

基于流量感知的无线接入网智能切片资源分配方法研究

赵晨 张铖 黄永明*

(东南大学移动通信全国重点实验室, 江苏南京 211189)

摘要: 为保障业务服务等级协议(Service Level Agreements, SLA)需求,使不同应用间差异化的服务质量需求同时得到良好满足,无线接入网(Radio Access Network, RAN)的动态网络资源配置十分关键。本文针对网络流量波动、网络状态快速变化的场景,提出一种结合时间序列预测和深度强化学习的智能化带宽分配策略,通过合理利用长短期记忆(Long Short-Term Memory, LSTM)网络和 Dueling 深度 Q 网络(Dueling Deep Q Network, Dueling DQN)以最大限度提高无线接入网切片的频谱效率和 SLA 满意度。通过使用 LSTM 网络对切片中数据流量进行预测,可以将深度强化学习算法的计算周期与实际切片配置周期有效解耦。同时为了在保障 LSTM 性能的同时降低其计算复杂度,以适应 RAN 中有限的计算资源,本文采用控制神经元连接比例的随机连接长短期记忆(Randomly Connected LSTM, RCLSTM)网络。此外, Dueling DQN 与传统 DQN 相比,能够提高切片策略学习过程的 Q 值估计精度,以提升收敛速度。仿真结果表明:与原始 DQN、优势 Actor-Critic(Advantage Actor Critic, A2C)和硬切片方法相比,所提 RCLSTM-Dueling DQN 方案可以通过提前感知网络性能变化,有效降低网络环境波动对密集流量场景下无线切片资源管理的影响,在具有三种不同流量波动模式及服务质量需求的 RAN 切片场景中,可以获得收敛速度、频谱效率和切片 SLA 满意率的明显提升。同时,10%连接比例的 RCLSTM 网络能在保持极低性能损失的前提下,将原始 LSTM 的计算时间降低约 11%。

关键词: 无线接入网络切片; 动态资源分配; 时间序列预测; 深度强化学习

中图分类号: TN929.5 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2024.04.010

引用格式: 赵晨,张铖,黄永明. 基于流量感知的无线接入网智能切片资源分配方法研究[J]. 信号处理, 2024, 40(4): 719-732. DOI: 10.16798/j.issn.1003-0530.2024.04.010.

Reference format: ZHAO Chen, ZHANG Cheng, HUANG Yongming. Research on traffic-aware intelligent slicing resource allocation for radio access network[J]. Journal of Signal Processing, 2024, 40(4): 719-732. DOI: 10.16798/j.issn.1003-0530.2024.04.010.

Research on Traffic-aware Intelligent Slicing Resource Allocation for Radio Access Network

ZHAO Chen ZHANG Cheng HUANG Yongming*

(National Mobile Communications Research Laboratory, Southeast University, Nanjing, Jiangsu 211189, China)

Abstract: Dynamic network resource allocation is crucial in a radio access network (RAN) to meet the requirements of service-level agreements (SLAs), while simultaneously ensuring that the different service-quality requirements of differ-

收稿日期: 2023-06-26; 修回日期: 2023-08-10

*通信作者: 黄永明 huangym@seu.edu.cn *Corresponding Author: HUANG Yongming, huangym@seu.edu.cn

基金项目: 国家自然科学基金杰出青年基金项目(62225107); 国家自然科学基金面上项目(62271140); 国家重点研发计划课题(2018YFB1800801)

Foundation Items: The National Natural Science Foundation of China (62225107, 62271140); The National Key R&D Program of China (2018YFB1800801)

ent applications are effectively met. This paper proposes an intelligent bandwidth-allocation strategy that combines time-series prediction and deep reinforcement learning to address the scenarios of network traffic fluctuation and rapid network state change by reasonably using a long short-term memory (LSTM) network and dueling deep Q network (Dueling DQN) to maximize the spectral efficiency and satisfy the SLA of radio access network slices. By using LSTM networks to predict traffic in slices, the computational cycle of the deep reinforcement learning algorithm can effectively be decoupled from the actual slice configuration cycle. In order to reduce the computational complexity of the LSTM while maintaining its performance to fit the limited computational resources of a RAN, a randomly connected LSTM (RCLSTM) network with a controlled neuronal connectivity ratio is used. In addition, the Dueling DQN can improve the accuracy of the Q-value estimation for the slicing strategy learning process compared to a conventional DQN to enhance the convergence speed. Simulation results showed that the proposed RCLSTM-Dueling DQN scheme could effectively reduce the impact of network-environment fluctuations on wireless slicing resource management in dense traffic scenarios by sensing network performance changes in advance, compared to the original DQN, advantage actor-critic (A2C), and hard slicing approaches. In RAN slicing scenarios with three different traffic fluctuation patterns and three different quality-of-service requirements, significant improvements in the convergence speed, spectral efficiency, and slice SLA satisfaction rates could be achieved. Also, a 10% connected RCLSTM network could reduce the computation time of the original LSTM by approximately 11% while maintaining a very low performance loss.

Key words: radio access network slicing; dynamic resource allocation; time series prediction; deep reinforcement learning

1 引言

无线通信网络所支撑的业务类型正随着垂直行业涌现而持续增加,这些业务在能耗、时延、速率和可靠性等方面有着不同甚至相互矛盾的需求。基于“一刀切”设计的传统网络因其在安全性、成本和可用性等方面的灵活性不足,无法提供满足如此多样化业务需求的服务。为实现按需定制、多业务支持的服务,5G提出了“网络切片(Network Slicing, NSL)”技术^[1]。通过在共享物理网络基础设施上“切分”出多个端到端的逻辑独立网络,以实现网络的按需定制^[2]。网络切片是具有可定制、可伸缩、可编程等性质的逻辑化网络,每个网络切片服务于一种业务类型^[3-4]。但网络切片的灵活性也对网络资源的管理带来了严峻挑战,如何设计动态切片方案,使不同应用间差异化的服务质量需求同时得到良好满足,具有重要研究意义。

完整的端到端网络切片通常由无线接入网(Radio Access Network, RAN)切片、传输网(Transport Network, TN)切片和核心网(Core Network, CN)切片构成(其基本形式称为DÉCOR^[5])。三部分切片之间相互独立,可以单独对每类切片进行资源管理,本文仅针对RAN切片进行相关研究。RAN侧切片资源管理与优化的研究仍存在较大完善空间:由于切片是面向用户需求的,而无线资源又相对稀缺,RAN切片面临的关键难点在于如何通过资源抽象、虚拟化和服务间分离来有效利用用户

需求进行无线资源配置,通过对时域、频域、码域和空间等无线资源进行自适应的分配和调整,以应对切片的实时需求。

深度强化学习(Deep Reinforcement Learning, DRL)^[6-7]在实现5G/6G智能化的过程中发挥着关键作用^[8],为NSL资源管理提供了除传统基于模型的方法之外的无模型技术。通过与环境交互深入学习,以满足与体验质量(Quality of Experience, QoE)相关的体验水平协议(Experience Level Agreements, XLA)与服务质量(Quality of Service, QoS)相关的服务水平协议(Service Level Agreements, SLA)^[9]。文献[10]提出了一种资源分配机制,使用DQN和软件定义网络(Software Defined Network, SDN)控制器,将无线资源分配给超高可靠低延迟通信(Ultra Reliable Low Latency Communication, URLLC)和增强型移动宽带(enhanced Mobile BroadBand, eMBB)切片,以最大化数据速率。DQN智能体的状态空间考虑了终端用户集、预分配的无线资源块、信道增益、最小数据速率以及最大延迟。但是,该方法并没有考虑切片QoS需求。文献[11]提出了一种平衡QoS满意度和切片资源利用率的策略,基于DQN在虚拟RAN中分配无线和回程资源。该策略将状态空间视为包含满意率和切片占用资源的向量,DQN智能体的动作则建模为资源分配百分比。文献[12]提出了一种基于DQN的动态框架,保留未使用的带宽资源并分配给虚拟RAN切片,以最大限度地提高QoS满意度和资源利用率。状态

空间考虑了分配的虚拟资源百分比和每个切片的平均资源利用率。不同于上述文献以最大化资源利用率为优化目标,文献[13]以提高虚拟网络供应商的长期资源利用率和收入为目标,提出了一种基于DQN的机制。将5G-RAN资源分配给服务于移动通信、视频通信和车辆通信的切片。该机制将状态空间建模为随机到达的带宽请求和能量消耗,动作空间则表示为移动虚拟网络运营商为实现利润最大化而选择的切片。

另一方面,从历史流量数据中获取隐藏模式,对网络流量进行预测,对于提高网络效率和优化网络性能具有至关重要的意义^[14]。事实上,流量学习与预测是移动通信网络架构设计的重要基础。根据实际流量需求动态配置网络资源,能够提高网络能量效率,准确预测未来流量负载可以在设计具有流量感知性的移动通信网络中发挥重要作用^[15]。

