

DOI:10.19651/j.cnki.emt.2209905

基于多智能体 Actor-Critic 算法的异构网络能效优化*

张茜茜¹ 李君^{1,2} 李正权^{3,4} 于心远¹(1.南京信息工程大学 南京 210044; 2.无锡学院 无锡 214105; 3.江南大学轻工过程先进控制教育部
重点实验室 无锡 214122; 4.北京邮电大学网络与交换技术国家重点实验室 北京 100876)

摘要: 为了最大限度地提高异构网络(HetNets)的能量利用率,本文将能效优化问题设计为一个多级决策问题,根据优化目标的资源分配,首先将初始问题分解为对参数几乎空白子帧比例和小区间范围扩展进行优化的两个子问题,采用多智能体 Actor-Critic(MAAC)算法对子问题进行求解,然后通过迭代各优化子问题的解,解决初始优化问题。在参数优化过程中,将单个小基站作为一个智能体,采用 MAAC 算法对各自 CRE 寻找最优解,实现小区间异步 CRE 优化。实验结果表明,该方法在保持能效稳定提高的前提下,相比较于表格格式 Q 学习的循环 Q 学习算法,其收敛速度提高了 40%,并且通过异步优化 CRE 的方式取得了基站间更均衡的负载。

关键词: 异构网络;eICIC;多智能体;MAAC

中图分类号: TN911 **文献标识码:** A **国家标准学科分类代码:** 510.5015

Energy efficiency optimization of heterogeneous networks based on multi-agent actor-critic algorithm

Zhang Xixi¹ Li Jun^{1,2} Li Zhengquan^{3,4} Yu Xinyuan¹(1. Nanjing University of Information Science & Technology, Nanjing 210044, China;
2. Wuxi University, Wuxi 214105, China; 3. Key Laboratory of Advanced Control of Light Industry Process, Ministry of Education, Jiangnan University, Wuxi 214122, China; 4. State Key Laboratory of Network and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In order to maximize energy efficiency in heterogeneous networks, the problem of energy efficiency optimization is designed as a multistage decision problem in this paper. According to the resource allocation of the optimization goal, the initial problem is decomposed into two sub-problems that optimize the parameters almost blank subframe ratio and cell range expansion. The multi-agent actor-critic algorithm is used to solve the sub-problem, and then the initial optimization problem is solved by iterating over the solution of each optimization sub-problem. In the process of parameter optimization, a single small base station is used as an agent, and the MAAC algorithm is used to find the optimal solution for each CRE, so as to realize the asynchronous CRE optimization between cells. Experimental results show that compared with the Turbo Q-learning algorithm, the convergence speed of the proposed method is increased by 40%, and the load between small base stations is more balanced through asynchronous optimization of CRE.

Keywords: HetNets; eICIC; multi-agent; MAAC

0 引言

现如今不断增长的流量需求给移动网络运营商在服务、覆盖、负载平衡和服务质量方面带来了前所未有的挑战^[1-3],为了有效的解决数据流量的爆炸性增长问题,缓解

宏基站负载问题,3GPP 提出增强型蜂窝小区间干扰协调(enhanced inter-cell interference coordination, eICIC)技术,引入小区间范围扩展(cell range expansion, CRE)和几乎空白子帧(almost blank subframe, ABS)策略。CRE 的提出,允许小基站通过向其参考信号接收功率(reference signal

收稿日期:2022-05-11

* 基金项目:国家自然科学基金(61571108)、网络与交换技术国家重点实验室(北京邮电大学)开放课题资助项目(SKLNST-2020-1-13)、南京信息工程大学无锡研究生创新实践项目(WXCX202101)资助

receiving power, RSRP) 添加正偏置值来增加小基站的覆盖范围,简单来说就是人为的放大用户终端所接受的来自小基站 (small base station, SBS) 的参考信号功率,使得更多的用户选择与 SBS 相关联,但同时也带来了严重的同频干扰问题。

为了缓解位于小基站扩展范围内用户的干扰问题,提高用户信道质量^[4-5], eCIC 提出在时域中配置固定 ABS 比例,使宏基站 (macro base station, MBS) 和 SBS 在不同子帧阶段使用无线电资源^[6]。Oh 在文献[7]中提出了一种基于期望用户数据速率即固定 ABS 比例的小区选择方案以提高平均用户速率,增加卸载用户数,然而他并没有考虑到各个基站之间的负载均衡问题。近年来随着用户数量的不断增加,基站间负载不均衡所导致部分基站过载等问题愈发突出。目前针对能效优化的 CRE 策略研究相对较少,现有 CRE 的研究^[8-12]大多忽略了对 ABS 比例进行优化,仅考虑在小区间同步优化 CRE 偏置值,并没有针对各个基站的不同特性进行优化策略以缓解负载均衡问题。文献[13-15]虽然考虑了 CRE 和 ABS 的联合优化问题,但其动作空间维数较大,也没有对各个基站的 CRE 进行单独优化策略。

