

DOI:10.11918/202204106

# 超密集网络中基于改进 DQN 的接入选择算法

唐 宏<sup>1,2</sup>, 刘小洁<sup>1,2</sup>, 甘陈敏<sup>1,2</sup>, 陈 榕<sup>1,2</sup>

(1. 重庆邮电大学 通信与信息工程学院, 重庆 400065; 2. 移动通信技术重庆市重点实验室(重庆邮电大学), 重庆 400065)

**摘要:** 在超密集网络环境中, 各个接入点密集部署在热点区域, 构成了复杂的异构网络, 用户需要选择接入合适的网络以获得最好的性能。如何为用户选择最优的网络, 使用户自身或网络性能达到最佳, 称为网络接入选择问题。为了解决超密集网络中用户的接入选择问题, 综合考虑网络状态、用户偏好以及业务类型, 结合负载均衡策略, 提出了一种基于改进深度 Q 网络 (deep Q network, DQN) 的超密集网络接入选择算法。首先, 通过分析网络属性和用户业务的偏好对网络选择的影响, 选择合适的网络参数作为接入选择算法的参数; 其次, 将网络接入选择问题利用马尔可夫决策过程建模, 分别对模型中的状态、动作和奖励函数进行设计; 最后, 利用 DQN 求解选网模型, 得到最优选网策略。此外, 为了避免 DQN 过高估计 Q 值, 对传统 DQN 的目标函数进行优化, 并且在训练神经网络时, 引入了优先经验回放机制以提升学习效率。仿真结果表明, 所提算法能够解决传统 DQN 的高估问题, 加快神经网络的收敛, 有效减少用户的阻塞, 并改善网络的吞吐能力。

**关键词:** 超密集网络; 接入选择; 深度 Q 网络 (DQN); 优先经验回放; 负载均衡

中图分类号: TN92

文献标志码: A

文章编号: 0367-6234(2023)05-0107-07

## Access selection algorithm based on improved DQN for ultra-dense networks

TANG Hong<sup>1,2</sup>, LIU Xiaojie<sup>1,2</sup>, GAN Chenmin<sup>1,2</sup>, CHEN Rong<sup>1,2</sup>

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Chongqing Key Lab of Mobile Communications Technology (Chongqing University of Posts and Telecommunications), Chongqing 400065, China)

**Abstract:** In the ultra-dense network environment, each access point is deployed in the hotspot area, which forms a complex heterogeneous network. Users need to choose the appropriate network to access, so as to achieve the best performance. Network selection problem is to choose the optimal network for the user, so that the user or network performance reaches the best. In order to solve the access selection problem of users in ultra-dense networks, we proposed an ultra-dense network access selection algorithm based on the improved deep Q network (DQN), considering network states, user preferences, and service types, and combining with load balancing strategies. First, by analyzing the influence of network attributes and user preferences on network selection, the appropriate network parameters were selected as the parameters of the access selection algorithm. Then, the problem of network access selection was modeled by Markov decision-making process, and the states, actions, and reward functions of the model were designed. Finally, the optimal network strategy was obtained by using DQN to solve the network selection model. In addition, the target function of traditional DQN was optimized to avoid overestimation of Q value by DQN, and a priority experience replay mechanism was introduced to improve learning efficiency. Simulation results show that the method could well solve the problem of overestimation of traditional DQN, accelerate the convergence of neural network, effectively reduce user congestion, and improve network throughput performance.

**Keywords:** ultra-dense network; access selection; deep Q network (DQN); priority experience replay; load balancing

随着互联网与物联网技术的飞速发展, 无线通信服务在近几年内发展迅猛。为了适应数据流量的迅速增加、设备的连接、新兴业务和应用的需要, 超密集网络应运而生<sup>[1]</sup>。超密集网络是 5G 技术中最具有发展前景的技术之一, 它可以将低功率基站大

规模部署到一些热点区域, 从而提高系统的承载能力<sup>[2]</sup>。由于超密集网络具有密集化、异构化的特性并且移动用户业务呈现多样性, 如何合理地选取合适的接入网络已成为一个值得深入探讨的问题<sup>[3]</sup>。为了保障用户服务质量 (quality of service, QoS) 和提

收稿日期: 2022-04-28; 录用日期: 2022-06-24; 网络首发日期: 2022-08-29

网络首发地址: <https://kns.cnki.net/kcms/detail/23.1235.T.20220826.1621.020.html>

基金项目: 长江学者和创新团队发展计划 (IRT\_16R72)

作者简介: 唐 宏 (1967—), 男, 教授, 博士生导师

通信作者: 刘小洁, 84551250@qq.com

高网络整体资源利用率,需要设计合适的接入选择算法。

针对异构网络中用户接入选择问题,国内外研究者进行了大量深入的研究。文献[4-7]以用户接收信号强度、信号干扰比或传输速率等作为判决因素,将用户接入到收益最大的网络中;但是这类方法考虑的因素层面单一,不能适应动态变化的超密集网络环境,用户接入的可能不是综合性能最好的网络。文献[8-11]从多个角度对网络选择的影响因素进行了研究,采用多属性决策判决法对用户的偏好和网络效用进行了分析,保证了用户的 QoS;然而,此类算法过于注重用户 QoS 而忽略了网络的承受能力,可能会将大部分用户接入到同一个网络中,加重某些网络的负载,导致负载不均衡。文献[12-14]将选网问题建模为马尔可夫决策过程,利用 Q 学习<sup>[15]</sup>自主学习的特点,为用户选择长期收益最大的网络,在一定程度上减少了切换次数;然而,在 Q 学习过程中,由于网络数量和用户数量的增多,使得系统的状态维数迅速增长,运算量也随之剧增,从而影响了算法的运行效率。

