·移动互联与通信技术· 文章编号：1000-3428(2021)05-0169-07 文献标志码：A 中图分类号：TN929.5

基于DQN的超密集网络能效资源管理

郑冰原[1,2,3]，孙彦赞[1,2,3]，吴雅婷[1,2,3]，王 涛[1,2,3]，方 勇[1,2,3]
（1.上海大学 上海先进通信与数据科学研究院，上海 200444；2.上海大学 特种光纤与光接入网重点实验室，上海 200444；
3.上海大学 特种光纤与先进通信国际合作联合实验室，上海 200444）

摘 要：小基站的密集随机部署会产生严重干扰和较高能耗问题，为降低网络干扰、保证用户网络服务质量（QoS）并提高网络能效，构建一种基于深度强化学习（DRL）的资源分配和功率控制联合优化框架。综合考虑超密集异构网络中的同层干扰和跨层干扰，提出对频谱与功率资源联合控制能效以及用户QoS的联合优化问题。针对该联合优化问题的NP-Hard特性，提出基于DRL框架的资源分配和功率控制联合优化算法，并定义联合频谱和功率分配的状态、动作以及回报函数。利用强化学习、在线学习和深度神经网络线下训练对网络资源进行控制，从而找到最佳资源和功率控制策略。仿真结果表明，与枚举算法、Q-学习算法和两阶段算法相比，该算法可在保证用户QoS的同时有效提升网络能效。

关键词：超密集网络；能效；资源分配；强化学习；功率控制；深度学习

开放科学（资源服务）标志码（OSID）：

中文引用格式：郑冰原，孙彦赞，吴雅婷，等. 基于DQN的超密集网络能效资源管理[J]. 计算机工程，2021，47(5)：169-175.

英文引用格式：ZHENG Bingyuan，SUN Yanzan，WU Yating，et al. DQN-based energy efficiency resource management for ultra-dense network[J]. Computer Engineering，2021，47(5)：169-175.

DQN-based Energy Efficiency Resource Management for Ultra-dense Network

ZHENG Bingyuan[1,2,3]，SUN Yanzan[1,2,3]，WU Yating[1,2,3]，WANG Tao[1,2,3]，FANG Yong[1,2,3]
（1.Shanghai Institute for Advanced Communication and Data Science，Shanghai University，Shanghai 200444，China；
2.Key Laboratory of Specialty Fiber Optics and Optical Access Networks，Shanghai University，Shanghai 200444，China；
3.Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication，Shanghai University，
Shanghai 200444，China）

【Abstract】Dense random deployment of small base stations will cause serious interferences and significant energy consumption problems. In order to reduce network interference，ensure users' network Quality of Service（QoS）and improve network Energy Efficiency（EE），this paper constructs a joint optimization framework based on Deep Reinforcement Learning（DRL）for resource allocation and power control. By comprehensively considering the same layer interference and cross layer interference in ultra-dense heterogeneous networks，the joint optimization of energy efficiency and user QoS for joint control of spectrum and power resources is proposed. To address the NP-hard characteristics of the joint optimization problem，a joint optimization algorithm based on DRL framework for resource allocation and power control is proposed，and the states，actions and return functions of joint spectrum and power allocation are defined. Then reinforcement learning，online learning，and offline training of deep neural network are used to control network resources，so as to find the best resource and power control strategy. Simulation results show that compared with the enumeration algorithm，Q-learning algorithm and two-stage algorithm，the proposed algorithm can effectively improve network energy efficiency while ensuring users' QoS.

【Key words】ultra-dense network；Energy Efficiency（EE）；resource allocation；Reinforcement Learning（RL）；power control；deep learning

DOI：10. 19678/j. issn. 1000-3428. 0057720

基金项目：国家自然科学基金（61501289）。

作者简介：郑冰原（1994—），男，硕士研究生，主研方向为超密集网络、能效优化、强化学习；孙彦赞、吴雅婷，副教授、博士；王 涛、方 勇，教授、博士。

收稿日期：2020-03-13 **修回日期：**2020-05-11 **E-mail：**bingyuanzheng@163.com

0 概述

随着智能终端数量的急剧增加和网络流量的指数级增长,目前蜂窝网络对网络容量和用户速率的需求面临着巨大挑战。超密集异构网络[1]作为5G移动通信的关键技术之一,可有效提高网络覆盖率和网络容量。由于基站的密集部署将会产生严重的干扰和能耗问题,从而导致网络性能下降[2],致使用户网络服务质量(Quality of Service,QoS)无法得到有效保障。