针对以上问题,为了能够最大限度地提高异构网络 (heterogeneous networks, HetNets) 的能量利用率,本文结合多级决策过程和多智能体 Actor-Critic (multi-agent actor-critic, MAAC) 算法优化网络资源配置。将能效优化问题设计为一个多级决策问题,根据优化目标的资源分配,对动作空间进行分解,将初始问题分解为对参数 ABS 比例和 CRE 进行优化的两个子问题,采用 MAAC 算法对子问题进行求解,有效的降低了动作空间的维数和大小。然后通过迭代各优化子问题的解,解决能效优化问题。在求解优化 CRE 子问题过程中,考虑到各个基站之间的负载均衡问题,将单个小基站作为一个智能体,采用 MAAC 算法寻找各自 CRE 的最优解,实现小区间异步 CRE 优化。

1 系统模型

本文所考虑的是如图 1 所示的两层 HetNets 场景, SBS 在 MBS 覆盖范围内随机部署, MBSs 和 SBSs 的集合

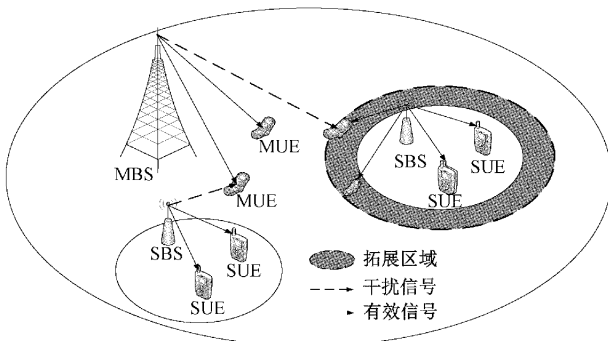


图 1 系统模型

分别用 M 和 S 表示。用户随机进入网络后,与 MBSs 关联的用户被记为宏基站服务用户 (MBS UEs, MUE),与 SBS 相连的被记为小基站服务用户 (SBS UEs, SUE)。

记 ABS 比例为 α , SBS 在整个子帧阶段中保持正常功率发送。MBS 在非 ABS (nABS) 子帧阶段以正常功率传输,在 ABS 子帧阶段中保持静默或以低功率传输。也就是说,在 ABS 子帧阶段,位于扩展范围内的用户将几乎不会接收到来自宏基站的干扰。因此,连接到 SBS k 的 UE n 获得的信干噪比 (signal to interference plus noise ratio, SINR) 可表示为:

$$\text{SINR}_{k,n} = \begin{cases} \frac{P_s^{k,n} G_{k,n}}{\sum_{j \in M, j \neq k} P_M^{j,n} G_{j,n} + \sum_{i \in S, i \neq k} P_S^{i,n} G_{i,n} + N_0}, & k \in S^{n\text{ABS}} \\ \frac{P_s^{k,n} G_{k,n}}{\sum_{j \in M, j \neq k} P_M^{j,n} G_{j,n} + \sum_{i \in S, i \neq k} P_S^{i,n} G_{i,n} + N_0}, & k \in S^{\text{ABS}} \end{cases} \quad (1)$$

其中, $P_s^{k,n}$ 是在 nABS 阶段,从 SBS k 到 UE n 的传输功率; $P_M^{j,n}$ 是指在 nABS 阶段,从 MBS j 到 UE n 的传输功率; $P_S^{i,n}$ 指在 ABS 阶段,从 MBS j 到 UE n 的传输功率; $G_{k,n}$ 表示从 SBS k 到 UE n 的信道增益; N_0 表示噪声功率; 分别用 $S^{n\text{ABS}}$ 和 S^{ABS} 表示 SBSs 在 nABS 和 ABS 期间的集合。

对于 MBSs, 连接到 MBS k 的 UE n 获得的 SINR 为:

$$\text{SINR}_{k,n} = \begin{cases} \frac{P_M^{k,n} G_{k,n}}{\sum_{j \in M, j \neq k} P_M^{j,n} G_{j,n} + \sum_{i \in S} P_S^{i,n} G_{i,n} + N_0}, & k \in M^{n\text{ABS}} \\ \frac{P_M^{k,n} G_{k,n}}{\sum_{j \in M, j \neq k} P_M^{j,n} G_{j,n} + \sum_{i \in S} P_S^{i,n} G_{i,n} + N_0}, & k \in M^{\text{ABS}} \end{cases} \quad (2)$$

$M^{n\text{ABS}}$ 和 M^{ABS} 表示在 nABS 和 ABS 子帧期间的 MBSs 集合,其中 M 和 m 代表 $M^{n\text{ABS}}$ 和 M^{ABS} 。

