基于深度强化学习的应急物联网切片资源预留算法

孙国林[1]，欧睿杰[1]，刘贵松[1,2]

（1. 电子科技大学计算机科学与工程学院，四川 成都 611731；2. 电子科技大学中山学院，广东 中山 528402）

摘 要：针对应急物联网（EIoT）超低时延服务需求，设计了面向超低时延传输应急物联网的多切片网络架构，提出 EIoT 切片资源预留和多异构切片资源共享与隔离的方法框架。所提框架采用深度强化学习方法实现实时异构切片间资源需求的自动预测与分配，切片内用户资源分配建模为基于形状的二维背包问题并采用启发式算法数值求解，从而实现切片内资源定制化。仿真结果表明，基于资源预留的方法能够使 EIoT 切片显式保留资源，提供了更好的安全隔离级别；深度强化学习能够保证资源预留的准确和实时更新，有效兼顾资源利用率和切片差异化服务质量要求。与 4 个已有算法对比表明，Dueling DQN 具有更好的性能优势。

关键词：应急物联网；深度强化学习；资源预留；超低时延通信

中图分类号：TN92

文献标识码：A

doi: 10.11959/j.issn.1000−436x.2020200

Deep reinforcement learning-based resource reservation algorithm for emergency Internet-of-things slice

SUN Guolin[1], OU Ruijie[1], LIU Guisong[1,2]

1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
2. Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan 528402, China

Abstract: Based on the requirements of ultra-low latency services for emergency Internet-of-things (EIoT) applications, a multi-slice network architecture for ultra-low latency emergency IoT was designed, and a general methodology framework based on resource reservation, sharing and isolation for multiple slices was proposed. In the proposed framework, real-time and automatic inter-slice resource demand prediction and allocation were realized based on deep reinforcement learning (DRL), while intra-slice user resource allocation was modeled as a shape-based 2-dimension packing problem and solved with a heuristic numerical algorithm, so that intra-slice resource customization was achieved. Simulation results show that the resource reservation-based method enable EIoT slices to explicitly reserve resources, provide a better security isolation level, and DRL could guarantee accuracy and real-time updates of resource reservations. Compared with four existing algorithms, dueling deep Q-network (DQN) performes better than the benchmarks.

Key words: emergency IoT, deep reinforcement learning, resource reservation, ultra-low latency communication

1 引言

5G 旨在提供千倍于 4G 的传输容量提高、至少千亿个物联网设备连接、高达 10 Gbit/s 的传输速率以及低至毫秒级的超低时延用户体验。除了人与人的通信之外，下一代移动互联网将实现人与机器、机器与机器之间的零距离连接，无线技术将以崭新的方式推动未来经济和社会的发展。因此，超低时

收稿日期：2020−06−18；修回日期：2020−08−19

基金项目：国家自然科学基金资助项目（No.61771098）；四川省科技计划基金资助项目（No.2020YFQ0025）

Foundation Items: The National Natural Science Foundation of China (No. 61771098), The Science and Technology Research Program of Sichuan Province (No.2020YFQ0025)

延传输被视为 5G/B5G 系统的主要技术特征之一，其目标是实现 1 ms 以下的端到端传输时延，从而支持人对机器、机器对机器的实时通信和远程控制应用。5G/B5G 除了在传输时延、可靠性和吞吐量方面提出了更高的要求之外，还对下一代移动互联网架构进行了重大变革。软件定义网络和网络功能虚拟化技术作为 5G/B5G 网络架构的创新技术，使基础设施网络可以切分为几个逻辑网络，允许多个差异化应用共享同一张物理网络和资源，即所谓的网络切片技术。每个独立切片可以调用在公共网络基础设施上运行的虚拟网络功能，并按需对其进行通信和计算资源的配置和调整，从而满足特定网络切片应用的特定业务需求[1-2]。通常，每个租户会与基础设施提供商签订服务水平协议。因此，通过自定义切片应用和功能，动态分配自定义切片的资源，公共移动网络可以支持特定的应急物联网切片，并保证该切片与其他移动网络切片的共存和安全隔离[3-4]。综上所述，面向应急物联网的应用业务需求，首先，需要保障单一应急物联网（EIoT, emergency Internet of things）切片的服务质量，允许租户管理其定制切片的网络性能；其次，需要考虑多异构切片共存的问题，通过复用切片流量实现基础架构的规模经济。

近年来，在资源切片方面已有大量研究工作，但是在异构混合数据流场景中仍然存在以下问题：1) 在无线资源有限的情况下，如何既保证所有切片的资源效率，又准确地满足切片需求；2) 如何根据服务水平协议（SLA, service level agreement）的要求为每个切片动态分配资源，以满足不同切片的服务质量（QoS, quality of service）要求；3) 在流量状态实时变化的高动态环境中，资源分配方案如何智能响应网络的变化特性并适应变化。本文基于虚拟化管理程序，如基于内核的虚拟机（KVM, kernel-based virtual machine），为托管在不同节点的多个虚拟基站分配资源，并为其调度相应的硬件物理资源和无线资源，从而实现频谱资源的共享和数据复用[5]。其中，物理资源块（PRB, physical resource block）作为最小粒度的无线资源被分配到不同虚拟基站节点。虚拟基站用来实现多网络切片间的资源共享和基于流量整形的隔离机制[6]。文献[7]提出了一种切片方案，通过配置切片和流调度器为切片提供资源。Cell-Slice 是基于数据面的网络切片方法，不需要修改基站的原有数据流调度算法，而是在网关采用流量整形机制自适应控制数据流速率[8]，这种控制方法可用于基于 WiMAX（world interoperability for microwave access）或 LTE（long-term evolution）标准的最大持续速率的调整机制[9-10]，但其只关注在保证速率的情况下为切片提供可用资源。文献[11]将费用开销定义为一个通用的目标函数，提出了一种基于凸优化模型和分布式交替方向乘子法（ADMM, alternating direction method of multiplier）求解的解决方法。然而，实际上不同切片可能具有不同的 QoS 要求，从而导致具有不同的优化目标函数。面向多租户异构云无线接入网场景，综合考虑多租户的优先级、服务质量和干扰水平限制、基带资源限制、前端和回程容量限制等因素，文献[12]提出了多个基于凸优化模型的动态网络切片方法，由于其工作捆绑了虚拟化资源分配和用户物理资源分配，因此无法实现异构切片的资源定制。文献[13]提出了一种全网范围的资源共享方案，该方案能够对存在于基站上的不同切片进行隔离，但 SLA 的严格 QoS 约束会阻碍用户在请求模式发生变化时实时满足 QoS 要求。文献[14]仅假设一个用于触觉通信的切片，并未专门针对混合流量处理资源切片。对于多异构切片共存场景，通过预测和估计切片资源需求，动态权衡用户 QoS 满意度和系统资源利用效率，自动地实时响应来自切片用户的动态资源请求是至关重要的。文献[15-16]将深度强化学习方法用于多切片资源分配问题，文献[15]主要针对移动车联网内容缓存资源，文献[16]仅考虑了 2 个切片实例，基于传统 DQN（deep Q-network）算法来实现。本文在文献[16]已有工作的基础上，针对混合流量自主资源配置和定制问题，提出了 Dueling DQN 算法，改进 Dueling 网络结构加速学习收敛，并采用自适应线性奖励机制自动平衡切片的资源利用率和 QoS 满意度，并且验证了安全隔离效果。

本文主要针对特定的应急物联网场景，研究一种通用的切片资源预留方法，同时考虑多个异构切片共存场景下多切片性能的动态安全隔离。针对特定的应急物联网，基于资源预留的方法可以提供严格的服务质量保证、切片间资源的保护和隔离，并提供资源可定制性和稳定性。所以，针对应急物联网应用，本文主要采用资源预留来保证端到端时延和可靠性，并为用户提供定制化物理资源，同时推广至多异构切片共存场景。本文的主要研究工作如下。

1) 面向应急物联网的多切片资源管理架构包括基于深度强化学习的切片资源预留模块、基于形状的物理资源块分配模块。面向差异化的异构网络切片需求，深度强化学习（DRL, deep reinforcement learning）智能体对切片的资源预留比例进行动态调整，输出结果是一个资源比例；物理资源分配模块将单一切片内基站的 PRB 分配给其关联用户。

