DOI:10.19652/j.cnki.femt.2405915

5G异构网络中基于多目标Actor-Critic的资源分配*

曾韦健[1] 李 晖[2]
(1.南京信息工程大学电子与信息工程学院 南京 210044;
2.中国航空研究院研究生院 扬州 225006)

摘 要:在5G异构网络(heterogeneous network,HetNet)中广泛部署小基站可以提高网络容量和用户速率,但密集部署也会产生严重干扰和更高能耗问题。为了最大化网络能量效率(energy efficiency,EE)并保证用户服务质量(quality of service,QoS),提出了一种在小蜂窝基站中嵌入能量收集器供电的资源分配方案。首先,针对网络系统的下行链路,将频谱和小基站发射功率分配问题建模为联合优化系统能效和用户满意度的多目标优化问题。其次,提出了基于深度强化学习的多目标演员-评论家(multi-objective actor-critic,MAC)资源分配算法求解所建立的优化模型。最后,仿真结果表明,相比于其他传统学习算法,能量效率提高了11.96%~12.37%,用户满意度提高了11.45%~27.37%。

关键词:5G异构网络;能量效率;用户满意度;多目标优化;深度强化学习
中图分类号:TN929.5 **文献标识码**:A **国家标准学科分类代码**:510.5030

Resource allocation based on multi-objective Actor-Critic in 5G heterogeneous networks

Zeng Weijian[1] Li Hui[2]
(1. College of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Graduate School of Chinese Aeronautical Establishment, Yangzhou 225006, China)

Abstract: The widespread deployment of small base stations in 5G heterogeneous networks (HetNet) can improve network capacity and user rates, but the dense deployment will also cause severe interference and higher energy consumption problems. In order to maximize the network energy efficiency (EE) and guarantee the user quality of service (QoS), this paper presents a resource allocation scheme that embeds energy harvester power supply in small cell base stations. Firstly, for the downlink of the network system, the spectrum and small base station transmit power allocation problem was modeled as a multi-objective optimization problem to jointly optimize system energy efficiency and user satisfaction. Secondly, a multi-objective actor-critic (MAC) resource allocation algorithm based on deep reinforcement learning was proposed to solve the established optimization model. Finally, simulation results show that compared with other traditional learning algorithms, the energy efficiency of the proposed algorithm is improved by 11.96%~12.37%, and the user satisfaction is improved by 11.45%~27.37%.

Keywords: 5G heterogeneous network; energy efficiency; user satisfaction; multi-objective optimization; deep reinforcement learning

0 引 言

随着5G(5th generation)技术的不断发展,日益增长的业务需求和海量数据使得传统网络难以满足[1]。为支持用户终端高速、海量的数据服务并提供更好的覆盖,5G蜂窝网络将广泛部署更密集的小蜂窝基站(small-cell base station,SBS),这种方法可以提高网络覆盖和容量,但大量部署SBS也带来了小区间干扰、资源浪费等新的问题,尤

收稿日期:2024-01-06
* **基金项目**:国家自然科学基金(61661018)、江苏省基础研究计划青年基金(BK20210064)、无锡市科技创新创业资金(WX03-02B0137-022200-34)项目资助

其是智能终端处理的任务数量庞大，将会带来巨大的网络能耗。同时，智能城市、智能交通等5G应用场景也需要更高的服务质量[2-3]。因此，在新一代的移动通信发展布局中，需要制定更为实用的分配方案。

针对上述问题，现有工作主要是研究资源分配（resource allocation，RA）算法来减少能耗。文献[4-5]从随机几何角度分析了由电网和可再生能源供电的多层蜂窝网络的性能，文献[6]在设备到设备（device-to-device，D2D）辅助蜂窝网络场景下，采取随机休眠/唤醒策略来动态关闭SBS以节省网络能源，文献[7-9]提出将基站分簇，用户分组，先聚类再分配的解决方案，这样有利于减轻集群内和集群间的干扰。但上述工作只以能量效率或频谱效率等单一网络性能作为优化目标进行研究，难以权衡其他的网络指标。

为满足5G网络同时对能量效率和用户满意度的需求，可以将资源分配问题建模为以能量效率和用户满意度为目标的多目标优化问题[10]，文献[11]提出了一种非支配排序遗传算法，将能量效率、频谱效率（spectral efficiency，SE）和功耗联合求解，文献[12-13]则分别是在HetNet的D2D和MIMO（multiple input and multiple output）应用场景下去权衡EE和SE。但这些解决方案都是将多目标优化问题转化为单目标优化问题，效率不高，尤其是当HetNet中基站密度进一步增大时，所建立的联合模型空间过大，使得问题的求解困难。