针对以上问题,提出一种新的基于 DQN 的接入选择算法,该算法综合考虑网络、用户和业务 3 个层面的因素,从而确保用户的 QoS,并有效地解决网络的高动态特性所造成的负载不均衡问题。同时,利用双 Q 学习<sup>[16]</sup>算法思想,将 DQN 中的目标函数进行优化,解决 DQN 算法在进行 Q 值估计时存在的过高估计问题,并且,针对传统的 DQN 算法由于收敛速度慢导致的算法处理时延过大的情况,采用优先级经验回放机制<sup>[17]</sup>加速了算法的收敛。

## 1 问题建模

接入选择过程主要分为 3 个阶段:网络发现、选网决策和接入执行<sup>[18]</sup>,研究重点集中在选网决策阶段。超密集网络场景中部署了多种不同类型的接入点,一个移动终端能够从各个网络中接收到基站发出的信号,并自由选择周围的网络。移动用户在时隙  $\tau$  内保持连接,并且只选择一个网络进行传输。用户需要在每个时隙  $\tau$  结束时做出决策,决策时段用  $T = \{\tau, 2\tau, \dots, d\tau\}$  表示,其中  $d\tau$  代表连接终止的时间。根据排队理论,将用户的到达和离开过程定义为服从参数  $\lambda_i$  和  $\mu_i$  的独立泊松分布。则在时隙  $\tau$  内,  $p$  个用户到达系统的概率可表示为  $P(p) = \frac{(\lambda_i)^p}{p!} e^{-\lambda_i\tau}$ ,  $q$  个用户离开系统的概率为  $P(q) = \frac{(\mu_i)^q}{q!} e^{-\mu_i\tau}$ 。

为保证基于改进 DQN 的接入选择算法顺利运行,首先需要建立强化学习模型,然后,运用深度学习的方法,来解决强化学习的问题。采用马尔可夫决策过程建立强化学习模型,用  $\langle S, A, P, \pi, r \rangle$  表示。其中,  $S$  为网络的离散状态空间,包含影响网络的各个参数、用户连接状态以及用户业务类型;  $A$  为动作空间,由候选网络集合表示;  $P$  表示网络状态由  $s_t$  转移到下一个状态  $s_{t+1}$  的概率;用户的动作策略  $\pi: S \rightarrow A$  表示动作  $A$  由状态  $S$  决定;奖励函数  $r: S \times A \rightarrow R$  表示用户执行动作后接收到的来自超密集网络环境的反馈。强化学习过程见图 1,在每个决策时刻  $t$ ,智能体根据观测到的环境状态  $s_t$  以及策略  $\pi$  来完成动作  $a_t$ ,环境会向智能体回送奖励  $r_t$ ,然后环境进入一个新的状态  $s_{t+1}$ 。

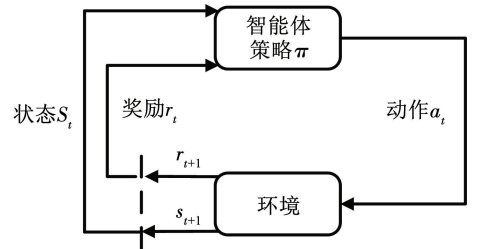


图 1 强化学习过程

Fig. 1 Reinforcement learning process

1) 状态和动作空间。假设在每个决策时刻  $t$ , 移动用户的候选网络数量为  $M$ 。选择代表用户 QoS 和网络状态的 5 个参数作为决策因素,分别为可用带宽( $B_m$ )、时延抖动( $D_m$ )、丢包率( $L_m$ )、价格( $C_m$ )和安全性( $S_m$ )。则某一时刻网络状态空间为

$$S = \{B_m, D_m, L_m, C_m, S_m, m, h, n\} \quad (1)$$

式中:  $m$  为用户终端连接到的网络,  $h$  为业务类型,  $n$  为当前用户终端所连接的网路中服务的用户数。

移动用户的接入动作由  $a$  表示,则动作空间  $A$  为

$$A = \{a | a \in \{1, 2, \dots, M\}\} \quad (2)$$

2) 奖励函数。由于考虑到多个网络参数,为了更好地均衡用户偏好和网络实际状态,将奖励函数定义为

$$R_t = r_{\text{load},i,t} + \sum_x \omega_x r_{x,i,t} \quad (3)$$

式中:  $\omega_x$  为第  $x$  个网络参数的权重,该权重由层次分析法结合熵权法求出;  $r_{x,i,t}$  为  $t$  时刻候选网络  $i$  的参数  $x$  的归一化奖励;  $r_{\text{load},i,t}$  为  $t$  时刻候选网络  $i$  的负载的归一化奖励。

针对效益型参数,如可用带宽、安全性等,  $r_{x,i,t}$  定义为

$$r_{x_{i,t}} = \begin{cases} 0, & x_{i,t} \leq x_{\min} \\ \frac{x_{i,t} - x_{\min}}{x_{\max} - x_{\min}}, & x_{\min} < x_{i,t} < x_{\max} \\ 1, & x_{i,t} \geq x_{\max} \end{cases} \quad (4)$$