传统的DRL算法在选择动作的过程中多数只基于当前的观测状态,如上一切片时间窗口的服务请求,而在RAN切片资源配置这类部分可观测马尔科夫决策过程中,连续时间点之间观测到的状态具有一定的相关性,而仅仅依据当前时刻的观测状态无法全面地描述环境特征,缺少状态序列变化过程中关键的隐藏信息。另外,每个切片配置时间窗口内,RAN切片数据包请求情况受到为业务类型本身特性所决定的数据包大小与请求时间规律的影响,不同窗口内的服务请求会产生波动,导致对环境的观测与切片资源配置方案实施之间存在时间差的问题。因此,仅仅依赖上一切片时间窗口的服务请求情况进行当前切片时间窗口的资源分配,难以得到最优的配置结果。由于循环神经网络(Recurrent Neural Networks, RNN)的变体LSTM^[16]已成功应用于诸如语音识别^[17]、机器翻译^[18]、图像/视频分类^[19]等序列预测任务,本文在DRL算法基础之上,引入循环神经网络,利用LSTM在时间序列预测方面的有效性,挖掘网络流量变化规律,提升RAN切片的性能。具体贡献如下。

针对网络流量波动、网络状态快速变化的场景,本文从历史流量数据获取隐藏模式,对未来无线网络流量进行预测,并结合DRL以确定最优RAN切片资源分配方案。具体而言,将LSTM与能够提高Q值估计精确度的Dueling DQN^[20]相结合,辅助智能体感知网络环境状态变化,提前适应变化的流量模式,从而解耦DRL算法的计算周期与实际切片配置周期,取得更好的资源分配效果。同时,

采取随机连接的LSTM(Random Connected LSTM, RCLSTM)网络架构^[21],在保障预测性能的同时有效降低传统LSTM网络的计算复杂度,以适配RAN中有限的计算资源。仿真结果表明:所提RCLSTM-Dueling DQN方案可以有效降低网络环境波动对RAN切片资源管理的影响。相比传统切片方案,如原始DQN、优势Actor-Critic(Advantage Actor Critic, A2C)和硬切片方法,所提RCLSTM-Dueling DQN方案在具有三种不同流量波动模式及服务质量需求的RAN切片场景中,以获得收敛速度、频谱效率和切片SLA满意率的明显提升。此外,10%连接比例的RCLSTM网络能在保持极低预测性能损失的同时,将原始LSTM的计算时间降低约11%,显著降低所提RCLSTM-Dueling DQN方案的部署代价。

2 系统模型

2.1 无线网络切片资源分配场景

本文考虑一个单基站多切片多用户的RAN侧无线资源分配场景,网络切片集合为 \mathcal{N} ,记为 $\{1, 2, \dots, N\}$ 。其中 U 个用户根据其业务类型的不同被划分为 N 个切片用户组 $\mathcal{U}_n, n \in [1, N]$ 。不同业务类型的用户具有不同的数据包大小、包到达时间间隔,以及不同的SLA需求。基站包含的 N 个切片共享该基站能够提供的所有无线资源。该资源分配场景如图1所示。

本文考虑分配的资源为频谱资源,由物理上的时域和频域构成,如图1右侧所示,其基本单位为长期演进(Long Term Evolution, LTE)标准中的单位资源块(Resource Block, RB)。具体地,一个RB资源块在时域上由连续的7个正交频分复用(Orthogonal Frequency-Division Multiplexing, OFDM)符号组成,每个OFDM符号长度为 $71.4 \mu\text{s}$;频域上由12个连续相互正交子载波组成,每个子载波之间间隔15 kHz。因此,一个RB资源块占用0.5 ms的时隙和180 kHz的频谱带宽。这些基本的RB共同构成了整个无线频谱资源,RAN切片资源分配即将这些RB组成的时频资源分配给不同类型的切片。

针对时隙传输模型,定义每个时隙长度为 $T(=0.5 \text{ ms})$,一个切片时间窗口包含 C 个时隙。所有切片共享总带宽 W ,在每个切片时间窗口,基站根据网络状态与切片SLA需求,为每个切片分配一定的无线网络资源(RB比例),定义为 $[w_1, w_2, w_3, \dots, w_N]$ 。在每个切片窗口内,采用预定MAC层调度方式,如

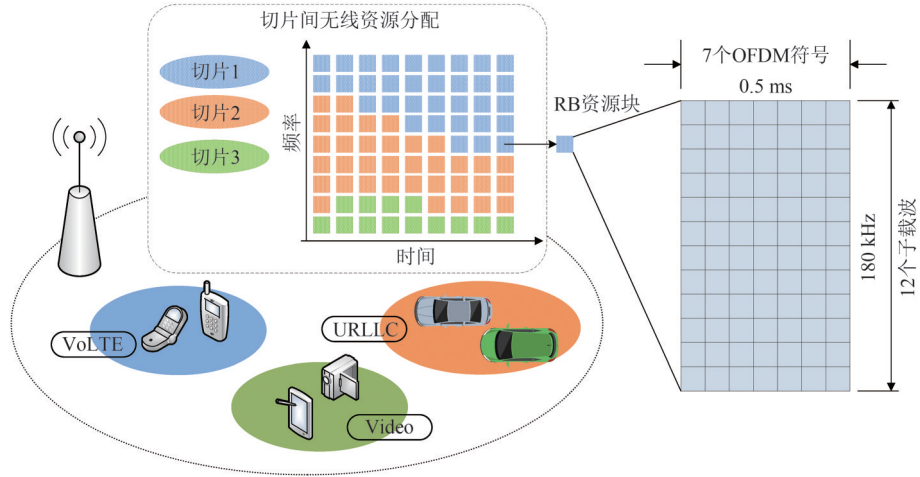


图1 无线接入网切片资源分配场景

Fig. 1 RAN slicing resource allocation scenario

轮询方式、优化参数的优先级调度方式等,将每个业务类型切片分得的资源分配给相应用户组中的用户。

2.2 切片满意度模型

对于网络切片 n , 满足其 SLA 需求所需要的最小数据速率记为 R_n^{req} , 最大时延记为 T_n^{req} . 定义用户 u_n 发送给基站的数据包集合为 \mathcal{X}_{u_n} , 则用户 u_n 发送给基站的数据包数量为 $|\mathcal{X}_{u_n}|$.

对于用户 u_n 的数据包 $x_{u_n} \in \mathcal{X}_{u_n}$, 大小记作 $c_{x_{u_n}}$, 时延记作 $t_{x_{u_n}}^d$, 包括等待时延 $t_{x_{u_n}}^w$ 与传输时延 $t_{x_{u_n}}^t$, 等待时延指从业务数据包生成至数据包传输, 后者为数据包传输时间。则数据包 $x_{u_n} \in \mathcal{X}_{u_n}$ 的时延和速率为:

$$t_{x_{u_n}}^d = t_{x_{u_n}}^w + t_{x_{u_n}}^t, \quad x_{u_n} \in \mathcal{X}_{u_n} \quad (1)$$

$$r_{x_{u_n}} = \frac{c_{x_{u_n}}}{t_{x_{u_n}}^d}, \quad x_{u_n} \in \mathcal{X}_{u_n} \quad (2)$$

切片 SLA 满足率 (SLA Satisfaction Ratio, SSR) 由切片时间窗口内切片的 SLA 满足情况定义, 主要包括传输速率满足率和时延满足率。对于第 n 个切片, 统计当前切片时间窗口内, 同时满足时延小于最大可容忍时延和速率大于最小需求速率的数据包比例, SSR 计算如式(3)所示:

$$\text{SSR}_n = \frac{\sum_{u_n \in \mathcal{U}_n} \sum_{x_{u_n} \in \mathcal{X}_{u_n}} I(r_{x_{u_n}} > R_n^{\text{req}}, t_{x_{u_n}}^d < T_n^{\text{req}})}{\sum_{u_n \in \mathcal{U}_n} |\mathcal{X}_{u_n}|} \quad (3)$$

其中, I 是一个指示函数:

$$I(x) = \begin{cases} 1, & x = 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

2.3 频谱效率模型

对于切片 n 中的用户 u_n , 在当前切片时间窗口, 将其与基站之间的平均传输数据速率定义为 r_{u_n} , 表示用户传输数据包流量与切片窗口时间窗口长度之比, 如式(5)所示:

$$r_{u_n} = \frac{\sum_{x_{u_n} \in \mathcal{X}_{u_n}} c_{x_{u_n}}}{CT}, \quad \forall u_n \in \mathcal{U}_n \quad (5)$$

将频谱效率 (SE) 定义为整个基站单位带宽的平均传输速率, 如式(6)所示:

$$\text{SE} = \frac{\sum_{n \in \mathcal{N}} \sum_{u_n \in \mathcal{U}_n} r_{u_n}}{W} \quad (6)$$

2.4 系统效益与优化目标

定义系统效益为频谱效率 SE 和切片 SSR 的加权和, α 为 SE 的重要性系数, $\beta = \{\beta_1, \beta_2, \dots, \beta_N\}$ 为 SSR 的重要性系数, 其中 β_n 为切片满意度 SSR_n 的重要性权重。则系统效益 J 可表示如式(7):

$$J = \alpha \text{SE} + \sum_{n \in \mathcal{N}} \beta_n \text{SSR}_n \quad (7)$$