由式 (1) 和 (2) 可得出连接到 BS k 的 UE n 的传输速率为:

$$R_{k,n} = \begin{cases} (1-\alpha)B \log(1 + \text{SINR}_{k,n}), & k \in M^{n\text{ABS}} \cup S^{n\text{ABS}} \\ \alpha B \log(1 + \text{SINR}_{k,n}), & k \in M^{\text{ABS}} \cup S^{\text{ABS}} \end{cases} \quad (3)$$

其中, B 是系统带宽。

根据式 (1)~(3) 可得知,在 HetNets 下行链路中的 UE 存在两种干扰模式。当 MBS 在 nABS 子帧阶段,用户同时受到来自 MBS 和 ABS 的干扰,当 MBS 在 ABS 子帧阶段,用户仅接收来自 SBS 发送的信号。本文综合考虑 ABS 和 CRE 等参数,将优化目标函数设定为系统的能源效率,由此,可以建立一个联合优化的能效问题:

$$\max_{\{x_{k,n}, \alpha\}_{k \in M \cup S, n \in U}} \sum_{k \in M \cup S} \sum_{n \in U} x_{k,n} \log\left(\frac{R_{k,n}}{\sum_{k \in M \cup S} P_{k,n}}\right) \quad (4)$$

$$s, t \ 0 \leq x_{k,n}$$

其中, $x_{k,n}$ 表示 BS k 和 UE n 之间的连接状态, 表示为:

$$x_{k,n} = \{0, 1\} \quad \forall k \in M \cup S, n \in U \quad (5)$$

$$\sum_{k \in M \cup S} x_{k,n} = 1 \quad \forall n \in U \quad (6)$$

$x_{k,n} = 1$ 表示在 BS k 和 UE n 之间建立连接, 否则为 0。式(6)表示小区中的每个 UE 只能连接到一个 BS。

在不同子帧时间从 BS k 到 UE n 的发射功率可以表示为:

$$P_{k,n} = \begin{cases} P_M^{k,n}, & k \in M^{n,ABS} \\ P_m^{k,n}, & k \in M^{ABS} \\ P_S^{k,n}, & k \in S \end{cases} \quad (7)$$

因此, 结合上述分析, 可以建立如下联合优化问题:

$$\max_{(\alpha, \beta)} f(\alpha, \beta) = \max_{(\alpha, \beta)} \sum_{k \in M \cup S} \sum_{n \in U} x_{k,n} \log \left(\frac{R_{k,n}}{\sum_{k \in M \cup S} P_{k,n}} \right) \quad (8)$$

$$s, t \ 0 \leq x_{k,n}$$

其中, β 表示 CRE 参数。

2 Actor-Critic 算法

由于系统优化问题是一个未知的非凸问题, 本文考虑将问题式(8)对应的参数优化问题建模为一个马尔可夫决策过程(Markov decision process, MDP), 在此基础上提出采用 AC 算法对 MDP 进行求解^[16]。设置一个有限离散时间序列 $t \in \{0, 1, \dots, T\}$ 。ABS 和 CRE 用 $\alpha \in A$ 和 $\beta \in B$ 分别表示。在每个时隙 t , Actor 网络接收到智能体局部观测环境状态 $s_t = s \in S$, 选择执行动作 $a_t = a \in A$, 其中 s 是基站覆盖范围内的用户位置, S 表示用户位置的集合, 动作 α 是降低小区干扰的最佳 α, β 参数配置, A 表示动作参数配置的集合。决策完成后 Actor 网络将策略 π_θ 输入 Critic 网络, Critic 网络根据智能体的决策动作和当前状态信息计算优势函数。本文采用 Native Critic^[17] 遵循的时序差分(temporal-difference, TD)优势函数更新策略梯度。由此可将该 MDP 用元组 (S, A, P, R) 表示, 其中:

1) 状态: 基站服务用户在 t 时刻的位置被认为是状态 $s_t = s, s \in S = \{s_1, s_2, \dots, s_t, \dots, s_T\}$ 表示在当前时刻智能体观察到的状态信息。

2) 动作: 在状态 s_t 下, $a_t = (\alpha, \beta)$, 其中 $\alpha \in A, \beta \in B$ 。 α 表示 ABS 比例, β 表示 CRE 偏置。动作空间 $X = |A| \times |B|$, 其中 $|\cdot|$ 表示集合的势。

3) 状态转移: 用户随机进入小区的某一个状态, 状态转移概率是随机的。

4) 奖励函数: 优化目标为最大化系统的能量效率, 由于能效较大在仿真实验中容易陷入局部最优, 因此本文将奖励 R_t 设置为能源效率的负倒数。

在 MDP 框架下, 在状态 s 下遵循策略 π_θ 的预期总奖励可用状态值函数 $V_{\pi(s)} = E(R_t | s_t = s)$ 体现。动作值函

数 $Q_{\pi(s)} = E(R_t | s_t = s, a_t = a)$ 则体现出在状态 s 下, 根据策略 π_θ 选择执行动作 a 的预期奖励。

在 Actor 网络下, 本文选择高斯正态分布近似策略分布, 其策略表示为:

$$\pi_\theta(a, s) = \frac{1}{\sqrt{2\pi}\sigma_\theta(s)} \exp\left(-\frac{(a-\mu_\theta(s))^2}{2(\sigma_\theta(s))^2}\right) \quad (9)$$