针对代价型参数,如时延抖动、丢包率、价格等, $r_{x_{i,t}}$ 定义为

$$r_{x_{i,t}} = \begin{cases} 1, & x_{i,t} \leq x_{\min} \\ \frac{x_{\max} - x_{i,t}}{x_{\max} - x_{\min}}, & x_{\min} < x_{i,t} < x_{\max} \\ 0, & x_{i,t} \geq x_{\max} \end{cases} \quad (5)$$

一个网络可同时为多个移动用户提供服务,但当连接用户超过负载上限时,接入请求失败,导致用户阻塞,设计负载均衡策略可以有效地减少阻塞率。 $t$ 时刻网络 $i$ 的负载情况可由负载系数 $l_{i,t}$ 表示

$$l_{i,t} = \frac{n_{i,t}}{N_i} \quad (6)$$

式中: $n_{i,t}$ 为 $t$ 时刻接入网络 $i$ 的用户数, $N_i$ 为网络 $i$ 能够同时服务的最大用户数。

根据负载系数,将负载的归一化奖励定义为

$$r_{\text{load}_{i,t}} = \frac{1}{l_{i,t} + 1} \quad (7)$$

## 2 基于改进DQN的接入选择算法

算法最终目标是获得使长期累积回报期望最大的动作,即选择长期综合效用最大的网络。因此,策略 $\pi$ 下的动作-价值函数( $Q$ 函数)被定义为状态 $s_t$ 下从状态空间 $S$ 中执行动作 $a_t$ 的期望累积回报,则 $Q$ 函数为

$$Q_{\pi}(s_t, a_t) = E[R_{t+1} + \gamma Q_{\pi}(s_{t+1}, a_{t+1})] = R(a_t | s_t) + \gamma \sum_{s_{t+1} \in S} P(s_{t+1} | s_t, a_t) \sum_{a_{t+1} \in A} \pi(a_{t+1} | s_{t+1}) \cdot Q_{\pi}(s_{t+1}, a_{t+1}) \quad (8)$$

式中: $R_t$ 为 $t$ 时刻的立即回报; $\gamma$ 为折扣因子,用来评价之后时刻立即回报的重要程度。

因此,贝尔曼方程为

$$Q^*(s_t, a_t) = \max_{\pi} Q_{\pi}(s_t, a_t) = R(a_t | s_t) + \gamma \sum_{s_{t+1} \in S} P(s_{t+1} | s_t, a_t) \cdot \max_{a_{t+1} \in A} Q^*(s_{t+1}, a_{t+1}) \quad (9)$$

对于任何状态 $s_t \in S$ ,其对应的最优动作就是使 $Q^*(s_t, a_t)$ 值最大的动作,则最优策略为

$$\pi^* \equiv \arg \max_{a_t \in A} Q^*(s_t, a_t), \forall s_t \in S \quad (10)$$

在 $Q$ 学习中, $Q$ 表中的每个状态-动作对( $s, a$ )都对应一个 $Q$ 值, $Q$ 值的更新公式为

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha (R_{t+1} + \gamma \max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \quad (11)$$

式中 $\alpha \in (0, 1)$ 为学习率。

### 2.1 传统DQN算法

$Q$ 学习中 $Q$ 表的维护既要耗费大量的运算资源,又易发生维度爆炸,因此可以采用一种非线性函数近似器来近似 $Q$ 函数。神经网络是一种常用的非线性函数近似器,利用DNN作为 $Q$ 函数近似器的 $Q$ 学习方法即为DQN。传统DQN算法重点有两部分,分别是构建目标网络和引入经验回放机制。

1) 构建目标网络。DQN使用神经网络(主网络)来逼近 $Q$ 函数,其参数化表示为 $Q_{\text{main}}(s_t, a_t; \theta_k)$ ,其中, $\theta_k$ 为第 $k$ 次迭代时神经网络的参数。主网络每次迭代的优化目标为

$$y = R_{t+1} + \gamma \max_a Q_t(s_{t+1}, a; \theta_k^-) \quad (12)$$

该优化目标由另一个与主网络结构相同的目标网络 $Q_t(s_{t+1}, a; \theta_k^-)$ 产生,其中 $a$ 为当前状态下所有可能的动作, $s_{t+1}$ 为网络下一个时刻的状态, $\theta_k^-$ 由主网络参数经过 $G$ 次迭代后复制而来。

在主网络的训练中,采用了一种基于随机梯度下降的算法,使损失函数最小,从而使神经网络的参数得到最大程度的更新,损失函数为

$$L_k(\theta_k) = E[(y - Q_{\text{main}}(s_t, a_t; \theta_k))^2] \quad (13)$$

由式(13)可得损失函数的梯度

$$\square_{\theta_k} L_k(\theta_k) = E[(y - Q_{\text{main}}(s_t, a_t; \theta_k)) \square_{\theta_k} Q_{\text{main}}(s_t, a_t; \theta_k)] \quad (14)$$

