



研究与开发

基于OS-MBRL的网络切片资源动态分配方法研究

严嘉辉¹, 钟玮轩¹, 董黎刚¹, 蒋献¹, 王广昌², 陆凌蓉²

(1. 浙江工商大学, 浙江 杭州 310018;

2. UT斯达康通讯有限公司, 浙江 杭州 310051)

摘要: 随着网络用户业务需求的增长, 如何实现网络切片动态和准确的资源分配是当下网络必须解决的问题。考虑传统无模型强化学习方法需要较长的模型训练时间, 提出了一种基于OS-MBRL (model based RL supported by online SVM) 的网络资源动态分配方法。该方法利用在线支持向量机算法构建了一个系统模型, 保证在分配较少资源的情况下产生较低的服务等级协议 (service level agreement, SLA) 违规次数。仿真实验结果表明, 与归一化优势函数 (normalized advantage function, NAF) 算法、深度Q网络 (deep Q-network, DQN) 算法和双延迟深度确定性策略梯度 (twin delayed deep deterministic policy gradient, TD3) 算法相比, 该方法能够最高减少80%的SLA违规次数, 同时降低9%的资源分配。

关键词: 网络切片; 资源分配; 强化学习; 在线支持向量机算法

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024228

Research on dynamic allocation of network slicing resources based on OS-MBRL

YAN Jiahui¹, ZHONG Weixuan¹, DONG Ligang¹, JIANG Xian¹, WANG Guangchang², LU Lingrong²

1. Zhejiang Gongshang University, Hangzhou 310018, China

2. UT Starcom Telecom Co., Ltd., Hangzhou 310051, China

Abstract: With the growth of business needs of network users, how to achieve dynamic and accurate resource allocation of network slicing is a problem that must be solved in the current network. Considering that traditional modelless reinforcement learning methods require a longer model training time, a dynamic resource allocation method based on OS-MBRL was proposed. The online support vector machines algorithm was utilized to construct a system model that could handle dynamically changing data streams and continuously update the model to adapt to new data, ensuring a

收稿日期: 2024-05-08; 修回日期: 2024-07-29

通信作者: 董黎刚, donglg@zjsu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62301488); 浙江省重点研发计划项目 (No.2021C01036); 浙江省高等教育学会高等教育研究重点立项课题项目 (No.KT2022017); 浙江省新型网络标准与应用技术重点实验室项目 (No.2013E10012)

Foundation Items: The National Natural Science Foundation of China (No.62301488), The Key Research and Development Program of Zhejiang Province (No.2021C01036), Zhejiang Higher Education Association Higher Education Research Key Project (No.KT2022017), Zhejiang Provincial Key Laboratory of New Network Standards and Application Technologies (No.2013E10012)



lower number of SLA violations when allocating fewer resources. Simulation experiment results show that compared with NAF algorithm, DQN algorithm, and TD3 algorithm, the proposed method can reduce SLA violations by up to 80% and resource allocation by 9%.

Key words: network slicing, resource allocation, reinforcement learning, online SVM algorithm

0 引言

近年来,网络切片被广泛视为解决网络资源利用率较低问题的关键技术之一^[1],通过在一个共同的物理基础设施上部署多个逻辑网络来实现资源共享。每个逻辑网络提供例如增强型移动宽带(enhanced mobile broadband, eMBB)、大规模机器类通信(massive machine type communications, mMTC)或超可靠低时延通信(ultra reliable and low latency communication, URLLC)等特定类型的服务,而这些逻辑网络个体被称为网络切片^[2]。这种创新确保了从关键业务到大众消费服务的各种应用场景都能获得优化的网络体验,并且增强了网络的灵活性和可扩展性。

在当今用户流量剧增的环境下,网络切片的资源动态分配能够有效地避免资源配置过度或配置不足,但大部分针对网络切片资源动态分配问题的研究都聚焦于传统的无模型强化学习算法。Albonda等^[3]提出一种基于离线强化学习的高效无线电接入网(radio access network, RAN)切片方案,目的是在保证资源可用性的同时,最大限度地提高资源利用率;Wang等^[4]提出一种基于强化学习的网络切片资源分配方法,并根据不同资源类型来定义效用函数;Li等^[5]研究了用深度强化学习(deep reinforcement learning, DRL)算法解决一些网络切片场景中的资源管理。其中文献[3]使用的Q学习(Q-learning)算法和文献[5]使用的深度Q学习(deep Q-learning, DQL)算法,都是为离散动作空间设计的,且随着网络切片数量的增加,分配动作数量也呈指数级增长,这使得其提出的方法变得不再可行。针对该问

题,更多解决方案被提出。其中,Guo等^[6]提出一种基于双延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient, TD3)算法的eMBB和URLLC穿刺预测公平性调度方案,该方案显著减少了URLLC穿刺实例,提高了eMBB吞吐量并成功实现了eMBB用户之间的公平性;归一化优势函数(normalized advantage function, NAF)是一种将深度Q网络(deep Q-network, DQN)算法扩展到连续动作空间的技术,在文献[7]中用于网络切片之间的资源分配。虽然上述强化学习算法可以获得更快的收敛速度,但其针对资源分配和服务等级协议(service-level agreement, SLA)实现方面的问题依然存在较大改进空间。基于此,本文提出一种基于OS-MBRL(model-based RL supported by online SVM)的网络切片资源动态分配方法,主要贡献和创新如下。

(1) 针对现有无模型强化学习算法中存在的低学习效率问题,以及现有基于模型的强化学习算法中资源分配效率与SLA低违规率不能同时满足的问题,本文考虑使用在线支持向量机(online support vector machine, online SVM)算法对现有切片资源分配方法进行优化。首次引入online SVM算法来构建一个系统模型,用于理解和模拟决策过程,旨在实现算法模型的在线学习功能,并降低算法模型在训练过程中对网络系统的影响。

(2) 本文提出一种基于OS-MBRL的网络切片资源动态分配方法。利用基于模型的强化学习算法解决网络中的资源分配问题,通过引入online SVM算法,实现基于OS-MBRL的系统模型

在线学习能力, 保证在网络切片低 SLA 违规率的同时降低网络资源的分配。

1 国内外研究现状

用户设备数量激增和需求多样化, 对网络切片服务的需求日益复杂。这给网络服务提供商带来了挑战: 必须快速且精准地处理业务请求, 并根据具体需求定制网络切片方案。因此, 实现网络切片动态和准确的资源分配显得尤为关键^[8]。

解决网络切片资源分配问题的主流观点是采用强化学习 (reinforcement learning, RL) 方法。Sohaib 等^[9]提出一种基于半监督学习和深度强化学习的架构, 对网络资源进行智能管理, 在保证 URLLC 可靠性要求的同时保持了较高的 eMBB 平均速率和较高的收敛率。Nassar 等^[10]提出一个基于, 如 Q 学习、SARSA (state-action-reward-state-action)、期望 SARSA 等, 各种强化学习算法的资源分配框架, 旨在最大化 5G 雾无线电接入网 (fog computing based radio access network, F-RAN) 中资源的利用率, 同时确保物联网应用的低延迟要求。Zhang 等^[11]在深度学习的基础上提出一种约束风险敏感深度强化学习 (constrained risk-sensitive deep reinforcement learning, CRS-DRL) 方法, 在涉及 eMBB 和 URLLC 的动态多路复用场景中进行联合调度。Sun 等^[12]在强化学习的基础上引入深度学习概念, 提出一种基于 DQN 算法的网络资源分配策略, 目的在于实现服务质量 (quality of service, QoS) 满意度和网络切片资源利用率的平衡。陈赓等^[13]提出一种基于双 DQN (duelling double deep Q-network, D3QN) 的双时间尺度网络切片资源分配算法, 以归一化频谱效率和不同切片用户指标 QoS 的加权和为优化目标, 有效提升了资源切片与调度联合的效率。

文献[9]一个共同的显著特征是采用了无模型

强化学习 (model-free reinforcement learning, MFRL) 方法, 这导致了较长的模型训练时间^[14], 且已有的研究模型在实际运行的网络上进行在线学习时存在局限性。在无模型强化学习算法训练过程中, 代理程序需要通过探索和尝试不同策略以积累足够的经验并找到一个最优的行动。这个过程中, 代理程序的探索行为可能会在短期内导致网络服务质量下降, 影响用户体验。

基于上述问题, 更多解决方案被提出。其中, 文献[15]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资源分配策略, 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电接入网切片的 SLA 满意度。文献[16]提出一种资源分配机制, 使用 DQN 算法和软件定义网络控制器, 将无线资源分配给 URLLC 和 eMBB 切片, 以最大化数据速率。

此外, 还有基于博弈模型解决网络资源分配问题的研究^[17]。文献[18]提出一种基于博弈论的自动化机制, 用来协助用户根据需求及时做出决策。文献[19]提出一种能收敛到博弈均衡的分布式算法, 有效地提高了基础设施提供者 (infrastructure provider, InP) 和用户设备 (user equipment, UE) 的收益。文献[20]提出一种基于匹配博弈的资源分配方法, 有效地降低了资源分配的响应时延和系统成本。

综上所述, 现有网络资源分配方案普遍以分配公平性作为目标, 未考虑用户业务差异, 同时缺少对资源分配与用户业务满意度共存问题的研究, 因此本文提出了一种基于 OS-MBRL 的网络资源动态分配方法, 该方法能够保证在分配较少资源的情况下产生较低的服务等级协议违规率。



2 问题描述与建模

本节将对网络场景下的切片资源分配问题进行描述与建模。

2.1 问题描述

网络切片是网络中的一项关键功能，其允许用户在一个共享物理网络上定制一个独立逻辑网络，从而满足用户特定的服务类型和功能需求。在网络根据用户需求创建网络切片后，网络服务提供商必须为每个网络切片分配必要的频谱、回程和计算资源，以满足与该切片用户达成的SLA^[21]。