为降低算法的计算复杂度，人工智能（artificial intelligence，AI）算法受到了广泛关注。文献[14]提出了一种能量感知在线算法，以最大限度地提高系统效用，包括吞吐量和公平性，同时考虑了系统的可持续性和稳定性。文献[15]提出了一种具有多移动设备的物联网雾计算系统动态分配方案，可以最大限度地降低移动设备延迟、能耗和系统成本。在AI算法中，强化学习（reinforcement learning，RL）通过智能体与环境的持续交互[16]，广泛应用于系统的决策制定[17]。深度强化学习（deep reinforcement learning，DRL）算法可以看作是在传统RL算法上的改进，对于处理资源分配[18]、优化等复杂问题可以提供更好的解决方案。

综上所述，以往研究工作在采用DRL框架优化网络能效时多是对单一资源进行优化，且很少考虑满足用户的需求，因此，本文在研究5G异构网络中采用DRL算法对网络的能量效率和用户满意度进行联合优化。首先，针对网络下行链路，提出一种面向两层5G异构网络的资源分配框架，即在保证QoS前提下，尽可能使用高效的新型能源，提高能效。其次，在满足QoS的同时，对异构网络中的能量效率和用户满意度联合优化，将优化问题建模为多目标资源优化模型，并引入基于DRL的多目标演员-评论家（multi-objective actor-critic，MAC）算法，采用MAC算法来解决资源的联合分配。最后，对所提算法进行仿真分析，并验证其收敛性。

1 系统模型及问题表述

1.1 系统模型

基于两层5G异构网络的资源分配模型如图1所示，网络中有一个宏基站和多个小基站，分别负责大规模监控和半静态监控。本文考虑5G异构网络的下行链路场景，网络系统中基站的集合表示为 $B=\{B_0,B_1,\cdots,B_n,\cdots,B_N\}$，其中 B_0 代表宏基站（Macro Base Station，MBS），$B_1\sim B_N$ 代表 N 个SBS。MBS作为系统的控制中心，对系统进行统一的用户调度和资源分配。每个SBS上都嵌入了能量收集装置和相应的可充电电池[19]，它们共享有限的带宽资源，并服务于一组 $M=\{1,2,\cdots,m,\cdots,M\}$ 的用户设备（user equipment，UE）。假设系统的总无线带宽被限制为 W，通信网络以正交频分多址（orthogonal frequency division multiple access，OFDMA）技术为基础，将频谱划分为 K 个资源块（resource block，RB），总资源块可表示为 $RB=\{RB_1,RB_2,\cdots,RB_k,\cdots,RB_K\}$，每个资源块带宽 $B_k=W/K$ 相同。

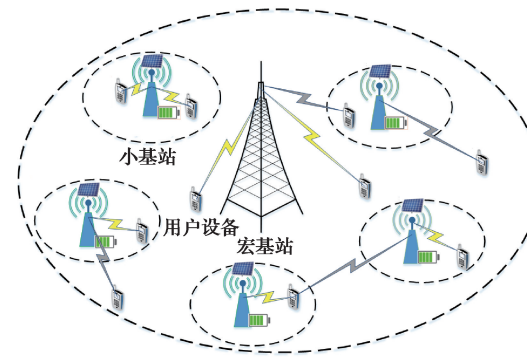

图1 系统模型

Fig. 1 System model

1.2 问题表述

设网络在每个时隙 t 用户设备随机出现，$P_{n,m}(t)$ 表示用户 m 所在的基站 n 在 t 时刻的发射功率。SBS使用电池中收集的能源，每个SBS电池当前的电量为 $E=\{e_1(t),e_2(t),\cdots,e_n(t),\cdots,e_N(t)\}$，电池容量为 E_M。假设SBS总有足够的能量来初始化能量收集装置并发送或接收控制信号，每个基站对用户的发射功率固定。

因此，t 时刻基站 n 向用户 m 传输数据，在资源块 RB_k 上的信干噪比（signal to interference plus noise ratio，SINR）为：

$$\xi_{n,m,k}(t)=\frac{g_{n,m,k}P_{n,m}(t)}{\sum_{i\in N,i\neq n}\sum_{m\in M}x_{i,m,k}g_{n,m,k}P_{n,m}(t)+\sigma^2(t)} \tag{1}$$

式中：$g_{n,m,k}$ 是基站 n 在 RB_k 上到用户 m 的信道增益；

$\sum_{i\in N,i\neq n}\sum_{m\in M}x_{i,m,k}g_{n,m,k}P_{n,m}(t)$ 是用户 m 在 RB_k 上受到其他基站的干扰总和；$\sigma^2(t)$是噪声功率。$x_{n,m,k}$ 是反映基站与用户连接状态的指示变量，$x_{n,m,k}=1$ 表示分配，否则不分配。