针对 RAN 切片无线资源动态分配问题, 设定其目标为在资源总数不变的限制下, 最大化系统效益。当用户数量、数据传输需求改变时, 同一类型切片在不同切片时间窗口内接收到的请求数量会产生波动, 使得资源调度问题动态化、复杂化。尤其是对于 SLA 要求较为严格的业务类型, 难以将切片的 SSR 保持在合适的水平。这时, 就需要考虑如何设计切片资源管理系统, 为服务于不同业务的切片灵活地分配带宽。为了在保障切片 SLA 需求的前提下, 尽可能提高频谱效用, 最大化系统效益, 因此将优化问题定义为式(8):

$$\begin{aligned}
\max_{w_n, n=1,2,\dots,N} J &= \alpha \text{SE} + \sum_{n \in \mathcal{N}} \beta_n \text{SSR}_n \\
&= \alpha \frac{\sum_{n \in \mathcal{N}} \sum_{u_n \in \mathcal{U}_n} r_{u_n}}{W} + \sum_{n \in \mathcal{N}} \beta_n \frac{\sum_{u_n \in \mathcal{U}_n} \sum_{x_{u_n} \in \mathcal{X}_{u_n}} I(r_{x_{u_n}} > R_n^{\text{req}}, t_{x_{u_n}}^d < T_n^{\text{req}})}{\sum_{u_n \in \mathcal{U}_n} |\mathcal{X}_{u_n}|} \\
\text{s.t.} \quad \sum_{n \in \mathcal{N}} w_n &= W
\end{aligned} \tag{8}$$

3 智能化动态RAN切片配置策略

3.1 问题建模

首先,将RAN侧切片的时频资源分配问题建模为马尔科夫决策过程(MDP)。MDP的表示形式为五元组 (S, A, T_a, R_a, γ) : S 表示环境的可达状态有限集,即状态空间; A 是一个有限动作集,表示Agent可以采取的行为; $T_a(s, s') = Pr(s_{t+1} = s' | s_t = s, a_t = a)$ 是在 t 时刻环境处于 s 状态下,Agent选取动作 a 后,在 $t+1$ 时刻,状态转移到 s' 的概率; $R_a(s, s')$ 是状态 s 下采取动作 a ,状态转移到 s' 后得到环境给予的reward; $\gamma \in [0, 1]$ 是折扣因子,表示未来奖励与当前奖励之间的重要性差距。在RAN切片资源分配问题中,重点关注MDP五元组中的状态 S ,动作 A ,以及奖励 R_a 。本文对三者的定义如下:

(1) 状态空间

由于DRL是根据当前时刻的观测状态来计算并做出最佳动作,从而在保障用户SLA需求的前提下,最大化整个系统的效益。因此状态 S 需要反映出当前网络中各切片流量请求。首先记 $D_n(t)$ 为切片 n 在第 t 个调度周期的切片数据包请求数量,接下来定义状态向量 S_t 为在第 t 个切片窗口所有 N 个切片数据包请求数量 $D_n(t)$ 的集合,即 $S_t = \{D_1(t), D_2(t), \dots, D_N(t)\}$ 。另外,为了更好地捕捉环境的波动情况,将状态向量组合成待处理状态序列,经过一定预处理后得到下一切片时间窗口流量预测结果。具体而言,在第 t 个切片窗口,将过去 T 个切片窗口的状态向量 S_t 进行拼接,得到此时的待处理状态序列 O_t ,也即:

$$O_t = \{S_{t-T}, S_{t-T+1}, \dots, S_{t-1}\} \tag{9}$$

(2) 动作空间

对于本文的无线网络切片资源分配问题,动作即如何为每个业务类型的切片分配频谱资源。因此,在第 t 个切片窗口,动作 a_t 定义为:

$$a_t = \{w_1, w_2, \dots, w_N\}, \text{ s.t. } W = \sum_{n \in \mathcal{N}} w_n \tag{10}$$

需要注意的是,带宽的颗粒度会影响动作空间的大小和资源分配结果的好坏。过大的带宽颗粒度会导致资源分配不灵活,还会造成一定程度的浪费,不仅难以适应环境的动态变化,而且难以达到较优的系统效益,但其优点在于动作空间小,模型收敛速度快。而过小颗粒度的资源分配,会导致动作空间随之呈指数型增长,模型不易收敛,训练时间长。本文也将在带宽颗粒度大小方面设计实验比较其对方案性能的影响。

(3) 奖励函数

奖励函数将引导Agent学习应对每一种状态应采取何种动作,从而找到最优策略。如式(9)所示,本文的目标是最大化系统效益,这与强化学习累积回报最大化的策略相符合,所以将系统效用 J 直接作为奖励是一种直观可行的方法。但是,在一些情况下,将问题的优化目标直接作为奖励函数并不能达到理想的学习效果,需要对奖励函数进行适应于问题模型的设计,即奖励塑形。对于RAN切片资源分配问题,资源分配方案既要满足切片SLA保障,又要最大化频谱效率。但网络带宽是有限的,不同切片的SSR与SE之间可能存在相互矛盾的关系,某一项的提高并不一定带来整体系统效益的提高。因此,为了使得Agent学习的最终结果与优化目标相符,考虑到本文所涉及的三类业务中,SSR受到用户流量波动和SLA影响最大的是URLLC业务,奖励塑形机制设计如下:首先,以指示函数 $I(\cdot)$ 表示是否函数内RAN切片的SSR都达到了一定要求,如果达到则 $I(\cdot) = 1$,否则 $I(\cdot) = 0$ 。在VoLTE和Video业务SSR达标的情况下,如果 $\text{SSR}_{\text{URLLC}}$ 没有达到要求,则奖励是与 $\text{SSR}_{\text{URLLC}}$ 正相关的函数 $f_1(\text{SSR}_{\text{URLLC}})$,促使Agent向满足URLLC业务需求的方向学习;如果三种业务类型切片的SSR都满足要求,则奖励为与SE成正相关的函数 $f_2(\text{SE})$,以提高频谱效率;而当 $\text{SSR}_{\text{VoLTE}}$ 或 $\text{SSR}_{\text{Video}}$ 任意一者没有达到要求,则奖励为一负常数 $-U$ 。上述设计具体如式(11)所示:

$$r_t = \begin{cases} f_1(\text{SSR}_{\text{URLLC}}), & \text{if } I(\text{SSR}_{\text{Video}}, \text{SSR}_{\text{VoLTE}}) = 1 \ \& \ I(\text{SSR}_{\text{URLLC}}) = 0 \\ f_2(\text{SE}), & \text{if } I(\text{SSR}_{\text{Video}}, \text{SSR}_{\text{VoLTE}}, \text{SSR}_{\text{URLLC}}) = 1 \\ -U, & \text{otherwise} \end{cases} \quad (11)$$

3.2 RCLSTM-Dueling DQN 网络框架

为适应网络流量波动场景,本文在 Dueling DQN 基础之上,引入 RCLSTM 挖掘切片流量变化规律(具体细节将在 3.3 小节介绍)。RCLSTM 的记忆功能可以有效地捕捉服务需求变化,两种算法结合可以有效辅助智能体感知网络环境,提前适应变化的流量模式,达到更优的资源分配效果。将 RCLSTM 与 DRL 算法 Dueling DQN 相结合,构建一个端到端的资源配置模型,简称为 RCLSTM-Dueling DQN 模型。整体的算法结构如图 2 所示。

3.3 基于 RCLSTM 的状态处理模块

在本文的无线网络流量感知方案中,将采用预测准确度较高的等时间间隔单步预测策略,并采用改进的 LSTM 模型降低计算时间与资源开销。为适应 RAN 中有限的计算资源,需要在 LSTM 的预测性能与计算复杂度之间权衡考虑,在性能损失相对较小的前提下减少计算量。因此考虑采用将原有 LSTM 神经元连接进行了稀疏化处理的随机连接 LSTM (Random connectivity Long Short-Term Memory, RCLSTM)。

传统 LSTM 单元内的神经元连接方式为全连接,而在 RCLSTM 结构中,通过 Kovalenko 随机图模型改变了 LSTM 神经元的连通性。图 3 的上半部分为

LSTM 对应的随机图,将 LSTM 单元的神经网络视为两层的 MLP,一层为 t 时刻的输入和 $t-1$ 时刻 LSTM 单元的隐藏状态,另一层为 LSTM 隐藏神经元,两层之间的连线即表示神经元之间的连接,是需要训练的参数。LSTM 单元和 RCLSTM 单元内部神经元连接情况的对比如图 3 下半部分所示。左侧是全连接的 LSTM 单元,右侧则是一种实例化的 RCLSTM 单元,具有一定的连通率,其稀疏程度可以由阈值参数 Th 控制。RCLSTM 可以有效改善 LSTM 单元计算速度慢的缺点,当 RCLSTM 的连通率降低到一定程度,其计算速度将由于矩阵的稀疏性而得到提升。

状态处理模块中维护一个缓存区 \mathcal{T} ,用于接收待处理状态序列 $O_t = \{S_{t-T}, S_{t-T+1}, \dots, S_{t-1}\}$ 作为输入,状态处理网络由全连接神经网络和 RCLSTM 网络构成,其结构如图 4 所示。首先将输入的状态向量通过全连接神经网络进行简单的特征提取,然后将全连接神经网络的输出按照时间顺序输入到 RCLSTM 网络中迭代运算。RCLSTM 网络通过过去的 T 个观测向量捕捉数据流量随时间的变化模式,辅助 Dueling DQN 决策模块提前部署网络资源以应对网络流量波动。RCLSTM 最后一个时间步的隐藏状态结果 \hat{S}_t 将用于后续 Dueling DQN 决策模块生成动作与价值。