超密集网络中的资源分配策略会影响网络性能和用户体验,针对异构网络和超密集网络中的资源分配问题已有广泛研究。文献[3-4]利用随机几何分析系统能效(Energy Efficient,EE)与基站(Base Stations,BSs)密度之间的关系。文献[5]提出利用联合功率控制和用户调度的策略对网络能效进行优化。文献[6]提出了基于分簇的高能效资源管理方案,并阐述了资源分配和功率分配分阶段优化方法可实现能效优化。文献[7]通过联合考虑功率分配及负载感知来优化网络能效。文献[8]考虑了用户QoS需求并对网络能效进行分析。上述研究主要是通过传统的优化算法对网络能效进行优化。而在超密集网络中,基站数量的增加使得上述算法的复杂度将会急剧增大。为降低算法的计算复杂度,基于强化学习的无线资源分配策略受到了广泛关注。

文献[9-11]表明,无模型强化学习框架可用来解决无线网络中的动态资源分配问题。文献[12]利用强化学习框架对网络的功率分配进行优化,在提升网络容量的同时保障了用户QoS。然而在超密集网络中,网络规模庞大且结构复杂,基于Q-学习的资源分配算法存在动作状态空间的爆炸问题,使得基于Q-学习的资源分配算法收敛缓慢且难以找到最优解。而深度强化学习(Deep Reinforcement Learning,DRL)作为一种新兴工具可有效克服上述问题。利用DRL进行资源分配具有允许网络实体学习和构建关于通信和网络环境的知识、提供自主决策以及学习速度快等优点。因此,DRL适合解决超密集网络中具有较大状态和动作空间的复杂资源管理优化分配问题。在目前基于DRL的无线通信网络资源管理研究中,多数采用的是深度Q-学习网络(Deep Q-learning Network,DQN)。DQN是一种新的DRL算法[13],其通过将RL与深度神经网络相结合[14]来解决Q-学习的局限性。文献[15]在多小区网络中利用深度强化学习框架对基站功率进行控制,实现了网络容量的优化。文献[16]采用DQN优化小型基站的ON/OFF策略,以有效提高能源效率。文献[17]阐述了基于DQN的频谱资源分配,以实现网络能效和频谱效率的平衡。在超密集网络中,基于DRL框架对网络能效进行优化时多数是通过对单一资源控制而实现的,且很少考虑满足用户的QoS需求。因此,本文研究基于DRL的资源分配和功率控制联合优化问题,并考虑用户的QoS需求,以实现网络能效的进一步提升。

1 系统模型及优化问题

1.1 系统模型

本文考虑由一个宏基站和N个毫微基站(Femto Base Station,FBS)组成的超密集异构网络下行链路场景,如图1所示。宏基站作为整个网络的信息中心可收集整个网络的信息,并决定整个网络的资源块分配和功率控制策略。宏基站和各个毫微基站之间共享整个频率资源。同一时刻,每个用户设备(User Equipment,UE)只能与一个基站相关联,而宏用户之间及各个毫微基站内用户之间使用正交频谱资源。宏用户设备(Macro User Equipment,MUE)与毫微用户设备(Femto User Equipment,FUE)之间及不同毫微基站下的毫微用户之间均可使用相同的频谱资源。

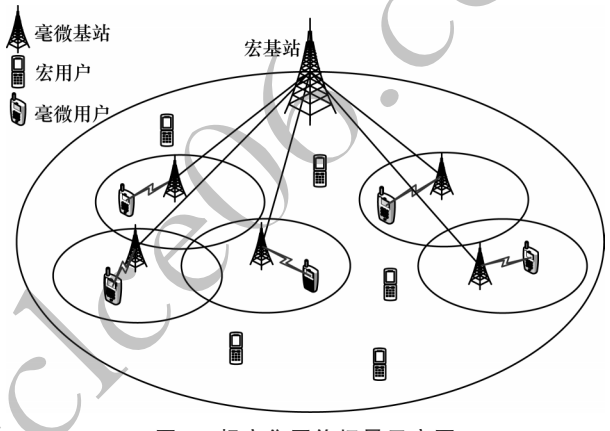

图1 超密集网络场景示意图

Fig.1 Schematic diagram of ultra-dense network scenario

1.2 问题的优化

在超密集异构网络中,FBS的集合可表示为$A=\{1,2,\cdots,N\}$。在每个时隙t,用户设备随机出现,并基于信号强度与相应的基站进行关联。为方便表示,将字母m和s作为下标,分别表示对应的宏基站和毫微基站。网络的总用户集合可表示为$\boldsymbol{U}^t=\{\boldsymbol{U}_{\mathrm{m}}^t,\boldsymbol{U}_{\mathrm{f}}^t\}$,其中$\boldsymbol{U}_{\mathrm{m}}^t$和$\boldsymbol{U}_{\mathrm{f}}^t=\sum_{n\in A}\boldsymbol{U}_{\mathrm{f},n}^t$分别表示宏基站用户集合和总的毫微基站用户集合。整个频谱资源被分为L个资源块(Resource Block,RB),而总资源块可表示为$B=\{1,2,\cdots,L\}$。两层网络及毫微基站之间共享所有资源块,同时每个资源块只能分配给一个用户。宏基站和各毫微基站的最大发射功率可分别表示为P_{m}^t和$P_{\mathrm{f},i}^t(i\in A)$。由于宏基站和毫微基站共享频谱资源,宏用户(MUE)会受到毫微基站的干扰,因此在时隙t内,MUE_k在资源块l上的信干噪比(Signal to