根据香农公式，系统的总容量为：

$$C(t)=\sum_{n\in N}\sum_{m\in M}\sum_{k\in K}x_{n,m,k}B_k lb(1+\xi_{n,m,k}(t)) \tag{2}$$

则系统的能量效率为：

$$EE(t)=C(t)/P(t) \tag{3}$$

式中：$P(t)$为基站的总功率，包括 MBS 的静态功率 $P_{MC}(t)$、SBS 的静态功率 $P_{SC}(t)$和 MBS 的发射功率 $P_M(t)$[20]，SBS 发射功率 $P_S(t)$的能量由电池提供，即 $P(t)=P_{MC}(t)+NP_{SC}(t)+MP_M(t)$。

由于每个用户一次只能发出一个请求，因此用户数 M 同时也表示用户向基站发出的最大请求数，计 $r_n(t)$为 t 时用户对基站 B_n 的业务请求数，$r_n(t)\in[0,M]$。假设每个请求都需要 c 个恒定数量的逻辑通道来建立连接并接收服务。因此，在特定时隙 t 中，基站 n 所需的信道数为：

$$C_n(t)=c\cdot r_n(t) \tag{4}$$

设维持一个信道进行数据传输所需的能量为 E_{tr}，则对于电池剩余量为 $e_n(t)$的基站，可提供的信道数：

$$B_c(t)=e_n(t)/E_{tr} \tag{5}$$

因此，可定义用户平均满意度为：

$$\delta(t)=\frac{1}{N}\sum_{n\in N}B_c(t)/C_n(t) \tag{6}$$

用户平均满意度反映了 t 时刻基站 n 对用户请求的满足程度。为联合优化网络能量效率和用户满意度，定义效用函数：

$$\eta(t)=\varepsilon\cdot EE(t)+(1-\varepsilon)\cdot\delta(t) \tag{7}$$

式中：ε 是平衡能效和用户满意度的参数，$\varepsilon\in[0,1]$。本文的优化目标是在保证 QoS 的前提下，最大化效用函数，则联合优化问题可表示为：

$$\max_{\{x_{n,m,k},P_{n,m}(t),e_n(t)\}}\eta(t) \tag{8}$$

$$\text{s.t.}\quad C1:x_{n,m,k}\in\{0,1\}$$

$$\forall m\in\mathbf{M},\forall n\in\{0,1,\cdots,N\},\forall k\in\{1,2,\cdots,K\}$$

$$C2:\sum_{k\in K}x_{n,m,k}\geqslant 1$$

$$C3:\sum_{m\in M}x_{n,m,k}\leqslant 1 \tag{9}$$

$$C4:0\leqslant P_{n,m}(t)\leqslant P_{max}$$

$$C5:0\leqslant e_n(t)\leqslant E_M\ \forall n\in\{1,\cdots,N\}$$

$C1\sim C3$ 表示每个用户可以分配多个带宽资源，且关联同一基站用户分配的带宽不同，$C4$ 表示 SBS 的发射功率数值上不能超过最大发射功率 P_{max}，$C5$ 表示 SBS 用的是电池中收集的能源，不消耗电网能量，存储能量 $e_n(t)$不超过电池容量上限 E_M。

2 算法设计

2.1 马尔可夫决策过程

根据所研究的 5G 异构网络场景，资源联合分配的决策过程可描述为马尔可夫决策过程（Markov decision process，MDP），可将该过程建模为 5 元组 $<\mathbf{S},\mathbf{A},P(s'\mid s,a),R,\gamma>$，其中 $P(s'\mid s,a)$ 是在状态 $s\in$S 选择动作 $a\in$A 并得到下一个状态 $s'\in\mathbf{S}$ 的转移概率；R 为奖励；γ 为折扣率，是用来衡量未来奖励对现在影响的超参数，$\gamma\in(0,1)$。采用强化学习算法解决 MDP 问题，同时考虑到 5G 异构网络中基站的部署和网络环境更为复杂，动作空间和状态空间大小急剧增加，为大幅提高计算速率，在 RL 框架中引入深度神经网络（deep neural networks，DNN），形成深度强化学习。本文采用基于 DRL 的多目标演员-评论家算法来优化效用函数。

标准的强化学习模型包括智能体、环境、状态、动作和奖励等基本要素，智能体基于策略选择动作与当前环境进行交互，以获取奖励，本文选取宏基站作为智能体，根据优化问题的决策变量、约束条件及优化目标定义智能体状态、动作和奖励。

1）状态（state，S），网络状态由每个用户的信干噪比 ξ 和每个 SBS 电池的当前电量 $e_n(t)$决定，因此状态空间定义为：

$$\mathbf{S}(t)=[\xi_1,\xi_2,\cdots,\xi_M,e_1(t),e_2(t),\cdots,e_N(t)] \tag{10}$$