本文研究场景考虑一个典型的蜂窝网络系统^[22]，由一个基站提供对属于 K 个网络切片的多个UE的访问，具体考虑在六边形单元上进行下行传输。当 $K=3$ 时，网络业务场景如图1所示。

基站和用户设备之间无线接口上的数据传输被结构化为帧，每帧又被划分为时间和频率上的分区：在时间维度上，帧被划分为传输时间间隔（transmission time interval, TTI），也称为子帧，是时间上连续的片段；在频率维度上，带宽被划分为子载波。在后续展开的研究场景

中，假设TTI持续时间为1 ms，子载波间隔为15 MHz，且多个网络切片能够在同一无线帧内并行共存。

每个网络切片均被分配了无线接口的一部分时间-频率资源，最小时频分配单元称为资源块（resource block, RB），由1个TTI和12个子载波组成，即 $1\text{ ms} \times 180\text{ MHz}$ 。参考Qi等的论文^[21-22]研究场景，本场景明确分配给一个网络切片的RB仅由该切片使用，从而确保切片之间的隔离。因此每个网络切片根据所交付服务类型的特点，运行其自身的调度程序，实现在每个TTI上为用户接入设备分配可调度的RB。

在本文研究的网络场景中，每个网络切片被分配的RB数量可以在周期性的时间点上进行调整，这些时间点被称为决策阶段，通常相隔数个帧。在每个决策阶段，控制代理程序会选择一种控制动作，指定每个切片在下一个决策阶段之前可使用的、独占的RB数量。在连续的决策阶段之间，控制代理程序会收集每个切片关于用户数据流量、信道质量状态以及SLA履行参数的观测数据，这些数据将被用于下一次决策。

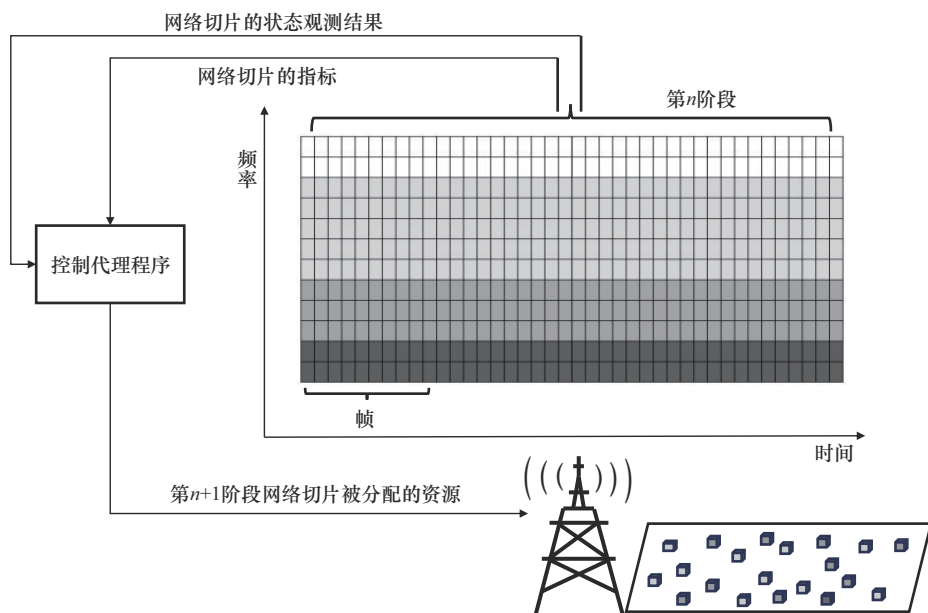


图1 网络业务场景

2.2 问题建模

针对前述的网络切片资源分配问题, 本文构建了一个问题模型。在本文研究中, 用 $\mathcal{K}$ 表示总数为 K 的网络切片集合, 这些切片共存于无线接口中, C 表示在网络切片之间进行分配的资源块总数, N 表示决策阶段总数。在每个决策阶段 $n=1, 2, \dots$, 控制代理程序接收自上一个决策阶段以来的每个切片的观测结果记为 $\mathbf{s}_{n-1}^{(i)}$, 其中 i 表示网络切片的索引号, $i \in \mathcal{K}$ 。在决策阶段 n 下, 所有切片观测结果的集合记为 $\mathbf{s}_n = (\mathbf{s}_n^{(1)}, \mathbf{s}_n^{(2)}, \dots, \mathbf{s}_n^{(K)})$, 作为对网络切片系统全局状态的观测。其中 $\mathbf{s}_n^{(i)}$ 是一个向量, 具体通过在决策阶段 $n-1$ 和 n 之间聚合或平均多个随机过程的实例得到。随机过程可以包括 UE 的接入和断开、每个 UE 使用的保证比特速率 (guaranteed bit rate, GBR) 流量或非 GBR 流量、每个 UE 缓冲数据或对每个信道上的信号与干扰加噪声比的测量。基于此, 本节构建的问题模型使用一个随机多维变量 $\mathbf{S}_n^{(i)}$ 表示在决策阶段 n 时切片 i 的观测结果, $\mathbf{s}_n^{(i)}$ 则表示 $\mathbf{S}_n^{(i)}$ 的一个特定样本。换言之, $\mathbf{s}_n^{(i)}$ 作为随机多维变量 $\mathbf{S}_n^{(i)}$ 中抽取的一个具体的观测值而存在。

在决策阶段 n 时, 将控制代理程序针对网络切片 i 的分配动作记为 $\mathbf{a}_n^{(i)}$, 其中 $i \in \mathcal{K}$, 表示系统选择分配给切片 i 的 RB 数量。所有分配动作的集合记为 $\mathbf{a}_n = (\mathbf{a}_n^{(1)}, \mathbf{a}_n^{(2)}, \dots, \mathbf{a}_n^{(K)})$, 表示第 n 个决策阶段的系统控制动作。由于控制代理程序的决策是由随机多维变量 $\mathbf{S}_n^{(i)}$ 决定的, 因此对每个切片的分配动作 $\mathbf{A}_n^{(i)}$ 也是随机多维变量, 而 $\mathbf{a}_n^{(i)}$ 作为一个具体的值存在, 与 $\mathbf{s}_n^{(i)}$ 类似。

此外对每个网络切片 $i (i \in \mathcal{K})$ 定义一个指示函数 $I^{(i)}$, 该函数向控制代理程序提供关于切片 SLA 履行情况的信息。具体而言, 如果切片 i 的 SLA 在决策阶段 n 和 $n+1$ 之间没有得到履行, 则指示函数 $I^{(i)}(\mathbf{s}_n^{(i)}, \mathbf{a}_n^{(i)}) = 1$; 反之 $I^{(i)}(\mathbf{s}_n^{(i)}, \mathbf{a}_n^{(i)}) = 0$ 。指示

函数 $I^{(i)}$ 包含 SLA 中定义的所有关键性能指标。例如, 在一个 eMBB 网络切片 i 中, 如果系统无法同时满足 GBR 流量和非 GBR 流量的 QoS 级别, 该切片的指示函数 $I^{(i)}$ 将返回 1。只有在观测结果中所有指定的 QoS 都得到满足时, 指示函数 $I^{(i)}$ 才会返回 0。

控制代理程序的控制动作由策略 π 决定, 该策略被定义为一个函数, 接收上一阶段系统观测结果 $\mathbf{s}_{n-1}$ 的同时提供每个切片的资源分配动作集合 $\mathbf{a}_n$ 。对于给定的随机系统动作以及网络切片的初始观测结果 $\mathbf{S}_0^{(1)}, \mathbf{S}_0^{(2)}, \dots, \mathbf{S}_0^{(K)}$, 策略 π 确定了切片 i 中观测-动作对的随机序列 $\mathbf{S}_0^{(i)}, \mathbf{A}_1^{(i)}, \mathbf{S}_1^{(i)}, \mathbf{A}_2^{(i)}, \mathbf{S}_2^{(i)}, \dots, i \in \mathcal{K}$, 该序列即系统的轨迹。在本问题模型中, 如果一个策略提供的所有分配动作满足 $\sum_{i \in \mathcal{K}} \mathbf{a}_n^{(i)} \leq C$ 且 $\mathbf{a}_n^{(i)} \geq 0$, 即所有切片所分配资源的和不超过资源总数 C , 且每个切片被分配的资源数不少于 0, 则该策略被视为可接受的, 所有可接受的策略集合记作 Π 。

基于上述定义内容, 为控制代理程序寻找有效策略的问题可以被表述为一个约束马尔可夫决策过程 (constrained Markov decision process, CMDP)。本问题模型研究的 CMDP 目标即找到一个策略 $\pi \in \Pi$, 使得分配的 RB 平均数量最小, 同时保持每个网络切片的 SLA 违规平均数在一个期望界限 δ 之下:

$$\begin{aligned} & \min_{\pi \in \Pi} \lim_{N \rightarrow \infty} \frac{1}{N} E_{\pi} \left[\sum_{n=1}^N \sum_{i=1}^K \mathbf{A}_n^{(i)} \right] \\ & \text{s.t. } \lim_{N \rightarrow \infty} \frac{1}{N} E_{\pi} \left[\sum_{n=1}^N I^{(i)}(\mathbf{S}_n^{(i)}, \mathbf{A}_n^{(i)}) \right] \leq \delta, i \in \mathcal{K} \quad (1) \end{aligned}$$

其中, E_{π} 表示在策略 π 下轨迹分布的期望值, 也就是执行策略 π 后可能出现的状态序列。通过只考虑可接受的策略 $\pi \in \Pi$, 隐式地包含了容量约束, 并且每个切片所有 QoS 目标都在其指示函数 $I^{(i)}$ 中得到体现。然而该 CMDP 无法被直接解决, 因为网络切片系统状态是未知的, 并且无法被直