2) 基于深度强化学习的资源切片策略。切片资源分配的目标是在保证用户 QoS 的前提下，最大化系统的资源利用效率。由于无线网络环境的时变性和动态性，DRL 智能体通过与无线网络环境的动态交互，能够根据当前的状态做出最优的动作，自动实时地调整切片的资源比例。

3) 基于形状的物理资源定制。针对多网络切片差异化服务质量需求，不同切片对速率和时延指标各有偏重。根据切片速率和时延需求，可以计算用户请求占用的频域和时域的 RB 数量，进而确定其占用的 RB 集合的形状。物理资源分配被建模成二维几何背包问题，其目标是最大化资源利用率，减少形状组合带来的资源浪费。

4) 系统仿真结果表明，综合考虑切片服务质量满意度和系统资源效率等评估指标，基于深度强化学习的切片资源预留算法具有很好的收敛性。与传统的 NVS（network virtualization substrate）和 NetShare 算法相比，所提 Dueling DQN 算法更佳，有效地平衡了异构共存切片的性能。

2　系统模型

2.1　网络模型

如图 1 所示，本文所提多切片网络架构采用软件定义网络（SDN, software defined networking）和网络功能虚拟化（NFV, network function virtualization）的网络架构，具体包括 SDN 控制器、终端用户设备（UE, user equipment）、网络切片、基站和频谱资源。SDN 控制器负责切片级的资源调度和决策，利用消息信令接口通知具体基站调整其切片的资源预留与分配数量等；基站为不同切片提供一定数量的 RB 资源；终端用户设备通过携带其所属切片识别信息发送资源请求，从某个关联基站获取和占用所属切片的 RB 资源。从资源方面，本文主要考虑频谱资源，即由时域和频域组成的 RB。本文主要考虑 4 种切片类型，分别为高清视频（HDTV, high-definition television）、海量终端物联网（MIoT, massive IoT）、EIoT 和 UEb（UE broadband）。

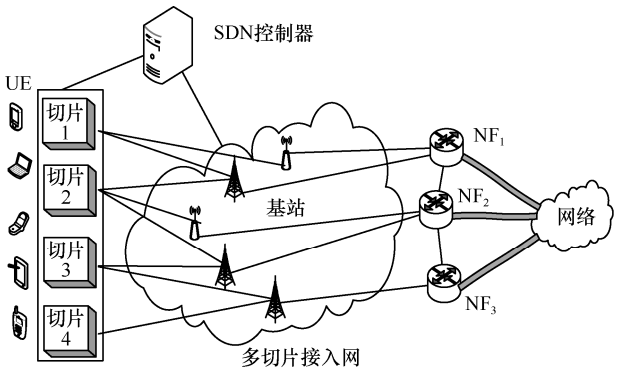

图 1　多切片网络架构

针对多异构切片共存场景，本文提出了一种基于深度强化学习的资源预留方法框架，如图 2 所示。其基本原理是，DRL 智能体与无线网络环境不断交互并获取环境的当前状态，智能体根据当前环境的状态选择一个动作执行，执行该动作之后会使环境从当前状态以某概率转移到另一个状态，同时环境反馈给智能体一个奖励或惩罚。智能体不断重复上

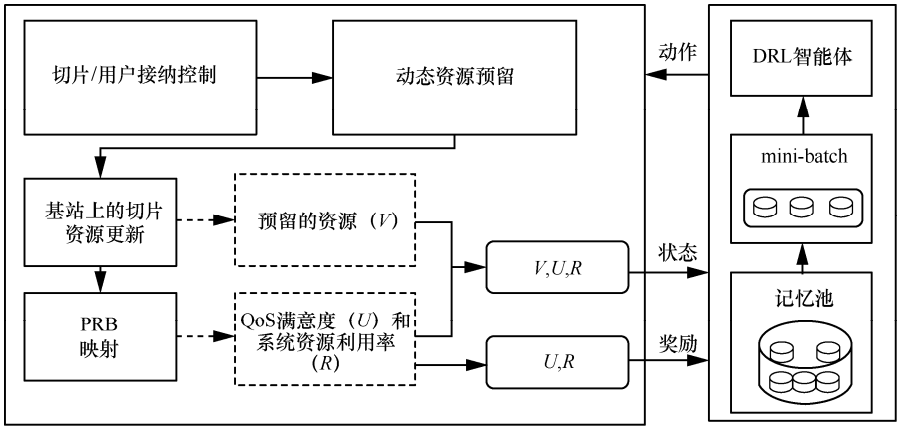

图 2　基于 DRL 的资源预留方法框架

述过程，以尽可能多地获得来自环境的奖励。

首先，资源预留分成初始资源预留和动态资源预留 2 个阶段。初始资源预留是指根据签订的 QoS 协定，每个基站给各个切片分配一个固定的资源比例。由于应急物联网切片在单个基站的资源是有限的，因此需要控制接纳用户的数量。通常，用户的接纳控制可以建模为 0-1 整数规划问题，并通过已有算法求解，其目标为在保证终端的速率和时延要求前提下，最大化物联网终端的接入数量。

第二，由于终端移动性和无线环境的时变性等固有特点，一旦单个基站应急物联网切片的终端数量发生变化，则可能出现资源不够或者资源过剩的问题。因此，需要根据应急物联网的负载状态进行预测，并动态调整切片的资源自适应于应急物联网切片资源需求的动态变化。进而，将切片资源的动态调整映射到不同的基站上，即动态资源预留。

第三，用户级物理资源分配由基站完成连接态用户数据到 PRB 集合的映射。当具体切片的预留资源 V 映射到基站资源时，基站需要为该切片的连接态终端分配 PRB 资源。然而，不同的多用户物理资源分配方法会造成系统资源利用率 R 和用户 QoS 满意度 U 的差异。在保证切片 QoS 满意度的前提下，如何最大化切片的资源效率，是单一切片内资源定制化研究的问题。

最后，DRL 智能体完成一次切片资源分配，终端即可获得相应的物理 RB 资源。进而，终端获得 QoS 满意度评估，并统计得到该切片的资源利用效率，从而反馈给智能体一个奖励或惩罚，无线网络环境转移更新至下一个状态。智能体将当前环境状态、资源分配动作、反馈奖励和环境下一个状态组成一个四元组，作为一个样本存储到记忆池。通过记忆回放机制，智能体会根据训练周期配置从记忆池随机选取 mini-batch 样本数据对智能体进行强化训练，从而不断更新神经网络的系数来降低损失。

2.2　时延模型

针对应急物联网场景，EIoT 切片对响应时延要求较高，而对速率要求可能较低。不同的应用服务切片对响应时延和传输速率的要求是不同的。因此，需要时延模型有效评估应急物联网基站对每个终端用户的服务时延。本文做如下假设：1)终端用户 u 发送每个数据分组到达的时间服从指数分布，均值为 $\frac{1}{\lambda_u}$，并且任意邻接的 2 个数据分组到达的时间间隔是相互独立的，λ_u 为终端用户 u 的数据分组到达率，其单位为 packet/s；2)某特定切片 s 所服务终端 u 的数据分组长度均为 L_{uk} bit，而不同切片应用的数据分组大小是相互独立的。因此，终端 u 发送一个数据分组至基站 k 的时间 t_{uk} 为

$$t_{uk}=\frac{L_u}{c_{uk}}=\frac{1}{c_{uk}^*} \tag{1}$$

其中，c_{uk} 为终端 u 从基站 k 实际获得的传输速率，单位为 bit/s；c_{uk}^* 则为归一化的实际传输速率，单位为 packet/s。基于上述假设，根据排队论 M/M/1 理论模型[17]，可以计算出用户 u 的数据分组的平均服务时延 τ_{uk} 为

$$\tau_{uk}=\frac{1}{\frac{1}{t_{uk}-\lambda_u}}=\frac{1}{a_{uk}c_{uk}^*-\lambda_u} \tag{2}$$