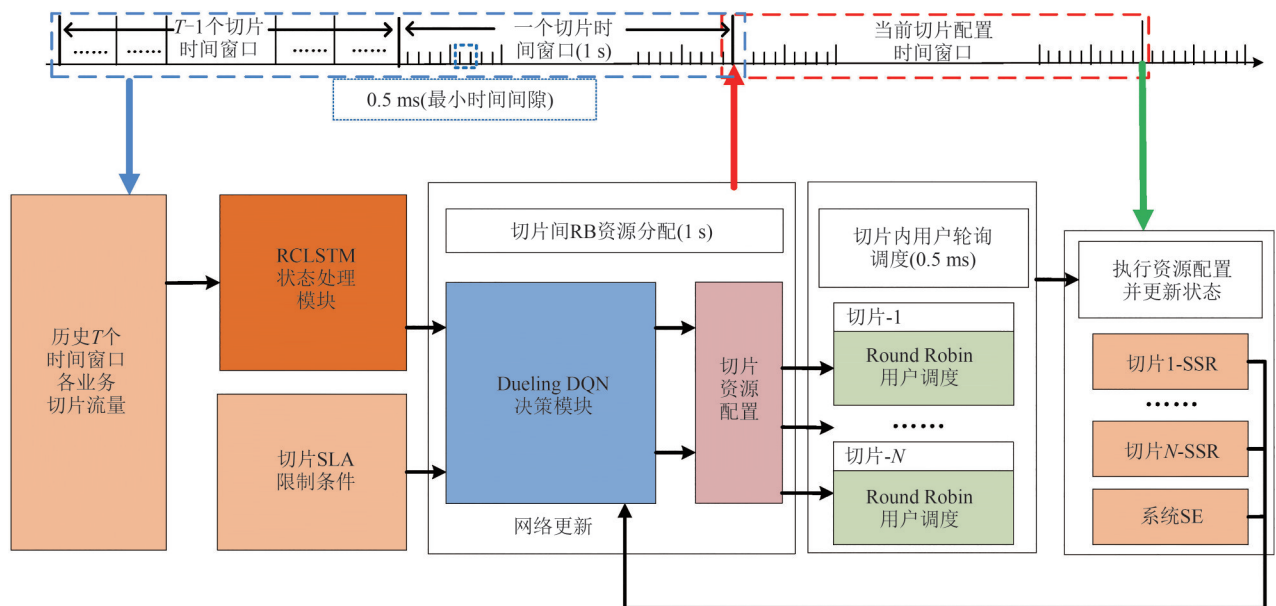


图 2 RCLSTM-Dueling DQN 算法框图

Fig. 2 Block diagram of RCLSTM-Dueling DQN algorithm

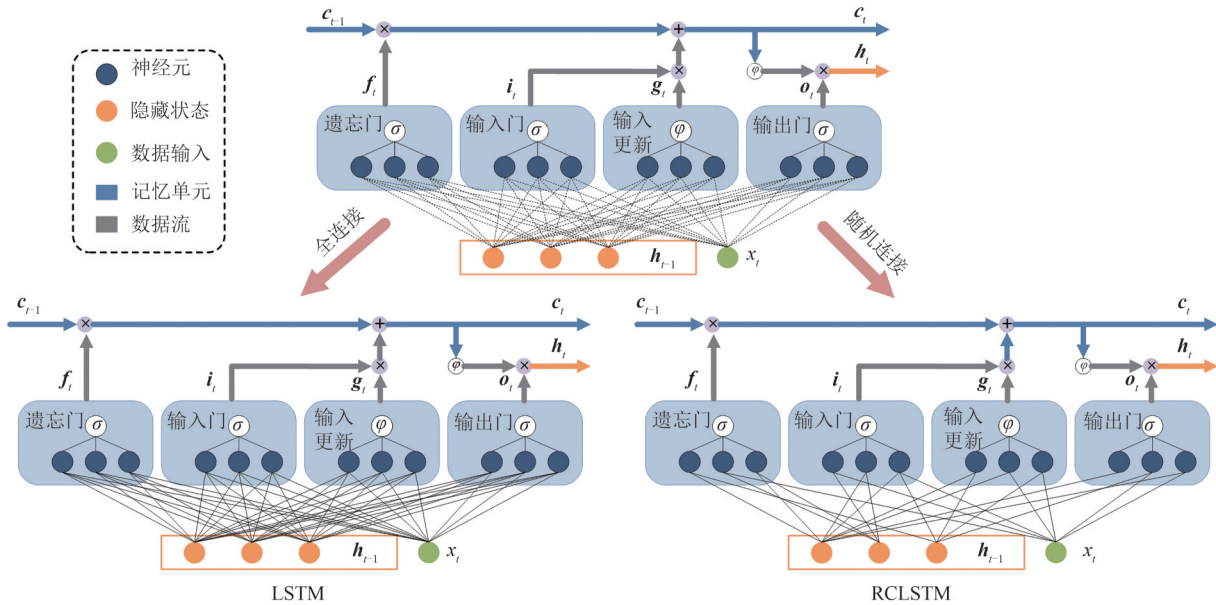


图 3 LSTM单元和RCLSTM单元的神经元连接方式

Fig. 3 Neural network connectivity of LSTM and RCLSTM units

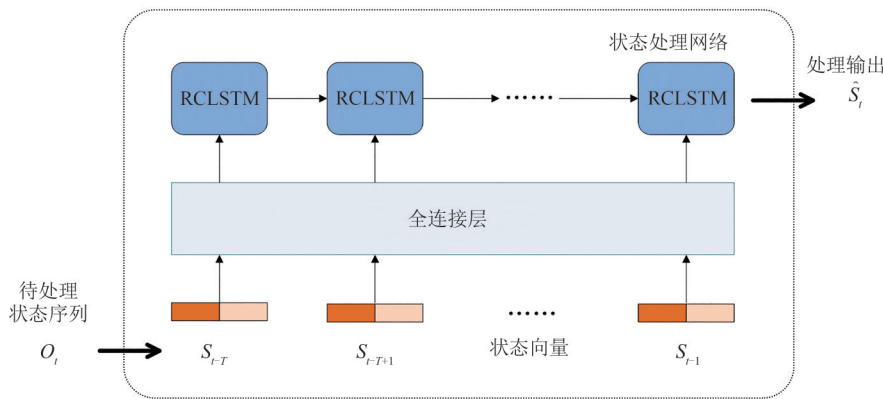


图 4 状态处理网络模型

Fig. 4 Network model for state processing

3.4 基于 Dueling DQN 的决策模块

深度强化学习的代表性免模型算法 DQN, 在解决高维计算问题和决策问题方面具有很大的优势。DQN 有一个当前策略 Q 网络和一个目标 Q 网络。目标 Q 网络的目标 Q 值为:

$$\text{Target}Q_d = r + \gamma \max_a Q_d(s', a'; \theta) \quad (12)$$

DQN 中的 Q 值更新公式记为:

$$Q_d(s, a; \theta) = Q(s, a; \theta) + \alpha (\text{Target}Q_d - Q(s, a; \theta)) \quad (13)$$

但是 DQN 也有一些缺点, 比如 DQN 只能根据某一状态下某一动作的价值 $Q(s, a)$ 对该状态进行评估, 但是当动作对状态的影响很小的时候, 可以将状态与动作进行一定程度的分离的 Dueling DQN

能够有更好的学习表现。当一个状态中有很大的动作空间时, 大部分动作可能不会对状态产生任何影响, 将状态的动作无关值和 Q 值解耦可能会使学习更健壮。Dueling DQN 提出了一种新的网络架构来实现这一想法^[20], 在上述情况下, Dueling DQN 比现有 DQN 架构能更精确地估计 Q 值。Dueling DQN 的网络结构如图 5 所示。

现在引入优势函数, 定义为价值函数和 Q 函数之间的差值。优势函数说明智能体执行的动作 a 与其他动作相比有多大的优势。因此, 状态值函数 V 表示状态的优劣, 而优势函数代表动作的优劣。把价值函数和优势函数结合起来, 即得到智能体在状态 s 下执行动作 a 的效果, 即原来的 Q 函数。如

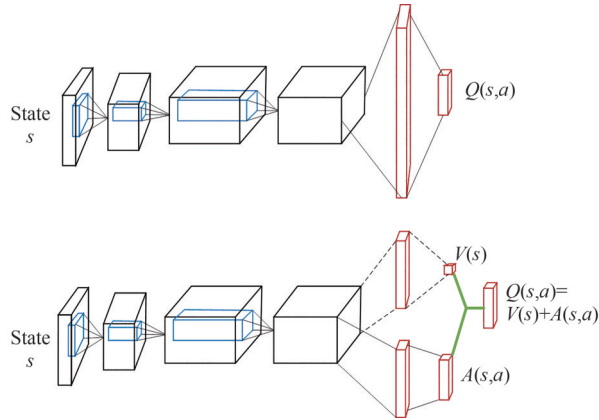


图5 DQN与Dueling DQN网络结构

Fig. 5 DQN and Dueling DQN architecture

式(14)所示:

$$Q(s, a) = V(s) + A(s, a) \quad (14)$$

Dueling DQN将这两部分表征分开:

$$Q(s, a; \theta, \theta_v, \theta_a) = V(s; \theta, \theta_v) + (A(s, a; \theta, \theta_a) - \max_{a'} A(s, a'; \theta, \theta_a)) \quad (15)$$

