年来，基于深度强化学习（DRL, deep reinforcement learning）算法的网络切片越来越受到关注。DRL 算法常被应用于解决资源分配等动态决策问题，相比传统优化算法具有更好的环境感知能力。文献[6]在单基站多切片的 RAN 场景中研究了最大系统 SE 和服务水平协议满意度问题，主要提出了生成对抗网络驱动的深度分布式 Q 网络（GAN-DDQN, generative adversarial network powered deep distributional Q network）以及 Dueling GAN-DDQN 算法来为每种切片分配合适的资源块（RB, resource block）并且分别验证了不同带宽分辨率对切片性能的影响。文献[7]在 RAN 多切片场景中研究了用户移动性对需求感知的影响，提出了把长短期记忆（LSTM, long short term memory）网络融入优势演员-评论家（A2C, actor-critic）算法中，使用 LSTM 网络从过去的状态中捕获服务请求的变化规律并在 Actor 网络和 Critic 网络中的损失函数

添加了熵正则化。文献[8]在多小区场景下提出了 2 种时间尺度的 RAN 切片机制来优化 URLLC 和 eMBB 服务的性能。在大时间尺度内，SDN 控制器采用探索和利用指数权重算法根据每种业务的需求为基站分配频谱资源；在小时间尺度内，每个基站利用多智能体强化学习（MADRL, multi-agent deep reinforcement learning）算法将其可用的资源调度到终端用户。文献[9]在多基站的场景下针对虚拟化 RAN 提出了一种粗资源分配方案和基于深度 Q 网络（DQN, deep Q network）的动态资源切片细化方案。考虑了资源预留问题可以满足切片中用户突然增加的情况，在用户满意度和资源利用率之间取得了平衡。文献[10]在多基站多切片的密集蜂窝网络场景中研究了基站之间的协作以及资源分配问题，为应对频繁的基站切换并满足不同业务之间的波动，提出了一种基于图注意力的 MADRL 算法。图注意力用来加强代理之间的时空合作，通过图注意力与 DRL 算法结合突出了基于图注意力的优越性。文献[11]考虑到切片数量变化的问题，使用了一种分布式优先经验回放方法，在一个切片作为一个代理的基础上管理资源，通过测试多种场景验证了既能满足切片需求也不会过度地分配资源。文献[12]考虑了无线资源以及计算资源的动态分配，以实现最佳的用户体验和最高的资源利用率，提出了一种具有两层范式的动态资源分配方案，在本地层采用惩罚函数辅助的线性规划来将切片中的无线资源分配给用户，在边缘层使用 DRL 算法将计算资源动态分配给边缘网络。文献[13]在车联网的场景中研究了天地一体化车载网络的资源切片和调度问题，以支持时延敏感服务和时延容忍服务，提出了一种基于双层强化学习求解方法。基于传统优化算法的网络切片也得到了广泛研究。文献[14]针对 5G 多基站场景提出了一种两级软切片方案，同时考虑了可靠性以及时延要求，在网络层面首先确定 eMBB 和 URLLC 所需要的资源，在基站层面的资源调度问题主要是实现 URLLC 服务之间的实时资源共享，采用启发式算法来获得接近最优的资源分配策略。文献[15]在多基站的场景下主要研究了切片的隔离，在满足切片内单个用户吞吐量的同时最大化切片吞吐量，通过采用李雅普诺夫优化算法解决。文献[16]考虑了多基站 RAN 场景，首先基础设施提供商（InP, infrastructure provider）将资源分配给移动虚拟网络运营商（MVNO, mobile virtual network

operator), 然后 MVNO 使用分配的资源为其终端用户提供服务。由于 MVNO 的利润和满足用户 QoS 需求方面存在矛盾, 提出了三层 Stackelberg 博弈来联合解决 InP 的频率和功率分配问题。文献[17]提出了一个统一的 RAN 切片框架, 最大限度地提高资源利用率的同时保证用户 QoS。上述研究工作很少考虑用户移动的场景, 多基站多切片用户移动的场景可能会加剧资源需求的波动, 并使切片间资源管理更具挑战性。因此, 在利用 DRL 算法解决多基站多切片的用户移动场景中如何合理地实现切片间资源分配; 如何最大化系统 SE; 如何保证不同切片用户 QoS 需求等方面仍然存在一定的挑战。

针对以上挑战, 本文主要的研究工作如下。

1) 建立了多基站多切片的下行链路传输场景, 以系统 SE 和不同切片用户 QoS 指数的加权和为优化目标, 建立了一个优化问题。该问题是一个非凸、非线性 NP-hard 问题。

2) 提出了一种双时间尺度的 DRL 求解方法。首先将研究问题分为大时间尺度的资源切片和小时间尺度的资源调度问题。这 2 种时间尺度是紧密耦合的。大时间尺度的资源切片对每个切片的资源调度造成资源约束, 资源调度的性能可以为切片决策提供反馈。

3) 在大时间尺度内, 由于无线网络环境的时变性和用户请求动态性, DRL 智能体能够根据当前的环境状态选择合适的切片策略, 实现实时地调整切片资源比例。在小时间尺度内, 基站根据动态网络状况和服务需求通过轮询调度算法将频谱资源调度到终端用户。

4) 仿真结果表明, 综合考虑保证切片用户 QoS 需求、SE 和系统效用等评估指标, 相比于其他 4 种基准算法, 所提算法具有优异的性能。

本文的创新性在于提出基于竞争双深度 Q 网络 (D3QN, dueling double deep Q network) 的双时间尺度网络切片算法解决多基站多切片用户移动场景下的切片间以及切片内的资源分配问题。大时间尺度内采用 D3QN 算法解决多基站多切片用户移动场景下的切片间资源分配问题; 小时间尺度内在同一切片采用轮询调度算法, 把同一切片内所有用户的优先级都视为相等, 所有用户可以被周期性地调度资源, 能够保证同一切片内每个用户被调度的概率相同。相比于现有的小时间尺度的分配算法, 本文算法在调度过程中考虑了用户之前被调度

的情况, 能够保证资源在切片内分配得更加公平。所提算法在系统效用和系统 SE 方面有一定程度的提升, 更重要的是该算法在稳定性和收敛性方面有了较大程度的提升。

1 系统模型和问题建模

1.1 系统模型

本文考虑支持 SDN 控制器的 RAN 多基站多切片正交频分多址下行链路传输场景, 由一组有限的基站组成, 其集合表示为 $\mathcal{M}=\{1,2,\dots,m,\dots,M\}$, 如图 1 所示。这些基站都是由 SDN 控制器直接管理的。切片表示为 $n \in \mathcal{N}=\{n_e, n_u, n_v\}$, 其中 e、u 和 v 分别代表 eMBB、URLLC 和 VoLTE。 $N=|\mathcal{N}|$ 表示切片的总数。所有基站共享聚合带宽 W , 每个 RB 的带宽为 B , 分配给基站 BS_m 的 RB 表示为 $\mathcal{K}_m=\{1_m, 2_m, \dots, k_m, \dots, K_m\}$ 。 u_{mn} 和 \mathcal{U}_{mn} 分别表示基站 BS_m 下切片 n 的用户和用户集合。 (m, n) 表示基站 BS_m 下切片 n 的服务, $A_{m,n}$ 表示分配给服务 (m, n) 的 RB 数量。具有多个复杂移动模式的用户随机分布在该区域中。每个用户 u 只能与一个基站相关联并属于一个切片。在大时间尺度内, SDN 控制器先收集整个系统内切片请求发送数据包数量, 然后对共享聚合带宽 W 进行切片并且根据基站不同切片的负载状态分配切好的资源。在小时间尺度内, 基站根据动态网络状况和服务需求将资源调度给终端用户。