其中，a_{uk} 是终端 u 与基站 k 之间的关联变量，如果用户 u 与基站 k 相关联，则 a_{uk} 为 1；否则为 0。

2.3　效用函数

效用函数主要用于表征终端对服务质量的满意程度。此外，它也是反馈给智能体的回报函数的一部分。不难理解，不同切片的服务类型不同，其对速率或者时延要求也不相同，即不同切片的满意度函数存在差异。例如，应急物联网切片的满意度计算主要依赖于时延，而 HDTV 切片主要依赖于传输速率等。假设切片 s 所服务终端 u 的最小速率需求为 $c_u^{\min}$，最大时延需求为 $\mathrm{Td}_u^{\max}$。在一个调度周期 T，每个终端根据获得的服务速率和时延自动计算服务质量满意度 Sat_u，然后对该切片的所有终端的满意度进行平均，即可得到该切片用户的平均满意度函数 Sat_s。

具体地，终端 u 对速率敏感的服务质量满意度为

$$\mathrm{Sat}_u^v=\frac{1}{1+\mathrm{e}^{-\beta_1(c_{uk}-c_u^{\min})}} \tag{3}$$

终端 u 对时延敏感的服务质量满意度为

$$\mathrm{Sat}_u^d=\frac{1}{1+\mathrm{e}^{-\beta_2(\mathrm{Td}_{uk}-\mathrm{Td}_u^{\max})}} \tag{4}$$

其中，β_1 和 β_2 为 Sigmoid 函数的斜率[18]。因此，通过式(3)和式(4)可计算切片 s 的平均满意度，其计算式为

$$\mathrm{Sat}_s = \frac{1}{|U_s|}\sum_{u\in U_s}\frac{\mathrm{Sat}_u^v + \mathrm{Sat}_u^d}{2} \tag{5}$$

其中，$|U_s|$ 表示切片 s 的终端用户数量，U_s 表示切片 s 的终端用户集合。

3 问题建模

3.1 基于深度强化学习的切片资源预留

面向应急物联网切片资源预留，需要对切片资源需求进行动态预测，该问题可以建模为一个马尔可夫决策过程，并通过深度强化学习算法来解决，从而实现多个异构切片的资源共享和隔离。下面以 Dueling DQN 算法为例，建立马尔可夫决策模型。智能体的目标是寻找一个最优策略 π^*，最大化未来预期的回报奖励[19]。

根据当前策略 π、状态 s、动作 a，可以得到 Q 值 $Q_\pi(s^t,a^t)$ 和状态值 $V_\pi(s^t)$。

$$Q_\pi(s^t,a^t) = E\{r^t \mid s^t = s, a^t = a, \pi\} \tag{6}$$

$$V_\pi(s^t) = E_{a^t-\pi(s^t)}[Q_\pi(s^t,a^t)] \tag{7}$$

则 Q 函数的最优方程可表示为

$$Q_{\pi^*}(s^t,a^t) = R(s^t,a^t) + \gamma\sum_{s'}P(s' \mid s^t,a^t)\cdot \max_{a^t\in A} Q_{\pi^*}(s^{t+1},a^{t+1})\forall s^t,a^t,\pi$$

$$V_\pi(s^t) = \max_{a^t\in A} Q_{\pi^*}(s^t,a^t) \tag{8}$$

其中，γ 为马尔可夫过程的衰减因子，P 为当前状态 s^t 转移到下一个状态 s' 的概率。

根据式(6)和式(7)，决策函数定义为

$$A_\pi(s^t,a^t) = Q_\pi(s^t,a^t) - V_\pi(s^t)A_\pi(s^t,a) \tag{9}$$

其中，状态值函数 V 用来衡量状态 s 的好坏，值函数 Q 用来评价在当前状态 s 下选择某个特定动作 a 的好坏。

$$E_{a^t-\pi(s^t)}[A_\pi(s^t,a^t)] = 0 \tag{10}$$

对于确定性策略 $a^* = \arg\max_{a^t\in A} Q(s^t,a^t)$，由于 $Q(s^t,a^t) = V(s)$，可得 $A(s^t,a^t) = 0$。

综上所述，Dueling DQN 的输出可表示为

$$Q(s^t,a^t;\theta,\varsigma,\xi) = V(s^t;\theta,\xi) + (A(s^t,a^t;\theta,\varsigma) \tag{11}$$

其中，θ 为卷积层参数，ς 和 ξ 分别为决策函数和价值函数的参数。然而，$Q(s^t,a^t;\theta,\varsigma,\xi)$ 可能是无法得到的，因为它仅是真实 Q 函数的参数化估计。因此，本文引入聚合层，分别为状态 s 对应的每个动作 a 生成 Q 值。

$$Q(s^t,a^t;\theta,\varsigma,\xi) = V(s^t;\theta,\xi) + \left(A(s^t,a^t;\theta,\varsigma) - \frac{1}{|A|}\sum_{a^{t+1}}A(s^t,a^{t+1};\theta,\varsigma)\right) \tag{12}$$

深度强化学习为异构切片资源需求预测和切片资源预留提供了一种通用的算法框架，包含状态空间 State、动作空间 Action 和奖励回报函数 Reward 这 3 个基本要素。针对应急物联网场景，定义如下。

1) State，表示应急物联网状态。应急物联网状态包含三方面信息，分别为当前切片预留资源数量、切片资源占用数量和切片的平均服务质量满意度，具体可用以下 3 个数值表示。切片的资源预留比例 V_s，指切片在整个系统资源的占比，而不是单个基站上的资源占比；切片的资源利用率 RU_s，指实际使用的资源与切片预留资源之间的占比；切片 QoS 满意度 Sat_s，指该切片所有终端的服务质量满意度的平均值。针对应急物联网多个异构切片共存场景，State 集合定义为 $[V_s,\mathrm{RU}_s,\mathrm{Sat}_s]$。

2) Action，表示所执行的动作集合。DRL 智能体每获取一个状态，便会根据贪心算法选取并执行一个动作。针对异构切片间的动态资源预留问题，动作操作就是动态调整切片资源的系统占比。也就是说，在原来的预留资源数量的基础上，增加或减少一定的比例。假设初始切片预留的资源比例为 V_s，所执行的动作为 a，则调整后的资源比例为 $V_s' = V_s(1+a)$。由于 DRL 智能体仅在离散动作空间选取动作，需要将连续的动作空间进行离散化处理。如果单切片场景的动作空间的维度为 M，N 个切片共存场景，则动作空间的维度为 MN。因此，针对异构切片共存场景，动作空间的离散程度和粒度大小对于收敛速度有较大的影响。

3) Reward，表示环境交互所反馈的奖励回报。在每次迭代中，智能体都会根据当前的环境状态选取并执行一个动作，然后环境转移至下一个状态并反馈给智能体一个回报奖励。一般来说，这个回报奖励应该反映选取的动作是否正确。针对应急物联网的多切片共存场景，回报奖励应与切片 QoS 满意度和切片资源利用率相关。假设切片 QoS 满意度为 Sat_s，切片资源利用率为 RU_s，则单个切片的奖励回报函数为

$$r_s = \alpha\mathrm{Sat}_s + \beta\mathrm{RU}_s \tag{13}$$

其中，$\alpha(0\leqslant\alpha\leqslant1)$ 为切片 QoS 满意度的权重，$\beta(0\leqslant\beta\leqslant1)$ 为切片资源利用率的权重。整个系统

的奖励回报函数定义为所有切片奖励回报函数之和。如果 β 与 α 引入线性关系，即 β=1−α，可以定义一种自适应的奖励回报模型，能够自动调整这 2 个权重值，自动平衡 2 个独立因素对奖励回报的影响[20]。采用基于分数的合并机制，使奖励模型能够自动学习和调整以适应新的场景。

$$\alpha = \sigma(\mathrm{Sat}_s - \mathrm{RU}_s) \tag{14}$$

其中，$\sigma(\cdot)$ 为 Sigmoid 函数，$\sigma(x) = \dfrac{1}{\mathrm{e}^{-x}}$。$\alpha \in [0,1]$ 表示每个奖励度量的重要性。Sigmoid 函数常被用作神经网络的激活函数，将变量映射到 0～1。由于满意度函数和资源利用率均为 0～1，因此式(14)中的 α 也为 0～1，从而保证式(13)的 r_s 为 0～1。自动切片资源预留算法流程如下。

1) 初始化记忆池容量 D 和 mini-batch 样本数 d。

2) 初始化输入状态和输出动作空间的维度，并随机初始化神经网络系数。

3) 设定 epsilon 贪心算法的概率控制参数 ε。

4) 根据当前状态 s 选取动作，具体动作的选取采用 epsilon 策略，即随机产生一个值 π，如果 $\pi < \varepsilon$，则从输出动作集合中随机选择一个动作 a，否则选择具有最大 Q 值的动作 a。