正态分布的均值和方差分别为 μ_θ 和 σ_θ 。在强化学习框架下, 通常策略梯度(policy gradient, PG)算法更新网络参数 θ , 本文采用梯度下降法对其进行更新:

$$\theta' = \theta + l_a \nabla_\theta \log \pi_\theta(a | s) A_\pi(s, a) \quad (10)$$

其中, l_a 为 Actor 网络的学习率, $A_\pi(s, a)$ 为优势函数, 通常由 $Q(s, a) - V(s)$ 表示。在 Critic 网络中引入 TD 误差 δ_t , 作为优势函数对网络参数 ξ 进行更新。其中 $\delta_t = V(s_t) - V_\xi(s_t) = R' + \gamma V_\xi(s_t) - V_\xi(s_t)$, $R_{t+1} = R'$, $V_\xi(s_{t+1}) = V_\xi(s_t)$, ξ 表示为 Critic 网络中神经网络全连接层的权重参数。因此式(10)可写为:

$$\theta' = \theta + l_a \nabla_\theta \log \pi_\theta(a_t | s_t) \delta_t \quad (11)$$

Critic 网络的优化目标为最小化 TD 误差 δ_t , 由此采用梯度下降法可对其参数进行更新:

$$\xi' = \xi + l_c \delta_t \nabla_\xi V_\xi(s_t) \quad (12)$$

其中, l_c 为 Critic 网络的学习率。

本文将初始目标优化问题分解为两个参数联合优化问题, 结合多级决策过程和 MAAC 算法对动作空间进行分解, 以降低动作空间的大小。在参数优化过程中结合多智能体技术提出在优化系统 ABS 参数的同时, 异步优化 CRE。

1) 固定 ABS, 优化 CRE

将宏基站覆盖范围内的各个小基站看作为一个智能决策体, 对于基站 k 来说, 动作为 $a_t = \beta_k$, 其中 $\beta_k \in B$, 形成一个联合动作集 $a^* = \{\beta_1, \dots, \beta_N\}$ 。动作空间大小为 $X = |B|$ 。与单智能体类似, 多智能体强化学习的目标是最大化奖励^[18]。在 MDP 框架下, 定义状态值函数 $V_{\pi(s)} = E(R_t | s_t = s)$ 是在状态 s 下, 遵循联合策略 π^* 的预期总奖励。动作值函数 $Q_{\pi^*(s, a^*)} = R(s, a^*) + \gamma \sum_{s' \in S} P_{s, s'}^{a^*} V_{\pi^*(s')}$ 定义了状态 s 下, 遵循联合策略 π^* 选择联合动作 a^* 的预期奖励。其网络策略更新算法如算法 1 所示。

算法 1: 固定 ABS, 优化 CRE

初始化: $a_t = (\alpha, B)$

目标: 优化各个小基站 CRE 偏置 $B, B = \{\beta_1, \dots, \beta_N\}$

1. 初始化: $n = 0$, 设置阈值 h_1 , 策略参数 $\theta_1 = \theta_{10}$, 值函数参数 $\xi_1 = \xi_{10}$, 状态 $s = s_0$, 小基站个数 N

2. While $n \leq h_1$ do

3. 在状态 s 下根据联合策略 π^* 执行动作 a^* ;

4. 执行动作 a^* 后,状态由 s 转移至 s' , 获得即时奖励 R' ;
5. 计算 TD 误差 $\delta_{1t} = R' + \gamma_1 V_{\xi_1}(s_t)' - V_{\xi_1}(s_t)$;
6. 根据式(11)和(12)更新 Actor 网络和 Critic 网络参数:
 $\theta_1' \leftarrow \theta_1 + l_{a_1} \nabla_{\theta_1} \log \pi_{\theta_1}(a_t^* | s_t) \delta_{1t}$;
 $\xi_1' \leftarrow \xi_1 + l_{c_1} \delta_{1t} \nabla_{\xi_1} V_{\xi_1}(s_t)$;
7. $n \leftarrow n + 1$
8. end while

2) 固定 CRE, 优化 ABS

将网络中所有宏基站看作为一个智能体。 t 时刻, s_t 状态下执行动作 $a_t = \alpha, \alpha \in A$, 动作空间大小为: $X = |A|$ 。其网络策略更新算法如算法 2 所示。