1.2 移动模型

URLLC 服务通常存在于流动性较高的场景(如自动驾驶), 然而 eMBB 服务广泛存在于中低流动性场景(如行人移动), VoLTE 服务大部分存在于静止的环境当中。由于损耗不同, 具有相同 QoS 需求的用户所需的 RB 也可能不同。因此, 应该将用户移动性特征考虑在内。假设 eMBB、URLLC 和 VoLTE 用户的移动速度分别服从均匀分布 $[1,5]$ m/s、 $[6,10]$ m/s 和 $[0,1]$ m/s。

1.3 通信模型

通过 RB $k_m \in \mathcal{K}_m$ 与基站 BS_m 相关联的用户 $u_{mn} \in \mathcal{U}_{mn}$ 可实现的数据速率为

$$r_{u_{mn}, k_m} = B \log \left(1 + \frac{p_{k_m}^m g_{u_{mn}, k_m}^m}{N_0 B} \right) \quad (1)$$

其中, B 是每个 RB 的带宽, $p_{k_m}^m$ 是基站 BS_m 对 RB $k_m \in \mathcal{K}_m$ 的传输功率, g_{u_{mn}, k_m}^m 是基站 BS_m 与其相关

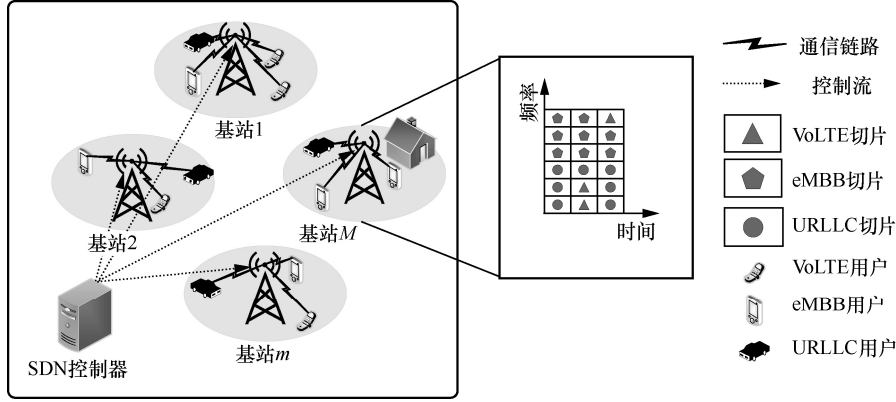


图 1 基于 SDN 控制器的 RAN 架构

联用户 $u_{mn} \in \mathcal{U}_{mn}$ 通过 RB $k_m \in \mathcal{K}_m$ 经过路径损耗和阴影衰落后的平均信道增益， N_0 是噪声功率谱密度。因此，与基站 BS_m 相关联的用户 $u_{mn} \in \mathcal{U}_{mn}$ 实现的数据速率为

$$r_{u_{mn}} = \sum_{k_m \in \mathcal{K}_m} x_{u_{mn},k_m} r_{u_{mn},k_m} \quad (2)$$

$$x_{u_{mn},k_m} = \begin{cases} 1, & \text{RB } k_m \in \mathcal{K}_m \text{ 分配给用户 } u_{mn} \in \mathcal{U}_{mn} \\ 0, & \text{其他} \end{cases} \quad (3)$$

因此系统 SE 表示为

$$SE = \frac{\sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} \sum_{u_{mn} \in \mathcal{U}_{mn}} r_{u_{mn}}}{W} \quad (4)$$

QoS 考虑在速率和时延都满足的情况下成功传输数据包与传输总数据包个数的比值，表示为

$$QoS_{mn} = \frac{\sum_{u_{mn} \in \mathcal{U}_{mn}} \sum_{q_{u_{mn}} \in Q_{u_{mn}}} x_{q_{u_{mn}}}}{\sum_{u_{mn} \in \mathcal{U}_{mn}} |Q_{u_{mn}}|} \quad (5)$$

$$x_{q_{u_{mn}}} = \begin{cases} 1, & r_{u_{mn}} > \overline{r_{u_{mn}}} \text{, 且 } l_{u_{mn}} < \overline{l_{u_{mn}}} \\ 0, & \text{其他} \end{cases} \quad (6)$$

其中， QoS_{mn} 表示基站 BS_m 中切片 n 用户 QoS， $|Q_{u_{mn}}|$ 表示基站 BS_m 发送给用户 u_{mn} 数据包的总和， $x_{q_{u_{mn}}} \in \{0,1\}$ 在速率 $r_{u_{mn}} > \overline{r_{u_{mn}}}$ 和时延 $l_{u_{mn}} < \overline{l_{u_{mn}}}$ 都满足的情况下为 1，否则为 0。 \overline{QoS}_n 表示整个系统内切片 n 用户平均 QoS，即

$$\overline{QoS}_n = \frac{\sum_{m \in \mathcal{M}} QoS_{mn}}{M} \quad (7)$$

1.4 问题建模

在 RAN 中 RB 分配的主要问题是如何为终端

用户实现最优分配，以满足在不同服务下用户数据速率和时延方面的 QoS 要求^[8]。系统效用定义为系统 SE 和不同切片用户 QoS 的加权和，并以系统效用作为优化目标。因此，RB 分配问题可以表示为

$$\begin{aligned} & \underset{x_{u_{mn},k_m}}{\text{maximize}} \quad J = \alpha SE + \beta_n \sum_{n \in \mathcal{N}} \overline{QoS}_n = \\ & \quad \alpha \frac{\sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} \sum_{u_{mn} \in \mathcal{U}_{mn}} r_{u_{mn}}}{W} + \beta_n \frac{\sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}} QoS_{mn}}{M} \\ & \text{s.t. } C_1 : x_{u_{mn},k_m} = \begin{cases} 1, & \text{RB } k_m \in \mathcal{K}_m \text{ 分配给用户 } u_{mn} \in \mathcal{U}_{mn} \\ 0, & \text{其他} \end{cases} \\ & \quad C_2 : \sum_{u_{mn} \in \mathcal{U}_{mn}} x_{u_{mn},k_m} \leq 1, \forall k_m \in \mathcal{K}_m \\ & \quad C_3 : \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} A_{m,n} \leq W \\ & \quad C_4 : x_{q_{u_{mn}}} = \begin{cases} 1, & r_{u_{mn}} > \overline{r_{u_{mn}}} \text{ \& } l_{u_{mn}} < \overline{l_{u_{mn}}} \\ 0, & \text{其他} \end{cases} \quad (8) \end{aligned}$$

其中， α 和 $\beta_n = [\beta_e, \beta_u, \beta_v]$ 分别表示系统 SE 和不同切片用户 QoS 的重要性系数。式(8)的优化目标是找到 RB 的分配策略使整个系统 SE 和切片用户 QoS 指数的加权和最大。约束条件 C_1 判断 RB $k_m \in \mathcal{K}_m$ 是否分配给用户 $u_{mn} \in \mathcal{U}_{mn}$ ，约束条件 C_2 限制每个 RB 一次只能分配给一个终端用户，约束条件 C_3 保证所分配的资源总和不超过系统的总带宽，约束条件 C_4 判断速率和时延是否满足切片的服务指标要求。