2) 经验回放机制。每个决策时刻智能体与环境的交互信息表示为一条经验 $e = (s_t, a_t, r_t, s_{t+1})$ ,将所有经验存储到序列 $D = \{e_1, e_2, \dots, e_N\}$ 中,建立经验回放机制。每次从 $D$ 中随机选取少量的经验样本进行训练。通过重复采样历史数据提高了数据的利用率,与此同时在一定程度上降低了数据之间的关联性。

### 2.2 改进DQN算法

由式(12)可以看出,动作的选择和评价都基于同一个参数 $\theta_k^-$ ,这将有一定几率造成 $Q$ 值的过高估计,在进行网络选择时将有可能导致用户无法接入到综合性能最好的网络。为了解决这个问题,参考双 $Q$ 学习算法思想,对DQN算法做出如下改进:将动作的评估由原来的主网络计算转为由目标网络计算,而动作的选择仍然由主网络操作。此时需要将DQN中优化目标进行改进,具体为

$$y = R_{t+1} + \gamma Q_t(s_{t+1}, a_{\text{main}}; \theta_k^-) \quad (15)$$

式中: $a_{\text{main}} = \arg \max_a Q_{\text{main}}(s_{t+1}, a; \theta_k)$ , $\theta_k$ 和 $\theta_k^-$ 分别为主网络和目标网络的参数。

传统DQN算法中经验转移从回放记忆中使用同等频率进行采样,对回放序列中经验的重要程度

缺乏考虑,从而导致抽样随机性较大、学习效率低等情况,当网络状态维度增大时,网络选择算法可能会因为处理时延增大而失效。

针对此问题,引入求和树结构的优先经验回放机制。优先经验回放的基本理念是:打破平均抽样,给有较高学习效率的样本以较大的抽样权重。以时间差分(temporal difference, TD)误差作为优先级指标,误差越大,对梯度学习的影响越大,故应该优先取样。利用求和树结构,实现按照优先级为 TD 误差的抽样。具体表示为:将每个经验的优先级值视为一个叶节点,两个节点作为一组,向上叠加,形成一个二叉树,根节点的值即为所有经验的优先级值的和。在对训练数据进行采样时,先将样本划分为多个区间,区间的大小为根节点的总优先级值除以每个批次中的样本数,每个区间中随机抽取一个数,从根节点开始由上至下搜索叶节点,确定最终的采样数据。TD 误差为

$$\delta_t = R_{t+1} + \gamma \max_a Q_t(s_{t+1}, \arg \max_a Q_{\text{main}}(s_{t+1}, a; \theta^-); \theta^-) - Q_{\text{main}}(s_t, a_t; \theta) \quad (16)$$

TD 误差的绝对值越大,提取样本进行训练的可能性就越大,但是这样容易造成网络的过拟合,对此,将第  $i$  条经验样本被提取的概率为

$$P(i) = \frac{p_i}{\sum p_i} \quad (17)$$

式中: $p_i = |\delta_i + \varepsilon| > 0$ ,  $\varepsilon$  是一个很小的值,用于保证 TD 误差为 0 的经验样本能被抽取到,从而避免神经网络的过拟合,保证算法的性能。

基于经验优先级采样的方法无需遍历经验池也能提取到优先级高的样本,减少运算资源,加快神经网络的学习速度。

针对以上两条对 DQN 的改进方法,设计基于改进 DQN 的接入选择算法的结构框图见图 2。

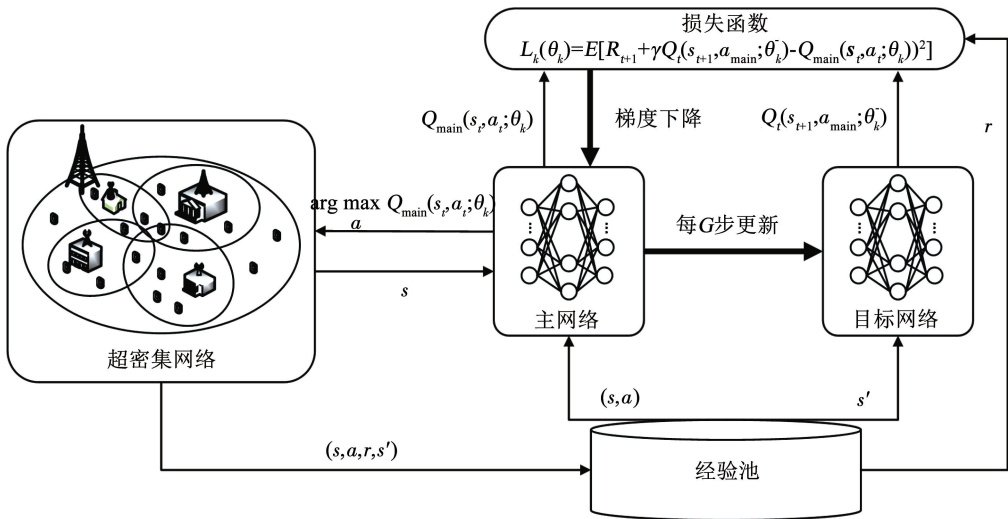


图 2 改进 DQN 接入选择算法结构框图

Fig. 2 Block diagram of improved DQN access selection algorithm

整个网络接入选择算法主要分为离线训练与在线决策两部分。离线学习阶段,用户终端读取网络状态信息作为状态  $s$ ,系统将状态  $s$  输入主网络,并在前向传播之后输出当前状态下所有状态-动作对的  $Q$  值,根据  $\varepsilon$ -greedy 策略选择动作  $a$ , $\varepsilon$ -greedy 策略见式(18)。训练的每一步都更新  $\varepsilon$ -greedy 策略,即  $\varepsilon \leftarrow \varepsilon - (\varepsilon_{\text{initial}} - \varepsilon_{\text{final}}) / \xi$ 。

$$\pi(a|s) = \begin{cases} 1 - \varepsilon, a^* = \arg \max_{a \in A} Q(s, a) \\ \varepsilon, \text{随机 } a \end{cases} \quad (18)$$