5) 执行动作 a，即增加或减少切片资源的系统占比，并将切片资源比例映射为基站资源比例，进而通过终端物理资源分配，生成系统反馈，即用户 QoS 满意度和资源利用率，并通过式(13)和式(14)计算生成奖励回报 r_s。

6) 统计切片在各个基站上的资源数量和比例，更新切片在系统资源的占比，产生下一个环境状态 s'。

7) 将四元组 tuple$<s,a,r,s'>$ 作为一个新样本存储到记忆池中。

8) 如果记忆池已满，则随机选一批数据作为 mini-batch 进行神经网络的训练。

9) 如果当前 episode 的索引值达到上限，则算法终止，否则跳到步骤 4)。episode 表示增强学习智能体在环境中执行某个策略从开始到结束这一过程。

上述流程中，步骤 5)采用自适应 Reward 函数的定义为所提算法的主要创新点，简化了人工参数配置，并能够自动完成参数配置，Dueling 网络结构加速了算法收敛，本文在多切片共存场景对自动切片资源预留算法进行了性能验证。

3.2　基于形状的用户资源定制

在既定的切片资源约束的前提下，基站会根据切片可利用的资源数量，为关联到该基站的连接态终端分配物理 RB 资源。因为每个基站的带宽是有限的，所以一个重要问题是在一个调度周期 T 内，基站如何协调调度更多的终端数据流最大化 RB 资源的利用率，即尽可能减少资源的空闲。又因为每个切片应用的服务质量要求是差异化的，所以需要对用户资源进行定制。综合以上两点需求，本文针对异构切片共存问题，采用基于形状的切片内物理资源分配模型。类似地，文献[21]将频谱资源建模为离散的二维时间频率网格，通过定义服务质量需求的效用函数，将物理 RB 分配建模为二维几何背包问题，采用一种启发式算法搜索和取舍不同组合的资源分配选项，并根据传输速率和服务时延等指标评估其算法性能，本文主要将其扩展至多异构切片场景。

针对应急物联网及多切片共存应用场景，假设部署了 K 个基站，对于任意基站 $k \in \{1,2,\cdots,K\}$，均部署了 S 个切片，而对于任意切片 $s \in \{1,2,\cdots,S\}$，各个切片都有不同的服务质量要求。终端均匀分布在基站周围，任意终端 $u \in \{1,2,\cdots,U_s\}$ 都可能请求切片 s 的服务。假设同一切片服务的所有终端的服务质量要求都相同，而最小速率要求和最大时延要求分别为 $c_u^{\min}$ 和 $\mathrm{Td}_u^{\max}$。针对无线接入网频谱资源，虚拟化的资源粒度可定义为时隙和带宽的乘积[22]，本文仅考虑虚拟化的资源粒度为 RB 级。假设基站的系统带宽为 B，频域资源离散化表示为 M 个连续的 RB，每个 RB 的带宽为 B_m；时域资源离散化表示为 T 个连续子帧，每个子帧的时长为 t_l，整个调度周期的时间长度为 Tt_l。因此，根据香农定理，用户 u 从基站 k 得到一个 RB (t,m) 可以获得平均传输速率为

$$c_{uk}^{(t,m)} = \frac{B_m}{T}\mathrm{lb}(1+\gamma_{uk}) \tag{15}$$

其中，γ_{uk} 为终端 u 和基站 k 之间信道传播的信干噪比。充分考虑切片用户之间的 QoS 差异化需求，基站需要为特定的切片用户调度定制化物理资源，并协调多切片用户在一个调度帧内的资源分配。例如，用户关联策略需要考虑基站回程可用资源的多少；在给定资源条件下，为异构切片用户协调选择恰当的传输时隙，满足其差异化的传输时延要求；为了保证所有切片用户 u 的实际等待时延满足其最大时延要求的上限，每个用户发送的 2 个连续相邻数据分组的时间间隔应小于 $\mathrm{Td}_u^{\max}$。

基于上述分析，本文提出根据切片用户的最小传输速率和最大等待时延要求，即更精细的 QoS 需求，计算用户发送数据流所需的时隙和频域 RB 分布的形状，并进行基于形状的物理资源分配。

$$n_u^h = \left\lceil \frac{Tt_l}{\mathrm{Td}_u^{\max}} \right\rceil \tag{16}$$

$$n_u^v = \left\lceil \frac{c_u^{\min}}{n_u^h c_{uk}^{(t,m)}} \right\rceil \tag{17}$$

其中，n_u^h 为所需时隙数量的最小值，n_u^v 为所需频域 RB 数量的最小值。因此，每个终端用户 u 在基站 k 上可以获得的实际传输速率为

$$c_{uk} = n_u^h n_u^v c_{uk}^{(t,m)} \tag{18}$$

由式(16)～式(18)可计算出每个切片用户实际需要的 PRB 数量为 $n_u^h n_u^v$，因此，该 PRB 分配问题可建模为一个二维几何背包问题。其目的是在有限资源约束条件下，最大化系统频谱资源利用率[23-24]。基于形状的 PRB 映射如图 3 所示，基站的整体 PRB 资源集合可以看作一个由时频域组成的资源网格 $\boldsymbol{G}$，此资源网格的 RB 数量是有限的。

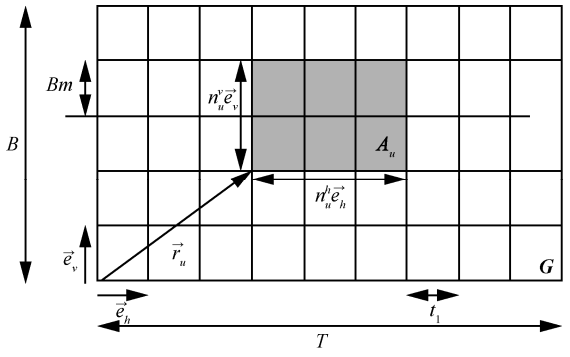

图 3 基于形状的 PRB 映射

假设一个调度周期 T 中，同一个 RB 只能被分配给一个终端用户，而不能重复分配，则终端用户所获得的 PRB 集合可看成一个矩形块 $\boldsymbol{A}_u$。$\boldsymbol{A}_u$ 包含位置信息，为一个向量，可以借助效用函数 UF(u) 来评估分配结果的好坏。一个数据流的 QoS 性能越好，则其分配的 $\boldsymbol{A}_u$ 的效用值越高；反之越低。因此，该二维背包问题的目标是最大化所有资源块的效用之和。目标函数定义为

$$\max \sum_{u \in U_{sk}} \mathrm{UF}(u) x_u$$

$$\text{s.t.} \quad \text{(a)}\ \boldsymbol{A}_u \subseteq G_s, \forall s \in S, u \in U_{sk}$$

$$\text{(b)}\ \boldsymbol{A}_u \bigcap \boldsymbol{A}_z = \varnothing, \forall u \neq z$$

$$\text{(c)}\ x_z \in \{0,1\}$$

$$G_s := \left\{ \overrightarrow{r_{\text{zero}}} + n_s^h \vec{e}_h + n_s^v \vec{e}_v \mid n_s^h, n_s^v \geqslant 0, n_s^h \leqslant \frac{T}{T_b^{\min}}, n_s^v \leqslant \frac{B_s}{B_{sc}^{\min}} \right\}$$

$$\boldsymbol{A}_u := \left\{ \begin{array}{l} \vec{r}_u + n_u^h \vec{e}_h + n_u^v \vec{e}_v \mid n_u^h, n_u^v \geqslant 0, \vec{r}_u = \beta_u^h \vec{e}_h + \beta_u^v \vec{e}_v, \\ \beta_u^h, \beta_u^v \geqslant 0, n_u^h + \beta_u^h \leqslant n_s^h, n_u^v + \beta_u^v \leqslant n_s^v \end{array} \right\} \tag{19}$$