一般来说，求解该目标函数 J 非常具有挑战性。这主要是由于优化变量 x_{u_{mn},k_m} 的二进制性质和定义的用户服务质量 QoS_{mn} 与优化变量 x_{u_{mn},k_m} 有直接关系，该目标函数 J 是非线性问题。因此该 RB 分配问题是一个 0-1 非线性规划问题，与背包问题类似也是典型的 NP-hard 问题。由于传统算法求解

背包问题时间复杂度较高，因此本文提出了一种 D3QN 算法解决 RB 的分配任务。

证明 目标函数 J 是 NP-hard 问题详见附录 1。

2 基于 D3QN 的双时间尺度网络切片算法

2.1 双时间尺度分配机制

考虑到动态服务请求到达率的影响，在大时间尺度内基于 D3QN 算法确定频谱资源切片比例，在小时间尺度内基于动态的网络条件和业务需求采用轮询调度的方式将可用 RB 分配给用户。双时间尺度网络切片模型如图 2 所示，其中 T_0 表示大时间尺度内共有 T_0 个小时隙， b 表示第 b 个小时隙。

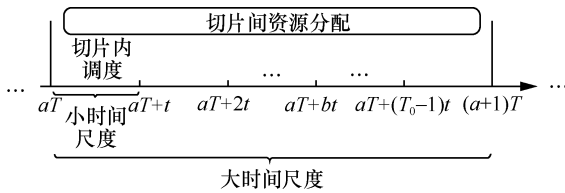


图 2 双时间尺度网络切片模型

大时间尺度的资源切片。在切片窗口 aT 开始时，SDN 控制器首先做出资源切片决策 $\varphi_n \in [0,1], \forall n \in \mathcal{N}$ ，即确定分配给切片 n 频谱资源的切片比例，再根据每个基站切片的用户数量分配切好的 RB。在整个切片窗口内，切片决策保持不变。由于服务请求模式的变化以及用户移动性的影响，服务需求和可用频谱资源随时间变化，因此在每个切片窗口结束时，SDN 控制器根据基站的反馈评估系统性能的差异，并在下一个切片窗口 $(a+1)T$ 调整资源切片决策。值得注意的是，不同切片窗口中的资源切片决策是独立的。

小时间尺度的资源调度。基于大时间尺度的资源切片决策，在每个时隙开始时进行资源调度。根据网络拓扑、用户移动性和业务需求的变化为各个终端用户分配频谱资源。

2.2 马尔可夫决策过程

对于目标函数 J 的网络切片资源分配模型，可将其转化为一个无模型的马尔可夫决策过程 (MDP, Markov decision process)。该 MDP 的状态空间与动作空间均具有较高的维度，可用一个四元组 $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$ 表示，其中， \mathcal{S} 表示智能体所有可能状态组成的状态空间， \mathcal{A} 表示智能体所有可能采取的动作集合， \mathcal{P} 表示转移概率， \mathcal{R} 表示奖励函数^[18]。

状态空间 \mathcal{S} 。 \mathcal{S} 表示整个系统的状态。SDN 控

制器充当智能体，将大时间尺度内所划分的 RB 分配给基站。切片窗口 aT 内每个切片请求发送的数据包数量作为状态。如果在切片窗口内没有该切片请求，则该切片的状态设置为 0。为了方便状态的处理，对其进行了标准化。packet _{n} 表示切片 n 在切片窗口内请求传输的数据包总和，mean _{n} 、std _{n} 分别表示切片 n 传输数据包请求到达间隔分布的平均值、标准差。因此，在切片窗口 aT 的状态 S^{aT} 简记为 $S = \{s_e, s_u, s_v\}$ 。

$$s_n = \frac{\text{packet}_n - \text{mean}_n}{\text{std}_n}, \forall n \in \mathcal{N} \quad (9)$$

动作空间 \mathcal{A} 。 \mathcal{A} 表示所执行动作的集合。SDN 控制器每获取一个状态，便会根据贪心策略选取并执行该动作。对于该网络切片资源分配问题，将频谱资源 RB 分配方案作为动作。在大时间尺度内，动作 A 共包含 3 个元素，记为 $A = \{A_e, A_u, A_v\}$ ，其中， $\varphi_n \in [0,1], \forall n \in \mathcal{N}$ 表示该大时间尺度内分配给切片 n 的切片比例。SDN 控制器确定切片比例 φ_n 之后，根据不同基站下切片所连接的用户数分配切好的资源，在小时间尺度基站在同一切片内采用轮询调度的方式分配给用户。由于在把带宽划分为 RB 的过程中可能会出现不能整除的情况，因此在划分 RB 时进行向下取整。使用 ALL _{n} 表示 SDN 控制器分配给切片 n 的 RB 数量，分配给服务 (m,n) 的 RB 数量 $A_{m,n}$ 可以表示为

$$\text{ALL}_n = \left\lfloor \frac{\varphi_n W}{B} \right\rfloor, \forall n \in \mathcal{N} \quad (10)$$

$$A_{m,n} = \frac{|U_{mn}|}{\sum_{m \in \mathcal{M}} |U_{mn}|} \text{ALL}_n, \forall n \in \mathcal{N} \quad (11)$$

转移概率 \mathcal{P} 。由于该 MDP 中的状态为每个切片请求的数据包数目，每个切片请求的数据包数目随机变化，用户平均 QoS 同无线环境中的噪声密切相关，因此无法直接量化其转移概率，所以该模型为无模型 MDP。

奖励函数 \mathcal{R} 。 \mathcal{R} 表示智能体与环境交互后所得到的奖励回报。通常来说，这个奖励回报能直接反映选取动作的好坏^[19]。系统 SE 和用户 QoS 都是评估指标，需要综合考虑。因此本文引入了分层即时奖励机制。为了尽可能满足不同切片用户 QoS 需求，eMBB、VoLTE 切片用户 QoS 阈值都设置为 0.98，URLLC 切片用户 QoS 阈值为 0.95。由于 eMBB、

VoLTE 切片指标要求比较容易实现,因此在设置奖励函数时这2种切片用户 QoS 必须同时大于或等于 0.98, 否则给予一个大的负奖励^[7]。在训练过程中注意到, 系统 SE 的最大值取值范围为 340~350 bit/(s·Hz), 最小值取值范围为 190~200 bit/(s·Hz), 因此本文对系统 SE 最大值和最小值取折中, 系统 SE 的指标要求为 280 bit/(s·Hz)。为了尽可能使奖励函数值在以 0 为中心的一定范围内波动并表现出分配策略的差异, 具体奖励函数设计步骤如下。

该奖励函数共由 3 个阶段组成, 由于优化目标是在满足不同切片用户 QoS 的同时最大化系统 SE, 因此把判断 $\overline{QoS_u}$ 、 $\overline{QoS_e}$ 以及 $\overline{QoS_u}$ 是否满足切片的服务指标要求放在前 2 个阶段, 第三阶段判断系统 SE 是否满足指标要求。第一阶段判断 $\overline{QoS_u}$ 以及 $\overline{QoS_e}$ 是否满足切片服务指标要求, 若不满足则给予一个大的负奖励 R_1 , 即

$$R_1 = -5 \quad (12)$$

否则进入第二阶段。第二阶段判断 $\overline{QoS_u}$ 是否满足切片服务指标要求, 若不满足则奖励函数为 R_2 , 即

$$R_2 = 10(\overline{QoS_u} - 0.7) \quad (13)$$

第三阶段判断系统 SE 是否满足系统 SE 的指标要求, 奖励函数 R_3 可以表示为

$$R_3 = 4 + \max\{0, 10\alpha(SE - 280)\} \quad (14)$$

其中, $\overline{QoS_u}$ 、 $\overline{QoS_e}$ 和 $\overline{QoS_u}$ 分别表示整个系统中 URLLC、VoLTE 和 eMBB 切片用户平均 QoS, 可以通过式(7)计算得出。由于前 2 个阶段是判断用户 QoS 是否满足服务指标要求, 为了使智能体能够在满足用户 QoS 的同时最大化系统 SE, 因此第三阶段的奖励函数值略大, 可以引导智能体调整其分配策略在满足用户 QoS 的同时得到更高的系统 SE。