θ_v 、 θ_a 为非共用部分全连接层两条流的参数, θ 表示共用全连接层的参数。Dueling DQN的架构本质上与DQN相同, 只是末端的全连接层被分为两个流。一个流计算价值函数, 另一个流计算优势函数。最后使用聚合层将两个流结合得到 Q 函数。此外, 文献[20]也提出用如下的平均值代替max, 以获得更好的稳定性:

$$Q(s, a; \theta, \theta_v, \theta_a) = V(s; \theta, \theta_v) + \left(A(s, a; \theta, \theta_a) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(s, a'; \theta, \theta_a) \right) \quad (16)$$

优势函数只需直接对优势进行平均, 而不需要追求最优优势。

3.5 RCLSTM-Dueling DQN算法流程

在深度强化学习中, 策略的评估与改进两部分交替进行, 策略的评估即为Agent与环境之间进行交互, 策略的改进则是对网络参数进行更新。在本文无线接入网切片资源配置的问题中, DRL进行探索与更新的示意图如图6所示。

在初始化RCLSTM和Dueling DQN网络参数后, 首先初始化环境, 观测到状态 S_0 。其次, 在前 T 个时间步, 需要进行多次与环境的随机交互得到待处理状态序列 $O_t = \{S_1, S_2, \dots, S_T\}$ 以填充状态处理网络的缓冲区 \mathcal{T} , 此后, 从第 $T+1$ 个调度周期开始, RCLSTM-Dueling DQN将通过与环境的不断交互学习最优资源分配方法, 具体过程如算法1所示:

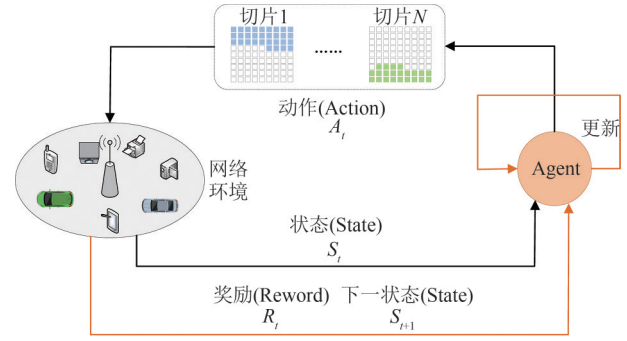


图6 RAN切片资源配置问题中智能体与环境的交互过程
Fig. 6 Interaction process between an intelligent agent and its environment in the RAN slicing resource allocation problem

算法1 RCLSTM-Dueling DQN

输入: RCLSTM网络连通率 P , 回放缓冲区容量 N , 奖励折扣因子 γ , 目标 Q 网络更新频率 ω , 学习率 α , 批大小 m , 贪婪策略因子 ϵ ;

输入: 清空回放缓存区 \mathcal{B} , 初始化主 Q 网络参数 θ_Q , RCLSTM网络参数 θ_L 。

1. 初始化目标 \hat{Q} 网络参数 $\hat{\theta}_Q \leftarrow \theta_Q$;
2. 初始化容量为 T 的空状态处理缓存区 \mathcal{T} , 时间步设为 $t = T+1$;
3. 初始化环境, 观测到状态 S_0 ;
4. **for** 时间步 $i = 1, 2, \dots, T$ **do**
5. 随机选择动作 A_i 并执行;
6. 在第 i 个调度周期结束, Agent观测到状态 S_i ;
7. 将 S_i 添加到状态处理缓存区末尾;
8. **end for**
9. **for** 时间步 $t = T+1, T+2, T+3, \dots$ **do**
10. 将状态处理缓存区的每一个观测拼接得到待处理状态序列 $O_t = \{S_{t-T}, S_{t-T+1}, \dots, S_{t-1}\}$;
11. 状态处理网络接收输入 O_t 并计算输出 \hat{S}_t ;
12. 将预测结果 \hat{S}_t 输入到决策网络模块;
13. 决策模块以 ϵ 的概率随机选择动作 A_t , 否则选择动作 $A_t = \arg \max_a Q(\hat{S}_t, a; \theta_Q)$;
14. 执行动作 A_t , 得到下一观测 S_t , 并得到奖励 R_t ;
15. 删除状态处理缓冲区中的第一个观测, 并将 S_t 添加到末尾;
16. 如果回合结束, 设置 $B_t = 1$, 否则设置 $B_t = 0$;
17. 将经验 $(\hat{S}_t, A_t, R_t, S_t, B_t)$ 保存到回放缓存区 \mathcal{B} ;
18. 从 \mathcal{B} 中随机采样 m 个元组 $(S'_i, A_i, R_i, S_i, B_i)$, $i = 1, 2, \dots, m$;
19. 如果 $B_t = 0$, 设置 $Y_t = R_t + \gamma \max_{a'} \hat{Q}(S_t, a'; \hat{\theta}_Q)$, 否则设置 $Y_t = R_t$;
20. 使用梯度下降算法更新 Q 网络参数 θ_Q , 损失函数为 $(Y_t - Q(\hat{S}_t, A_t; \theta_Q))^2$;
21. 每 ω 时间步更新目标 \hat{Q} 网络参数;
22. 如果回合结束, 跳出循环;
23. **end for**

4 仿真结果

本小节搭建了一个RAN切片场景,验证所提出的RCLSTM-Dueling DQN方案在资源分配问题中的性能。该场景中存在一个基站,服务于三种不同业务类型的切片:VoLTE、Video和URLLC。在以基站为中心的50 m范围内,随机分布100名用户。三种业务类型对应服务用户数目之比为3:6:1。网络切片环境的物理参数基于3GPP TR 36.814标准^[22]设置,具体如表1所示。

另外基于TS 22.261标准^[23],本文分别根据VoIP模型参数设置VoLTE标准业务流量,根据视频流模型参数设置Video业务流量,以及根据FTP2模型参数设置URLLC切片业务流量,具体数据如表2所示。不同类型的切片接收相应业务类型用户的服务请求,请求到达间隔与数据包大小服从一定概率分布。

系统总带宽为18 MHz,由于带宽分配粒度的不同会影响动作空间大小,并影响最终资源分配效果,因此本文设置两种带宽分配粒度分别为1.8 MHz和180 kHz,并将展示两种情况的仿真结果。在一个切片配置窗口内,以0.5 ms的粒度作为在每个切片内执

$$r_i = \begin{cases} (SSR_{URLLC} - 0.6) \times 10, & \text{if } SSR_{Video} \geq 0.97 \ \& \ SSR_{VoLTE} \geq 0.97 \ \& \ SSR_{URLLC} < 0.9 \\ 5 + \alpha SE, & \text{if } SSR_{Video} \geq 0.97 \ \& \ SSR_{VoLTE} \geq 0.97 \ \& \ SSR_{URLLC} \geq 0.9 \\ -4, & \text{otherwise} \end{cases} \quad (17)$$

4.1 RCLSTM 仿真结果

为了验证RCLSTM模型在预测准确性与计算资源节约方面的优势,搭建RCLSTM网络设计仿真实验,对上一小节所述无线网络流量数据集进行预测。所使用的计算机硬件设备包括Intel Core i5-12400F处理器、NVIDIA GeForce GTX显卡,使用PyTorch深度学习框架,模型的训练阶段采用GPU加速计算,测试阶段则采用CPU。划分训练集和测试集的比例为7:3,构建数据集所用滑动时间窗口长度为100。

表1 无线网络环境仿真参数设置

Tab. 1 Wireless network environment simulation parameter settings

参数名称	参数值
系统带宽	18 MHz
RB颗粒度	1.8 MHz, 180 kHz
小区半径	50 m
基站发射功率	46 dBm
接收天线增益	20 dB
载波频率	2 GHz
高斯白噪声功率谱密度	-174 dBm/Hz
信道	瑞利衰落

行用户调度的周期,即每0.5 ms按顺序将每个切片的带宽分配给该切片内的活跃用户。而一个切片配置窗口包含2000个调度时隙,即1 s,在这1 s内,分配给每个切片的带宽资源保持不变。因此,智能体每秒一次更新其神经网络。

本文的优化目标是在保障所有切片满意度的同时尽可能提高系统效益,因此对奖励函数进行塑形是很有意义的。具体奖励塑形函数设计如式(17)所示:

为了验证RCLSTM模型在预测准确性与计算资源节约方面的优势,首先考察RCLSTM网络的预测结果,在RCLSTM网络中,每个门包含400个神经元,即有400个隐藏状态。神经元之间的连通率可以调节,连通率是性能评估方面的重要参数。RCLSTM网络以RMSE为损失函数,使用Adam算法对权重参数进行优化。这里以Video服务类型的训练结果为例进行展示,RCLSTM连通率设为40%。

图7(a)直观地显示了RCLSTM网络预测所得

表2 RAN切片流量关键参数设置

Tab. 2 RAN slicing traffic key parameter settings

业务类型	VoLTE	Video	URLLC
切片资源配置周期		1 s	
切片内调度方式		Round robin (0.5 ms)	
用户数量	30	60	10
用户数据包到达时间间隔	均匀分布 (Min=0, Max=160 ms)	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)
用户数据包大小	常数 (40 Byte)	截断帕累托分布 (指数系数=1.2, Mean=100 Byte, Max=250 Byte)	可变常数: {6.4, 12.8, 19.2, 25.6, 32} KByte
SLA时延需求	51 Kbps	100 Mbps	10 Mbps
SLA速率需求	10 ms	10 ms	1 ms