接观测。即便没有这些限制，大维度的状态及控制空间也会导致数值方法失效。

为了使用强化学习算法解决网络切片的资源分配问题，需要将式（1）中的CMDP改写为一个马尔可夫决策过程（Markov decision process, MDP），具体通过去除SLA约束，并将之间接地融入目标函数中实现。参考Mei等的方法^[22-24]，在目标函数中添加一个SLA违规计数器作为惩罚项，再乘以一个权重因子 λ 以平衡奖励与惩罚，得到式（2）所示的MDP：

$$\min_{\pi \in \Pi} \lim_{N \rightarrow \infty} \frac{1}{N} E_{\pi} \left[\sum_{n=1}^N \sum_{i=1}^K [A_n^{(i)} + \lambda I^{(i)}(S_n^{(i)}, A_n^{(i)})] \right] \quad (2)$$

其中， $\sum_{i=1}^K [A_n^{(i)} + \lambda I^{(i)}(S_n^{(i)}, A_n^{(i)})]$ 可以被解释为系统在决策阶段 n 时所产生的总成本，且 $\sum_{i=1}^K \lambda I^{(i)}(S_n^{(i)}, A_n^{(i)})$ 代表惩罚项。

式（2）是一个平均成本的MDP，其目标为最小化每个决策阶段的平均成本。然而针对网络资源分配问题，长期平均成本的优化可能会忽略短期内重要决策的影响，所以相较于平均成本MDP，折扣MDP更适合本文应用场景。此外，折扣因子 γ 的存在，使得远期奖励对当前决策的影响变小，在防止函数无限增长的同时，能够帮助算法更快地收敛到最优策略，大多数强化学习方法都是针对折扣MDP。

由于在折扣MDP中的目标是最大化折扣回报，欲将平均成本MDP转化为折扣MDP，首先需要将式（2）中的成本项 $\sum_{i=1}^K [A_n^{(i)} + \lambda I^{(i)}(S_n^{(i)}, A_n^{(i)})]$ 记作 $-r_n(s_n, a_n)$ ，即考虑用负成本作为正奖励，成本项越大奖励就越少；其次，引入折扣因子 $\gamma \in (0, 1)$ 来衡量控制代理程序对未来奖励的重视程度，为了将未来奖励转换为当前价值，具体方法是将奖励乘以折扣因子 γ 的 n 次方，即 $\gamma^n r_n(s_n, a_n)$ 。此外，折扣因子 γ 的引入，使得未

来奖励随着时间推移被逐渐贬值，不再需要 $1/N$ 来均衡每个决策阶段的奖励。基于上述处理，式（2）可以转化为如下折扣MDP：

$$\max_{\pi \in \Pi} \lim_{N \rightarrow \infty} E_{\pi} \left[\sum_{n=1}^N \gamma^n r_n(s_n, a_n) \right] \quad (3)$$

式（3）目标是收益最大化，具体指控制代理程序沿系统轨迹获得的折扣奖励之和，其中 (s_n, a_n) 代表在决策阶段 n 下系统轨迹中的观测-动作对。

3 基于OS-MBRL的网络资源分配方法

针对前述问题模型，本节提出了基于OS-MBRL的网络资源分配方法及其模块设计。

3.1 基于模型的问题解决策略

在基于OS-MBRL的网络资源分配方法中所使用的环境模型并不追求对真实转移函数的高度拟合，而是学习这些动态分配动作对切片SLA指示函数 $I^{(i)}$ 的影响。在决策阶段 n 下，为每个切片 i 建立一个预测模型 $h_n^{(i)}$ ，根据前一阶段结束时收到的观测结果 $s_{n-1}^{(i)}$ ，预测控制代理程序给出的分配动作 $a_n^{(i)}$ 是否能够满足切片SLA，并用 $\mathcal{H}_n = \{h_n^{(1)}, h_n^{(2)}, \dots, h_n^{(K)}\}$ 表示模型集合。当 $h_n^{(i)}$ 输出为0时，代表模型预测分配动作 $a_n^{(i)}$ 能够满足切片SLA；相反，当 $h_n^{(i)}$ 输出为1时，则代表模型预测分配动作 $a_n^{(i)}$ 不能满足切片SLA。对于给定模型 $h_n^{(i)}$ ，可以将式（1）中的CMDP近似地处理为一个一步前瞻控制问题，得到一个模型预测控制器^[25]，其中观测-动作对必须根据 $h_n^{(i)}$ 满足每个切片SLA。然而该方法没有考虑边界 δ ，可能导致过多的SLA违规现象。因此，需要定义一个函数 $e^{(i)}$ 来表示切片 i 在给定 $(s_n^{(i)}, a_n^{(i)})$ 情况下，模型 $h_n^{(i)}$ 给出的预测结果为假阴性的概率，即将原本违反切片SLA的分配动作预测为满足切片SLA的概率：

$$e^{(i)}(h_n^{(i)}, s_{n-1}^{(i)}, a_n^{(i)}) = P(I^{(i)}(s_n^{(i)}, a_n^{(i)}) = 1 | h_n^{(i)}(s_{n-1}^{(i)}, a_n^{(i)}) = 0) \quad (4)$$

其中，将零假设 ($h_n^{(i)}(s_{n-1}^{(i)}, a_n^{(i)}) = 0$) 即模型根据上一阶段观测结果 $s_{n-1}^{(i)}$ 预测分配动作 $a_n^{(i)}$ 能够满足 SLA，与在决策阶段 n 中存在 SLA 违规的情况即 $I^{(i)}(s_n^{(i)}, a_n^{(i)}) = 1$ 联合计算条件概率。基于此，函数 $e^{(i)}$ 可以表示预测结果为假阴性的概率。

基于上述内容，问题模型被转化为一个一步前瞻控制问题，其中每一个控制动作 a_n 都应当是可接受的，并且 a_n 中每一个元素 $a_n^{(i)}$ 都应根据预测模型 $h_n^{(i)}$ 满足切片 SLA 的要求，错误概率的边界为 δ 。

$$\begin{aligned} \min_{a_n^{(1)}, \dots, a_n^{(K)}} \quad & \sum_{i=1}^K a_n^{(i)} \\ \text{s.t.} \quad & e^{(i)}(h_n^{(i)}, s_{n-1}^{(i)}, a_n^{(i)}) \leq \delta \\ & a_n^{(i)} \geq 0, i \in \mathcal{K} \\ & \sum_{i \in \mathcal{K}} a_n^{(i)} \leq C \end{aligned} \quad (5)$$

3.2 基于 online SVM 的在线学习

本节研究目标是实现控制代理程序的即插即用能力，同时对模型在学习和适应过程中产生的遗憾值提出要求。遗憾值是一个衡量在线学习算法性能的关键指标，用于比较在线学习算法的决策性能与某个最优基准策略的性能差异。

online SVM 算法是传统 SVM 算法的一种变体，能够处理动态变化的数据流，并不断更新模型以适应新数据。得益于其较低的计算复杂度，这种方法可以加快训练速度，适用于大规模和流式数据场景，满足本研究对于网络切片场景下在线学习模型的需要。

而 OSVM_SP 是一种基于最小封闭球问题的新型 online SVM 算法，专为处理流式数据而设计，能够在数据流的单次传递过程中完成学习和更新^[26]。OSVM_SP 算法如下。

输入：训练样本和标签 (x_i, y_i) , $i = 1, 2, \dots, N$, 正则化参数 C , 核函数 $k_\phi(x_i, x_j)$ 及其核参数 σ

输出：拉格朗日乘子 $\alpha \in R^{N+1}$, 偏置项 b

初始化：高维空间 z_i

循环开始，对于 $i = 1, 2, \dots, N$ 的每个样本，执行以下操作

根据 $z_i = \phi(x_i) = [y_i \phi(x_i); e_i / \sqrt{C}]$ 将 (x_i, y_i)

映射到高维空间 z_i

如果接受的样本 $i = 1$

对算法进行初始化，设置 $\alpha_1^1 = 1, \beta_1^1 = 1, \gamma^1 = NaN, r^1 = 0$

如果接受的样本 $i \geq 2$

计算 z_i 到最小封闭球中心 c^{i-1} 的距离 $d(z_i, c^{i-1})$

如果 $d(z_i, c^{i-1}) > r^{i-1}$ ，即 z_i 在当前封闭球外

如果 $\gamma^{i-1} = NaN$

如果 $\gamma^{i-1} \neq NaN$

计算 z_i 到当前最小封闭球直径两端点的距离 $d(z_i, p^{i-1})$,

$d(z_i, q^{i-1})$

如果 $d(z_i, p^{i-1}) \leq d(z_i, q^{i-1})$

互换 β 与 γ 的值

对所有 $j = 1, \dots, i$ ，计算 $\alpha_j^i, \beta_j^i, \gamma^i, r^i$

如果 $d(z_i, c^{i-1}) \leq r^{i-1}$ ，即 z_i 在当前封闭球内

将 $\alpha_j^i, \beta_j^i, \gamma^i$ 设置为 0

结束循环

找到非边界支持向量实例及其标签 (x_k, y_k) 并

计算偏置项 $b = y_k - \sum_{i=1}^N \alpha_i^i y_i k_\phi(x_i, x_k)$

返回 $\alpha = [\alpha_1^N, \alpha_2^N, \dots, \alpha_N^N]$ 和 b

获取返回值后，利用计算式 $\text{sign}(\sum_{i=1}^N \alpha_i y_i k_\phi(x_i, x) + b)$ 生成决策函数，对新输入的数据点 x 进