2.3 基于 D3QN 双时间尺度网络切片算法

D3QN 算法结合了 Double DQN 和 Dueling DQN 算法的思想, 进一步提升了算法的性能。D3QN 算法的其他方面与 DQN 一致, 算法框架如图 3 所示。

Dueling DQN 将 Q 网络分成 2 个部分。第一部分仅与状态 S 有关, 表示状态的好坏程度, 这部分为价值函数, 记作 $f_{\text{value}}(S; w, \xi)$ 。第二部分同时与状态 S 和动作 A 有关, 表示该状态下某个动作相对于其他动作的好坏程度, 即采取该动作的优势, 这部

分为优势函数, 记作 $f_{\text{adv}}(S, A; w, \zeta)$ 。 $f_{\text{value}}(S; w, \xi)$ 和 $f_{\text{adv}}(S, A; w, \zeta)$ 之和表示在该状态下确定的某个动作的值, 即

$$Q(S, A; w, \xi, \zeta) = f_{\text{value}}(S; w, \xi) + f_{\text{adv}}(S, A; w, \zeta) \quad (15)$$

其中, w 、 ξ 、 ζ 分别是网络公共部分、价值函数、优势函数的网络参数。不同的动作有不同的偏差, 价值函数是一个标量, 直接将 $f_{\text{value}}(S; w, \xi)$ 和 $f_{\text{adv}}(S, A; w, \zeta)$ 相加会导致学习效果不佳。为了改进这种方法, 通常使用优势函数的平均值进行计算, 即

$$Q(S, A; w, \xi, \zeta) = f_{\text{value}}(S; w, \xi) + (f_{\text{adv}}(S, A; w, \zeta) - \frac{1}{|A|} \sum_{A' \in A} f_{\text{adv}}(S, A'; w, \zeta)) \quad (16)$$

由于 DQN 在计算目标 Q 值时, 每次都选择下一状态中最大 Q 值, 这会使 Q 值出现高估计的问题。为了解决这一问题, Double DQN 不再在目标 Q 网络里面找各个动作中的最大 Q 值, 而是在当前 Q 网络中先找出最大 Q 值对应的动作^[20], 即

$$A^{\max} = \operatorname{argmax}_A Q(S', A; w) \quad (17)$$

然后利用 A^{\max} 在目标 Q 网络中计算目标 Q 值, 即

$$y^{\text{Double}} = R + \gamma Q(S', A^{\max}; w') \quad (18)$$

其中, y^{Double} 表示目标 Q 网络中计算的 Q 值, R 表示奖励回报, γ 表示折扣因子。

结合式(17)、式(18), 计算目标 Q 网络中的 Q 值为

$$y^{\text{Double}} = R + \gamma Q(S', \operatorname{argmax}_A Q(S', A; w); w') \quad (19)$$

损失函数定义为

$$L^{\text{Double}}(w) = \mathbb{E}[y^{\text{Double}} - Q(S, A; w)^2] \quad (20)$$

其中, $\mathbb{E}[\cdot]$ 表示期望运算。在选取动作时采用 ε -贪心策略。 ε -贪心策略是指每次尝试时以 ε 的概率进行探索, 以 $1-\varepsilon$ 的概率选择 Q 值最大的动作进行利用^[21]。

$$A = \begin{cases} \operatorname{argmax}_A Q, & 1-\varepsilon \\ \text{随机}, & \varepsilon \end{cases} \quad (21)$$

大时间尺度基于 D3QN 的网络切片算法如算法 1 所示。

算法 1 大时间尺度基于 D3QN 的网络切片算法
输入 系统环境参数、切片参数和 D3QN 算法参数

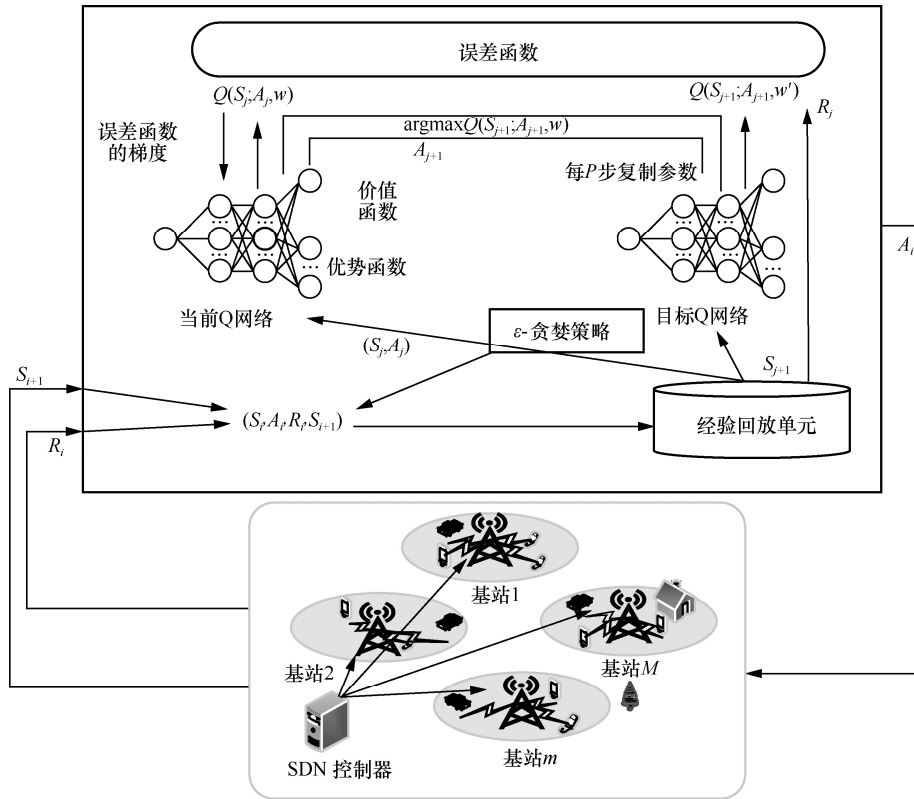


图 3 D3QN 算法框架

输出 系统 SE、切片用户 QoS 和系统效用

- 1) 初始化 当前 Q 网络参数 w 、目标 Q 网络参数 w' 、经验回放单元
- 2) 根据式(9)获得观测值 $S = \{s_e, s_u, s_v\}$
- 3) for iterations = 1: I
- 4) 状态 S_i 输入当前 Q 网络, 计算出各个动作对应的 Q 值, 根据式(21)选择当前 S_i 下对应的动作 A_i , 确定切片比例 ϕ_n
- 5) for bs = 1: BS_NUM
- 6) 根据式(10)、式(11)计算分配给该基站每个切片 RB 数量
- 7) 基站向用户分配 RB 并计算数据速率 计算 SE 以及每个切片用户 QoS
- 8) end for
- 9) 计算系统平均 SE、切片用户 QoS 和系统效用
- 10) 根据式(12)~式(14)综合判断奖励 R_i 并且得到新的状态 S_{i+1} , 统计下一回合每个基站每个切片连接的用户数量
- 11) 将 (S_i, A_i, R_i, S_{i+1}) 存入经验回放单元
- 12) 从经验回放单元中随机采样 m 个样本, $(S_j,$