算法 2: 固定 CRE, 优化 ABS

初始化: $a_t = (\alpha, B)$

目标: 优化网络 ABS 比例 α

1. 初始化: $n = 0$, 设置阈值 h_2 , 策略参数 $\theta_2 = \theta_{20}$, 值函数参数 $\xi_2 = \xi_{20}$, 状态 $s = s_0$, 小基站个数 N
2. While $n \leq h_2$ do
3. 在状态 s 下策略 π_{θ} 执行动作 a ;
4. 执行动作 a 后, 状态由 s 转移至 s' , 获得即时奖励 R' ;
5. 计算 TD 误差 $\delta_{2t} = R' + \gamma_2 V_{\xi_2}(s_t)' - V_{\xi_2}(s_t)$;
6. 根据式(11)和(12)更新 Actor 网络和 Critic 网络参数:
 $\theta_2' \leftarrow \theta_2 + l_{a_2} \nabla_{\theta_2} \log \pi_{\theta_2}(a_t | s_t) \delta_{2t}$;
 $\xi_2' \leftarrow \xi_2 + l_{c_2} \delta_{2t} \nabla_{\xi_2} V_{\xi_2}(s_t)$;
7. $n \leftarrow n + 1$
8. end while

综上, 采用算法 3 即可解决原始问题式(8)。由此可看出, 本文成功的将动作空间由最初的 $X = |A| \times |B|$ 转化为 $X = |A| + |B|$, 大大降低了动作空间的大小。

算法 3: MAAC 的优化问题

初始化: $\alpha \in A, \beta \in B$, 奖励 R , Actor 网络学习率 $l_a = \{l_{a_1}, l_{a_2}\}$, Critic 网络学习率 $l_c = \{l_{c_1}, l_{c_2}\}$, 折扣因子 $\gamma = \{\gamma_1, \gamma_2\}$, 策略参数 $\theta = \{\theta_1, \theta_2\}$, 值函数参数 $\xi = \{\xi_1, \xi_2\}$

目标: 优化网络 ABS 比例 α , 小基站最优 CRE 参数 $\{\beta_1, \dots, \beta_N\}$

1. 初始化: $n = 0$, 设置阈值 h_3 , 策略参数 $\theta_2 = \theta_{20}$, 值函数参数 $\xi_2 = \xi_{20}$, 状态 $s = s_0$, 小基站个数 N , 用户位置 p_i , ABS 比例 α , CRE 参数 $\{\beta_1, \dots, \beta_N\}$

2. While $n \leq h_3$ do

3. 修正 ABS 比例 α , 根据算法 2 计算 CRE 参数集 $\{\beta_1, \dots, \beta_N\}$ 。将求解的 CRE 参数集 $\{\beta_1, \dots, \beta_N\}$ 传递给步骤(4)。

4. 修正 CRE 参数集 $\{\beta_1, \dots, \beta_N\}$, 根据算法 3 计算 ABS 比例 α 。将求解的 ABS 比例 α 传递给步骤 3。

5. $n \leftarrow n + 1$

6. end while

3 仿真实验

实验模拟参数根据 3GPP LTE-A HetNets 框架设置, 无线信道建模为确定性路径损耗衰减和随机阴影衰落模型。在这一部分中, 宏基站固定坐标位置为 $(0, 0)$, 在 MBS 覆盖范围内随机部署了两个 SBS, 并根据负载条件选择工作模式。MBS 和 SBS 的覆盖半径分别为 300 m 和 100 m。设置用户进入随机小区状态 $S = \{s_1, s_2, \dots, s_t, \dots, s_t\}$, 其中 $l = 100$, 环境内存在一个理论最优动作。热噪声功率为 -176 dBm, 系统频谱带宽为 10 MHz, MBSs、SBSs 和 UE 的天线增益分别为 14、5 和 0 dBi。MBS 和 SBS 的最大传输功率分别设置为 46 dBm 和 30 dBm。在 eICIC 下, ABS 模式具有 8 个子帧的周期性, ABS 比例 $\alpha \in [0, 1]$, 表示为集合: $\{0/8, \dots, 7/8\}$ 。CRE 由 $\beta \in B$ 表示, 其中 $B = \{0, 3, 6, 9, 12, 15, 18\}$ 。具体参数如表 1 所示。

表 1 仿真参数列表

仿真参数	仿真值
MBS 覆盖半径	300 m
SBS 覆盖半径	100 m
系统带宽	10 MHz
MBS 发射功率	正常: 46 dBm; 休眠 12 dBm
SBS 发射功率	30 dBm
从 MBS 到 UE 的路径损耗	$128.1 + 37.6 \times \lg(R[\text{km}])$, R 是 MBS 到 UE 的距离
从 SBS 到 UE 的路径损耗	$149.1 + 37.6 \times \lg(R[\text{km}])$, R 是 MBS 到 UE 的距离