行预测。通过结合 OSVM_SP 算法对第 3.3 节中提到的在线预测模型 $h_n^{(i)}$ 和误差概率估计量 $\hat{e}_n^{(i)}$ 进行训练, 网络切片系统能够实时更新模型权重, 使模型在其在线学习阶段累计较低的 SLA 违规次数, 更新分配策略 $a_n^{(i)}$ 和切片的安全边界 $m_n^{(i)}$, 更好地满足用户需求。

3.3 基于 OS-MBRL 的系统模型

结合基于模型的问题解决策略和 OSVM_SP 算法, 本节提出了基于 OS-MBRL 的网络切片系统模型。基于 OS-MBRL 的网络切片资源分配模型结构如图 2 所示, 该模型包含在线预测学习机、在线误差估计学习机和控制器 3 个功能模块。

(1) 在线预测学习机

在线预测学习机上同时运行 K 个在线学习预测模型, 每个模型对应集合 $\mathcal{H}_n$ 中的一个 $h_n^{(i)}$, 用于对网络切片 i 在控制器给出的分配动作 $a_n^{(i)}$ 下 SLA 的违规情况进行预测分类, 结果记为 $\hat{y}_n^{(i)}$, 其值根据 $h_n^{(i)}$ 的输出得到。如果 $h_n^{(i)}(s_{n-1}^{(i)}, a_n^{(i)})$ 预测不存在 SLA 违规 (即输出为 0), $\hat{y}_n^{(i)}=0$; 相反, 如果 $h_n^{(i)}(s_{n-1}^{(i)}, a_n^{(i)})$ 预测 SLA 不会被满足, 则 $\hat{y}_n^{(i)}=1$ 。在学习过程中的每个决策阶段 n , 这些预测模型基于上一阶段末的

系统状态观测结果 $s_{n-1}=(s_{n-1}^{(1)}, s_{n-1}^{(2)}, \dots, s_{n-1}^{(K)})$ 、系统控制动作 $a_n=(a_n^{(1)}, a_n^{(2)}, \dots, a_n^{(K)})$ 和实际观测到的所有切片 SLA 违规情况 $y_n=(y_n^{(1)}, y_n^{(2)}, \dots, y_n^{(K)})$ 进行迭代, 其中 $y_n^{(i)}$ 值是指示函数 $I^{(i)}$ 的输出结果。如果 $I^{(i)}(s_n^{(i)}, a_n^{(i)})=0$, 则 $y_n^{(i)}=0$, 表示在决策阶段 n 时存在 SLA 违规; 相反, 如果 $I^{(i)}(s_n^{(i)}, a_n^{(i)})=1$, 则 $y_n^{(i)}=1$, 表示在决策阶段 n 时不存在 SLA 违规。

在每个决策阶段 n , 在线预测学习机接收来自网络切片 i 的观测-动作对 $(s_{n-1}^{(i)}, a_n^{(i)})$ 以及实际观测到的切片 SLA 违规情况 $y_n^{(i)}$, 其中 $i \in \mathcal{K}$ 。当 $y_n^{(i)}=0$ 时, 意味着决策阶段 n 中系统对切片 i 的分配动作满足其 SLA 的需求。由此可以进一步推导出: 当任何分配动作 $a > a_n^{(i)}$ 时, $y_n^{(i)}=0$ 。因此可以通过生成额外实例 $x_n=(s_{n-1}^{(i)}, a)$ 来增加在线学习过程中每个阶段的训练样本数量, 其中 $a=a_n^{(i)}, \dots, a_{\max}^{(i)}$ 。同样的, 当 $y_n^{(i)}=1$ 时, 考虑利用额外实例 x_n 对在线预测模型进行训练, 其中 $a=0, \dots, a_n^{(i)}$ 。

(2) 在线误差估计学习机

在线误差估计学习机的功能是在决策阶段 n

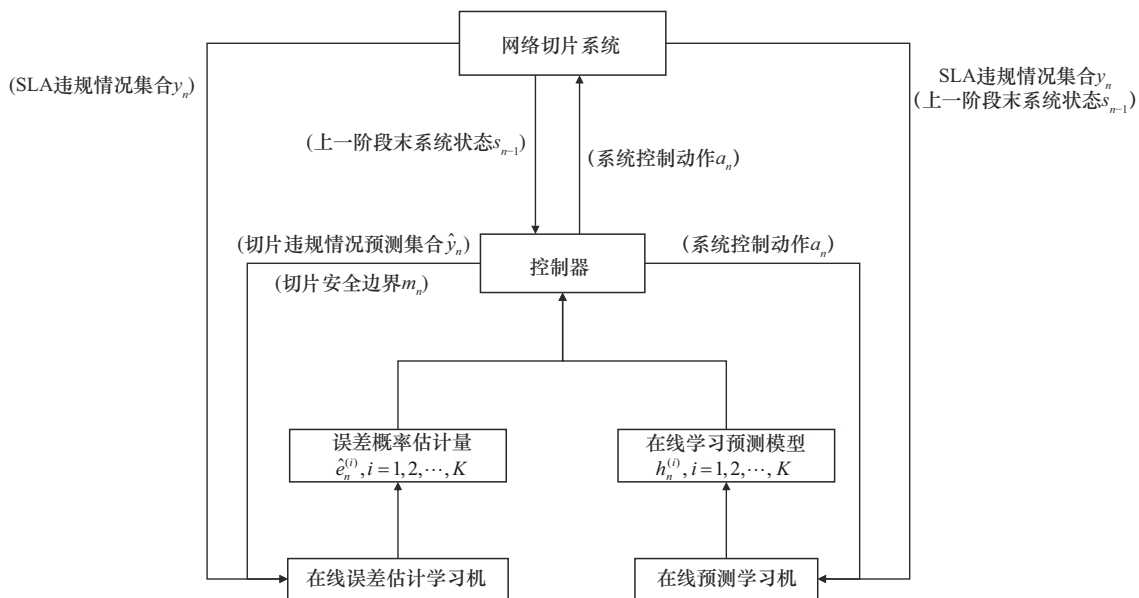


图2 基于 OS-MBRL 的网络切片资源分配模型结构

中完成对 $\mathcal{E}_n$ 中的 K 个误差概率估计量的更新, 其中 $\mathcal{E}_n = \{\hat{e}_n^{(1)}, \hat{e}_n^{(2)}, \dots, \hat{e}_n^{(K)}\}$ 为误差概率估计量 $\hat{e}_n^{(i)}$ 的集合。在学习过程中, 基于控制器对决策阶段 n 中所有网络切片 SLA 违规情况的预测向量 $\hat{\mathbf{y}}_n$ 、描述切片安全边界的集合 $\mathbf{m}_n = (\mathbf{m}_n^{(1)}, \mathbf{m}_n^{(2)}, \dots, \mathbf{m}_n^{(K)})$ 以及实际观测到的切片 SLA 的违规情况 $\mathbf{y}_n$, 对 $\hat{e}_n^{(i)}$ 进行迭代, 实现对在线预测学习机中训练的预测分类器准确性的监控。

给定一个网络切片 i 的初始状态 $\mathbf{s}_{n-1}^{(i)}$, 如果两个分配动作 $\mathbf{a}$ 和 $\mathbf{a}'$ 均能满足切片 SLA 的需求且 $\mathbf{a} < \mathbf{a}'$, 则存在 $e^{(i)}(h_n^{(i)}, \mathbf{s}_{n-1}^{(i)}, \mathbf{a}') \leq e^{(i)}(h_n^{(i)}, \mathbf{s}_{n-1}^{(i)}, \mathbf{a})$ 。基于此, 在决策阶段 n 时如果网络切片 i 存在一个最小分配动作 $\mathbf{a}$, 使得预测结果 $\hat{\mathbf{y}}_n^{(i)} = 0$, 则称 $\mathbf{a}'$ 存在一个 $\mathbf{a}' - \mathbf{a}$ 的安全边界。对于分配动作 $\mathbf{a}_n^{(i)}$ 的安全边界定义为:

$$\mathbf{m}_n^{(i)} = \mathbf{a}_n^{(i)} - \min \{ \mathbf{a} : h_n^{(i)}(\mathbf{s}_{n-1}^{(i)}, \mathbf{a}) = 0 \} \quad (6)$$

随着安全边界的增加, 预测误差的概率 $e^{(i)}$ 相应减小。通过计算决策阶段 n 下切片 i 的所有安全边界 $\mathbf{m} = 0, 1, \dots, \mathbf{a}_{\max}^{(i)}$ 对应的平均预测误差, 得到对误差概率 $e^{(i)}$ 的误差概率估计量 $\hat{e}_n^{(i)}$ 。

因为针对网络切片 i 与分配动作 $\mathbf{a}$ 的误差概率 $e^{(i)}$ 表示当切片的 SLA 实际未得到履行而预测结果 $\hat{\mathbf{y}}_n^{(i)} = 0$ 的概率, 所以只有当 $\hat{\mathbf{y}}_n^{(i)} = 0$ 时, $\hat{e}_n^{(i)}$ 才会进行迭代。对于给定的安全边界 $\mathbf{m}$, 迭代后的误差概率估计量表示为 $\hat{e}_{n+1}^{(i)}(\mathbf{m})$, 近似 $e^{(i)}(h_n^{(i)}, \mathbf{s}_{n-1}^{(i)}, \mathbf{a}')$, 其中 $\mathbf{a}'$ 以 $\mathbf{m}$ 为安全边界。

$$\hat{e}_{n+1}^{(i)}(\mathbf{m}) = (1 - \beta) \hat{e}_n^{(i)}(\mathbf{m}) + \beta \mathbf{y}_n^{(i)} \quad (7)$$