其中，约束条件(a)表示为用户 u 分配的矩形块 $\boldsymbol{A}_u$ 的大小不能超过其所属切片 s 时频网格 G_s 的边界范围，即为用户分配的物理资源块是有限制的；约束条件(b)表示 2 个用户资源 $\boldsymbol{A}_u$ 之间不能交叠，即相互隔离并独立；约束条件(c)表示用户 u 分配矩形资源块的决策变量，0 表示不分配，1 表示分配。值得注意的是，所有的矩形块 $\boldsymbol{A}_u$ 不能进行旋转操作，即矩形块的边必须和资源网格的边保持平行。时频资源网格的资源块的填充目的是在保证用户 QoS 满意度的前提下，最大化频谱资源利用率。本文采用左下对齐填充（BLP, bottom left-justified packing）算法对上述模型进行数值求解，其目标是最小化矩形块填充的高度[25]。

4 仿真结果

4.1 实验配置

本文系统仿真的场景配置参考了 5G 接入网相关标准，主要系统参数配置如表 1 所示。4 个基站均匀部署于 700 m×700 m 的范围内，基站的覆盖半径为 150 m，每 2 个相邻基站保持 120 m 的固定距离。针对无线传播环境，采用的路损模型为

$$\mathrm{PL} = 20\lg(d) + 20\lg(f) - 27.55$$

其中，d 为用户与基站间的距离，f 为信道频率，PL 单位为 dB。

针对多异构切片共存，本文定义了 4 个不同类型的切片实例，每一个具体切片提供特定的服务，其 QoS 需求各不相同。1) EIoT 切片具有最高优先级，其最大时延需求为 10 ms，最小速率需求为 10 kbit/s，数据分组大小为 120 bit，分组到达率为 100 packet/s[26]；2) HDTV 切片最小速率需求为 500 kbit/s，最大时延需求为 120 ms，数据分组大小为 4 000 bit [27]；3) MIoT 切片最大时延需求为 105 ms，最小速率需求为 12 kbit/s，数据分组大小为 500 bit，数据分组到达服从指数分布，平均为 100 packet/s；4) UEb 切片最小速率需求为 100 kbit/s，最大时延需求为 100 ms，其数据分组大小为 400 bit。仿真实验共持续 420 s，即用户持续传输数据分组

1 000 s。UEb 切片和 HDTV 切片的数据分组被建模为指数分布到达，平均为 100 packet/s。

表 1　　系统参数配置

参数	值
切片类别	4
基站数量/个	4
用户数量/个	189～435
系统带宽/ MHz	5
基站发射功率/ dBm	30
子帧长度/ms	1
帧长度/ms	10
基站的子信道个数/个	25
用户分组到达率/(packet·s^{-1})	[UEb:100,HDTV:100,MIoT:100,EIoT:100]
切片最小速率需求/(kbit·s^{-1})	[UEb:100,HDTV:500, MIoT:12, EIoT:10]
切片最大时延需求/ms	[UEb:100,HDTV:120,MIoT:105,EIoT:10]
数据分组大小/bit	[UEb:400,HDTV:4 000,MIoT:500,EIoT:120]
数据分组到达方式	EIoT:均匀分布,其他:指数分布
动态切片周期/ms	200

算法参数配置如下，DRL 算法学习率为 0.01，epsilon-greedy 值为 0.07，记忆池的大小为 8 000 条样本记录，每个 mini-batch 包含 32 条数据记录样本。基于现有文献调研，本文方法与 4 个已有算法（即 Q-leaning[14]、NVS[7]、NetShare[12] 和 DQN[16]）进行仿真对比分析。

1) Q-learning

文献[14]针对 5G 网络的一种特定应用（触觉通信）进行动态资源切片和定制。切片策略基于强化学习（Q-learning）技术，该技术将资源分配给具有不同需求的不同切片，并寻求最佳解决方案。切片策略根据流量需求估计为切片提供资源。然而，资源切片是在 RB 级别完成的，会使状态空间变得非常大，并导致维数灾难。由于 Q-learning 无法解决复杂的机器学习问题，因此 Q-table 无法收敛，并且 Hap-SliceR 采用 Q-learning 强化学习技术，无法为不同种类流量的资源切片问题找到最佳解决方案。

2) NVS

文献[7]将全局视图设置称为静态切片资源配置，也称为 NVS。这种方案假设切片的每个用户信道状态预先已知，即不考虑重新关联，考虑各个切片权重，并对资源进行统计配置。因此，资源配置仅基于网络切片的权重。

$$\varpi_s = \sum_{u=1}^{U_s} R_u, \forall u \in U_s \tag{20}$$

其中，ϖ_s 为整个网络中切片 s 的权重，由其所有用户的总数据速率需求定义；U_s 为 s 的用户数。根据文献[7]的静态切片资源配置，该切片的网络资源共享的固定权重为

$$\mathrm{W\_rat}_s = \frac{\varpi_s}{\sum_{s=1}^{S} \varpi_s}, \forall s \in \{1,2,\cdots,S\} \tag{21}$$

通过 $\mathrm{W\_rat}_s$ 计算切片的资源配置，即利用式(20)和式(21)确定基站之间的资源分配。在 NVS 中，切片资源份额是通过初始切片中切片的资源需求比例计算的。NVS 有 2 个缺点：首先，这种跨网络切片的总资源利用受到静态的每个基站资源预留的影响；其次，NVS 不考虑实时和非实时的流量类别。

3) NetShare

文献[11]提出的 NetShare 认为切片的资源部分在系统级别具有最大和最小的资源限制。NetShare 为每个切片设置基站级资源分配的上限和下限，假设一个基站的所有资源都被所有切片分配完全。根据比例公平原则，在基站上通过最大化按资源分配比例缩放的切片需求比例的效用函数，可以在 NetShare 中周期性地确定切片的动态资源分配。NetShare 为特定切片保留的资源在所有基站之间动态分配。

4) DQN

文献[15]针对雾接入网缓存资源切片划分和模式选择问题，提出了基于深度强化学习的解决办法。文献[16]针对异构切片无线资源切片划分问题，提出了基于 DQN 的资源需求动态预测算法，并采用 2 个切片实例来验证有益效果。本文主要在文献[16]的基础上，扩展为 4 个异构切片实例。

4.2　算法收敛性

本节对基于 DRL 的切片资源预留算法的收敛性进行对比。仿真实验运行了 3 000 个 episode，每个 episode 时长为 200 ms，每 50 个 episode 取点并绘制 Reward 曲线，如图 4 所示。Q-learning 算法的状态数量为 128 个，Reward 函数定义如式(13)所示，β 表示资源利用率的权重，本文设 β 为 0 或 1。当 $\beta=1$ 时，Dueling DQN 和 DQN 从 episode=500 开

始收敛，其系统 Reward 达到最大并归一化为 0.95。Q-learning 约从 episode=2 100 开始收敛，其系统 Reward 为 0.9。当 $\beta=0$ 时，Dueling DQN 和 DQN 同样从 episode=500 开始收敛，但是其最大系统 Reward 为 0.88。Q-learning 从 episode=2 100 后开始收敛，其最大系统 Reward 为 0.75。基于 Dueling DQN 的资源预留算法比 DQN 和 Q-learning 算法的收敛速度更快。

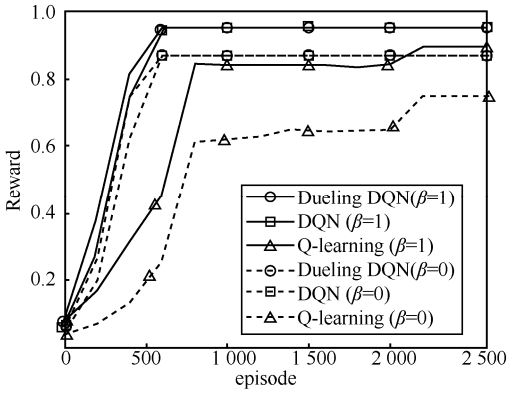

图 4　Reward 曲线

4.3　切片级资源预留对比

综合考虑多个异构切片共存的场景，本节基于 Dueling DQN、DQN 和 Q-learning 的切片资源分配结果进行比较。在资源视图中，定义了切片预留的资源 Reserved、切片分配的资源 Allocated，以及切片实际使用的资源 Used。通常预留的资源结果往往大于实际分配的资源结果。如果某一切片的用户数量增加，剩下的未使用资源可以重新分配给其他切片，从而保证了切片之间的安全隔离。本节配置 UEb、HDTV、EIoT 和 MIoT 的用户数量最大值分别为 60、11、240 和 124。当各个切片用户数量不断增加时，DRL 智能体会自动调整各切片间的资源分配，并将切片的资源比例动态映射到每个基站，最后进行用户 PRB 分配。