Interference plus Noise Ratio，SINR）可表示为：

$$\mathrm{SINR}_{\mathrm{m},l,k}^{t}=\frac{p_{\mathrm{m},l}^{t}h_{\mathrm{m},l,k}^{t}}{\sum_{i\in A}p_{\mathrm{f},i,l}^{t}h_{\mathrm{f},i,l,k}^{t}+\sigma^{2}} \tag{1}$$

其中，$p_{\mathrm{m},l}^{t}$为宏基站在资源块l上的发射功率，$h_{\mathrm{m},l,k}^{t}$为资源块l上宏基站到MUE_k的信道增益，$p_{\mathrm{f},i,l}^{t}$为毫微基站i在资源块l上的发射功率，$h_{\mathrm{f},i,l,k}^{t}$为资源块l上毫微基站i到MUE_k的信道增益，σ^2为噪声功率。

同理，毫微用户会受到宏基站和毫微基站的干扰。在时隙t内，与毫微基站i关联的毫微用户k在资源块l上的SINR可表示为：

$$\mathrm{SINR}_{\mathrm{f},i,l,k}^{t}=\frac{p_{\mathrm{f},i,l}^{t}h_{\mathrm{f},i,l,k}^{t}}{p_{\mathrm{m},l}^{t}h_{\mathrm{m},l,k}^{t}+\sum_{i\neq j,j\in A}p_{\mathrm{f},j,l}^{t}h_{\mathrm{f},j,l,k}^{t}+\sigma^{2}} \tag{2}$$

因此通过香农公式可得到在资源块l上MUE和FUE的速率为：

$$R_{\mathrm{m},l,k}^{t}=W\,\mathrm{lb}(1+\mathrm{SINR}_{\mathrm{m},l,k}^{t}) \tag{3}$$

$$R_{\mathrm{f},i,l,k}^{t}=W\,\mathrm{lb}(1+\mathrm{SINR}_{\mathrm{f},i,l,k}^{t}) \tag{4}$$

其中，W表示用户带宽。二进制指示变量$x_{\mathrm{m},l,k}^{t}$表示资源块l是否通过宏基站分配给UE_k，如果为1表示分配，否则不分配。同理，$x_{\mathrm{f},i,l,k}^{t}$表示资源块l是否通过毫微基站i分配给UE_k。本文分别定义$\boldsymbol{X}_{\mathrm{m}}^{t}$和$\boldsymbol{X}_{\mathrm{f}}^{t}$为宏用户和毫微用户的RB分配集合。因此，与宏基站和毫微基站i关联的用户速率可分别表示为：

$$R_{\mathrm{m},k}^{t}=\sum_{l\in B}x_{\mathrm{m},l,k}^{t}W\,\mathrm{lb}(1+\mathrm{SINR}_{\mathrm{m},l,k}^{t}) \tag{5}$$

$$R_{\mathrm{f},i,k}^{t}=\sum_{l\in B}x_{\mathrm{f},i,l,k}^{t}W\,\mathrm{lb}(1+\mathrm{SINR}_{\mathrm{f},i,l,k}^{t}) \tag{6}$$

总的系统容量可表示为：

$$R^{t}=\sum_{k\in\boldsymbol{U}_{\mathrm{m}}^{t}}\sum_{l\in B}R_{\mathrm{m},l,k}^{t}+\sum_{i\in A}\sum_{k\in\boldsymbol{U}_{i}^{t}}\sum_{l\in B}R_{\mathrm{f},i,l,k}^{t} \tag{7}$$

每个基站的功率包括发射功率和电路固定运行功率两个部分。本文定义毫微基站的发射功率集合为P_f^t，在下行链路传输中，总功耗可定义为：

$$P^{t}=\sum_{k\in\boldsymbol{U}_{\mathrm{m}}^{t}}\sum_{l\in B}x_{\mathrm{m},l,k}^{t}P_{\mathrm{m},l}^{t}+P_{\mathrm{m},c}+\sum_{i\in A}\sum_{k\in\boldsymbol{U}_{i}^{t}}\sum_{l\in B}x_{\mathrm{f},i,l,k}^{t}P_{\mathrm{f},i,l}^{t}+P_{\mathrm{f},c} \tag{8}$$