$A_j, R_j, S'_j), j = 1, 2, \dots, m$, 根据式(19)计算目标 Q 值

- 13) 根据式(20)计算误差并反向更新参数 w
- 14) 更新状态 $S_i = S_{i+1}$
- 15) 每隔 P 步复制目标 Q 网络参数, $w' = w$
- 16) end for

根据算法 1 中步骤 6) 的分配结果, 小时间尺度内利用轮询调度执行步骤 7)。小时间尺度轮询调度算法如算法 2 所示。

算法 2 小时间尺度轮询调度算法

输入 分配给基站每个切片 RB 数

输出 与该基站相关联用户的数据速率

- 1) 初始化 每个切片的服务调度索引
- 2) for $n = 1: N$
- 3) for $t = 1: T_0$
- 4) 找出符合条件的用户, 并统计该小时间尺度内需要传输数据包用户的数量
- 5) 计算出为每个用户平均分配的 RB 数
- 6) 计算剩余的 RB 数量 RB_left
- 7) 找出用户索引大于服务调度索引的用户个数 left_no

- 8) if left_no > RB_left
- 9) 把剩余 RB 分配给用户索引大于服务调度索引的用户
- 10) else
- 11) 取出该服务前 RB_left 个用户, 把剩余 RB 分配给用户
- 12) 记录分配给该切片最大用户索引, 更新服务调度索引和分配给用户的 RB 数
- 13) 根据式(2)计算实现的数据速率
- 14) end if
- 15) end for
- 16) end for

3 仿真实验

3.1 仿真场景与仿真参数设置

本文考虑一个大小为 240 m×240 m 的蜂窝网络环境, 包括由 SDN 控制器直接管理的 4 个半径为 40 m 的基站以及 1 200 个移动用户^[7]。为了验证本文所提出的基于 D3QN 的双时间尺度网络切片算法, 仿真实现使用 Python 平台和 Pytorch 工具执行。总带宽设置为 40 MHz, 带宽分配分辨率为 0.5 MHz, 每个 RB 的带宽为 180 kHz。仿真了 5G 的 3 种典型服务, 即用于语音通信的 VoLTE、超高清视频传输的 eMBB 和工业级应用的 URLLC。根据文献[22-23]设置仿真参数, 如表 1 所示。

D3QN 算法由两层全连接隐藏层组成, 每层含有 200 个神经元。采用 ReLU 作为激活函数, Adam 优化器以 0.001 的学习率更新网络参数。每 50 回合更新目标 Q 网络参数, 训练探索率 ϵ 一直保持 0.2。为了使代理更偏向于未来的奖励, 折扣因子 γ 设为

0.99。最小批量大小 m 和经验回放单元大小分别为 512 和 100 000。

3.2 仿真结果分析

为了更好地评估提出的基于 D3QN 的双时间尺度网络切片算法在系统 SE 和用户 QoS 加权和优化问题上的性能, 将本文所提出的基于 D3QN 的双时间尺度网络切片算法与其他 4 种基准算法进行了比较, 即 Double DQN 算法、Dueling DQN 算法^[24]、DQN 算法^[9]和 LSTM-A2C 算法^[7]。

1) 系统效用的对比

图 4 说明了系统效用随训练迭代次数的变化。SE 的重要性系数 α 设置为 0.01, QoS 的系数 β_n 设置为 [1,1,1], 总共迭代训练 5 000 次。可以观察到 D3QN 算法在大约 1 300 次收敛, 系统效用稳定在 6.47 左右。Double DQN 在当前训练次数下波动幅度较大, 最终未达到收敛。Dueling DQN 算法在收敛速度方面占有明显的优势, 系统效用最终收敛在 6.3 左右, 整体系统效用略差。DQN 算法只有个别取值能够达到同 D3QN 的效果且最终没有达到收敛的状态。LSTM-A2C 算法最终能够达到收敛状态, 但整体系统效用稳定在 4.8 左右。由于采用集中控制的方式可能导致动作数量过多, LSTM-A2C 算法容易陷入局部最优, 不适用于动作较多的场景, 因此性能较差。整个分配过程说明基于 D3QN 的网络切片算法能够在有限的训练次数下找到较合适的分配策略, 更适用于解决在多基站多切片用户移动场景下的优化系统 SE 和用户 QoS 问题。所提出的基于 D3QN 网络切片算法与基于 Double DQN、Dueling DQN、DQN 和 LSTM-A2C 算法相比, 系统效用分别提升了 3.22%、3.81%、7.48% 和 21.14%。

表 1 仿真参数设置

参数	eMBB	URLLC	VoLTE
带宽/MHz	40	40	40
轮询调度/ms	0.5	0.5	0.5
用户数量	400	600	200
用户移动速度/(m·s ⁻¹)	均匀分布[1,5]	均匀分布[6,10]	均匀分布[0,1]
数据包请求到达间隔分布	截断帕累托分布 [指数系数=1.2, mean=6 ms, max=12.5 ms]	指数分布 [mean=180 ms]	均匀分布 [min= 0 ms, max=160 ms]
数据包的大小分布	截断帕累托分布 [指数系数=1.2, mean=100 B, max=250 B]	0.3 MB	40 B
切片要求	速率	10 Mbit/s	51 kbit/s
	时延/ms	3	10

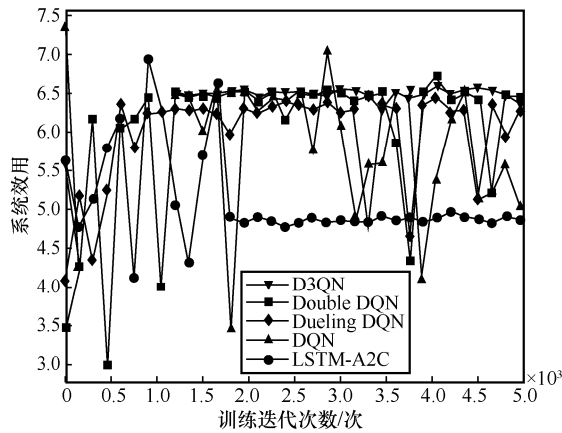


图 4 系统效用随训练迭代次数的变化

2) 系统 SE 的对比

图 5 给出了系统 SE 随训练迭代次数的变化。

从图 5 可以看出存在个别高于收敛值的异常点，这是因为在该状态下没有找到合适的分配策略，在分配过程中为某个切片分配过多的 RB，导致分配给其他切片的 RB 数量过少，此时分配 RB 数量过少的切片用户 QoS 很大程度上不能满足服务指标要求，所以此次的奖励回报很小，接下来的训练过程中在该状态下会避免采取该动作。因此，本文算法系统 SE 稳定在 350 bit/(s·Hz)左右。相比于 Dueling DQN 算法的 340 bit/(s·Hz)以及 LSTM-A2C 算法的 190 bit/(s·Hz)有了一定程度的提升。