图 2 和 3 纵坐标准确率表示为在当前训练步数下, 迭代过程中选择理论最优动作的比例, 其中一个训练步长表示为迭代 500 次。由此可知, 图 2 和 3 分别显示了表格格式 Q-Learning 的循环 Q-Learning (turbo Q-Learning, TQL) 算法和 MAAC 算法用户数设置为 50、100、150、200 时的准确率和训练步长的关系。通过图中可以看出 TQL 算法在不同负载条件下经过训练 700 步即 $500 \times 700 = 350\,000$ 次迭代后收敛。提出的 MAAC 算法在训练步数为 500 步即 $500 \times 500 = 250\,000$ 次迭代后收敛, 与 TQL 算法相比, MAAC 算法的收敛速度提高了本算法的 40%。

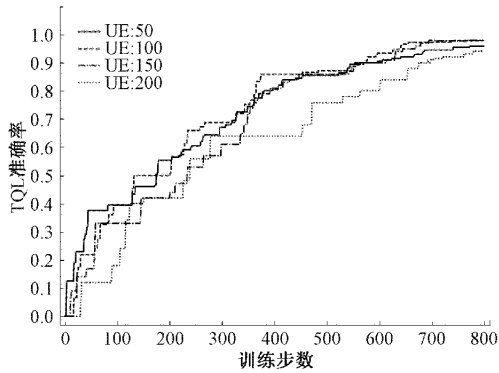


图 2 不同负载下的 TQL 算法收敛图

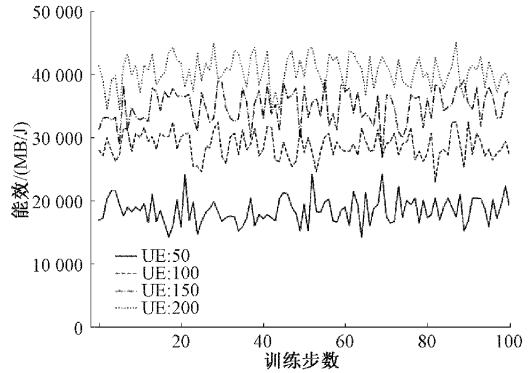


图 5 不同负载下的能效

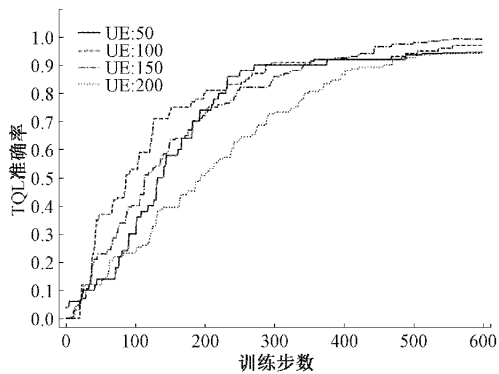


图 3 不同负载下的 MAAC 算法收敛图

数的动作选择。根据图 7~9 可以看出,随着训练步数的增加,ABS 比例趋于为 7/8,小基站 1 和小基站 2 CRE 偏置值分别为 15 和 6。

图 4 和 5 体现了采用 MAAC 算法准确率达到收敛之后训练 100 步,其功耗和能效的走向。此时的训练步长同图 3,表示为训练一步即迭代 500 次。由图 4 和 5 可以看出,当 MAAC 算法收敛之后其系统功耗趋于稳定,随着用户数的增加,能效也有显著提高。