图 5～图 7 为基于 DRL 的 3 种算法收敛时的资源分配情况。从图 5 可以看出，在高负载情况下，Dueling DQN 的 Used 和 Allocated 很接近，但是远小于 Reserved，且其 Allocated 比例之和最大为 0.752。从图 6 可以看出，DQN 造成 HDTV 切片的 Allocated 和 Used 差距较大，并且其 Allocated 比例之和最大为 0.824。从图 7 可以看出，Q-learning 造成 Reserved、Allocated 和 Used 分配异常。在轻负载时，HDTV、MIoT 和 EIoT 切片 Used 接近 Allocated，并且其 Allocated 比例之和最大为 0.95。综合上述结果可知，相比 Q-learning 和 DQN，基于

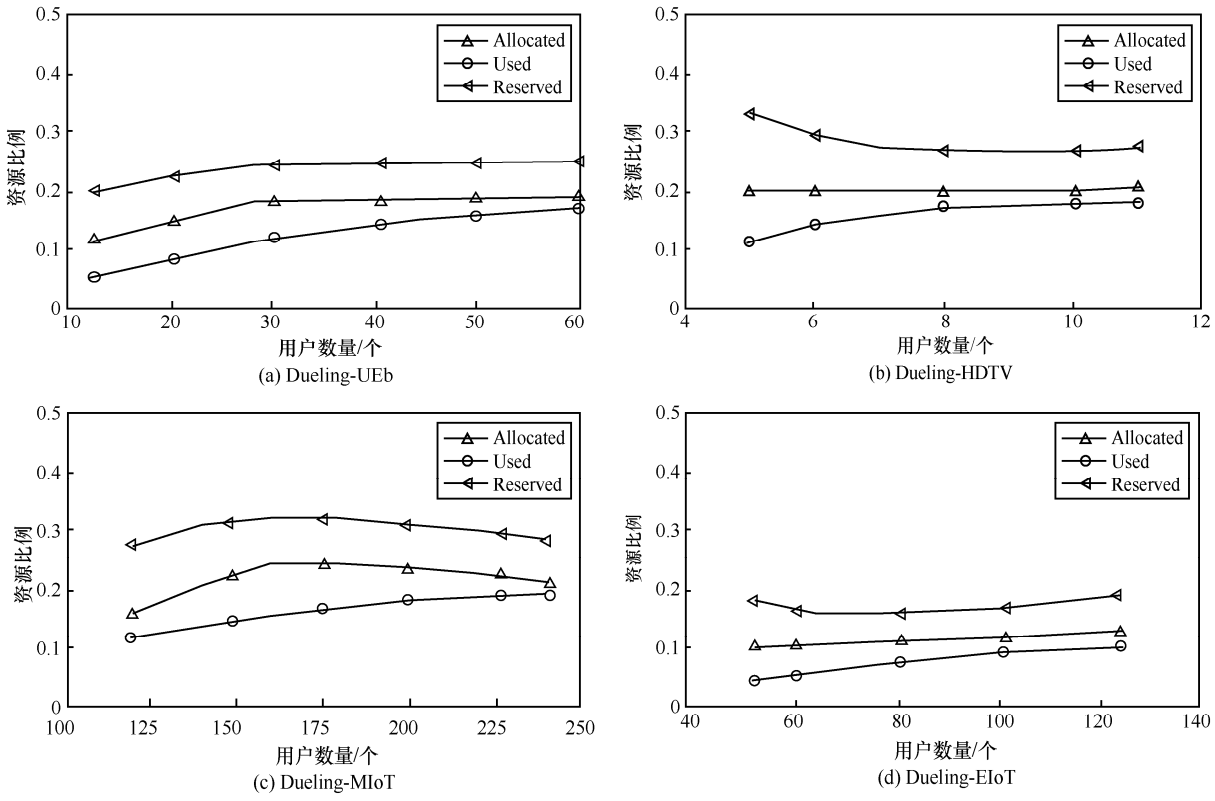

图 5　Dueling DQN 的资源分配情况

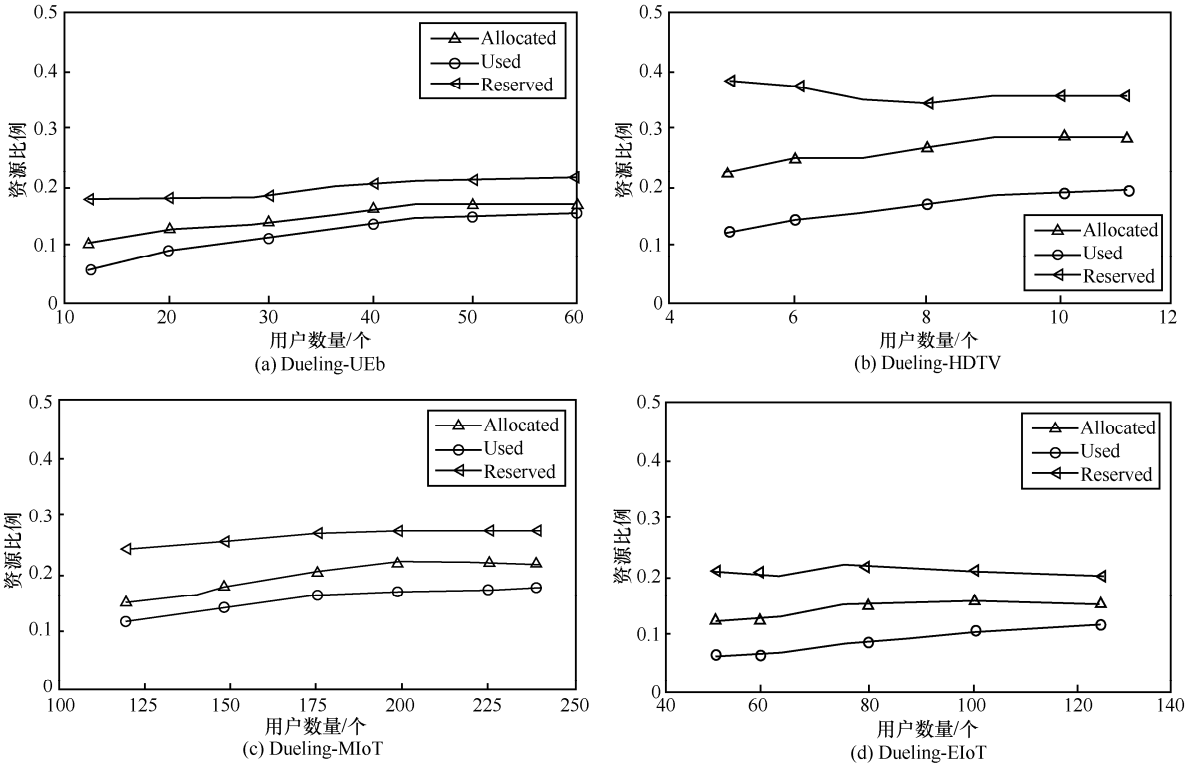

图 6　DQN 的资源分配情况

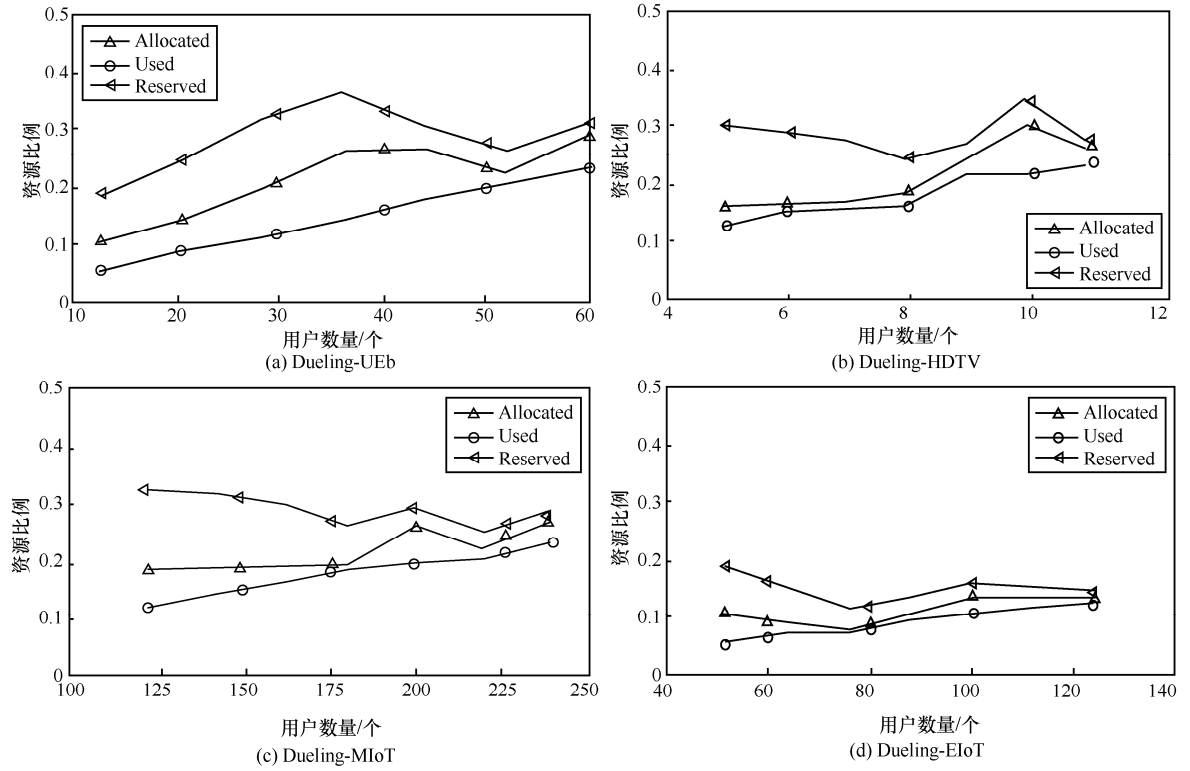

图 7　Q-learning 的资源分配情况

Dueling DQN 的资源需求预测和预留结果更加准确可靠，更节省资源，即能以最少的资源准确满足异构切片用户的差异化需求。

4.4　切片级性能对比

通常，切片资源分配的不同造成切片满意度和切片资源利用的性能不同。本节对 DRL（含 DQN

和 Dueling DQN）、NVS 和 NetShare 进行比较，评估 DRL 切片资源分配的性能。图 8 和图 9 分别给出了 4 种方法切片满意度性能和切片资源利用率性能的对比。从图 8 可以看出，MIoT 切片在用户数量为 200 时，NetShare 方法造成切片满意度降至 0.5 以下；NVS 方法造成 MIoT 切片和 EIoT 切片都存在切片满意度小于 0.5 的情况；针对 DQN 分配结果，当 UEb 切片在用户数量为 52 时，切片满意度降至 0.5 以下；Dueling DQN 所有切片的满意度都保持在 0.5 以上。类似地，从图 9 可以看出，NVS 和 NetShare 造成部分切片的满意度低于 0.5 时，其资源利用率为 1，从而证明了切片需求预测和资源预留的不准确，即其 Allocated 不足。Dueling DQN 能够保证 4 个切片的资源利用率都保持在可接受的水平。可以说明，多异构切片共存情况下，基于 Dueling DQN 的资源分配方法具有最佳的性能，可以自动平衡切片满意度和资源利用率的折中。

4.5　切片间资源隔离

针对异构切片共存场景，除了用户满意度和资源利用率指标外，还需要对切片间安全隔离效果进行评估。切片间的安全隔离是指，当某个切片遭受安全攻击时，如 DDoS（distributed denial of service），