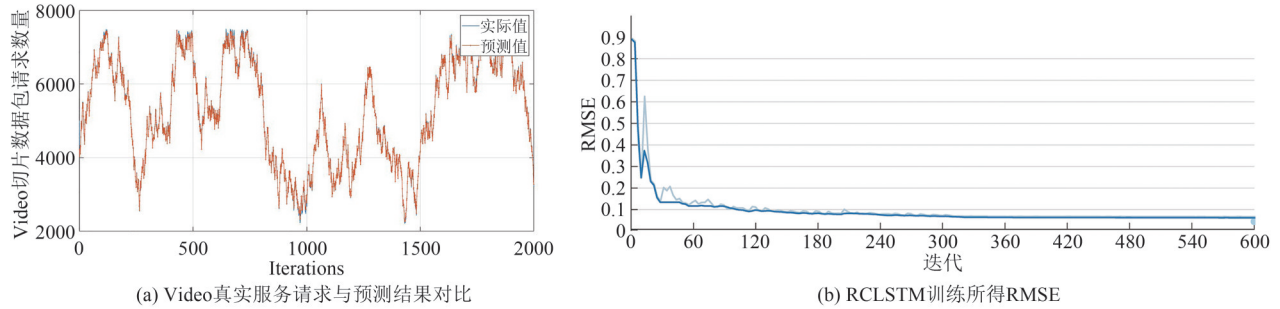


图7 Video服务请求及模型训练结果

Fig. 7 Video service request and model training results

结果与真实数据请求的对比关系,可以看出此时预测结果能够有效跟踪业务请求的变化趋势、捕捉其波动特征,尤其是在时间步为200、800、1500左右,真实流量发生较大的增长或下降时,RCLSTM模型也能很好的跟踪其变化趋势。图7(b)显示了RCLSTM训练所得RMSE,可以看出模型能够快速收敛并达到较高的预测准确度。

接下来验证RCLSTM连通率对预测准确性和计算时间的影响,仿真结果如图8所示。从运算时间上看(如图8紫色曲线所示),预测所需的计算时间总体上随RCLSTM神经元连通率的增加而变长。当神经元连通率为1%,计算时间约为6.23 s,此后,随着神经元连通率增加,计算时间增长较快,直到连通率为40%时出现转折,此时计算时间约为8.79 s,而在这之后计算时间几乎保持不变,当RCLSTM的连通率为100%(即包含400个隐藏神经元的传统LSTM)时,计算时间约为最长的8.81 s。综合来看,神经元之间的连通率对于计算速度有显著影响,连通率为10%的RCLSTM比连通率为100%的RCLSTM计算时间减少了11%。在连通率为40%时出现转折的原因是,当连通率降到40%及更低的水平时,较低的连通使得RCLSTM权重矩阵的稀疏程度降低到足以使稀疏计算发挥优势,从而降低了计算时间。另一方面,如图8蓝色曲线所示,连通率较低的神经元在提升RCLSTM模型运算速度的同时,并没有明显减损模型预测的准确度。当RCLSTM的连通率为100%(即传统LSTM),其预测误差RMSE在0.049左右,随着连通率的降低,准确度有所降低;这种趋势在连接比例为70%时出现转折,此后连通率对RMSE的影响变弱。与100%神经元连接的RCLSTM相比,1%的连通率以仅38%的性能损失带来了99%的训练参数减少。

综上所述,RCLSTM模型在预测准确度和计算时间方面是一种极具优势的无线网络流量预测解决方案。RCLSTM模型被证明可以提前感知网络

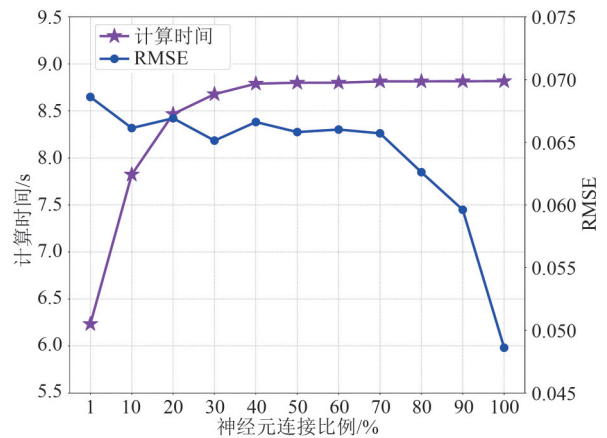


图8 不同RCLSTM连通率下计算时间与预测误差
Fig. 8 Computation times and prediction errors under different RCLSTM connectivity ratios

状态,辅助Dueling DQN提升资源配置的时效性,使整个RAN侧切片资源配置达到更优的效果。

4.2 RCLSTM-Dueling DQN 仿真结果

对于RCLSTM状态处理模块,输入的历史时间窗口观测向量为 $T=20$ 个,SE的权重系数 α 设为0.01,SSR系数 β 设为 $[1, 1, 1]$,连通率选取上文实验中在运算时间与误差之间表现较好折中的10%。为了充分评估RCLSTM-Dueling DQN算法的性能,本文将其与两种经典强化学习算法,即DQN和A2C算法,以及非动态性的为每类切片均匀分配资源的方法进行比较。

首先考虑资源分配粒度为1.8 MHz,对所有算法进行10000轮迭代,图9(a)至图9(d)分别对比了DQN算法、A2C算法、RCLSTM-Dueling DQN方法和均等分配四种方案在三种RAN切片的SSR及频谱效率上的性能表现。首先,在SSR性能方面。由于VoLTE业务对吞吐量和时延需求都不高,因而由图9(a)可见其在任何方法下都可以达到100%的SSR。URLLC业务由于奖励塑形函数中针对 SSR_{URLLC} 的设计而如

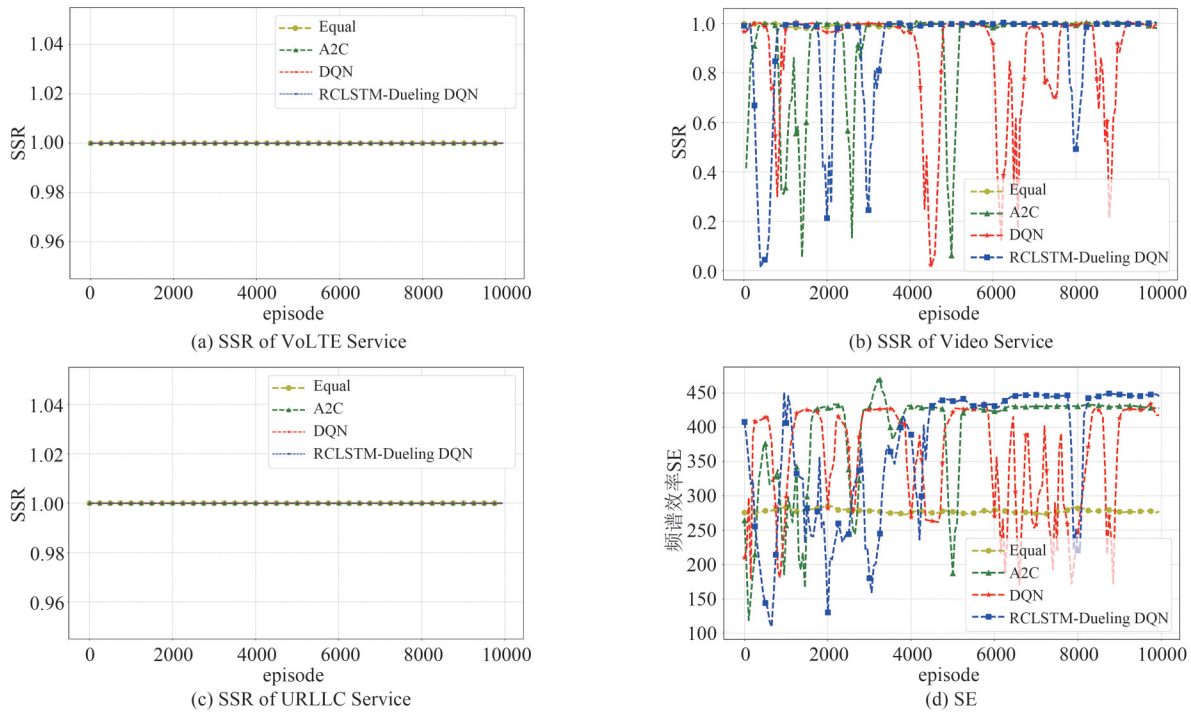


图 9 资源分配粒度为 1.8 MHz 时不同算法的 SSR 与频谱效率 SE

Fig. 9 Spectral efficiencies and SSRs of different algorithms at a resource allocation granularity of 1.8 MHz