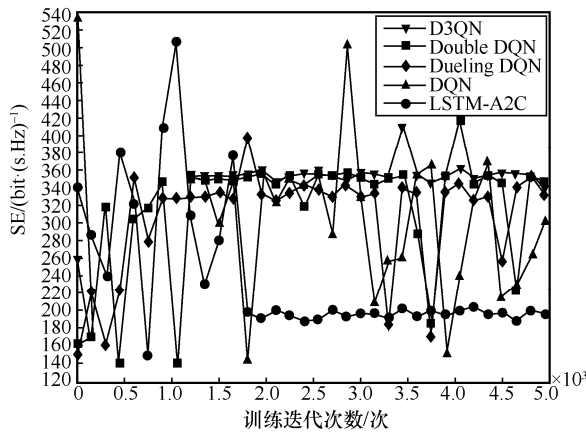


图 5 系统 SE 随训练迭代次数的变化

图 6~图 8 分别给出了 VoLTE、eMBB 和 URLLC 切片用户 QoS 随训练迭代次数的变化。从图 6 可以看出，VoLTE 切片用户 QoS 稳定在 1.00 可满足 VoLTE 切片服务要求。从图 7 中可以看出 eMBB 切片用户 QoS 都保持得较稳定，D3QN、Double DQN、Dueling DQN 和 DQN 算法基本能够稳定在 0.99 甚至能够达到 1.00，LSTM-A2C 算法可以达到 0.98。

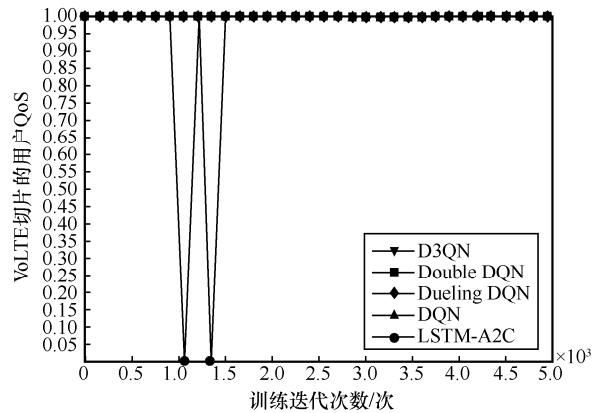


图 6 VoLTE 切片用户 QoS 随训练迭代次数的变化

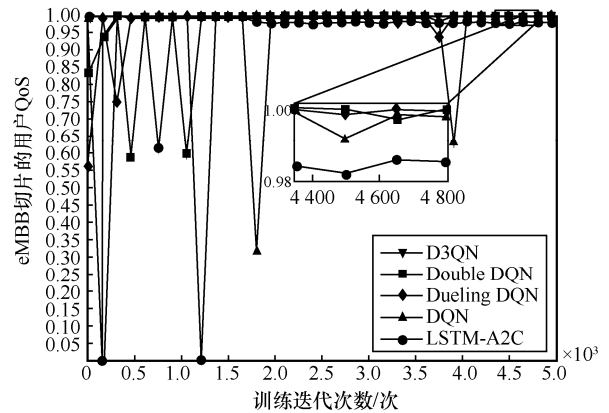


图 7 eMBB 切片用户 QoS 随训练迭代次数的变化

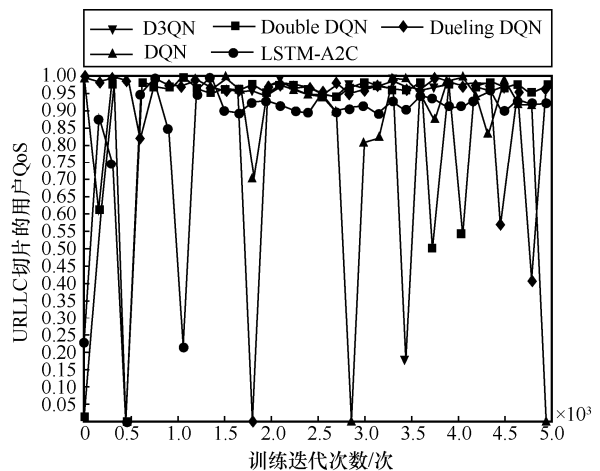


图 8 URLLC 切片用户 QoS 随训练迭代次数的变化

由于 URLLC 切片的服务要求最高，各算法的差异很容易表现出来。D3QN、Double DQN 和 Dueling DQN 算法能够稳定在 0.95 左右，LSTM-A2C 算法稳定在 0.91 左右。

3) 单时间尺度和双时间尺度性能的对比

图 9 为在相同环境下单时间尺度和双时间尺度的系统效用比较。从图 9 可以看出，单时间尺度的

基于 Double DQN、DQN 算法稳定性略差，而基于 Dueling DQN、LSTM-A2C 算法收敛后的稳定性较强，大致稳定在 6.3 左右。

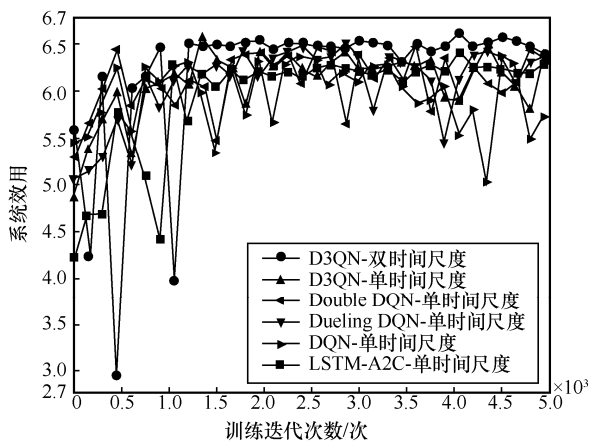


图 9 单时间尺度和双时间尺度的系统效用比较

由于 LSTM-A2C 算法实现了 A2C 和 LSTM 的融合，因此该算法具有长时记忆功能，具有较好的稳定性。单时间尺度性能较差原因在于该区域中的用户在整个系统中随机移动，导致一些基站过载而 RB 远远不足，而另一些基站的 RB 则会有剩余。相比于单时间尺度基于 D3QN、Double DQN、Dueling DQN、DQN 和 LSTM-A2C 网络切片算法，所提出算法在系统效用方面分别提升了 1.33%、1.01%、1.62%、3.59%和 4.00%。

图 10 为系统 SE 在不同时间尺度上随训练迭代次数的变化。在系统 SE 方面，双时间尺度分配算法明显高于单时间尺度分配算法。在单时间尺度下基于 LSTM-A2C 算法在稳定性方面优于其他算法，但系统 SE 只能约为 325 bit/(s·Hz)。所提算法相比于单时间尺度 D3QN、Double DQN、Dueling DQN、DQN 和 LSTM-A2C 算法在系统 SE 方面分别提升了 4.03%、3.57%、4.47%、6.91%和 6.26%。

图 11~图 13 显示了 3 种切片在双时间尺度和单时间尺度下用户 QoS 比较。从图 11 可以看出，双时间尺度分配算法和单时间尺度分配算法的 VoLTE 切片用户 QoS 都稳定在 1.0 附近，能够满足该切片的服务指标要求。图 12 为不同时间尺度下 eMBB 切片用户 QoS 比较。从图 12 可知，在单时间尺度下 D3QN 算法与 LSTM-A2C 算法的性能较好，用户 QoS 基本能够稳定在 0.99 甚至能够达到 1.0。基于 Double DQN 和 DQN 的算法能够满足服务指标要求但是稳定性略差。由于 VoLTE 切片对数据速率和时延的要求较低，服务需求很容易满足。