其中，$P_{\mathrm{m},c}$和$P_{\mathrm{f},c}$表示宏基站和毫微基站的电路固定运行功率。

在时隙t内，能效可表示为[18]：

$$\mathrm{EE}^{t}=\frac{R^{t}}{P^{t}} \tag{9}$$

根据文献[19]，用户k的流量延迟可定义为传输用户数据所需时间。基站的延迟可定义为服务用户的流量延迟之和。如果UE的数据要求为M bit，则毫微基站i在时隙t内的总流量延迟为：

$$D_{i}^{t}=\sum_{k\in U_{i}^{t}}\frac{M}{R_{\mathrm{f},i,k}^{t}} \tag{10}$$

在时隙t内，总的时间延迟可定义为：

$$D^{t}=\sum_{i\in A}D_{i}^{t} \tag{11}$$

为联合优化网络能源效率及用户服务质量，效用函数可定义为：

$$\eta(t)=\varepsilon\mathrm{EE}^{t}-(1-\varepsilon)D^{t} \tag{12}$$

同时考虑能源效率和QoS，其中，ε是为了平衡能效和时延的参数。本文的优化目标是在保证用户QoS需求的前提下，实现能源效率最大化，则联合优化问题可表示为：

$$\begin{aligned}&\eta(t)=\varepsilon\mathrm{EE}^{t}-(1-\varepsilon)D^{t}\\&\text{s.t.C1: }x_{m,l,k}^{t}=\{0,1\},\forall t,\forall l\in B\\&\quad\text{C2: }x_{f,i,l,k}^{t}=\{0,1\},\forall t,\forall i\in A,\forall l\in B\\&\quad\text{C3: }P_{m,l,k}^{t}>0,\forall t,\forall l\in B\\&\quad\text{C4: }P_{f,i,l,k}^{t}>0,\forall t,\forall i\in A,\forall l\in B\\&\quad\text{C5: }\sum_{k\in U_m}\sum_{l\in B}P_{m,l,k}^{t}\leqslant P_{m}^{t},\forall t\\&\qquad\sum_{k\in U_{f,i}}\sum_{l\in B}P_{f,i,l,k}^{t}\leqslant P_{f,i}^{t},\forall t,\forall i\in A\\&\quad\text{C6: }\sum_{k\in\boldsymbol{U}_m}x_{m,l,k}^{t}\leqslant 1,\sum_{k\in U_{f,i}}x_{f,i,l,k}^{t}\leqslant 1,\forall t,\forall i\in A\end{aligned} \tag{13}$$

其中，C1，C2，C6约束表示一个RB只能分配给一个用户，C3，C4表示基站的发射功率为正值，C5表示基站总的发射功率约束。该问题是一个非凸的多目标优化问题，且为NP-Hard问题，利用传统的求解方法存在算法复杂度较高的问题。

2　基于DQN的联合资源和功率分配框架

上述UDN场景下的联合资源分配问题可表示为马尔科夫决策过程（Markov Decision Processes，MDP）。采用强化学习技术可有效解决MDP问题，然而超密集网络规模庞大且拓扑结构复杂，使得算法的计算复杂度难以控制。DRL作为强化学习的升级，网络实体经过不断交互可学习和构建关于网络环境的知识，并进行自主决策，同时DNN的引入可大幅提高学习速度，在具有较大状态和动作空间的优化问题求解上有显著优势。因此，本文提出基于DRL的联合资源分配框架以优化网络能效。本节首先给出了强化学习的基本要素，并分别定义了联合资源分配和功率控制的状态、动作空间以及回报函数。其次提出了集中式的DRL算法以解决上述联合资源分配和功率控制的优化问题。

2.1　强化学习基本要素

在强化学习问题中，智能体（代理）基于策略选择动作与环境进行交互。强化学习框架中有状态空间、动作和回报3个要素。针对本文考虑的超密集异构网络以宏基站作为智能体，定义了基于强化学习框架的状态空间、动作和回报。具体描述如下：

1）状态空间：动作的选择由智能体决定，因此智能体需要整个网络信息。为了保证用户的QoS，同时优化网络能效，智能体需要获取网络中用户的QoS需求、时延、占用RB及各个基站功率等信息。则在时隙t内，智能体的状态可表示为：

$$s_t=\{\boldsymbol{M}^t,\boldsymbol{D}^t,\boldsymbol{X}_m^t,\boldsymbol{X}_f^t,\boldsymbol{P}_f^t\} \tag{14}$$

2)动作空间:为联合优化资源分配和功率控制,智能体需要决定每个用户的RB分配情况和毫微基站的发射功率。同时为了减少动作空间的大小,对基站的发射功率进行离散化并分为S个等级。因此,在时隙t内,智能体的动作可表示为:

$$\begin{aligned}a_t=&\{\{\boldsymbol{X}_{m,l,k}^t\in\{0,1\}|k\in\boldsymbol{U}_m^t,l\in B\},\\&\{\boldsymbol{X}_{f,i,l,k}^t\in\{0,1\}|i\in A,k\in\boldsymbol{U}_{f,l}^t,l\in B\},\\&\{\boldsymbol{P}_{f,i}^t=P_s|i\in A,s=1,2,\cdots,S\}\}\end{aligned} \tag{15}$$