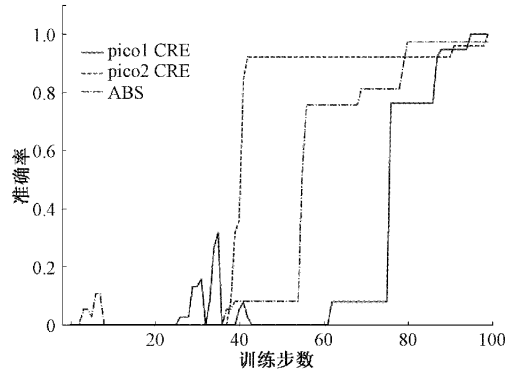


图 6 各参数训练收敛图

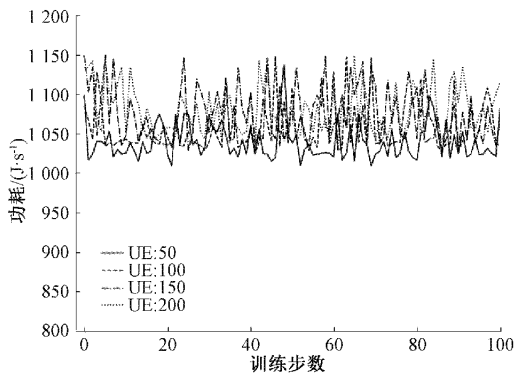


图 4 不同负载下的功耗

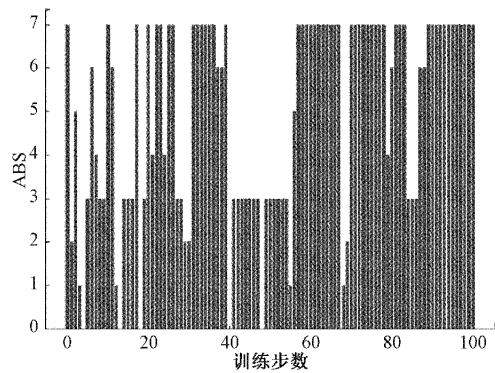


图 7 ABS 动作选择

图 6 显示出了用户总数为 50 时,算法 1 的阈值 h_1 设置为 800,算法 2 中阈值 h_2 设置为 500 时的各参数训练收敛图横坐标为训练步数,对于 ABS 优化来讲,每一步的迭代次数为 500;对于 CRE 优化来说,单步训练为迭代 800 次。纵坐标为准确率,准确率表示为在当前训练步数下,迭代过程中选择理论最优动作的比例。图 7~9 是在当前步

图 10~12 为随机生成用户总数为 50 之后,基站和用户位置保持不变,根据图 7~9 所示的动作选择,设置动作分别为(7/8,15,6),(7/8,6,6),(7/8,15,15)时用户关联情况。由图 10~12 可得知两个小基站关联用户个数分别为(22,22),(24,18),(29,15)。由此可见当基站选择相同的 CRE 时,很容易出现部分基站负载较多的情况,本实验通过异步优化 CRE,有利于避免该现象的出现,促使基站间关联用户数更加均衡。