2）动作（action，A），智能体的动作将实现具体的分配策略，包括用户的分配，以及在时隙 t 为每个用户分配资源块和功率，为减小动作空间复杂度，将发射功率离散化为 Z 个等级。因此，在时隙 t 执行的动作空间可定义为：

$$\mathbf{A}(t)=\{m\in\{0,1,\cdots,M\},x_{n,m,k}\in\{0,1\},P_{n,m}(t)\in\{0,P_1,P_2,\cdots,P_Z\}\} \tag{11}$$

3）奖励（reward，R），当系统处于 $s(t)$时，执行一个动作 $a(t)$后，就会获得一个即时奖励 $r(t)$，奖励代表优化目标，本文是联合优化能量效率和用户满意度，因此可以用效用函数表示奖励：

$$R(t)=\eta(t) \tag{12}$$

RL 方法的最终目标是找到最优策略 π^*，以最大化预期的累积奖励值，在 RL 中定义了价值函数 $V_\pi(s)$，表示从当前状态开始在策略 π 下累积奖励的期望值。根据贝尔曼方程[22]，价值函数迭代关系为：

$$V_\pi(s)=R(s,a)+\gamma\sum_{s'\in S}P(s'\mid s,a)V_\pi(s') \tag{13}$$

式(13)表示当前状态的价值函数与未来状态的价值函数之间的关系，这种关系可以分解为在当前状态 s 下选取动作 a 所获取的奖励与未来的折扣奖励之和，这个过程可以一直迭代下去，直到价值函数收敛，通过求解方程可以得到最优策略。

对应的最优策略为：

$$\pi^*(s)=argmax_{a\in A}\{R(s,a)+\gamma\sum_{s'\in S}P(s'\mid s,$$

$a)V_{\pi}^{*}(s')\}$ (14)

即最优策略是在价值函数取得最大值时所选取的动作。

2.2 Actor-Critic 算法

Actor-Critic 算法是一种结合策略梯度(policy gradient,PG)和时序差分(temporal difference,TD)的 RL 算法。典型的 AC 由两部分组成[23]。

1)Actor 部分,Actor 是指策略函数 $\pi_{\theta}(a|s)$,即学习一个策略以得到尽可能高的回报。

2)Critic 部分,Critic 是指价值函数 $V_{\omega}(s)$,对当前策略的价值函数进行评估。

Actor-Critic 算法将基于策略和基于价值的 RL 方法相结合,如图 2 所示,Actor 频繁地观察环境状态,然后通过参数化的策略来给出动作,而 Critic 通过使用参数化的价值函数和从环境中获得的奖励来评估所采取的行动。Critic 的输出用于计算 TD 误差,然后更新 Actor 和 Critic 的参数。

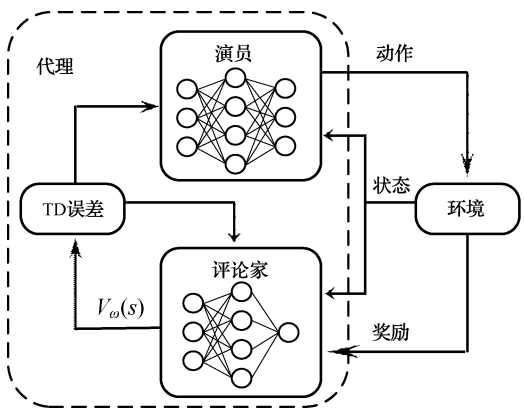

图 2 演员-评论家模型

Fig. 2 Actor-critic model

值函数的估计可以更好地引导策略更新,使得 Actor-Critic 算法在稳定性方面高于纯策略梯度算法。此外,为进一步平滑学习过程,增加样本利用率,在实际操作中可以使用经验回放技巧。相比传统的值函数方法,Actor-Critic 算法收敛速度更快,与神经网络结合可以处理高维状态空间,但 Actor-Critic 算法易于陷入局部最优解,这与超参数(学习率、折扣率等)的取值有关,在后续的仿真实验中,可以通过重复实验,仔细调整参数来得到最佳效果。

本文分别使用两个 DNN 来近似模拟 Actor 的策略函数 $\pi_{\theta}(a|s)$和 Critic 的价值函数 $V_{\omega}(s)$,θ 和 ω 分别为这两函数的参数。

在 Critic 网络中损失函数是目标值的均方误差,定义为:

$$L(\omega) = E[R(s(t),a(t)) + \gamma V_{\omega}(s(t+1)) - V_{\omega}(s(t))]^2 \quad (15)$$

为了最小化损失函数,可使用梯度下降法将值函数的参数 ω 进行更新:

$$\omega = \omega + \alpha_c \delta(t) \nabla_{\omega} V_{\omega}(s(t)) \quad (16)$$