动作选择结束后,用户终端在超密集网络环境中执行该动作,之后环境会返回一个奖励  $r$  和下一个状态  $s'$ ,此时元组  $(s, a, r, s')$  就作为一条经验存入经验池,重复这个过程直到经验条数满足训练要求

为止。当经验池存满之后,利用优先经验回放机制,从经验池中随机采样,将采样得到的奖励  $r$  和下一个状态  $s'$  送入目标网络,计算优化目标  $y$ ,并将  $y$  输入主网络计算损失值,从而更新主网络。之后用户终端再与环境交互,产生经验放入经验池,再从经验池中采样更新主网络,直到神经网络完全收敛。此外,训练过程中,每隔  $G$  步更新目标网络参数  $\theta^- = \theta$ 。

神经网络训练完成之后,进入在线决策阶段,用户终端可以根据当前业务类型以及超密集网络环境状态选择长期收益最大的网络接入,由于算法设计时考虑了网络、用户以及业务 3 个层面的因素,并且构建了负载均衡策略,可以实现用户 QoS 的提升以

及负载均衡,降低用户阻塞率。

### 3 仿真实验及分析

为了验证本文提出的改进 DQN 算法的性能,将其与 DQN、Q 学习以及 SAW 算法进行比较分析。首先对比改进 DQN 和传统 DQN 的 Q 值变化曲线和累积回报的收敛性,然后将这 4 种算法在相同条件下分别进行 100 轮仿真实验,并对这 4 种算法在平均用户接入阻塞率、系统吞吐量以及处理时延方面进行对比分析。

#### 3.1 仿真参数设置

仿真场景中考虑由一个家庭基站 ( femtocell base station, FeBS )、两个微微基站 ( picocell base station, PeBS ) 和两个 WIFI 热点组成的超密集异构网络。FeBS 覆盖半径为 200 m,带宽为 15 MHz,有 75 个资源块可以分配;PeBS 覆盖半径为 120 m,带宽为 5 MHz,有 30 个资源块可以分配,两种基站均采用 64QAM 调制,编码速率为 0.93。每个 FeBS 用户可以分配 5 个资源块,每个 PeBS 用户可以分配 3 个资源块,每个资源块由 12 个连续子载波组成。WIFI 热点覆盖半径为 70 m,且遵循 802.11n 协议,最大数据速率为 16 Mbit/s 和 12 Mbit/s,可以同时为 5 个用户提供服务。同时,用户的业务类型考虑会话类、流媒体类、背景类和交互类。在任意决策周期  $\tau$  内,每个用户只选择一种网络和一种业务类型进行通信。考虑到实际场景中用户的移动性,设置每个用户在时隙  $\tau$  内保持连接不变,并且随机产生移动方向,移动速度为 3 m/s。候选网络的参数设置见表 1。

| 表 1 候选网络及其参数                             |                                  |             |               |          |           |
|------------------------------------------|----------------------------------|-------------|---------------|----------|-----------|
| Tab. 1 Candidate networks and parameters |                                  |             |               |          |           |
| 候选网络                                     | 可用带宽/<br>(Mbit·s <sup>-1</sup> ) | 时延抖动/<br>ms | 丢包率/%         | 价格/<br>元 | 安全<br>性/% |
| FeBS                                     | 0 ~ 65                           | 3 ~ 10      | 0.002 ~ 0.006 | 0.9      | 95        |
| PeBS1                                    | 0 ~ 20                           | 10 ~ 15     | 0.002 ~ 0.006 | 0.4      | 85        |
| PeBS2                                    | 0 ~ 20                           | 10 ~ 15     | 0.002 ~ 0.006 | 0.4      | 85        |
| WIFI1                                    | 0 ~ 16                           | 15 ~ 25     | 0.004 ~ 0.008 | 0.2      | 70        |
| WIFI2                                    | 0 ~ 12                           | 15 ~ 25     | 0.004 ~ 0.008 | 0.2      | 70        |

#### 3.2 算法收敛性对比分析

首先将改进后的 DQN 算法与传统 DQN 算法在 Q 值估计方面进行比较,通过在 Q 值变化趋势上的对比,表明所提改进 DQN 算法在离线训练过程中的优势。图 3 为两种算法下 Q 值的变化趋势对比。由图 3 可以看出,在相同迭代次数下,传统 DQN 算