其中, $0 < \beta < 1$ 是学习率, 用于控制 $\mathbf{y}_n^{(i)}$ 对更新过程的影响程度。同样在学习过程中对误差概率估计量 $\hat{e}_n^{(i)}$ 采用样本增强策略, 通过安全边界 $\mathbf{m} = \mathbf{m}_n^{(i)}, \dots, \mathbf{a}_{\max}^{(i)}$ 更新 $\hat{e}_n^{(i)}$ 。

(3) 控制器

控制器的主要任务是在每个决策阶段 n , 根据上一阶段末的系统状态观测结果 $\mathbf{s}_{n-1}$ 、 $\mathcal{H}_n$ 和

$\mathcal{E}_n$, 生成一个系统控制动作 $\mathbf{a}_n = (\mathbf{a}_n^{(1)}, \mathbf{a}_n^{(2)}, \dots, \mathbf{a}_n^{(K)})$ 。此外控制器提供两个额外向量: 控制器对该决策阶段 n 中所有网络切片的 SLA 违规情况的预测 $\hat{\mathbf{y}}_n = (\hat{\mathbf{y}}_n^{(1)}, \hat{\mathbf{y}}_n^{(2)}, \dots, \hat{\mathbf{y}}_n^{(K)})$, 以及每个切片的安全边界 $\mathbf{m}_n = (\mathbf{m}_n^{(1)}, \mathbf{m}_n^{(2)}, \dots, \mathbf{m}_n^{(K)})$ 。这两个向量对于更新误差估计模型 $\mathcal{E}_n$ 至关重要。

在决策阶段 n 时, 控制器针对每个网络切片 i , 找到一个让误差概率估计量 $\hat{e}_n^{(i)}(\mathbf{m})$ 的预测误差概率不大于 $1 - \bar{\delta}$ 的边界最小值记为 $\mathbf{m}_n^{(i)}$, 其中 $\bar{\delta}$ 代表误差估计模型准确性要求的可靠因子。初始化 $\mathbf{a}_n^{(i)}$ 的值为 -1 , 切片 SLA 违规情况预测结果 $\hat{\mathbf{y}}_n^{(i)} = 1$ 。在达到资源分配上限 $\mathbf{a}_{\max}^{(i)}$ 前, 不断增加分配的 RBs 数量, 并且持续对切片 i 的 SLA 违规情况进行预测, 直到找到一个 $\mathbf{a}_n^{(i)}$ 使得预测模型 $h_n^{(i)}(\mathbf{s}_{n-1}^{(i)}, \mathbf{a}_n^{(i)})$ 的预测结果 $\hat{\mathbf{y}}_n^{(i)} = 0$ 。此时取 $\mathbf{m}_n^{(i)}$ 与 $\mathbf{a}_{\max}^{(i)} - \mathbf{a}_n^{(i)}$ 的较小者为安全边界 $\mathbf{m}_n^{(i)}$, 取 $\mathbf{a}_n^{(i)} + \mathbf{m}_n^{(i)}$ 的值为分配策略 $\mathbf{a}_n^{(i)}$ 。此外, 一旦所有网络切片分配的 RB 数量之和超过网络切片系统的 RB 总数 C , 即 $\sum_i \mathbf{a}_n^{(i)} > C$, 则按比例减少所有切片分配的资源块数量, 以确保资源使用总量不超过 C 。

3.4 基于 OS-MBRL 的网络切片资源动态分配方法

通过在线预测学习机、在线误差估计学习机以及控制器功能模块的协同工作, 实现了 OS-MBRL 算法, 完成了针对网络切片资源动态分配方法的研究。该算法首先初始化在线预测模型和在线误差估计模型, 并且在循环开始前将系统的初始状态 $\mathbf{s}_0$ 提供给功能模块。循环开始后, 在每个决策阶段都会执行以下步骤: 控制器根据上一阶段在线预测模型和在线误差估计模型提供的输出, 以及阶段末的系统观测结果, 输出 $\hat{\mathbf{y}}_n$ 、 $\mathbf{m}_n$ 和 $\mathbf{a}_n$; 在线预测学习机根据输入的 $\mathcal{H}_n$ 、 $\mathbf{y}_n$ 、 $\mathbf{s}_{n-1}$ 和 $\mathbf{a}_n$, 进行计算并输出 $\mathcal{H}_{n+1}$; 在线误差估计学习



机根据输入 $\mathcal{E}_n$ 、 $\mathbf{m}_n$ 、 $\hat{\mathbf{y}}_n$ 和 y_n ，进行计算并输出 $\mathcal{E}_{n+1}$ 。通过重复执行以上步骤，实时更新模型来适应网络状态的变化，动态地管理和分配网络资源，在减少 SLA 违规次数的同时降低分配给网络切片的资源。具体算法如下。

输入：网络切片系统 RB 总数 C ，可靠因子 $\bar{\delta}$ ，每个网络切片的最大 RB 数量限制 $\mathbf{a}_{\max} = (\mathbf{a}_{\max}^{(1)}, \mathbf{a}_{\max}^{(2)}, \dots, \mathbf{a}_{\max}^{(K)})$

初始化：在线预测模型 $\{\mathcal{H}_1 = h_1^{(1)}, h_1^{(2)}, \dots, h_1^{(i)}\}$ ，在线误差估计模型 $\mathcal{E}_1 = \{\hat{e}_1^{(1)}, \hat{e}_1^{(2)}, \dots, \hat{e}_1^{(i)}\}$

观测系统初始状态 s_0

循环开始，在每一个决策阶段 $n=1, 2, \dots$ 执行以下操作

控制器根据 $\mathcal{H}_n, \mathcal{E}_n$ 以及上一阶段末的系统状态观测结果 s_{n-1} ，输出切片的 SLA 违规情况预测 $\hat{\mathbf{y}}_n$ 、安全边界 $\mathbf{m}_n$ 以及系统控制动作 $\mathbf{a}_n$

部署系统控制动作 $\mathbf{a}_n$ 并观测切片的 SLA 实际违规情况 y_n 和系统状态 s_n

预测学习机的在线 SVM 根据输入 $\mathcal{H}_n, y_n, s_{n-1}, \mathbf{a}_n$ ，进行模型迭代，输出 $\mathcal{H}_{n+1}$

误差估计学习机的在线 SVM 根据输入 $\mathcal{E}_n, \mathbf{m}_n, \hat{\mathbf{y}}_n, y_n$ ，进行模型迭代，输出 $\mathcal{E}_{n+1}$

结束循环

4 实验与分析

本节将对 OS-MBRL 算法的性能进行仿真验证。为了验证提出算法的有效性，将根据系统中 SLA 违规数量和配置的资源块数量，与 NAF、DQN 和 TD3 算法在相同仿真环境下得到的实验数据进行对比，分析算法得出的系统性能。

4.1 实验场景

本实验基于 Python 开发环境，模拟在网络场景下多个 RAN 切片之间分配 RB 的情况。网络切片专用于一种通信类型，包括 eMBB 和 mMTC，二者在流量模型和 SLA 上存在区别。实验所需模拟参数及具体数值参照文献[17,27]配置。

eMBB 切片的流量生成器分别模拟了 GBR 流量和非 GBR 流量的 UE 接入与断开，这些 UE 通过使用恒定比特速率流或变化比特速率流进行区分，eMBB GBR 流量模型参数集见表 1，eMBB 非 GBR 流量模型参数集见表 2。

表 1 eMBB GBR 流量模型参数集

参数	参数值
UE 到达	到达速率为 2 user/min 的泊松过程
UE 连接时间	均值为 30 s 的指数分布
比特率	0.5 Mbit/s

表 2 eMBB 非 GBR 流量模型参数集

参数	参数值
UE 到达	到达速率为 5 user/min 的泊松过程
UE 连接时间	均值为 30 s 的指数分布
爆发到达	到达速率为 1 burst/s 的泊松过程
爆发长度	均值为 500 packet 的指数分布
爆发比特率	1 Mbit/s

mMTV 流量模型参数集见表 3，mMTC 切片的流量生成器模拟了 1 000 个设备，每个设备有传输周期和数据包重传次数两个参数。针对每个特定设备，这两个参数值从表 3 的集合中随机选取。

表 3 mMTV 流量模型参数集

参数	参数值
mMTV 设备	1 000
传输周期	{10, 50, 100, 150, 200, 250, 500, 1 000}s
数据包重传	{2, 4, 8, 16, 32, 64, 128}
UE 到达	到达速率为 5 user/min 的泊松过程

实验模拟的无线信道参数以及 UE 标称功率参考 TR 36.942^[28]所描述的城市宏基站传播模型。通过从包含衰落轨迹的数据集中抽取样本模拟频率选择性衰落,具体采用 ns-3 模拟器数据集,该数据集描述了常见的衰落/移动模型^[29]。针对每个新接入的 UE,模拟器将随机选择一个数据集,并生成一个随机整数作为从选定数据集中抽取样本的起始索引。

在实验场景中,每 50 个 TTI 更新一次网络切片之间的资源分配,每个 TTI 的持续时间为 1 ms。因此,相邻两个决策阶段的间隔时间为 50 ms。具体考虑了如下的 3 个子场景。

- 子场景一: 5 个 eMBB 切片共享每个 TTI 下的 200 个 RB。
- 子场景二: 3 个 eMBB 和 2 个 mMTC 切片共享每个 TTI 下的 150 个 RB。
- 子场景三: 1 个 eMBB 和 4 个 mMTC 切片共享每个 TTI 下的 100 个 RB。

网络 eMBB 切片的 SLA 见表 4, 网络 mMTC 切片的 SLA 见表 5。

表 4 网络 eMBB 切片的 SLA

参数	参数值
GBR 授权容量	20 RB/子帧
非 GBR QoS 兼容容量	30 RB/子帧
GBR 用户最大平均队列	100 kbit/UE
非 GBR 用户最大平均队列	150 kbit/UE