式中:α_c 为 Critic 网络的学习率,$\delta(t)$为 TD 误差。

$$\delta(t) = R(s(t),a(t)) + \gamma V_{\omega}(s(t+1)) - V_{\omega}(s(t)) \quad (17)$$

在 Actor 网络中目标函数定义为价值函数的均值:

$$J(\pi_{\theta}) = \sum_{s\in S} d_{\pi}(s) \sum_{a\in A} \pi_{\theta}(a|s) Q_{\pi}(s,a) \quad (18)$$

式中:$d_{\pi}(s)$是策略 π 下的状态分布函数[24];$\pi_{\theta}(a|s)$是策略分布。为了最大化目标函数,可以更新策略参数 θ:

$$\theta = \theta + \alpha_a \delta(t) \nabla_{\theta} ln \pi_{\theta}(a|s) \quad (19)$$

式中:α_a 为 Actor 网络的学习率。算法 1 为 Actor-Critic 的训练过程。

算法 1 基于多目标 Actor-Critic 的联合资源分配

输入:系统参数、状态空间 S、动作空间 A、转移概率 $P(s'|s,a)$、奖励 $R(s,a)$、折扣率 γ、学习率 α_a 与 α_c

输出:最优资源分配策略 π^*

1. 初始化网络参数 θ、ω、经验池 D
2. for $episode=1:max\_episode$
3. 初始化系统状态 $s(0)$,初始奖励 $R(0)=0$
4. for $t=1:T$
5. 在状态 $s(t)$下根据策略 $\pi_{\theta}(a|s)$执行动作 $a(t)$
6. 根据式(12)获得即时奖励 $R(t)$
7. 智能体得到新状态 $s(t+1)$,并将经验组 $<\mathbf{S},\mathbf{A},P(s'|s,a),R,\gamma>$ 存储到经验池 D
8. 估计状态值 $V_{\omega}(s)$,并计算 TD 误差 $\delta(t)$
9. 根据式(16)和(19)更新网络参数 ω、θ:
$\omega' \leftarrow \omega + \Delta\omega, \theta' \leftarrow \theta + \Delta\theta$
10. end
11. end

3 仿真实验

3.1 仿真参数

本文在 5G 异构网络场景下进行仿真分析,所提算法通过 Python 平台的 Pytorch 框架实现,选用 Ubuntu 18.04 LTS 作为操作系统平台,以保证仿真过程的稳定性。仿真实验在一台配备 Intel Core i7-8700K 处理器(主频 3.7 GHz)、32 GB DDR4 内存、NVIDIA GeForce RTX 3060Ti 显卡和 1TB SSD 存储的台式计算机上进行,以保证有充足的计算资源和存储空间来支持深度强化学习的训练和实验。

仿真参数参考 3GPP[25],如表 1 所示。MBS 设在网络中心,与周围的 SBS 共同关联用户设备,SBS 与 UE 在网络系统中服从泊松点分布。为了与所提算法进行比较,本文在相同环境下分别选取了在强化学习中具有代表性的深度 Q 网络算法(deep Q-learning,DQN)[26]、Q 学习算法(Q-learning)[27]以及经典的贪婪算法(greedy)[28]进行

对比。

本文 MAC 算法包括 Actor 网络和 Critic 网络，设两个神经网络均为 3 层结构，输入层相同，输入数据大小即为状态空间的大小；中间隐藏层的神经元个数也相同，都使用 Tanh 激活函数。不同的是 Actor 网络的输出层，输出的是动作的概率分布，大小等于动作空间大小，选用 Softmax 激活函数；而 Critic 网络输出层输出的是值函数，大小为 1，没有激活函数。

表 1　仿真参数

Table 1　Simulation Parameters

参数	参数值
基站覆盖半径 r/m	{500，50}
小基站个数 N	40
用户设备个数 M	100
资源块个数 K	20
系统带宽 W/MHz	10
基站最大发射功率/dBm	{46，30}
基站发射功率/dBm	{32，20，25，30}
基站静态功率/dBm	{8，5}
小基站电池容量 E_M/AH	500
逻辑通道个数 c	100
阴影衰落/dB	8
宏基站路径损耗	128.1+37.6lg(D[km]) D：MBS 到 UE 距离
小基站路径损耗	140.7+36.7lg(D[km]) D：SBS 到 UE 距离
热噪声功率谱密度/(dBm/Hz)	−174
折扣率 γ	0.9