动作空间随基站的增加呈指数级增长,动作空间的爆炸将是一个重要且困难的问题。每个动作都影响一个状态,这意味着状态空间的数量也很大。

3)回报函数:回报奖励代表框架的目标。为优化网络能效并同时保证用户的QoS,本文将优化问题式(13)作为最终优化目标。因此,回报函数可定义为:

$$r_t=\eta(t)=\eta(t|s_t,a_t) \tag{16}$$

2.2 基于DQN的联合资源和功率分配策略

智能体的目标是学习一个选择策略π,基于当前的状态s_t选择下一个动作$a_t=\pi(s_t)$,并得到即时回报r_t,然后得到下一个状态s_{t+1},持续该过程以得到最大预期累积回报。本文定义累积折扣奖励$V^\pi(s_t,a_t)$为:

$$V^\pi(s_t,a_t)=E_\pi\left[\sum_{t=1}^{T}\lambda^t\eta(t|s=s_t,a=a_t)\right] \tag{17}$$

其中,λ为折扣因子,$\eta(t|s=s_t,a=a_t)$为在状态s_t执行相应动作a_t的即时回报。

强化学习的目标是通过在线训练找到最优选择策略π^*,对于任意的选择策略π都满足$V^{\pi^*}(s_t,a_t)>V^\pi(s_t,a_t)$。在强化学习中,最典型的算法是Q-学习。Q-学习是解决马尔科夫过程的经典方法[20]。在Q-学习中,内部维护一个值函数可表示为$Q(s_t,a_t)$,其代表在状态s_t执行动作a_t的累积折扣奖励。智能体通过与环境相交互,利用反馈信息不断在线训练更新值函数,最终得到最优策略。根据贝尔曼方程,Q值的更新过程可表示为:

$$Q(s_t,a_t)=(1-\alpha)Q(s_t,a_t)+\alpha[r_t+\lambda\max_{a_{t+1}}Q(s_{t+1},a_{t+1})] \tag{18}$$

其中,α为学习率。

在超密集异构网络中,由于基站密集部署且网络环境更加复杂,使得状态、动作空间大小随基站数量呈指数级增加,很难通过查找Q值表的方式找到最优策略。为解决在复杂环境下Q-学习状态空间较大的问题,将深度神经网络引入到RL框架中以形成深度强化学习。DQN是DRL中较为经典的方法。通过RL在线学习和DNN网络的线下训练,可有效解决状态空间爆炸问题。在DQN中,通过强化学习技术产生训练数据,再利用DNN线下训练拟合出最佳值函数$Q(s_t,a_t)$。对于主深度神经网络输出Q值可表示为$Q(s_t,a_t|\theta)$,其中,θ为主神经网络参数。智能体基于神经网络输出的Q值选择相应的动作,最优选择策略可表示为:

$$\pi^*(s)=\operatorname{argmax}Q^*(s_t,a_t|\theta) \tag{19}$$

其中,$Q^*(s_t,a_t|\theta)$是通过DNN逼近的最佳Q值。为使$Q(s_t,a_t|\theta)$更为稳定,需要对目标Q值进行误差计算。逼近的目标Q值$\tilde{Q}(s_t,a_t|\theta^-)$可定义为:

$$\tilde{Q}(s_t,a_t|\theta^-)=\eta(t|s=s_t,a=a_t)+\gamma Q(s_{t+1},\pi^*(s_{t+1})|\theta^-) \tag{20}$$

其中,$\tilde{Q}(s_t,a_t|\theta^-)$为目标$Q$网络,$\theta^-$为目标$Q$网络参数,目标$Q$网络和主深度神经网络具有相同的神经网络结构。在每一步训练中,通过最小化损失函数来更新神经网络参数,且损失函数可表示为:

$$L(\theta)=E[(\tilde{Q}(s_t,a_t|\theta^-)-Q(s_t,a_t|\theta))^2] \tag{21}$$

在线学习阶段中为了防止目标策略陷入局部最优,本文在该阶段引入ε-贪婪策略进行动作的选择。这将存在$1-\xi$的概率可根据式(19)选择动作a_t和有ξ的概率随机选择动作。在初始阶段,智能体通过收集网络环境信息,得到当前网络的状态s_t。根据ε-贪婪策略选择动作a_t,该动作决定了网络中用户的RB分配及功率分配情况,执行动作即实施具体的资源和功率分配,并得到即时奖励r_t,同时网络转变为下一个状态s_{t+1}。接下来将经验向量(s_t,a_t,r_t,s_{t+1})存储到经验池中,并通过不断交互产生线下训练数据。

在线下训练阶段,利用DNN对在线学习产生的数据进行训练,并拟合出最佳值函数。当使用非线性函数逼近器时,强化学习算法得到的平均报酬可能不稳定甚至是发散的。这是因为一个小的Q值变化可能会显著影响政策。因此,数据分布和Q值与目标值$\tilde{Q}(s_t,a_t|\theta^-)$之间的相关性是多种多样的。为解决该问题,本文引入了经验重放和目标Q网络这2种机制。