表 5 网络 mMTC 切片的 SLA

参数	参数值
最大 UE 上下文数	1 000
每个用户最高时延	300 ms

4.2 实验方法

为了评估基于 OS-MBRL 的系统模型,实验针对每个子场景进行系统模型训练。训练包括两个连续阶段,分别为学习阶段和推理阶段。为了使模型有充分的学习时间以达到更好的预测结

果,本文选取持续 40 000 个决策阶段的学习阶段。在学习阶段中,基于 OS-MBRL 的系统模型将从与网络切片系统的交互中学习并迭代模型。开始学习时,模型必须通过与切片系统的交互,逐步学习和理解切片系统的行为和规律。这种情况相当于创建一个新网络切片,或者更新现有切片 SLA。

对学习阶段的评估至关重要,因为提出的基于 OS-MBRL 的系统模型目标是在网络切片系统运行时进行在线学习,同时将在线学习对所提供服务的负面影响降到最低。由于模型在学习阶段已经进行了充分的学习并且实验结果基本趋于稳定,在推理阶段,系统持续运行 10 000 个决策阶段来充分模拟经历学习阶段的 40 000 个决策阶段后系统保持稳定的情况。此时,OS-MBRL 系统模型不再继续学习,而是凭借前一阶段训练完成的模型在每个决策阶段给出 RB 的分配动作,并且模型已经对切片系统有了一定的理解,能够在切片系统中做出相对较优的决策。

模型性能参照以下指标进行衡量。

(1) 每个决策阶段的切片 SLA 平均违规数量。例如,针对 3 个子场景而言,该指标可能是从 0 到 5 的任意数字。

(2) 在学习阶段累计的切片 SLA 违规次数。该指标相当于遗憾值,代表模型未能选择最优决策导致的性能损失总量。

(3) 分配的 RB 数量。该指标衡量算法的频谱效率,分配资源总量的高低反映了模型对网络资源的利用效率。

4.3 结果与分析

算法实验首先测试预测学习机中的在线学习预测模型的准确率。通过记录预测模型在学习阶段的前 30 000 个决策阶段中,可靠因子 δ 为 0.97 和 0.99 时的预测准确率来分析模型性能。同时分别考虑了权重因子 λ 和折扣因子 γ 的值,发现在学



习阶段中当 $\lambda=1\,000$ 且 $\gamma=0.697$ 时能够更有效地避免SLA违规。实验将准确率为95%的置信区间以灰色区域表示,其下界是通过取准确率的平均值减去其标准误差的1.95倍来计算,上界则通过取准确率的平均值加上其标准误差的1.95倍来计算。

由实验结果可知,从子场景一到子场景三,由于mMTC切片的增加和总资源的减少,预测模型波动逐渐下降,同时可靠因子 δ 为0.99的预测模型准确率通常高于可靠因子 δ 为0.97的预测模型。子场景一、二、三的分类准确率变化分别如图3~图5所示。

不同子场景下准确率和波动性的差异可以归因于网络切片类型的不同组合以及其对资源的不同需求。eMBB切片间的高带宽竞争可能导致较大的准确率波动,而mMTC切片则可能带来更稳定的资源需求模式。通过在3个子场景下的实验,验证了基于在线支持向量机的在线学习预测算法在网络切片资源分配场景下的可用性以及较高准确率。

DQN算法是在强化学习的基础上引入深度学习的概念,目的在于实现服务质量QoS满意度和网络切片资源利用率的平衡;NAF算法是一种将DQN扩展到连续动作空间的无模型强化学习算法;TD3算法是一种非策略深度参与者-批评者

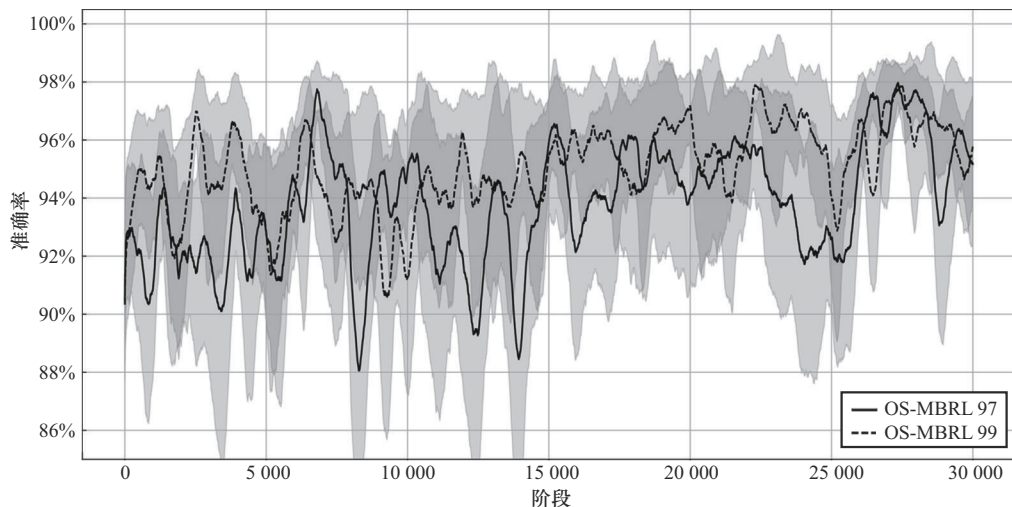


图3 子场景一的模型分类准确率变化

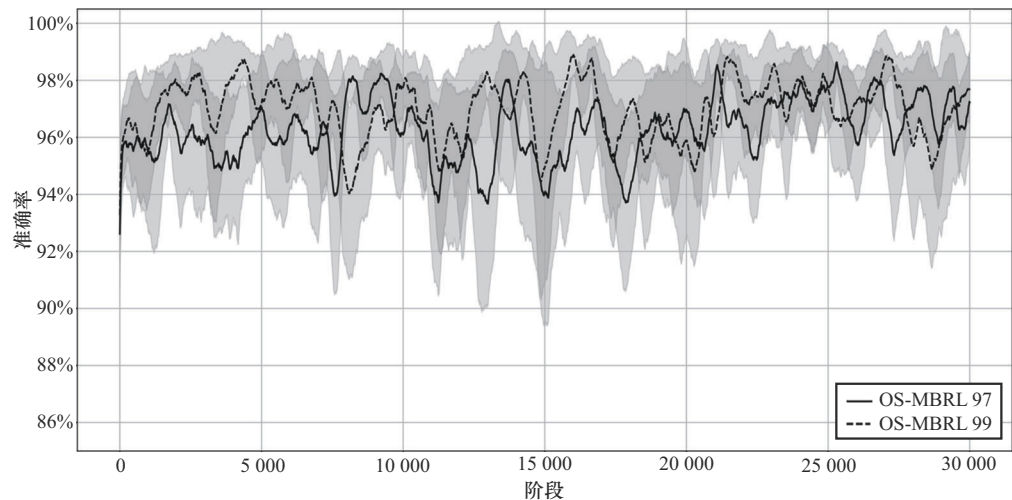


图4 子场景二的模型分类准确率变化

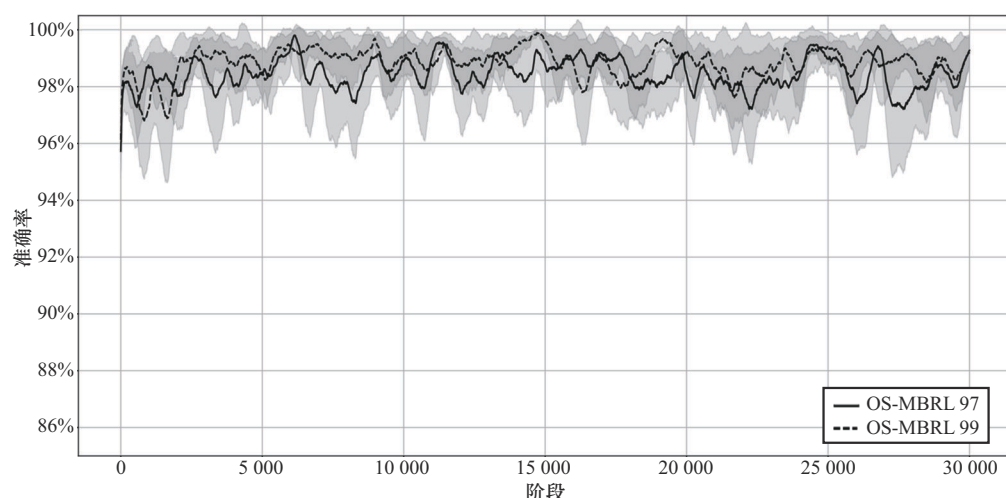


图5 子场景三模型分类准确率变化

算法, 是对深度确定性策略梯度算法 DDPG 算法的改进。这 3 种算法都是现有强化学习中先进的基线。本文将上述算法在仿真实验中得到的结果作为基准, 同 OS-MBRL 系统模型在相同场景下得到的结果进行对比。

与测试在线学习预测模型的方法相同, 在仿真实验中记录学习阶段的前 30 000 个决策阶段中, 记录可靠因子 δ 分别为 0.97 和 0.99 的 OS-MBRL 系统模型以及上述 3 种算法模型的 SLA 各项数据指标, 并将指标数据的 95% 的置信区间以灰色区域表示。

在子场景一的仿真实验中, 学习阶段的实验结果如下。

(1) 子场景一各模型学习阶段的网络切片 SLA 平均违规数量 (违规率) 如图 6 所示。其余算法模型的切片 SLA 平均违规数量均高于 OS-MBRL 系统模型。在 OS-MBRL 系统模型中, 由于可靠因子越大, 对所测量事物的稳定性也就越高, 因此可靠因子 δ 为 0.99 的算法模型的 SLA 平均违规数量普遍低于 δ 为 0.97 的算法模型。