图 9(c) 所示相对容易达到 100% 的 SSR。而 Video 业务, 由于其相对较高的吞吐量要求, 经过大约 3500 轮迭代, RCLSTM-Dueling DQN 算法能够收敛并达到 100% 的 SSR。注意在训练后期存在异常的突然性能下降(如图 9(b) 中第 8000 左右轮次迭代时)是由于 ϵ -greedy 探索策略中微小的探索概率引起的。另外, A2C 方法在经过大约 5000 轮迭代后, 基本能达到 100% 的 SSR, DQN 方法收敛状态不明显。其次, 在 SE 性能方面, 从图 9(d) 中可以看到, 随着训练轮次的增加, 平均分配方式的 SE 始终保持在一个较低的水平, 约 275 bps/Hz, DQN 方法的 SE 一直处于波动状态, 最大值为 425 bps/Hz 左右, A2C 方法在 5000 轮之后性能基本收敛, SE 稳定在 430 bps/Hz 左右, 而 RCLSTM-Dueling DQN 算法表现最优, 经过了约 4000 轮迭代后, 其 SE 基本稳定在将近 450 bps/Hz, 相较于 A2C 算法, 其收敛时间更快, 频谱效率也更高。

接下来分析上述结果出现的原因。由式 (12) 可知, 在奖励函数的设计中, 需要优先满足 URLLC 切片的 SSR, 其次是另外两种切片。在不同切片资源配置周期, 用户服务请求数量存在较大波动。由于 DQN 方法和 A2C 方法无法捕捉用户服务请求变化规律, 为了获得更高的奖励, 这两者更倾向于将较多资源分配给 URLLC 业务, 并追求三种业务切片 SSR 最大化, 但这导致了整个系统 SE 的降低。

然而, RCLSTM-Dueling DQN 算法则通过 RCLSTM 状态处理模块, 提前对环境进行状态感知, 有效捕获环境变化过程中隐含的时间相关性, 并以此作为基础进行灵活的切片资源配置决策, 最终实现在满足三种切片 SLA 的同时尽可能提高频谱效率。另外, RCLSTM-Dueling DQN 方案的收敛速度要快于 DQN 方法和 A2C 方法, 这说明通过 RCLSTM 捕捉环境状态变化规律有助于加速训练进程。

图 10 展示了四种方法的系统效用随迭代轮次的变化。由蓝色曲线可以看出, RCLSTM-Dueling DQN 方法的系统效用在前三千轮训练中存在明显的上升趋势, 之后维持在一个相对较高的水平。

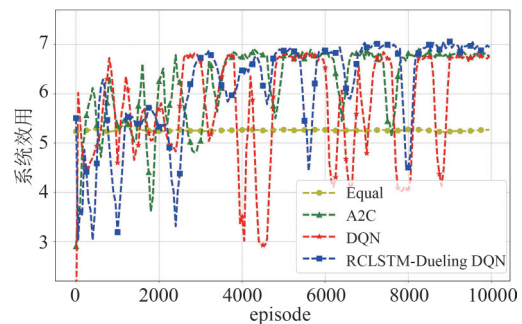


图 10 资源分配粒度为 1.8 MHz 时不同算法的系统效用
Fig. 10 System utilities of different algorithms at a resource allocation granularity of 1.8 MHz

A2C方法也能够收敛并达到较高的系统效用,但其最终表现略差于RCLSTM-Dueling DQN。DQN方法得到的结果依然波动较大,没有明显收敛趋势。总之,在资源分配粒度为1.8 MHz的情况下,三种智能化动态RAN切片配置方法达到的系统效用都明显优于均匀分配方案,并随着迭代次数的增加最终可以达到相近的性能。

在后续的实验中,将资源分配粒度设为180 kHz,以考察更细颗粒度的分配方式和较大动作空间对各种方案的性能影响。为实现较大动作空间下算法的收敛,相应仿真迭代轮次由10000增加到20000。首先,VoLTE的SSR情况如图11(a)可见,由于资源颗粒度的细化导致动作选择多样化,DQN算法结果由原本1.8 MHz粒度下稳定为1,变为极不稳定,而A2C方法和RCLSTM-Dueling DQN方法则几乎保持100%的SSR。图11(b)和图11(c)中,对于Video和URLLC切片,三种算法都保持了较好的性能,只有个别轮次出现一些波动。图11(d)展示了SE的变化情况,几种算法的趋势与更大带宽分配粒度下的结果相似,但当三种基于强化学习的方法收敛之后,RCLSTM-Dueling DQN方法与另外两种方法之间的性能差距与分配粒度为1.8 MHz时相比有所增大。这说明Dueling DQN的引入使得在较大动作空间下,对动作优势值和状态

值的分别估计可以带来明显的性能提升。

图12展示了资源分配粒度为180 kHz时四种方案的系统效用情况,其整体趋势与SE相似。通过与图10对比可见,RCLSTM-Dueling DQN方法拉大了与其他方法在系统效用上的差距,说明其在资源得到更加精细的划分之后,能够更加合理的利用资源,提高利用率,减少浪费。而DQN算法则受到动作空间增大的负面影响更大,系统效用有所降低,资源精细化并没有抵消动作空间过大的影响。

图13(a)至图13(c)展示了资源分配粒度为180 kHz时三种智能方法的资源分配策略。可以看出,在迭代初期阶段,为了迎合奖励塑形函数的设计,三种算法都为URLLC切片分配较多的资源,但是随着迭代的进行,由于RCLSTM-Dueling DQN方法能够很好地感知环境变化,提前预知三种切片的流量变化情况,该方法将更多资源分配给流量需求较大的Video切片,这样可在满足切片SLA的同时获得更高的频谱效率,进而提升整体的系统效用。具体而言,例如从图12中RCLSTM-Dueling DQN方法的曲线和图13(c)中RCLSTM-Dueling DQN方法的曲线对照可见,在5000轮次左右和12500轮次左右,RCLSTM-Dueling DQN进行两次较大的切片资源配置调整使得对应的两个时间点的系统效用得到很大提升。

接下来就奖励塑形对算法学习的影响进行研

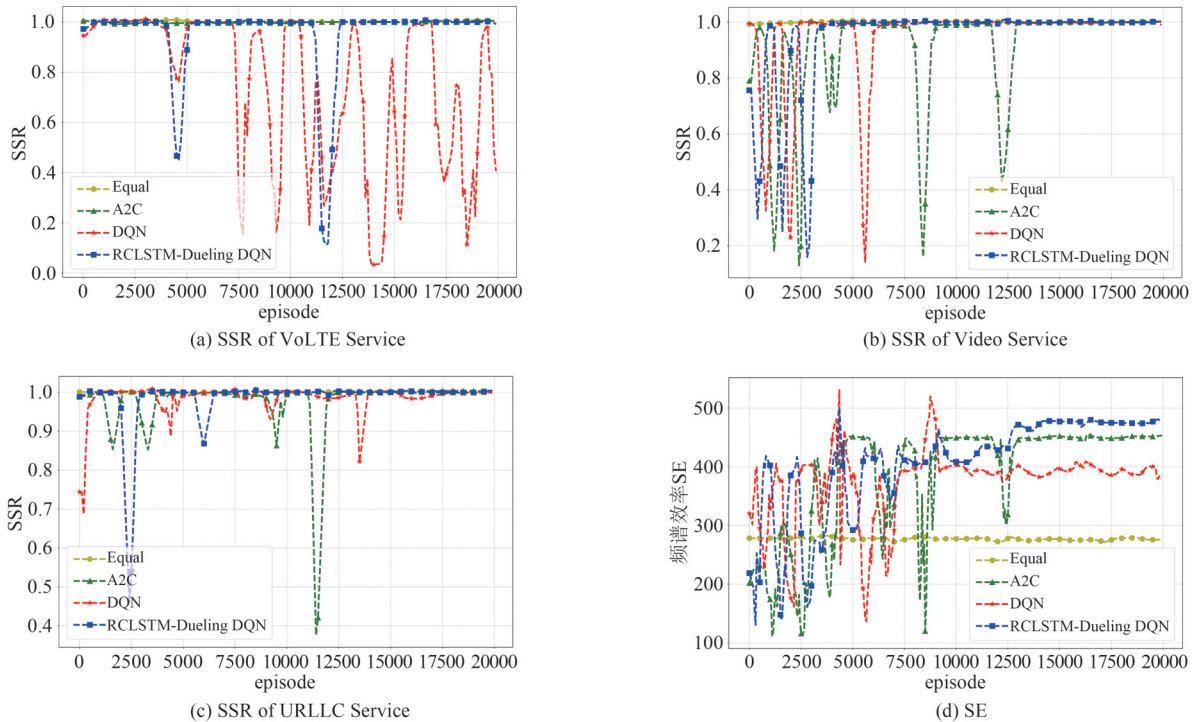


图11 资源分配粒度为180 kHz时不同算法的SSR与频谱效率SE

Fig. 11 Spectral efficiencies and SSRs of different algorithms at a resource allocation granularity of 180 kHz

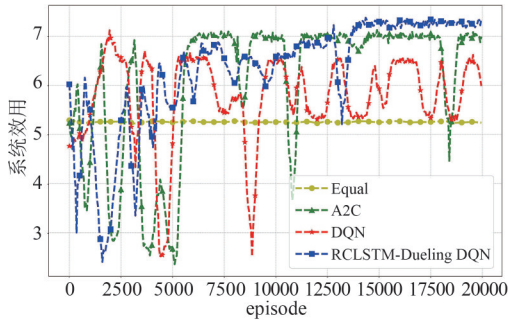


图 12 资源分配粒度为 180 kHz 时不同算法的系统效用
Fig. 12 System utilities of different algorithms at a resource allocation granularity of 180 kHz