从图 12 可以看出，eMBB 切片用户 QoS 采用双时间尺度分配算法稳定性较强。图 13 为不同时间尺度下 URLLC 切片用户 QoS 比较。从图 13 可知，相比于单时间尺度 Dueling DQN 和 LSTM-A2C 算法，所提算法的收敛值略低，但可以稳定在 0.95 左右，能够满足服务指标要求。目的是能够在满足 URLLC 切片用户 QoS 前提下，尽可能地分配给 eMBB 切片更多 RB，以提高系统 SE。

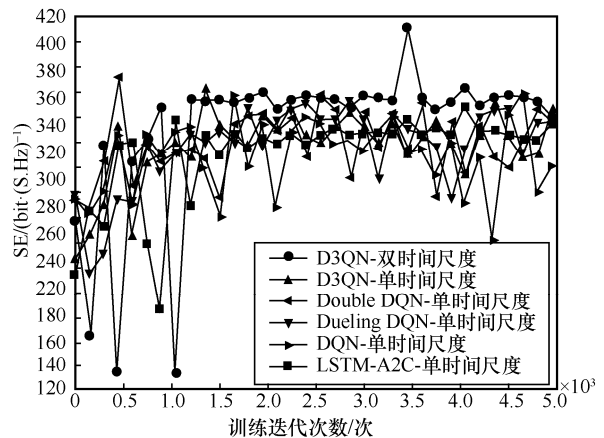


图 10 系统 SE 在不同时间尺度上随训练迭代次数的变化

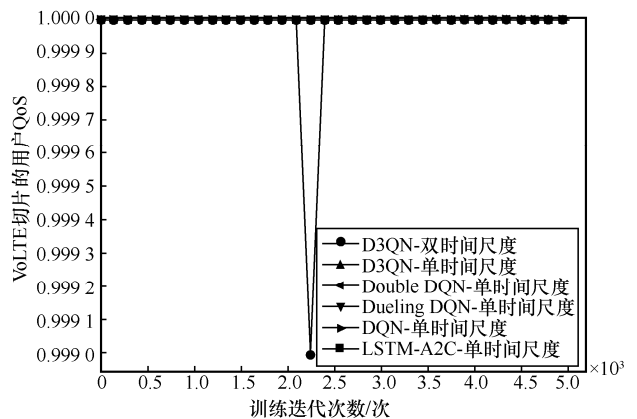


图 11 单时间尺度和双时间尺度下 VoLTE 切片用户 QoS 比较

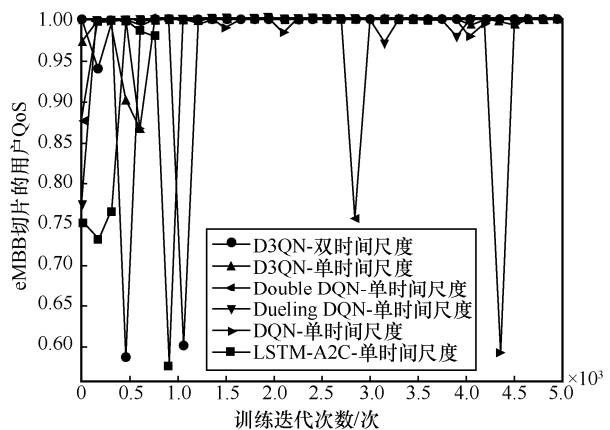


图 12 单时间尺度和双时间尺度下 eMBB 切片用户 QoS 比较

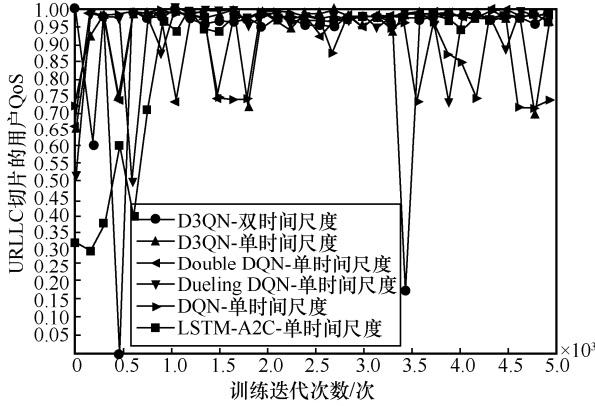


图 13 单时间尺度和双时间尺度下 URLLC 切片用户 QoS 比较

综上所述，本文所提算法能够更好地进行资源分配，在满足用户 QoS 需求、系统 SE 和系统效用方面具有较好的性能，证明了所提算法的有效性。

4 结束语

本文面向 B5G 多业务场景中的网络切片资源分配进行了研究。将系统 SE 和不同切片用户 QoS 指数的加权和作为优化目标，提出了一种基于 D3QN 的双时间尺度网络切片算法。该算法在大时间尺度内通过基于 D3QN 的强化学习算法确定资源切片比例，再根据每个基站的负载状况分配切好的资源，在小时间尺度同一切片内采用轮询调度算法解决 RB 分配给终端用户问题。将基于 D3QN 的双时间尺度网络切片算法与基于 Double DQN、Dueling DQN、DQN 和 LSTM-A2C 算法进行了对比。仿真结果表明，所提出的基于 D3QN 双时间尺度网络切片算法在满足用户 QoS 需求、系统 SE 以及系统效用等方面具有更好的性能表现。与基准算法相比具有一定的优势，更适合网络切片的智能分配，为网络切片资源管理提供了一种具有前景的解决方案。

附录 1 目标函数 J 是 NP-hard 问题的证明

背包的定义。假设有一个背包，其容量是 capacity，给定一组 Z 件物品，每件物品都有它的价值 p_z 和重量 w_z 。每个物品可以放入或不放入背包（相应的取值为 1 或 0）。该问题的目标是找到一个物品子集 $Z' \subseteq Z$ 使总价值 $\sum_{z \in Z'} p_z$ 最大，并且所选物品的总重量小于或等于背包容量，即 $\sum_{z \in Z'} w_z \leq \text{capacity}$ 。

考虑该问题中一个简化问题，在场景中只有一种切片 n

（ n 是一个定值）情形，目标函数变为

$$\begin{aligned} \text{maximize}_{x_{u_{mn}, k_m}} J_m &= \alpha \text{SE} + \beta_n \overline{\text{QoS}}_n = \\ &= \alpha \frac{\sum_{m \in \mathcal{M}} \sum_{u_{mn} \in \mathcal{U}_{mn}} r_{u_{mn}}}{W} + \beta_n \frac{\sum_{m \in \mathcal{M}} \text{QoS}_{mn}}{M} \\ \text{s.t. } C_1: x_{u_{mn}, k_m} &= \begin{cases} 1, \text{RB } k_m \in \mathcal{K}_m \text{ 分配给用户 } u_{mn} \in \mathcal{U}_{mn} \\ 0, \text{其他} \end{cases} \\ C_2: \sum_{u_{mn} \in \mathcal{U}_{mn}} x_{u_{mn}, k_m} &\leq 1, \forall k_m \in \mathcal{K}_m \\ C_3: \sum_{m \in \mathcal{M}} A_{m,n} &\leq W, \forall n \in \mathcal{N} \\ C_4: x_{q_{mn}} &= \begin{cases} 1, r_{u_{mn}} > \overline{r_{u_{mn}}} \ \& \ l_{u_{mn}} < \overline{l_{u_{mn}}} \\ 0, \text{其他} \end{cases} \end{aligned} \quad (22)$$