1)固定目标Q网络。在训练过程中Q值会发生偏移。因此,如果使用一组不断变化的值来更新主深度神经网络,那么值估计可能会失控,这将导致算法不稳定。为解决该问题,本文使用目标Q网络频繁而缓慢地更新主深度神经网络的值。即在训练时只训练主深度神经网络,经过多次在线训练后将主深度神经网络的参数更新到目标Q网络中。该做法会使得目标与估计Q值间的相关性显著降低,有效提高算法的稳定性。

2)经验重放策略。在线下训练阶段中,为使学习更加稳定,本文引入了经验重放策略。该算法首先初始化回放经验$\boldsymbol{D}$,即经验池。智能体通过与环境交互产生经验向量(s_t,a_t,r_t,s_{t+1})并存入经验池。其次,算法随机选取样本,即从经验池中随机抽取小批量的样本到DNN中进行训练。经过训练的DNN获得的Q值将用于获得新的经验,即这种机制允许DNN通过使用新旧经验更有效地训练网络。此外,通过使用经验重放可有效转换独立和恒等分布,从而消除观测之间的相关性。当经验池中有足够多的数据时,从经验池中随机抽取批量数据进行DNN网络训练,并定时更新神经网络参数θ。

本文所提基于DQN的联合资源和功率分配算法流程如算法1所示。

算法1　基于DQN的联合资源和功率分配算法

输入　强化学习状态:用户QoS,时延,RB及功率

输出　最优策略(RB及功率分配)、能效和时延折中

1.初始化经验池D,权重为θ的神经网络

2.for episode = 1: M

3.初始化超密集异构网络,初始状态s_t

4.for t = 1: T

5.由状态s_t利用ε贪婪策略选择动作a_t

6.执行动作a_t调整用户RB分配和基站发射功率,由式(16)得出回报r_t

7.接收下一个状态s_{t+1}

8.存储经验(s_t, a_t, r_t, s_{t+1})到经验池D

9.if D的容量大于N

10.从经验池D中随机抽取批量经验样本B

11.根据式(20)计算样本目标$Q(s_t, a_t|\theta)$

12.通过式(21)最小化损失函数,更新网络权重θ

13.更新选择策略$\pi^*(s) = \text{argmax}\, Q^*(s_t, a_t|\theta)$

14.end

15.end

3　仿真与结果分析

本节对所提算法进行仿真分析,以验证本文算法在保证用户QoS的前提下,在降低网络干扰和优化UDHN能源效率方面的有效性。在实验选择的场景中,毫微基站和宏用户都均匀地部署在覆盖区域。为了简化分析,本文设置一个毫微基站关联一个用户,同时将毫微基站的发射功率进行离散化处理,并分为3个等级,可取值为$p=\{20,25,30\}$。深度神经网络使用包含3个隐藏层的反馈神经网络,第1层包含400个神经元,第2层包含800个神经元,第3层包含300个神经元。本文利用瑞利衰落来模拟基站和用户之间的信道以及路径损失模型,其他仿真参数如表1所示。

表1　仿真参数设置

Table 1　Simulation parameters setting

参数	设置值
信道带宽/M	10
小基站密度/(个·m^{-2})	0.000 4~0.000 8
小基站最大发射功率/dBm	30
基站固定运行损耗/dBm	40
网络大小/m^2	100×100
资源块数	8
瑞利衰落参数	8
折中因子	0.000 01
学习率	0.5
折扣因子	0.9

为更好地分析本文所提DQN算法的性能,实验将DQN算法与最优能效枚举算法、基于Q-学习算法及两阶段算法[6]这3种算法进行对比。图2给出了当用户速率需求M分别为0.1、0.5、1.0时,本文所提DQN算法的网络能效随基站密度的变化情况。从图2可以看出,当用户速率需求一定时,随着基站密度的增大,网络能效逐渐减小。当基站密度一定时,随着用户速率需求的增大,网络需要更高的发射功率满足用户需求,网络能效呈下降趋势。因此,本文所提DQN算法可以根据用户QoS动态调整网络状态,优化网络能效。

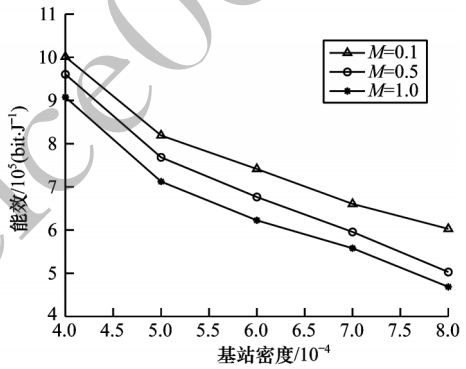

图2　不同用户速率需求下网络的总能效

Fig.2　Total energy efficiency of the network under different user rate requirements