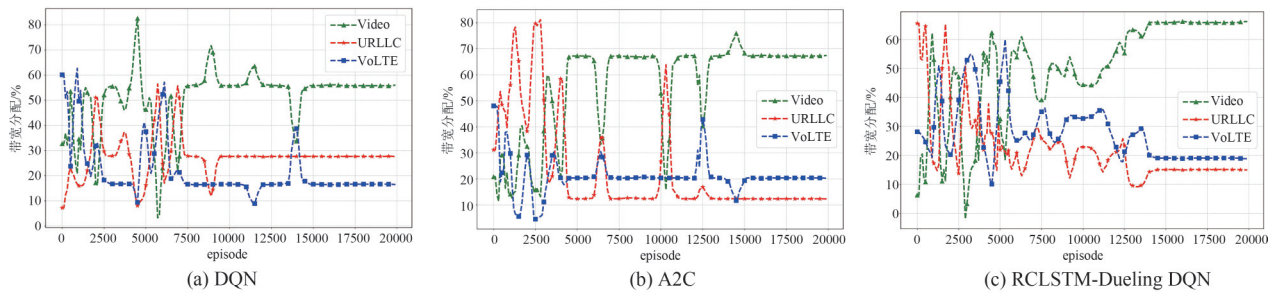


图 13 资源分配粒度为 180 kHz 时不同方案资源分配结果
Fig. 13 Resource allocation results of different algorithms at a resource allocation granularity of 180 kHz

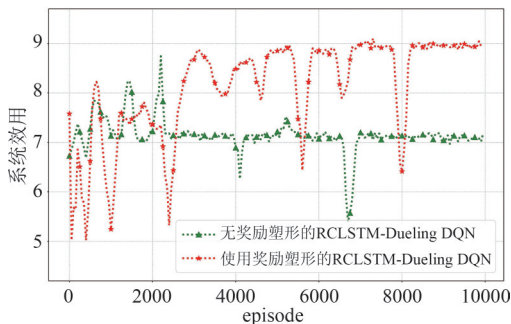


图 14 有无奖励塑形下 RCLSTM-Dueling DQN 算法的系统效益
Fig. 14 System utilities of RCLSTM-Dueling DQN algorithm with and without reward shaping

5 结论

本文为移动无线通信场景中的 RAN 切片资源管理设计了一个智能化解决方案,为保障切片 SLA 需求,将动态无线资源分配问题建模为马尔科夫决策过程,提出使用 RCLSTM 和 Dueling DQN 算法结合起来最大化 RAN 切片的 SLA 满意率和频谱效率。RCLSTM 网络通过控制传统 LSTM 网络中神经元连接比例,在保障预测准确性的同时降低传统 LSTM 网络的计算复杂度,以适应 RAN 中有限的计算资源。10% 连接比例的 RCLSTM 网络能在保持

图 14 显示了当带宽分配颗粒度为 1.8 MHz 时,有无奖励塑形的 RCLSTM-Dueling DQN 在学习过程中得到的系统效用差异。当不使用奖励塑形机制时,RCLSTM-Dueling DQN 直接采用系统效用作为奖励函数,这种方法没有分阶段的考虑 SSR 和 SE 之间可能存在的大小关系矛盾,并且一体化的加权求和使得系统效用在数值上存在较大变化,因而无法有效指导智能体的学习。从图 14 中可以看出,RCLSTM-Dueling DQN 在有奖励塑形情况下的性能明显优于没有奖励塑形的情况。仿真结果验证了 RCLSTM-Dueling DQN 方案中奖励塑形设计的有效性。

极低性能损失的同时将原始 LSTM 的计算时间降低约 11%。Dueling DQN 能够提高切片策略学习过程中 Q 值的估计精确度以提升收敛速度。智能决策方法 RCLSTM-Dueling DQN 的设计能够在动态网络切片环境中提前感知网络性能变化、获取历史流量数据中的隐藏模式,有效降低网络环境波动对密集流量场景下无线切片资源管理的影响。实验结果表明,在具有三种不同流量波动模式及服务需求需求的 RAN 切片场景中,与传统经典 DQL 类算法如 DQN、A2C 以及硬切片方法相比,RCLSTM-Dueling DQN 方法可以获得更快的收敛速度、更高的频谱效率和 100% 的切片 SLA 满意率。

参考文献

- [1] ROST P, MANNWEILER C, MICHALOPOULOS D S, et al. Network slicing to enable scalability and flexibility in 5G mobile networks [J]. IEEE Communications Magazine, 2017, 55(5): 72-79.
- [2] YOU Xiaohu, WANG Chengxiang, HUANG Jie, et al. Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts [J]. Science China Information Sciences, 2021, 64(1): 110301.
- [3] AFOLABI I, TALEB T, SAMDANIS K, et al. Network slicing and softwarization: A survey on principles,

- enabling technologies, and solutions[J]. IEEE Communications Surveys & Tutorials, 2018, 20(3): 2429-2453.
- [4] SHAYEA I, ERGEN M, HADRI AZMI M, et al. Key challenges, drivers and solutions for mobility management in 5G networks: A survey[J]. IEEE Access, 2020, 8: 172534-172552.
- [5] 3GPP. Architecture enhancements for dedicated core networks[S]. 3GPP TR 23.707 version 13.0.0 Release 13, 2014.
- [6] LI Hongjia, WEI Tianshu, REN Ao, et al. Deep reinforcement learning: Framework, applications, and embedded implementations[EB/OL]. 2017: arXiv: 1710.03792. <https://arxiv.org/abs/1710.03792>.
- [7] CASAS-VELASCO D M, RENDON O M C, DA FONSECA N L S. DRSIR: A deep reinforcement learning approach for routing in software-defined networking[J]. IEEE Transactions on Network and Service Management, 2022, 19(4): 4807-4820.
- [8] JIANG Wei, HAN Bin, HABIBI M A, et al. The Road towards 6G: A comprehensive survey[J]. IEEE Open Journal of the Communications Society, 2021, 2: 334-366.
- [9] KHAN L U, YAQOOB I, TRAN N H, et al. Network slicing: Recent advances, taxonomy, requirements, and open research challenges[J]. IEEE Access, 2020, 8: 36009-36028.
- [10] 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.
- [11] SUN Guolin, XIONG Kun, 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.
- [12] SUN Guolin, GEBREKIDAN Z T, BOATENG G O, et al. Dynamic reservation and deep reinforcement learning based autonomous resource slicing for virtualized radio access networks[J]. IEEE Access, 2019, 7: 45758-45772.
- [13] XI Ranran, CHEN Xin, CHEN Ying, et al. Real-time resource slicing for 5G RAN via deep reinforcement learning[C]//2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS). Tianjin, China. IEEE, 2020: 625-632.
- [14] AZZOUNI A, PUJOLLE G. A long short-term memory recurrent neural network framework for network traffic matrix prediction[EB/OL]. 2017: arXiv: 1705.05690. <https://arxiv.org/abs/1705.05690>.
- [15] LI Rongpeng, ZHAO Zhifeng, ZHENG Jianchao, et al. The learning and prediction of application-level traffic data in cellular networks[J]. IEEE Transactions on Wireless Communications, 2017, 16(6): 3899-3912.
- [16] GHOJOGH B, GHODSI A. Recurrent neural networks and long short-term memory networks: tutorial and survey[EB/OL]. 2023: arXiv: 2304.11461. <https://arxiv.org/abs/2304.11461>.
- [17] CHUNG J, KASTNER K, DINH L, et al. A recurrent latent variable model for sequential data[EB/OL]. 2015: arXiv: 1506.02216. <https://arxiv.org/abs/1506.02216>.
- [18] XU K, BA J, KIROS R, et al. Show, attend and tell: Neural image caption generation with visual attention[EB/OL]. 2015: arXiv: 1502.03044. <https://arxiv.org/abs/1502.03044>.
- [19] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. 2014: arXiv: 1409.0473. <https://arxiv.org/abs/1409.0473>.
- [20] WANG Ziyu, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48. New York, NY, USA. New York: ACM, 2016: 1995-2003.
- [21] HUA Yuxiu, ZHAO Zhifeng, LI Rongpeng, et al. Deep learning with long short-term memory for time series prediction[J]. IEEE Communications Magazine, 2019, 57(6): 114-119.
- [22] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects[S]. 3GPP TR 36.814 version 9.0.0 Release 9, 2010.
- [23] 3GPP. Service Requirements for Next Generation New Services Markets[S]. 3GPP TS 22.261 version 15.0.0 Release 15, 2017.

作者简介



赵晨女, 1998年生, 辽宁铁岭人。东南大学硕士生, 主要研究方向为无线接入网切片。

E-mail: 220200779@seu.edu.cn



张铖男, 1988年生, 安徽望江人。博士, 东南大学副研究员, 博士生导师, 主要研究方向为无线通信系统中的空时信号处理、机器学习辅助的无线通信智能优化技术等。

E-mail: zhangcheng_seu@seu.edu.cn



黄永明男, 1977年生, 江苏吴江人。博士, 网络通信与安全紫金山实验室普适通信研究中心主任, 东南大学信息学院副院长, 国家杰出青年基金获得者, 博士生导师, 主要研究方向为智能5G/6G移动通信、毫米波无线通信等。

E-mail: huangym@seu.edu.cn

(责任编辑: 边熙淳)