3.2　仿真与分析

学习率对算法奖励的影响如图 3 所示。对于演员-评论家模型，分别设 Actor 网络和 Critic 网络的学习率依次为 0.001 和 0.005。随着迭代次数的增加，系统的平均奖励随之增加。在训练的前 100 回合，奖励显著增加，其收敛速度随着学习率的增大而加快。此外，所提算法在迭代 300 次之后趋于收敛，逐渐得到最佳策略。从图 3 还可以发现，当两个网络的学习率都增加到 0.005 时，算法 1 开始收敛速度很快，但并没有达到最好的奖励，这是因为当学习率过大时，算法容易陷入局部最优解，出现过拟合情况。因此，在设置学习率时要权衡收敛速率和最优奖励。

为了观察算法的稳定性，对 Actor 网络和 Critic 网络隐藏层的神经元数，都分别设置为 32、64 和 128，如图 4 所示。

从图 4 可以看到，改变网络的神经元数对系统的平均奖励没有造成太大影响，这是因为除了演员-评论家模型本身具有较好收敛性质外，本文还对该模型引入了经验回放技术，以增强训练的效率和稳定性。当然，神经元数的

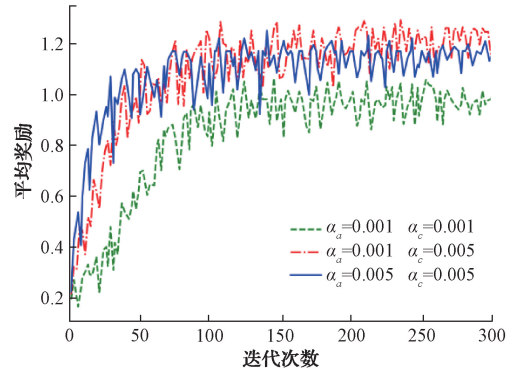

图 3　学习率对平均奖励的影响

Fig. 3　Influence of learning rate on average reward

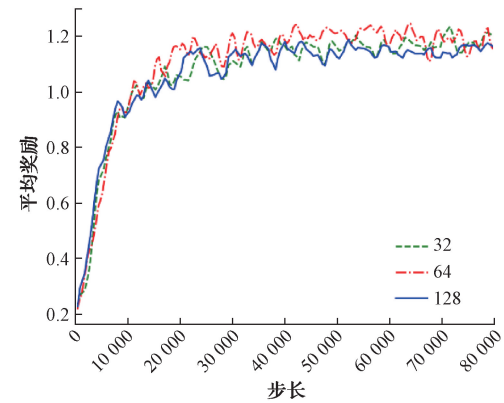

图 4　算法的稳定性

Fig. 4　Stability of the algorithm

增加会造成神经网络训练时间和计算成本的增加，也同样有可能出现过拟合情况，因此，当系统性能满足需求时，神经元数不需要设置过大。

不同算法 SBS 数量对能效的影响如图 5 所示，随着基站数的增加，系统的能效一开始都逐渐提高，但是，当基站数增大到一定限度时，能效的增加逐渐放缓，这是因为一开始基站的增加使得系统的能源得到充分利用。基站的增加同样带来了更多的干扰，当基站数趋于饱和时，干扰的影响变得更为明显，能效难以进一步提高。

此外，在选取算法中，本文 MAC 算法获得的能效最高，优于同为强化学习的 DQN 和 Q-Learning，而这两个算法也明显优于传统的 Greedy 算法。当小基站数为 40 时，MAC 的能效分别比 DQN 和 Q-Learning 提高了 11.96%和 12.37%。

不同算法下 SBS 数量对用户平均满意度的影响如图 6 所示。一方面，基站数的增加可以同时关联更多的用户设备，处理更多的业务请求，使得用户的平均满意度得到提高。

另一方面，本文所提算法仍表现出良好的效果，当小基站数为 20 时，MAC 的平均满意度分别比 DQN 和

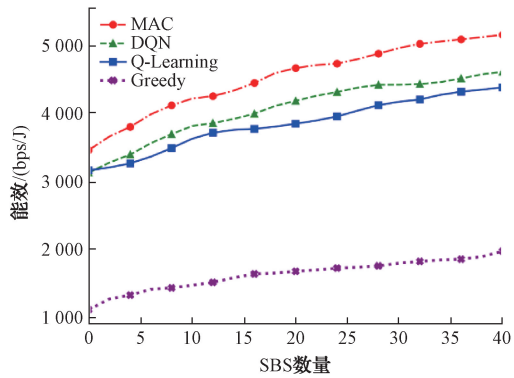

图 5 小基站数量对能量效率的影响

Fig. 5 Influence of the number of small base stations on energy efficiency

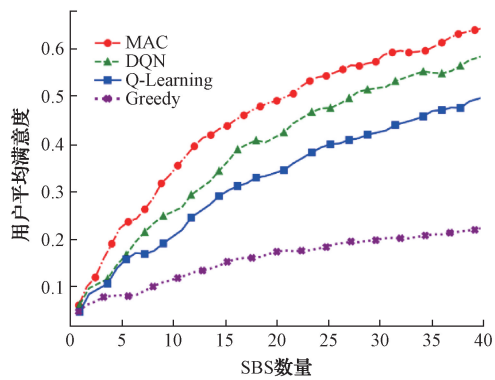

图 6 小基站数量对平均满意度的影响

Fig. 6 Influence of the number of small base stations on average satisfaction

Q-Learning 提高了 11.45%和 27.37%。当然，随着基站数量的不断增加，可提供的服务更多，用户满意度将会进一步提高，但能耗和干扰也会随之增加，所以需要考虑好需求和成本间的平衡。