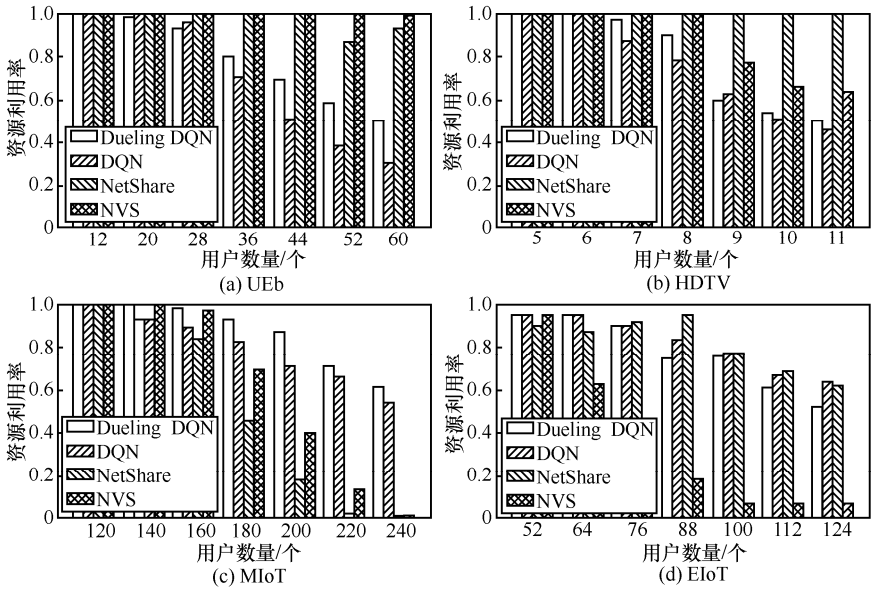

图 8　切片满意度性能对比

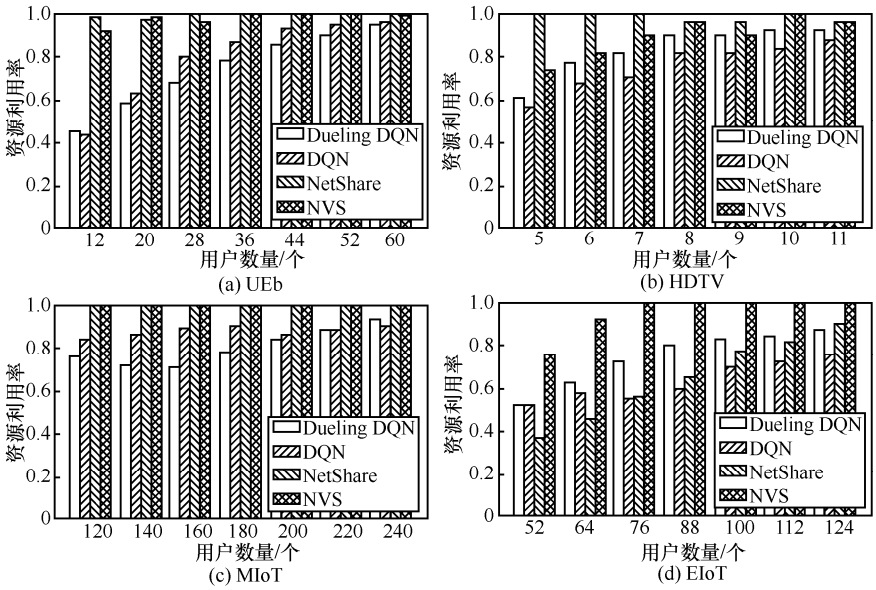

图 9　切片资源利用率性能对比

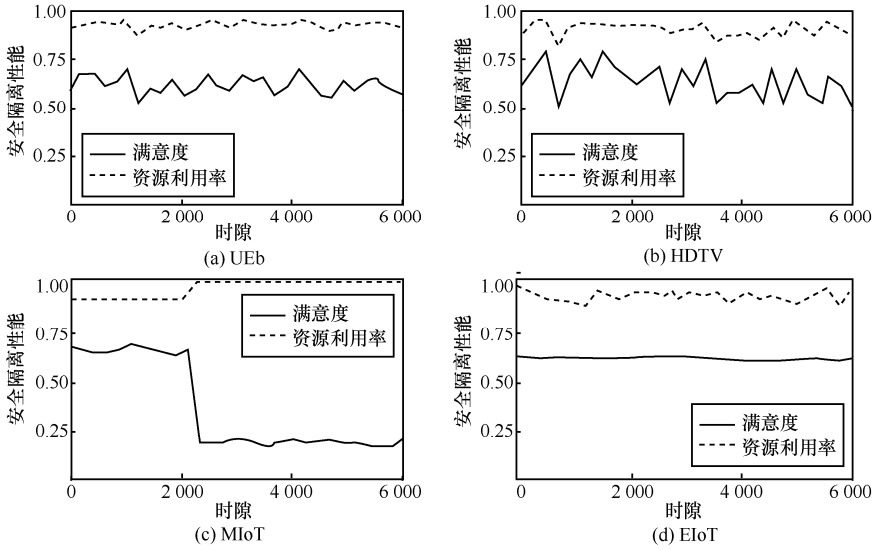

图 10　切片的资源隔离

其他与之共存的切片的性能不受到影响。所以，安全隔离性能是保障用户满意度和系统资源利用率的前提。本节设置 UEb、HDTV、MIoT 和 EIoT 切片的用户数量分别为 60、11、240 和 124。决策周期为 6 000 个时隙，每个时隙长度为 1 ms。根据最后一个 episode 的 4 个切片的资源分配结果，其 Allocated 的资源比例分别为 0.188、0.211、0.216 和 0.137。从第 2 000 个时隙开始，MIoT 的用户数量增加至 420，图 10 给出了 Dueling DQN 的切片性能隔离结果。从图 10 可以看出，UEb 和 HDTV 的切片满意度和资源利用率指标随着时隙的增加呈现一定的波动，但其切片资源均未用完。而 EIoT 的切片满意度和资源利用率则一直保持稳定水平。从第 2 000 个时隙开始，MIoT 切片的用户数量突然增加至 420，其切片满意度下降至 0.25 以下，同时资源利用率上升至 1，但是并没有导致其他 3 个切片的性能大幅下降。

5　结束语

针对应急物联网切片资源智能调度分配问题，本文提出了基于深度强化学习的资源预留和切片间的资源比例动态调整策略，以保证切片 QoS 满意度为前提，最大化各个切片的资源利用率，并保证切片间的性能安全隔离。针对异构切片差异化服务质量要求，物理资源定制问题被建模成一个二维背包问题，使用 BLP 算法进行求解，尽可能减少资源的浪费。系统仿真表明，基于 DRL 的资源预留策略的各方面性能均优越于 NVS 和 Netshare。

参考文献：

[1] FOUKAS X. Network slicing in 5G: survey and challenges[J]. IEEE Communications Magazine, 2017, 55(5): 80-87.

[2] KALOXYLOS A. A survey and an analysis of network slicing in 5G networks[J]. IEEE Communications Standards Magazine, 2018, 2(1): 60-65.

[3] YOUSAF F Z, SCIANCALEPORE V, LIEBSCH M, et al. MANOaaS: a multi-tenant NFV MANO for 5G network slices[J]. IEEE Communications Magazine, 2019, 57(5): 103-109.

[4] COSTA-PEREZ X, SWETINA J, GUO T, et al. Radio access network virtualization for future mobile carrier networks[J]. IEEE Communications Magazine, 2013, 51(7):27-35.

[5] ZAKI Y, ZHAO L, GOERG C, et al. LTE wireless virtualization and spectrum management[C]//Wireless & Mobile Networking Conference. Piscataway: IEEE Press, 2010: 1-6.

[6] BHANAGE G, SESKAR I, MAHINDRA R, et al. Virtual base station: architecture for an open shared WiMAX framework[C]//Second ACM SIGCOMM Workshop on Virtualized Infrastructure Systems and Architectures. New York: ACM Press, 2010:1-10.

[7] KOKKU R, MAHINDRA R, ZHANG H, et al. NVS: a substrate for virtualizing wireless resources in cellular networks[J]. IEEE/ACM Transactions on Networking, 2012, 20(5): 1333-1346.

[8] KOKKU R, MAHINDRA R, ZHANG H, et al. Cell-Slice: cellular wireless resource slicing for active RAN sharing[C]//Fifth International Conference on Communication Systems and Networks. Piscataway: IEEE Press, 2013:1-10.

[9] PETERS S W, HEATH R W. The future of WiMAX: multi-hop relaying with IEEE 802.16j[J]. IEEE Communications Magazine, 2009, 47(1): 104-111.

[10] HOLMA H, TOSKALA A. LTE for UMTS-OFDMA and SC-FDMA based radio access[M]. New Jersey: John Wiley & Sons, 2009.

[11] MAHINDRA R, KHOJASTEPOUR M A, ZHANG H, et al. Radio access network sharing in cellular networks[C]//21st IEEE International Conference on Network Protocols. Piscataway: IEEE Press, 2013: 1-10.

[12] LIANG C, YU F R. Distributed resource allocation in virtualized wireless cellular networks based on ADMM[C]//2015 IEEE Conference on Computer Communications Workshops. Piscataway: IEEE Press, 2015: 360-365.