法估计 Q 值起点较高,经过改进后的 DQN 算法由于对目标函数进行了优化,改善了 Q 值的高估问题。

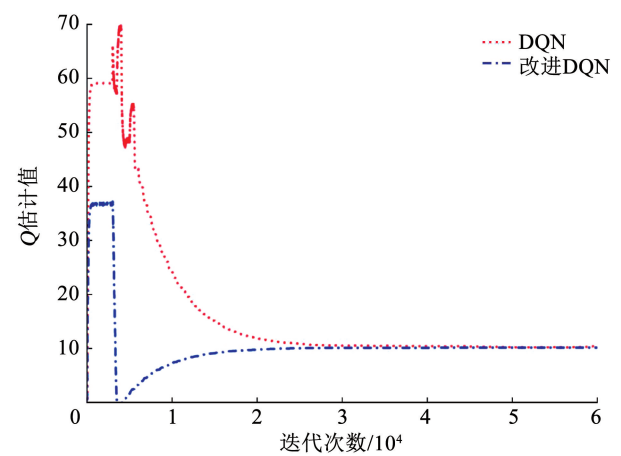


图 3 Q 估计值变化趋势对比

其次将改进 DQN 算法与传统 DQN 算法在累积回报方面进行比较,通过累积回报变化趋势对比,表明所提改进 DQN 算法在离线训练过程中的优势。图 4 为两种算法下累积回报的变化趋势。由图 4 可以看出,在相同迭代次数下,传统 DQN 算法获得的累积回报低于改进 DQN 算法,并且收敛速度较慢,这是因为改进 DQN 算法引入了优先经验回放机制,算法收敛加快。

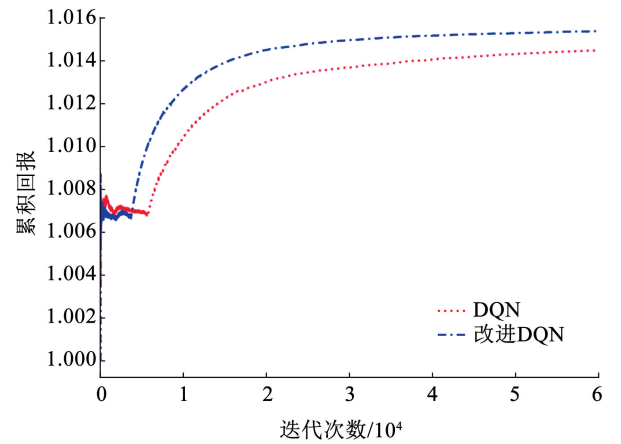


图 4 累积回报变化趋势对比

Fig. 4 Comparison of variation trend of cumulative return

#### 3.3 网络性能指标对比分析

1) 用户接入阻塞率对比分析。图 5 是在业务到达率为 0 ~ 21 时,改进 DQN、DQN、Q 学习和 SAW 这 4 种算法的接入阻塞率对比。从图中可以看出,当用户到达率小于 6 时,4 条曲线的差异不明显,因为到达率较低时,网络中的用户数量少,资源足够所有用户使用。当到达率大于 6 时,改进 DQN 算法的接入阻塞率平均比 DQN、Q 学习和 SAW 分别降低

了 5.4%、23.4%、15.8%，这是因为 SAW 算法只根据网络的当前状态进行接入决策，没有考虑到网络的长期收益。

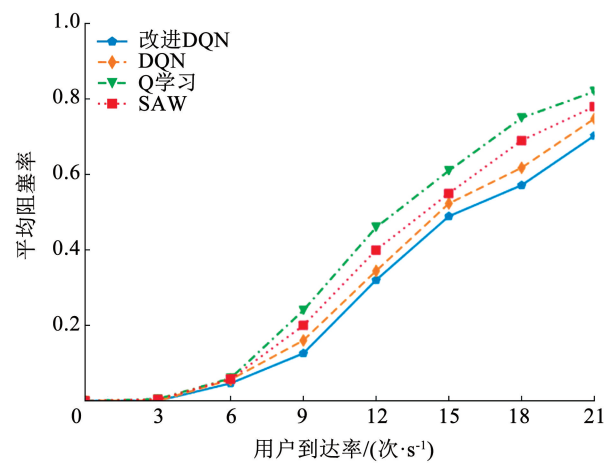


图 5 用户接入阻塞率对比

Fig. 5 Comparison of user access blocking rate

2) 系统吞吐量对比分析。图 6 为在业务到达率为 0~21 时,改进 DQN、DQN、Q 学习和 SAW 这 4 种算法的系统吞吐量对比。从图中可以看出,改进 DQN 算法下系统的吞吐量平均比 DQN、Q 学习和 SAW 分别提高了 10.4%、47.1%、26.2%，这是因为改进 DQN 算法具有自学习能力,能够通过自主学习自动调整接入策略,适应网络的动态变化。

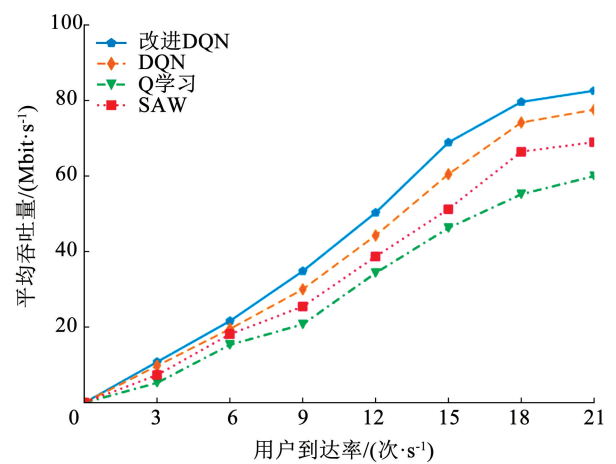


图 6 系统吞吐量对比

Fig. 6 Comparison of system throughput

3) 平均处理时延对比分析。图 7 为改进 DQN、DQN、Q 学习和 SAW 这 4 种算法的平均处理时延对比。此处处理时延表示仿真实验过程中用户完成一次选网动作程序运行的时间。从图中可以看出,改进 DQN 对比 Q 学习处理时延显著降低了 54.9%，这是因为 Q 学习的计算复杂度会随着状态空间的增加而急剧增大。DQN 算法由于抛弃了 Q 学习的查表方式选择用神经网络逼近 Q 函数,相对于 Q 学