网络的总能效随基站密度变化如图3所示,此时用户的速率需求为0.5M。从图中可知,随着网络中毫微基站的密度增大,所有算法的网络整体能效都呈下降趋势。这是由于随着毫微基站数量的增加,网络干扰和能耗更加严重,导致网络性能下降。与典型的Q-学习算法及两阶段算法相比,所提DQN算法具有更好的能效,与最优的能效遍历算法比较接近。这是由于在两阶段算法中,将RB分配和功率控制分为两步分别优化,然而RB分配阶段虽然避免了一部分网络干扰,但进行功率控制时,RB分配策略已经确定,制约着整体性能的提升。随着基站密度的增大,对网络性能影响越大。在DQN中,智能体不断与环境交互,将RB的分配策略以及相应的功

率分配策略同时作为网络动作优化网络性能，综合考虑了RB分配和功率分配的相互影响。智能体通过不断尝试与探索，逐步找到最佳的选择策略。同时，智能体经过DNN的训练后可根据网络环境变化自适应调整网络的资源分配策略。因此，相较于Q-学习和两阶段算法，本文算法具有更好的网络性能。由于DQN算法中加入用户QoS约束，且随着基站密度的增加网络中干扰加剧，并且需要更高的发射功率以保证用户速率，因此随着基站密度减小，本文所提DQN算法与枚举算法的差距逐渐减小。

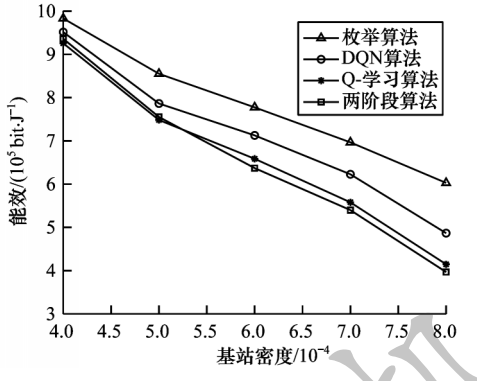

图3 4种算法在不同基站密度下的总能效

Fig.3 Total energy efficiency of four algorithms under different base station densities

当用户速率需求为0.5M时，网络中用户总时延随基站密度变化如图4所示。从图4可以看出，本文所提DQN算法相比其他算法具有更好的总用户时延性能。随着基站密度的增加，网络中用户基数增大，网络干扰加剧，且总的用户时延逐渐增大。由于枚举算法以最优能效为优化目标，基站密度增加会导致个别用户速率下降，导致整个网络总时延增大，因此枚举算法的时延会更大。而本文所提DQN算法将用户总时延作为回报函数的一部分，通过将RB分配和功率分配策略作为执行动作对RB和功率进行联合优化，可有效降低网络干扰，保证用户速率。结合图3和图4可知，DQN算法在提升网络能效的同时，可有效保证用户的QoS。

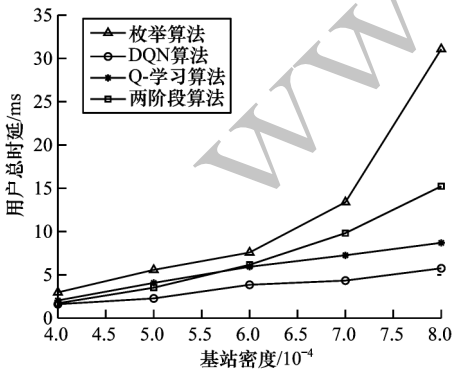

图4 4种算法在不同基站密度下的用户总时延

Fig.4 Total user delay of four algorithms under different base station density

本文所提DQN和Q-学习算法的迭代收敛曲线如图5所示。从图5可以看出，算法在经过近100次迭代后逐渐收敛，且在前50次迭代中，DQN算法的表现比Q-学习算法差。这是因为在前50次迭代中，Q-学习算法可从开始的反馈中学习，而DQN算法只是随机选择动作并将反馈信息存储在回放经验池中。而在100次迭代后，DQN和Q-学习算法都趋于稳定，且DQN算法的性能比Q-学习算法好。与典型的Q-学习算法相比，本文所提DQN算法不仅收敛更快，而且具有更好的性能指标。

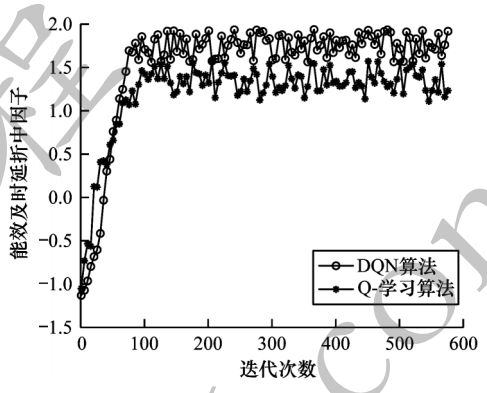

图5 2种算法的迭代收敛曲线

Fig.5 Iterative convergence curves of two algorithms

4 结束语

为降低超密集异构网络的同层和跨层干扰，并提高网络的能效，本文联合考虑用户QoS，提出联合RB分配和功率控制的优化问题。针对传统算法复杂度较高的问题，引入DQN框架并定义了优化网络能效和确保用户QoS的奖励函数。仿真结果表明，与典型Q-学习算法、两阶段算法及枚举算法相比，本文所提DQN算法可有效保证用户的QoS，且性能更优。下一步将研究基于多智能体的分布式资源管理问题，利用多智能协作减小网络干扰，进一步提升网络能效。

参考文献

[1] KAMEL M, HAMOUDA W, YOUSSEF A. Ultra-dense networks: a survey[J]. IEEE Communications Surveys & Tutorials, 2016, 18(4): 2522-2545.