在 5G 网络中，用户设备的数量同样影响网络资源分配，如图 7 所示，在用户设备数量较少时，资源还比较充足，设备间的干扰不大，能效尚保持在较高水平，分配方案间优势还没完全体现。当 UE 数量逐渐增大时，所有方案的 UE 能效性能均会下降，除了资源的有限外，过高的 UE 数量同样会产生严重的干扰。可以看出，MAC 算法在 UE 的能效中获得了最佳性能，相比之下，MAC 对 UE 的能效分别比 DQN 和 Q-Learning 平均高出了 9.1%和 15.5%，对传统的 Greedy 算法优势明显，这也得益于 MAC 的评价机制。

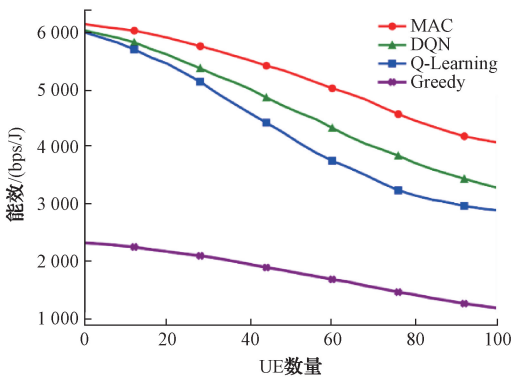

图 7 用户设备数量对能量效率的影响

Fig. 7 Influence of the number of user devices on energy efficiency

4 结 论

本文探讨了 5G 异构网络下行链路的资源分配问题，为尽可能地降低能耗，一方面，对基站本身的硬件条件做了扩展，使得基站可以利用自然资源，直接减少了电网能耗；另一方面，改进了以往的资源分配算法，采用更为灵活实际的 MAC 算法，实现了多目标的联合优化，仿真结果表明，相比于其他经典强化学习算法，所提算法在动态复杂的网络环境中有着良好的鲁棒性，在保证用户 QoS 的前提下，能更好地权衡系统其他性能。

在后续工作中，将进一步研究基于多智能体强化学习的多目标资源分配策略，考虑分布式资源管理问题，以适应移动通信场景多样化、终端智能化和数据海量化的发展趋势。

参 考 文 献

[1] HASNAT M A, RUMEE S T A, RAZZAQUE M A, et al. Security study of 5G heterogeneous network: Current solutions, limitations & future direction [C]. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2019: 1-4.

[2] YU P, YANG M, XIONG A, et al. Intelligent-driven green resource allocation for industrial Internet of things in 5G heterogeneous networks [J]. IEEE Transactions on Industrial Informatics, 2020, 18(1): 520-530.

[3] YU M. Construction of regional intelligent transportation system in smart city road network via 5G network [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(2): 2208-2216.

[4] ARDANUC M, BASARAN M, HMAMOUCHE Y, et al. Energy efficiency analysis in heterogeneous networks: A stochastic geometry perspective [J]. IEEE Open Journal of Vehicular Technology, 2023, 4: 438-443.

[5] LAM T T, DI RENZO M. On the energy efficiency of heterogeneous cellular networks with renewable energy sources-A stochastic geometry framework [J]. IEEE Transactions on Wireless Communications,

2020，19(10)：6752-6770.

[6] PANAHI F H，PANAHI F H，OHTSUKI T. Energy efficiency analysis in cache-enabled D2D-aided heterogeneous cellular networks[J]. IEEE Access，2020，8：19540-19554.

[7] LIANG L，WANG W，JIA Y，et al. A cluster-based energy-efficient resource management scheme for ultra-dense networks[J]. IEEE Access，2016，4：6823-6832.

[8] 金明录，郭楠. 基于 Thomas 簇过程的异构蜂窝网络能量效率分析[J]. 通信学报，2019，40(10)：149-156.
JIN M L，GUO N. Energy efficiency analysis of heterogeneous cellular networks based on Thomas cluster processes [J]. Journal of Communications，2019，40(10)：149-156.

[9] 王雪，刘京，孙佳妮，等. 基于谱聚类的异构蜂窝超密集网络高能效资源分配算法[J]. 通信学报，2021，42(7)：162-175.
WANG X，LIU J，SUN J N，et al. Energy efficient resource allocation algorithm for heterogeneous cellular ultra-dense networks based on spectral clustering [J]. Journal of Communications，2021，42(7)：162-175.