习来说,DQN 的计算复杂度受到状态空间维度的影响并不大,改进后的 DQN 在 DQN 基础上引入了优先经验回放机制,使算法的收敛性得到了一定的提高,降低了 14.8% 的处理时延。SAW 算法只是对参数进行简单的加权,相比于 SAW 算法,改进 DQN 算法为了使用户 QoS 和网络性能得到提高,牺牲了一定的处理时延。

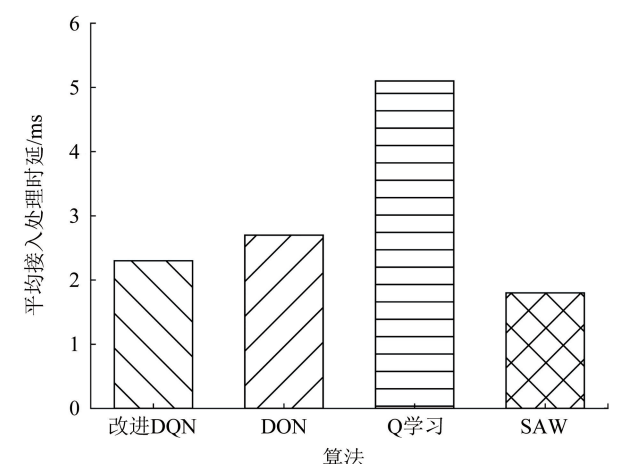


图 7 平均处理时延对比

Fig. 7 Comparison of average delay

4 结 论

提出了一种基于改进 DQN 的超密集网络接入选择算法,该算法既能通过修改目标函数解决 DQN 算法高估问题,引入优先经验回放机制又能够加快传统 DQN 算法的收敛速度。仿真实验证明,相较于传统 DQN、Q 学习、SAW 这 3 种算法,所提算法能够有效地提升系统吞吐量,同时确保用户 QoS,降低接入阻塞率,并且在一定程度上降低时间复杂度。在之后的研究中,将考虑同一环境中其他用户的网络选择行为对当前用户选网的影响,设法解决多用户同时竞争网络资源时的选网问题。

参考文献

[1] ARBI A, O'FARRELL T. Energy efficiency in 5G access networks: small cell densification and high order sectorisation[C]//2015 IEEE International Conference on Communication Workshop. London: IEEE, 2015: 2806. DOI: 10.1109/ICCWW.2015.7247604

[2] LIU Junyu, SHENG Min, LI Jiandong. Improving network capacity scaling law in ultra-dense small cell networks[J]. IEEE Transactions on Wireless Communications, 2018, 17(9): 6218. DOI: 10.1109/TWC.2018.2856766

[3] DAI Linglong, WANG Bichai, YUAN Yifei, et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends[J]. IEEE Communications Magazine, 2015, 53(9): 75. DOI: 10.1109/MCOM.2015.7263349

[4] CHANG C, HSEH C, CHEN Y. A preference value-based cell selection scheme in heterogeneous wireless networks [C]//2010 IEEE Wireless Communication and Networking Conference. Sydney:

IEEE, 2010; 2. DOI: 10.1109/WCNC.2010.5506518

[5] BORST S, KAYA A Ö, CALIN D, et al. Dynamic path selection in 5G multi-RAT wireless networks[C]//IEEE INFOCOM 2017-IEEE Conference on Computer Communications. Atlanta; IEEE, 2017; 3. DOI: 10.1109/INFOCOM.2017.8057228

[6] KAN Shanlei, CHEN Hongyuan, ZHU Yuanping, et al. Aggregation based cell selection methods for multi-RAT HetNet[C]//2016 8th International Conference on Wireless Communications & Signal Processing. Yangzhou; IEEE, 2016; 2. DOI: 10.1109/WCSP.2016.7752572

[7] NGUYEN D D, NGUYEN H X, WHITE L B. Performance of adaptive RAT selection algorithms in 5G heterogeneous wireless networks[C]//2016 26th International Telecommunication Networks and Applications Conference. Dunedin; IEEE, 2016; 72. DOI: 10.1109/ATNAC.2016.7878785

[8] YU Hewei, ZHANG Biao. A heterogeneous network selection algorithm based on network attribute and user preference[J]. Ad Hoc Networks, 2018, 72; 75. DOI: 10.1016/j.adhoc.2018.01.011

[9] YU Hewei, ZHANG Biao. A hybrid MADM algorithm based on attribute weight and utility value for heterogeneous network selection[J]. Journal of Network and Systems Management, 2019, 27(3); 765. DOI: 10.1007/s10922-018-9483-y