[2] NAM W, BAI D, LEE J, et al. Advanced interference management for 5G cellular networks[J]. IEEE Communications Magazine, 2014, 52(5): 52-60.

[3] REN Qi, FAN Jiancun, LUO Xinmin, et al. Analysis of spectral and energy efficiency in ultra-dense network[C]// Proceedings of 2015 IEEE International Conference on Communication Workshop. Washington D. C., USA: IEEE Press, 2015: 2812-2817.

[4] AN Lu, ZHANG Tiankui, FENG Chunyan. Stochastic geometry based energy-efficient base station density optimization in cellular networks [C]//Proceedings of

2015 IEEE Wireless Communications and Networking Conference. Washington D. C. , USA: IEEE Press, 2015: 1614-1619.

[5] SAMARAKOON S, BENNIS M, SAAD W, et al. Energy-efficient resource management in ultra dense small cell networks: a mean-field approach[C]//Proceedings of 2015 IEEE Global Communications Conference. Washington D. C. , USA: IEEE Press, 2015: 1-6.

[6] LIANG Liang, WANG Wen, JIA Yunjian, et al. A cluster-based energy-efficient resource management scheme for ultra-dense networks[J]. IEEE Access, 2016, 4: 6823-6832.

[7] WU Shie, ZENG Zhimin, XIA Hailun. Load-aware energy efficiency optimization in dense small cell networks[J]. IEEE Communications Letters, 2016, 21(2): 366-369.

[8] COSKUN C C, AYANOGLU E. Energy-spectral efficiency tradeoff for heterogeneous networks with QoS constraints[C]//Proceedings of 2017 IEEE International Conference on Communications. Washington D. C. , USA: IEEE Press, 2017: 1-7.

[9] GAO Yang, CHEN Shifu, LU Xin. Research on reinforcement learning technology: a review[J]. Acta Automatica Sinica, 2004, 30(1): 86-100. (in Chinese)
高阳，陈世福，陆鑫. 强化学习研究综述[J]. 自动化学报，2004，30(1)：86-100.

[10] SIMSEK M, BENNIS M, CZYLWIK A. Dynamic inter-cell interference coordination in HetNets: a reinforcement learning approach[C]//Proceedings of 2012 IEEE Global Communications Conference. Washington D. C. , USA: IEEE Press, 2012: 5446-5450.

[11] ZHAO Nan, LIANG Yingchang, PEI Yiyang. Dynamic contract incentive mechanism for cooperative wireless networks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(11): 10970-10982.

[12] AMIRI R, MEHRPOUYAN H, FRIDMAN L, et al. A machine learning approach for power allocation in HetNets considering QoS[C]//Proceedings of 2018 IEEE International Conference on Communications. Washington D. C. , USA: IEEE Press, 2018: 1-7.

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

[14] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[15] ZHANG Yong, KANG Canping, MA Tengteng, et al. Power allocation in multi-cell networks using deep reinforcement learning [C]//Proceedings of the 88th Vehicular Technology Conference. Washington D. C. , USA: IEEE Press, 2018: 1-6.

[16] LI Han, GAO Hui, LÜ Tiejun, et al. Deep Q-learning based dynamic resource allocation for self-powered ultra-dense networks [C]//Proceedings of 2018 IEEE International Conference on Communications Workshops. Washington D. C. , USA: IEEE Press, 2018: 1-6.

[17] LIU Zhiyong, CHEN Xin, CHEN Ying, et al. Deep reinforcement learning based dynamic resource allocation in 5G ultra-dense networks[C]//Proceedings of 2019 IEEE International Conference on Smart Internet of Things. Washington D. C. , USA: IEEE Press, 2019: 168-174.

[18] HAN F, SAFAR Z, LIU K J R. Energy-efficient base-station cooperative operation with guaranteed QoS[J]. IEEE Transactions on Communications, 2013, 61(8): 3505-3517.

[19] LEE G, SAAD W, BENNIS M, et al. Online ski rental for scheduling self-powered, energy harvesting small base stations [C]//Proceedings of 2016 IEEE International Conference on Communication. Washington D. C. , USA: IEEE Press, 2016: 1-6.

[20] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3/4): 279-292.

编辑　刘继娟