[13] LEE Y L, LOO J, CHUAH T C, et al. Dynamic network slicing for multitenant heterogeneous cloud radio access networks[J]. IEEE Transactions on Wireless Communications, 2018, 17(4): 2146-2161.

[14] AIJAZ A. Hap-SliceR: a radio resource slicing framework for 5G networks with haptic communications[J]. IEEE Systems Journal, 2018, 12(3): 2285-2296.

[15] XIANG H, YAN S, ANDPENG M. A realization of fog-RAN slicing via deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2020, 19(4): 2515-2527.

[16] SUN G, 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(1): 45758-45772.

[17] ROSS S. Introduction to probability models, 11th ed[M]. Salt Lake City: Academic Press, 2014.

[18] TANG L, ZHANG Y, LIANG R, et al. Virtual resource allocation algorithm for network utility maximization based on network slicing[J]. Journal of Electronics & Information Technology, 2017, 39(8): 1812-1818.

[19] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2017, 40(1):1-27.
LIU Q, ZHAI J W, ZHANG Z Z, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.

[20] LIU F, REN X, LIU Y, et al. simNet: stepwise image-topic merging network for generating detailed and comprehensive image captions[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2018: 137-149.

[21] GONZALEZ A, KUEHLMORGEN S, FESTAG A, et al. Resource allocation for block-based multi-carrier systems considering QoS requirements[C]//2017 IEEE Global Communications Conference. Piscataway: IEEE Press, 2017:1-7.

[22] RICHART M, BALIOSIAN J, SERRAT J, et al. Resource slicing in virtual wireless networks: a survey[J]. IEEE Transactions on Network and Service Management, 2016, 13(3): 1-15.

[23] 马康, 高尚. 分布估计算法求解矩形件排样优化问题[J]. 电子设计工程, 2017, 25(2): 49-54.
MA K, GAO S. Solution to optimize cutting pattern in rectangular packing problem based on estimation of distribution algorithm[J]. Electronic Design Engineering, 2017, 25(2):49-54

[24] 曾兆敏, 王继红, 管卫利. 二维板材切割下料问题的一种确定性算法[J]. 图学学报, 2016, 37(4): 471-475.
ZENG Z M, WANG J H, GUAN W L. A deterministic algorithm for solving the problem of two-dimensional sheet cutting stock[J]. Journal of Graphics, 2016, 37(4):471-475.

[25] CHAZELLE B. The bottom-left bin-packing heuristic: an efficient implementation[J]. IEEE Transactions on Computers, 1983, C-32(8): 697-707.

[26] HELMERSSON K W, ANSARI J. Ultra-reliable and low-latency communication for wireless factory automation: From LTE to 5G[C]// IEEE 21st International Conference on Emerging Technologies and Factory Automation. Piscataway: IEEE Press, 2016: 1-8.

[27] MAMMAN M, HANAPI Z H, ABDULLAH A, et al. Quality of service class identifier (QCI) radio resource allocation algorithm for LTE downlink[J]. PLoS One, 2019,14(1): e0210310.

[作者简介]

孙国林（1978− ），男，河北唐山人，博士，电子科技大学副教授、硕士生导师，主要研究方向为人工智能、区块链和移动智能系统等。

欧睿杰（1989− ），男，四川成都人，电子科技大学博士生，主要研究方向为人工智能、区块链、移动网络资源管理等。

刘贵松（1973− ），男，山东临沂人，博士，电子科技大学教授、博士生导师，主要研究方向为类脑计算、机器学习和模式识别与智能系统等。