(2) 子场景一各模型学习阶段的网络切片 SLA 累计违规数量如图 7 所示。由于加入了 on-line SVM 算法, 该算法能够使系统实时更新模型权重, 在其在线学习阶段累计较低的 SLA 违规次

数, 更新分配策略 $a_n^{(i)}$ 和切片的安全边界 $m_n^{(i)}$, 因此其余算法模型的切片 SLA 累计违规数量与其积累的速率都远高于 OS-MBRL 系统模型。同时, 在 OS-MBRL 系统模型中, 可靠因子 δ 为 0.99 的算法模型展现了更优的性能。

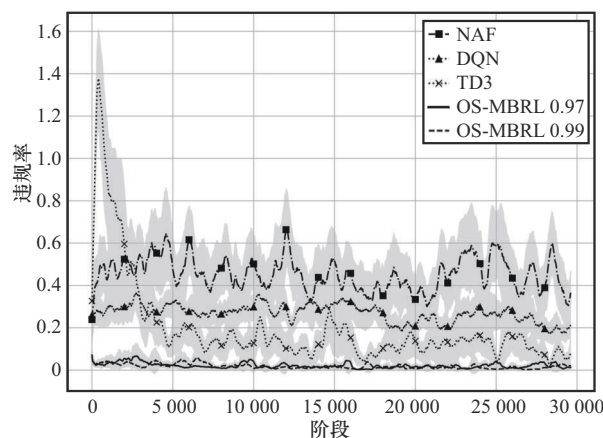


图6 子场景一各模型学习阶段的网络切片 SLA 平均违规数量

(3) 子场景一各模型学习阶段配置的资源块数量如图 8 所示。其余算法模型配置的 RB 数量稳定在 150 到 165 之间。普遍高于 OS-MBRL 系统模型配置的资源块数量。在 OS-MBRL 系统模型中, 可靠因子 δ 为 0.99 的算法模型通常配置更多的 RB 数量, 因为牺牲了部分资源以获得更高的测量稳定性, 使系统获得更低的 SLA 违规率。

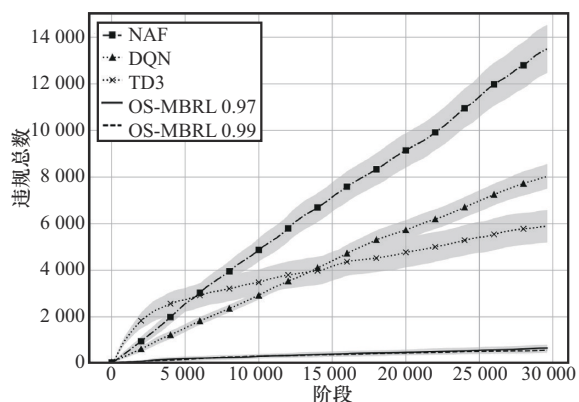


图7 子场景一各模型学习阶段的网络切片SLA累计违规数量

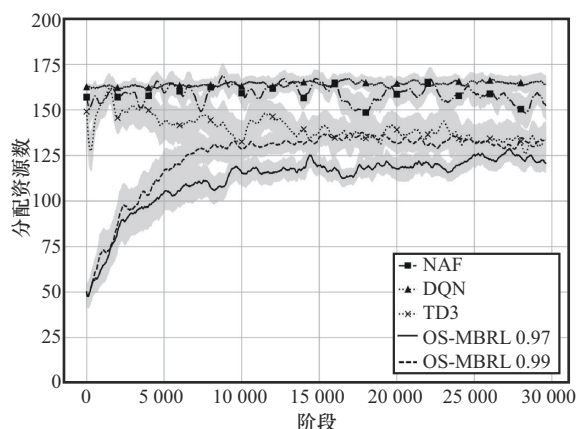


图8 子场景一各模型学习阶段配置的资源块数量

在子场景二的仿真实验中,学习阶段的实验结果如下。

(1) 子场景二各模型学习阶段的网络切片SLA平均违规数量如图9所示。其余算法模型的切片SLA平均违规数量远高于OS-MBRL系统模型。OS-MBRL系统模型的切片SLA平均违规数量在绝大部分阶段均低于0.25,其中可靠因子 δ 为0.99的算法模型表现相对较优。

(2) 子场景二各模型学习阶段的网络切片SLA累计违规数量如图10所示。其余算法模型的切片SLA累计违规数量与其积累的速率都远高于OS-MBRL系统模型。在OS-MBRL系统模型中,可靠因子 δ 为0.99的算法模型展现了更优的性能。

(3) 子场景二各模型学习阶段配置的资源块

数量如图11所示。其余算法模型配置的RBs数量稳定在120左右。在OS-MBRL系统模型中,可靠因子 δ 为0.99的算法模型通常配置更多资源块数量。

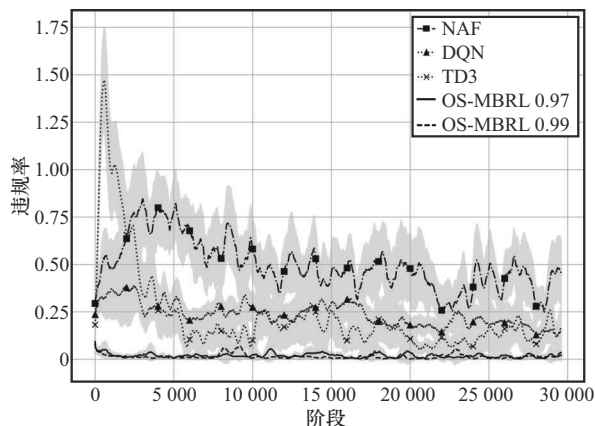


图9 子场景二各模型学习阶段的网络切片SLA平均违规数量

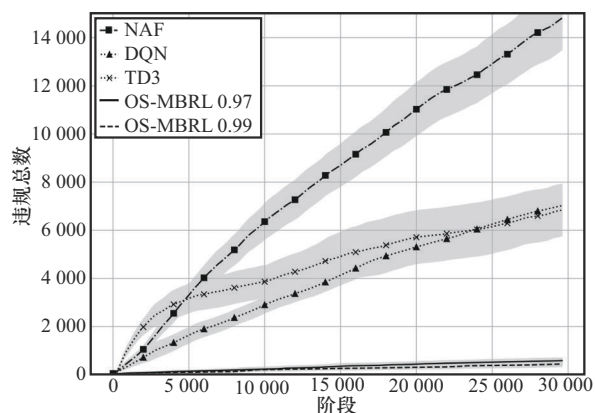


图10 子场景二各模型学习阶段的网络切片SLA累计违规数量

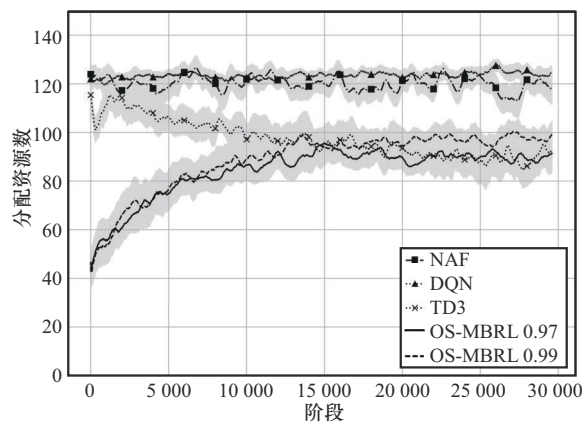


图11 子场景二各模型学习阶段配置的资源块数量

在子场景三的仿真实验中,学习阶段的实验结果如下。

(1) 子场景三各模型学习阶段的网络切片 SLA 违规率如图 12 所示。其余算法模型的切片 SLA 平均违规数量远高于 OS-MBRL 系统模型。在 OS-MBRL 系统模型中,可靠因子 δ 为 0.99 的算法模型的 SLA 平均违规数量在大部分阶段低于 δ 为 0.97 的算法模型。

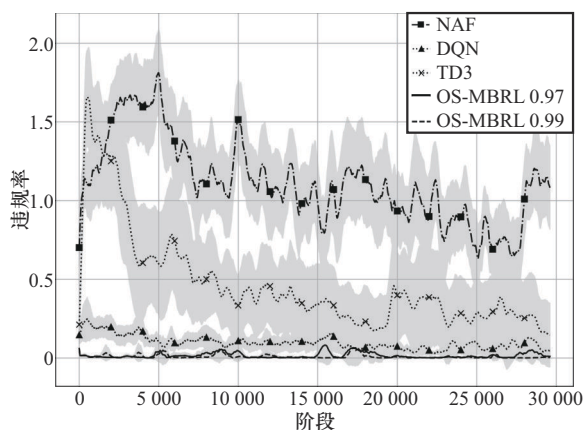


图 12 子场景三各模型学习阶段的网络切片 SLA 违规率

(2) 子场景三各模型学习阶段的网络切片 SLA 累计违规数量如图 13 所示。其余算法模型的切片 SLA 累计违规数量与其积累的速率都远高于 OS-MBRL 系统模型。在 OS-MBRL 系统模型中,可靠因子 δ 为 0.99 的算法模型展现了更优的性能,但此时二者的差距小于前两个子场景。

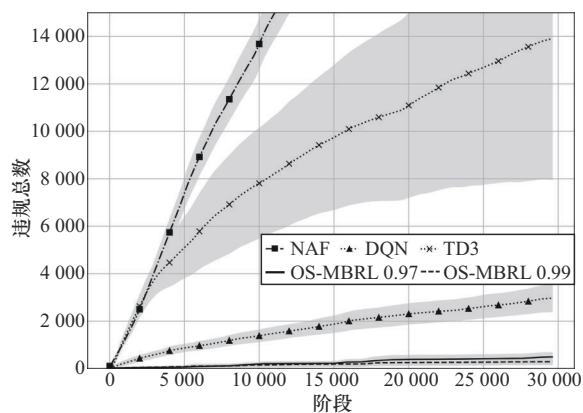


图 13 子场景三各模型学习阶段的网络切片 SLA 累计违规数量

(3) 子场景三各模型学习阶段配置的资源块数量如图 14 所示。其余算法模型配置的 RB 数量稳定在 80 左右。在 OS-MBRL 系统模型中,可靠因子 δ 为 0.99 的算法模型通常配置更多的 RB 数量,在大部分阶段稳定在 60 左右。