其中，式(22)的目的是找到集合 $\mathcal{K}_m, m \in \mathcal{M}$ 实现对切片 n 终端用户集 \mathcal{U}_n 的 RB 分配策略，在满足用户 QoS 的同时最大化系统 SE。约束条件 C_1 判断 RB $k_m \in \mathcal{K}_m$ 是否分配给用户 $u_{mn} \in \mathcal{U}_{mn}$ ，约束条件 C_2 限制每个 RB 每次仅分配给一个终端用户，约束条件 C_3 表示预分配给所有基站的资源之和不超过总带宽，约束条件 C_4 判断是否满足切片速率和时延要求。

将 0-1 背包问题映射到该目标函数 J_m ，物品数 Z 对应于 RB 数量，价值 p_z 是该切片实现的 SE 和 QoS 之和，物品的重量 w_z 对应于分配给每个基站 RB 的数量，总容量的限制为分配给基站 RB 的数量之和不能超过总带宽，即 $\sum_{m \in \mathcal{M}} A_{m,n} \leq W$ 。

由于该目标函数 J_m 显然是能够在多项式时间内完成的，并且 0-1 背包问题是 NP-hard 问题，因此该问题的简化问题 J_m 是一个 NP-hard 问题。由此可以得出结论，目标函数 J 是 NP-hard 问题。

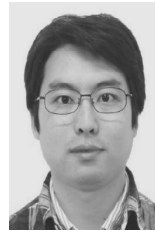
证毕。

参考文献：

- [1] DOGRA A, JHA R K, JAIN S. A survey on beyond 5G network with the advent of 6G: architecture and emerging technologies[J]. IEEE Access, 2021, 9: 67512-67547.
- [2] MLIKA Z, CHERKAOUI S. Massive IoT access with NOMA in 5G networks and beyond using online competitiveness and learning[J]. IEEE Internet of Things Journal, 2021, 8(17): 13624-13639.
- [3] ELAYOUBI S E, JEMAA S B, ALTMAN Z, et al. 5G RAN slicing for verticals: enablers and challenges[J]. IEEE Communications Magazine, 2019, 57(1): 28-34.
- [4] FILALI A, ABOUAOMAR A, CHERKAOUI S, et al. Multi-access edge computing: a survey[J]. IEEE Access, 2020, 8: 197017-197046.
- [5] SHU Z G, TALEB T. A novel QoS framework for network slicing in 5G and beyond networks based on SDN and NFV[J]. IEEE Network, 2020, 34(3): 256-263.
- [6] HUA Y X, LI R P, ZHAO Z F, 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.

- [7] LI R P, WANG C J, ZHAO Z F, et al. The LSTM-based advantage actor-critic learning for resource management in network slicing with user mobility[C]//Proceedings of IEEE Communications Letters. Piscataway: IEEE Press, 2020: 2005-2009.
- [8] FILALI A, MLIKA Z, CHERKAOUI S, et al. Dynamic SDN-based radio access network slicing with deep reinforcement learning for URLLC and eMBB services[J]. IEEE Transactions on Network Science and Engineering, 2022, 9(4): 2174-2187.
- [9] SUN G L, XIONG K, BOATENG G O, et al. Autonomous resource provisioning and resource customization for mixed traffics in virtualized radio access network[J]. IEEE Systems Journal, 2019, 13(3): 2454-2465.
- [10] SHAO Y, LI R P, HU B, et al. Graph attention network-based multi-agent reinforcement learning for slicing resource management in dense cellular network[J]. IEEE Transactions on Vehicular Technology, 2021, 70(10): 10792-10803.
- [11] ABIKO Y, SAITO T, IKEDA D, et al. Flexible resource block allocation to multiple slices for radio access network slicing using deep reinforcement learning[J]. IEEE Access, 2020, 8: 68183-68198.
- [12] YANG G, LIU Q, ZHOU X W, et al. Two-tier resource allocation in dynamic network slicing paradigm with deep reinforcement learning[C]//Proceedings of 2019 IEEE Global Communications Conference. Piscataway: IEEE Press, 2019: 1-6.
- [13] WU H Q, CHEN J Y, ZHOU C H, et al. Learning-based joint resource slicing and scheduling in space-terrestrial integrated vehicular networks[J]. Journal of Communications and Information Networks, 2021, 6(3): 208-223.
- [14] SHI W S, LI J L, YANG P, et al. Two-level soft RAN slicing for customized services in 5G-and-beyond wireless communications[J]. IEEE Transactions on Industrial Informatics, 2022, 18(6): 4169-4179.
- [15] PAPA A, JANO A, AYVAŞIK S, et al. User-based quality of service aware multi-cell radio access network slicing[J]. IEEE Transactions on Network and Service Management, 2022, 19(1): 756-768.
- [16] HU J Z, ZHENG Z J, DI B Y, et al. Multi-layer radio network slicing for heterogeneous communication systems[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(4): 2378-2391.
- [17] SUN Y, QIN S, FENG G, et al. Service provisioning framework for RAN slicing: user admissibility, slice association and bandwidth allocation[J]. IEEE Transactions on Mobile Computing, 2021, 20(12): 3409-3422.
- [18] ZHAO N, LIANG Y C, NIYATO D, et al. Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks[J]. IEEE Transactions on Wireless Communications, 2019, 18(11): 5141-5152.
- [19] 孙国林, 欧睿杰, 刘贵松. 基于深度强化学习的应急物联网切片资源预留算法[J]. 通信学报, 2020, 41(9): 8-20.
- SUN G L, OU R J, LIU G S. Deep reinforcement learning-based resource reservation algorithm for emergency Internet-of-things slice[J]. Journal on Communications, 2020, 41(9): 8-20.
- [20] 元伟敬, 宋清洋, 郭磊. 面向软件定义多模态车联网的双时间尺度 RAN 切片资源分配[J]. 通信学报, 2022, 43(4): 60-70.
- QI W J, SONG Q Y, GUO L. Dual time scale resource allocation for RAN slicing in software-defined oriented polymorphic IoV[J]. Journal on Communications, 2022, 43(4): 60-70.
- [21] 喻鹏, 张俊也, 李文璟, 等. 移动边缘网络中基于双深度 Q 学习的高能效资源分配方法[J]. 通信学报, 2020, 41(12): 148-161.
- YU P, ZHANG J Y, LI W J, et al. Energy-efficient resource allocation method in mobile edge network based on double deep Q-learning[J]. Journal on Communications, 2020, 41(12): 148-161.
- [22] 3GPP. Further advancements for E-UTRA physical layer aspects: TR36. 814 (V9.0.0)[S]. 2010.
- [23] 3GPP. Service requirements for 5G system (Release 15): TS 22.261 V18.3.0[S]. 2021.
- [24] SUN G L, XIONG K, BOATENG G O, et al. Resource slicing and customization in RAN with dueling deep Q-network[J]. Journal of Network and Computer Applications, 2020, 157: 102573.

[作者简介]



陈贻 (1984-), 男, 山东潍坊人, 博士, 山东科技大学副教授、硕士生导师, 主要研究方向为异构网络、泛在网络和软件定义移动网络、无线资源管理和优化算法。

齐书虎 (1998-), 男, 山东聊城人, 山东科技大学硕士生, 主要研究方向为网络切片、资源分配。

沈斐 (1983-), 女, 江苏南京人, 博士, 中国科学院上海微系统与信息技术研究所研究员、博士生导师, 主要研究方向为无线通信、边缘计算和雾计算的资源优化。

曾庆田 (1976-), 男, 山东潍坊人, 博士, 山东科技大学教授、博士生导师, 主要研究方向为 Petri 网、过程挖掘和知识管理。