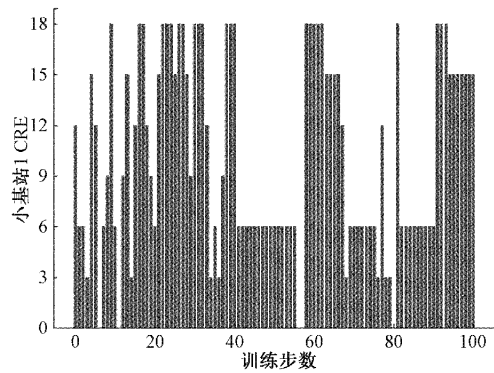


图 8 小基站 1 CRE 动作选择

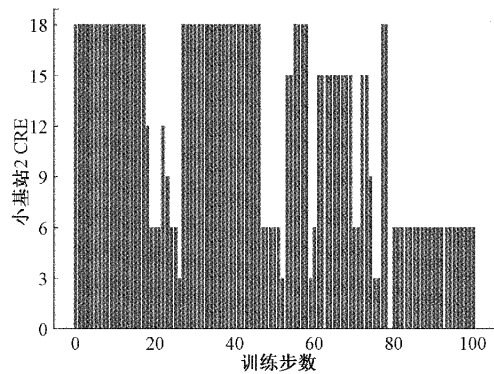


图 9 小基站 2 CRE 动作选择

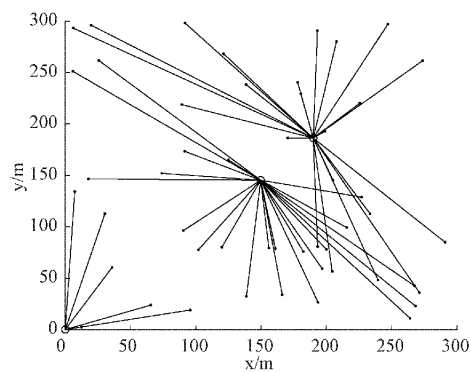


图 10 ABS=7/8, CRE=(15,6)时用户关联图

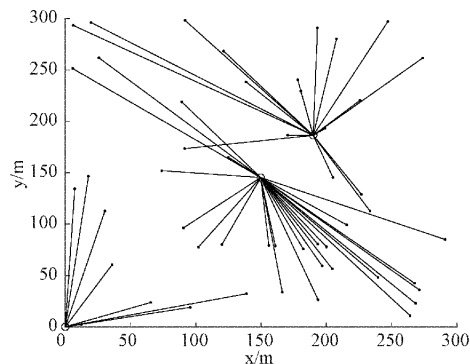


图 11 ABS=7/8, CRE=(6,6)时用户关联图

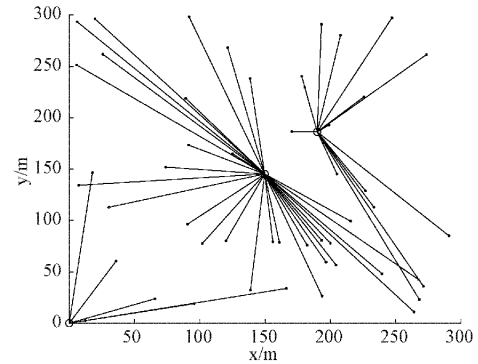


图 12 ABS=7/8, CRE=(15,15)时用户关联图

4 结 论

本文结合多级决策过程和 MAAC 算法对动作空间进行分解,成功的将动作空间由最初的 $X = |A| \times |B|$ 转化为 $X = |A| + |B|$, 大大降低了动作空间的大小。仿真结果证明本文提出的算法在保持能效稳定提高的情况下,与文献[6]中所提出的 TQL 算法进行对比,收敛速度提高了 40%,且本文通过异步优化 CRE,实现了基站间更加均衡的负载。然而本文仅考虑了在两个小基站基础上优化 CRE 偏置,后续将考虑增加小基站个数,将小基站负载均衡和功率预测作为双变量联合优化。

参考文献

- [1] KURIBAYASHI H P, et al. Particle swarm-based cell range expansion for heterogeneous mobile networks[J]. IEEE Access, 2020, 8: 37021-37034.
- [2] 王晓昌, 吴璠, 孙彦赞, 等. 基于联邦深度强化学习的车联网资源分配[J]. 电子测量技术, 2021, 44(10): 114-120.
- [3] 胡文建, 杨阳, 刘保安, 等. 基于粒子群算法的 5G 大规模 MIMO 通信系统效率评估[J]. 国外电子测量技术, 2022, 41(2): 46-52.
- [4] 彭艺, 付晓霞, 刘煜恒, 等. 高原山区 D2D 网络能量效率优化算法研究[J]. 电子测量与仪器学报, 2021, 35(9): 150-156.
- [5] 王海东, 谭国平, 周思源. 基于时延约束的 D2D 异构网络资源动态分配[J]. 电子测量技术, 2020, 43(17): 130-136.
- [6] WANG X M, LI L, LI J, et al. A turbo Q-learning(TQL) for energy efficiency optimization in heterogeneous networks[J]. Entropy, 2020, 22(9): 957.
- [7] OH J, HAN Y. Cell selection for range expansion with almost blank subframe in heterogeneous networks in proc[C]. 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC), 2012: 653-657.
- [8] YE Q Y, RONG B Y, CHEN Y D, et al. User

- association for load balancing in heterogeneous cellular networks [C]. IEEE Transactions on Wireless Communications, 2013, 12(6): 2706-2716.
- [9] SINGH S, ANDREWS J G. Joint resource partitioning and offloading in heterogeneous cellular networks[C]. IEEE Transactions on Wireless Communications, 2014, 13(2): 888-901.
- [10] NIE X, WANG Y, ZHANG J, et al. Coverage and association bias analysis for backhaul constrained HetNets with eICIC and CRE [J]. Wireless Pers. Commun., 2017, 97(4): 4981-5002.
- [11] HATTAB G, CABRIC D. Rate-based cell range expansion for downlink massive MIMO heterogeneous networks [C]. IEEE Wireless Communications Letters, 2018, 7(3): 296-299.
- [12] ANDREWS J G, SINGH S, YE Q Y, et al. An overview of load balancing in HetNets: Old myths and open problems [C]. IEEE Wireless Communications, 2014, 21(2): 18-25.
- [13] 陈蓓蓓, 吴呈瑜, 占敖. 基于时域 eICIC 的小区覆盖扩展策略能效优化[J]. 无线电通信技术, 2019, 45(5): 481-487.
- [14] TANG W J, FENG S L, LIU Y, et al. Joint low-power transmit and cell association in heterogeneous networks [C]. 2015 IEEE Global Communications Conference(GLOBECOM), 2015: 1-6.
- [15] ALI M S, COUCHENEY P, COUPECHOUX M. Load balancing in heterogeneous networks based on distributed learning in near-potential games[J]. IEEE Trans. on Wireless Communications, 2016, 15(7): 5046-5059.
- [16] 秦爽, 董星辰, 冯钢. 超密集蜂窝网络智能干扰协调算法[J]. 中兴通讯技术, 2020, 26(5): 45-52.
- [17] SU J Y, ADAMS C S, BELING A P. Value-decomposition multi-agent actor-critics [C]. AAAI 2021: Proceedings of the 35th Conference on Artificial Intelligence, Virtual Event; AAAI, 2021: 11352-11360.
- [18] 臧嵘, 王莉, 史腾飞. 基于注意力消息共享的多智能体强化学习方法[J]. 计算机应用, 2022, 6(3): 1-11.

作者简介

张茜茜, 硕士, 主要研究方向为机器学习, 无线通信。

E-mail: 20201249245@nuist.edu.cn

李君(通信作者), 博士, 教授, 主要研究方向为无线通信、资源分配、机器学习、编码译码等。

E-mail: 07a0303105@cjlu.edu.cn

李正权, 博士, 教授, 主要研究方向为无线通信、信号处理、信道编码译码等。

E-mail: lzq722@jiangnan.edu.cn