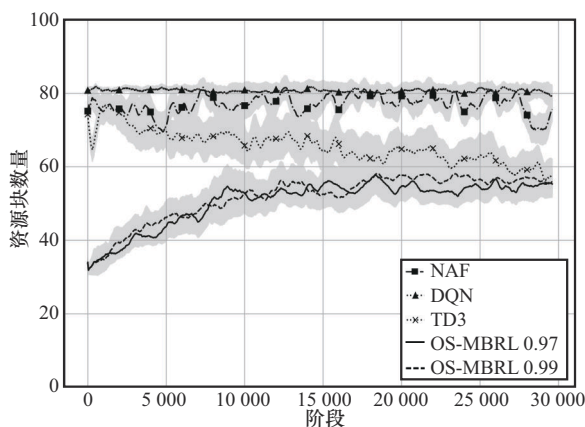


图 14 子场景三各模型学习阶段配置的资源块数量

综合 3 个子场景的仿真实验,可以得出结果:在在线学习阶段中,由于在线支持向量机算法可以实时更新模型,使模型始终保持一个较低的遗憾值,相较于其他 3 种算法模型,基于 OS-MBRL 的系统模型能够将网络切片的 SLA 平均违规次数最高降低 80%,同时减少 9% 的资源分配。实验验证了基于 OS-MBRL 的网络资源分配方法在管理网络切片的 SLA 违规方面有较高的性能,并能够提高网络资源的利用率。

5 结束语

本文针对网络的切片资源分配方法展开研究,提出了一种基于 OS-MBRL 的网络资源分配方法;引入一个基于在线支持向量机的系统模型对网络状态进行非线性分类,参与并辅助制定网络资源的分配决策,实现更高的样本效率,同时平衡地实现资源的最有效利用和切片服务等级协议的履行;通过在不同网络场景下的仿真实验,将此模型与 NAF、DQN 和 TD3 算法进行了比较。结果表



明, 相较于以上 3 种算法模型, 本文提出的系统模型均实现了显著的提升, 并且最高能够降低 80% 的 SLA 平均违规次数, 减少 9% 的资源分配, 验证了其有效性和较优的性能。

通过文献查阅、学术研究与仿真实验, 本文就网络的一大关键问题, 提出了针对性的解决方法。在验证 OS-MBRL 算法模型性能时, 本文提出的模型相较于其他所选强化学习算法模型有着明显优势, 期待后续研究能将 OS-MBRL 算法模型同更多强化学习算法模型进行对比, 同时探讨该模型在其他优化问题下的性能表现。

参考文献:

- [1] 何承卓. 基于 DAG 认证的网络切片场景切换算法[J]. 科学技术创新, 2024(5): 70-73.
HE C Z. Network slicing scene switching algorithm based on DAG authentication[J]. Scientific and Technological Innovation, 2024(5): 70-73.
- [2] KAMAL M A, RAZA H W, ALAM M M, et al. Resource allocation schemes for 5G network: a systematic review[J]. Sensors, 2021, 21(19): 6588.
- [3] ALBONDA H D R, PEREZ-ROMERO J. An efficient RAN slicing strategy for a heterogeneous network with eMBB and V2X services[J]. IEEE Access, 2019, 7: 44771-44782.
- [4] WANG Z, WEI Y, YU F R, 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.
- [5] LI R, ZHAO Z, SUN Q, et al. Deep reinforcement learning for resource management in network slicing[J]. IEEE Access, 2018, 6: 74429-74441.
- [6] GUO J, NIE G, TIAN H, et al. Puncture-predictive fairness scheduling scheme for eMBB and URLLC based on TD3 algorithm[C]//Proceedings of the CIC International Conference on Communications in China (ICCC). Piscataway: IEEE Press, 2023: 1-6.
- [7] QI C, HUA Y, LI R, et al. Deep reinforcement learning with discrete normalized advantage functions for resource management in network slicing[J]. IEEE Communications Letters, 2019, 23(8): 1337-1341.
- [8] RODRIGUEZ V Q, GUILLEMIN F, BOUBENDIR A. 5G E2E network slicing management with onap[C]//Proceedings of the 2020 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN). Piscataway: IEEE Press, 2020: 87-94.
- [9] SOHAIB R M, ONIRETI O, SAMBO Y, et al. Intelligent resource management for eMBB and URLLC in 5G and beyond wireless networks[J]. IEEE Access, 2023.
- [10] NASSAR A, YILMAZ Y. Reinforcement learning for adaptive resource allocation in fog RAN for IoT with heterogeneous latency requirements[J]. IEEE Access, 2019, 7: 128014-128025.
- [11] ZHANG W, DERA KHSHANI M, ZHENG G, et al. Constrained risk-sensitive deep reinforcement learning for eMBB-URLLC joint scheduling[J]. IEEE Transactions on Wireless Communications, 2024.
- [12] SUN G, 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.
- [13] 陈赓, 齐书虎, 沈斐, 等. 面向 B5G 多业务场景基于 D3QN 的双时间尺度网络切片算法[J]. 通信学报, 2022, 43(11): 213-224.
CHEN G, QI S H, SHEN F, et al. Dual time scale network slicing algorithm based on D3QN for B5G multi-service scenarios[J]. Journal on Communications, 2022, 43(11): 213-224.
- [14] HURTADO S J A, CASILIMAS K, CAICEDO R O M. Deep reinforcement learning for resource management on network slicing: a survey[J]. Sensors, 2022, 22(8): 3031.
- [15] 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.
ZHAO C, ZHANG C, HUANG Y M. Research on traffic-aware intelligent slicing resource allocation for radio access network[J]. Journal of Signal Processing, 2024, 40(4): 719-732.
- [16] 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.
- [17] 王再见, 谷慧敏. 基于联合优化的网络切片资源分配策略[J]. 通信学报, 2023, 44(5): 234-245.
WANG Z J, GU H M. Network slicing resource allocation strategy based on joint optimization[J]. Journal on Communica-

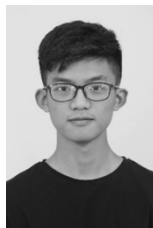
tions, 2023, 44(5): 234-245.

- [18] LIETO A, MALANCHINI I, MANDELLI S, et al. Strategic network slicing management in radio access networks[J]. IEEE Transactions on Mobile Computing, 2022, 21(4): 1434-1448.
- [19] TRAN T D, LE L B. Resource allocation for multi-tenant network slicing: a multi-leader multi-follower stackelberg game approach[J]. IEEE Transactions on Vehicular Technology, 2020, 69(8): 8886-8899.
- [20] RAVEENDRAN N, GU Y, JIANG C, et al. Cyclic three-sided matching game inspired wireless network virtualization[J]. IEEE Transactions on Mobile Computing, 2021, 20(2): 416-428.
- [21] 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.
- [22] QI C, HUA Y, LI R, et al. Deep reinforcement learning with discrete normalized advantage functions for resource management in network slicing[J]. IEEE Communications Letters, 2019, 23(8): 1337-1341.
- [23] LI R, ZHAO Z, SUN Q, et al. Deep reinforcement learning for resource management in network slicing[J]. IEEE Access, 2018, 6: 74429-74441.
- [24] MEI J, WANG X, ZHENG K, et al. Intelligent radio access network slicing for service provisioning in 6G: a hierarchical deep reinforcement learning approach[J]. IEEE Transactions on Communications, 2021, 69(9): 6063-6078.
- [25] BERTSEKAS D. Reinforcement learning and optimal control[M]. Cambridge: Athena Scientific, 2019.
- [26] HU L, HU C, HUO Z, et al. Online support vector machine with a single pass for streaming data[J]. Mathematics, 2022, 10(17): 3113.
- [27] AL-ALI M, YAACOUB E. Resource allocation scheme for eMBB and URLLC coexistence in 6G networks[J]. Wireless Networks, 2023, 29(6): 2519-2538.
- [28] 3GPP. Evolved universal terrestrial radio access(E-UTRA); radio frequency(RF) system scenarios, document 36.942: Version 15.0.0[S]. 2018.
- [29] BAUMGARTNER M, JUHAR J, PAPAJ J. Simulation of 5G and LTE-A access technologies via network simulator NS-3[C]// Proceedings of the 2021 44th International Conference on Telecommunications and Signal Processing(TSP). Piscataway: IEEE Press, 2021: 77-80.

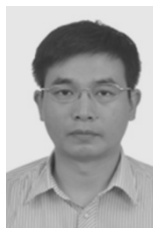
[作者简介]



严嘉辉 (2000-), 男, 浙江工商大学在读, 主要研究方向为通信与网络。



钟玮轩 (1999-), 男, 浙江工商大学在读, 主要研究方向为通信与网络。



董黎刚 (1973-), 男, 博士, 浙江工商大学信息与电子工程学院院长、教授、硕士生导师, 主要研究方向为新一代网络和分布式系统。



蒋献 (1988-), 男, 浙江工商大学讲师, 主要研究方向为智慧网络与智慧教育。



王广昌 (1970-), 男, UT斯达康通讯有限公司高级工程师, 主要研究方向为计算机网络、网络管理系统、SDN管控系统等。



陆凌蓉 (1972-), 女, UT斯达康通讯有限公司高级工程师, 主要研究方向为软件定义网络等。