[10] 靳冬慧，陈硕，王占刚. 密集异构网络中基于多目标优化的资源分配策略 [J]. 电讯技术，2023，63 (4)：466-474.
JIN D H，CHEN SH，WANG ZH G. Resource allocation strategy based on multi-objective optimization in dense heterogeneous networks [J]. Telecommunications Technology，2023，63 (4)：466-474.

[11] MASOUDI M，ZAEFARANI H，MOHAMMADI A，et al. Energy and spectrum efficient resource allocation in two-tier networks：A multi objective approach[C]. IEEE Wireless Communications and Networking Conference. IEEE，2017：1-6.

[12] HAO Y，NI Q，LI H，et al. Robust multi-objective optimization for EE-SE tradeoff in D2D communications underlaying heterogeneous networks[J]. IEEE Transactions on Communications，2018，66 (10)：4936-4949.

[13] HAO Y，NI Q，LI H，et al. Energy and spectral efficiency tradeoff with user association and power coordination in massive MIMO enabled HetNets[J]. IEEE Communications Letters，2016，20(10)：2091-2094.

[14] WU H，LYU X，TIAN H. Online optimization of wireless powered mobile-edge computing for heterogeneous industrial internet of things[J]. IEEE Internet of Things Journal，2019，6(6)：9880-9892.

[15] CHANG Z，LIU L，GUO X，et al. Dynamic resource allocation and computation offloading for IoT fog computing system [J]. IEEE Transactions on Industrial Informatics，2020，17(5)：3348-3357.

[16] LI H，WEI T，REN A，et al. Deep reinforcement learning：Framework，applications and embedded implementations [C]. IEEE/ACM International Conference on Computer-Aided Design. IEEE，2017：847-854.

[17] FADLULLAH Z M，TANG F，MAO B，et al. State-of-the-art deep learning：Evolving machine intelligence toward tomorrow's intelligent network traffic control systems[J]. IEEE Communications Surveys & Tutorials，2017，19(4)：2432-2455.

[18] LIU Y，YANG C，JIANG L，et al. Intelligent edge computing for IoT-based energy management in smart cities[J]. IEEE Network，2019，33(2)：111-117.

[19] WEI Y，ZHANG Z，YU F R，et al. Power allocation in Hetnets with hybrid energy supply using actor-critic reinforcement learning [C]. IEEE Global Communications Conference. IEEE，2017：1-5.

[20] LI D，ZHANG H，LONG K，et al. User association and power allocation based on Q-learning in ultra dense heterogeneous networks [C]. IEEE Global Communications Conference. IEEE，2019：1-5.

[21] CHEN S J，CHIU W Y，LIU W J. User preference-based demand response for smart home energy management using multi objective reinforcement learning[J]. IEEE Access，2021，9：161627-161637.

[22] ALMAHAMID F，GROLINGER K. Reinforcement learning algorithms：An overview and classification[C]. IEEE Canadian Conference on Electrical and Computer Engineering. IEEE，2021：1-7.

[23] YANG M，LIU N，ZUO L，et al. Dynamic charging scheme problem with actor-critic reinforcement learning[J]. IEEE Internet of Things Journal，2020，8(1)：370-380.

[24] DO Q V，KOO I. Dynamic bandwidth allocation scheme for wireless networks with energy harvesting using actor-critic deep reinforcement learning [C]. 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE，2019：138-142.

[25] HENRY S，ALSOHAILY A，SOUSA E S. 5G is real：Evaluating the compliance of the 3GPP 5G new radio system with the ITU IMT-2020 requirements[J]. IEEE Access，2020，8：42828-42840.

[26] WU Y C, DINH T Q, FU Y, et al. A hybrid DQN and optimization approach for strategy and resource allocation in MEC networks[J]. IEEE Transactions on Wireless Communications, 2021, 20(7): 4282-4295.

[27] WANG J, JIANG C, ZHANG K, et al. Distributed Q-learning aided heterogeneous network association for energy-efficient IIoT[J]. IEEE Transactions on Industrial Informatics, 2020, 16(4): 2756-2764.

[28] YI C, HUANG S, CAI J. Joint resource allocation for device-to-device communication assisted fog computing [J]. IEEE Transactions on Mobile Computing, 2021, 20(3): 1076-1091.

作 者 简 介

曾韦健，硕士研究生，主要研究方向为深度强化学习，无线通信。
E-mail:3382325058@qq.com

李晖(通信作者)，博士，教授，主要研究方向为无线通信网络、空间通信、海洋通信，AI在通信中应用等。
E-mail:hitlihui1112@163.com