[10] TANG Liangrui, JI Shiyu, YAN Jiangyu. A heterogeneous network access selection algorithm based on attribute dependence[J]. Wireless Personal Communications, 2017, 92(3); 1165. DOI: 10.1007/s11277-016-3600-6

[11] ZHU Anqi, MA Mingfang, GUO Songtao, et al. Adaptive multi-access algorithm for multi-service edge users in 5G ultra-dense heterogeneous networks[J]. IEEE Transactions on Vehicular Technology, 2021, 70(3); 2810. DOI: 10.1109/TVT.2021.3060573

[12] CHEN Jialing, YIN Mingxi, DUAN Xiaohui, et al. Q-learning based selection strategies for load balance and energy balance in heterogeneous networks[C]//2020 5th International Conference on Computer and Communication Systems. Shanghai; IEEE, 2020; 729. DOI: 10.1109/ICCCS49078.2020.9118518

[13] KHODMI A, REJEB S B, NASSER N, et al. MDP-based handover in heterogeneous ultra-dense networks[C]//2021 International Conference on Information Networking. Jeju Island; IEEE, 2021; 350. DOI: 10.1109/ICIN50884.2021.9334024

[14] 王月平, 徐涛. 基于演化博弈的用户接入机制[J]. 计算机应用, 2020, 40(5); 1394

WANG Yueping, XU Tao. User access mechanism based on evolutionary game[J]. Journal of Computer Applications, 2020, 40(5); 1394. DOI: 10.11772/j.issn.1001-9081.2019112024

[15] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3); 280. DOI: 10.1007/BF00992698

[16] VANHASSELT H. Double Q-learning[J]. Advances in Neural Information Processing Systems, 2010, 23(2); 2615

[17] SCHAUL T, QUAN J, ANTONOGLOU I, et al. Prioritized experience replay[C]//Proceedings of the 4th International Conference on Learning Representations. San Juan; [s. n.], 2016; 325. DOI: 10.48550/arXiv.1511.05952

[18] 俞鹤伟, 梁根. 异构无线网络接入选择算法综述[J]. 哈尔滨工业大学学报, 2017, 49(11); 179

YU Hewei, LIANG Gen. Overview of access selection algorithms for heterogeneous wireless networks[J]. Journal of Harbin Institute of Technology, 2017, 49(11); 179. DOI: 10.11918/j.issn.0367-6234.201605125

(编辑 苗秀芝)

(上接第 77 页)

[9] SOMASUNDARAM S D. Linearly constrained robust Capon beamforming[J]. IEEE Transactions on Signal Processing, 2012, 60(11); 5845. DOI: 10.1109/TSP.2012.2212889

[10] NAI S E, SER W, YU Z L, et al. Iterative robust minimum variance beamforming[J]. IEEE Transactions on Signal Processing, 2011, 59(4); 1601. DOI: 10.1109/TSP.2010.2096222

[11] 黄磊. 非理想条件下的自适应波束形成算法研究[D]. 合肥: 中国科学技术大学, 2016

HUANG Lei. Adaptive beamforming algorithms under nonideal conditions[D]. Hefei: University of Science and Technology of China, 2016

[12] MALLIPEDDI R, LIE J P, RAZUL S G, et al. Robust adaptive beamforming based on covariance matrix reconstruction for look direction mismatch[J]. Progress In Electromagnetics Research Letters, 2011, 25(1); 37. DOI: 10.2528/PIERL11040104

[13] GU Yujie, LESHEM A. Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation[J]. IEEE Transactions on Signal Processing, 2012, 60(7); 3881. DOI: 10.1109/TSP.2012.2194289

[14] SHEN Feng, CHEN Fengfeng, SONG Jinyang. Robust adaptive beamforming based on steering vector estimation and covariance matrix reconstruction[J]. IEEE Communications Letters, 2015, 19(9); 1636. DOI: 10.1109/LCOMM.2015.2455503

[15] GU Yujie, GOODMAN N A, HONG Shaohua, et al. Robust adaptive beamforming based on interference covariance matrix sparse reconstruction[J]. Signal Processing, 2014, 96; 376

[16] HUANG Lei, ZHANG Jing, XU Xu, et al. Robust adaptive beamforming with a novel interference-plus-noise covariance matrix reconstruction method[J]. IEEE Transactions on Signal Processing, 2015, 63(7); 1643. DOI: 10.1109/TSP.2015.2396002

[17] YUAN Xiaolei, GAN Lu. Robust adaptive beamforming via a novel subspace method for interference covariance matrix reconstruction[J]. Signal Processing, 2017, 130(1); 234. DOI: 10.1016/j.sigpro.2016.07.008

[18] CHEN Peng, YANG Yixin, WANG Yong, et al. Robust adaptive beamforming with sensor position errors using weighted subspace fitting-based covariance matrix reconstruction[J]. Sensors, 2018, 18(5); 1476. DOI: 10.3390/s18051476

[19] YANG Long, YANG Yixin, YANG Jie. Robust adaptive beamforming for uniform linear arrays with sensor gain and phase uncertainties[J]. IEEE Access, 2019, 7(1); 2678. DOI: 10.1109/ACCESS.2018.2886405

[20] CHEN Peng, YANG Yixin, WANG Yong, et al. Adaptive beamforming with sensor position errors using covariance matrix construction based on subspace bases transition[J]. IEEE Signal Processing Letters, 2019, 26(1); 19. DOI: 10.1109/LSP.2018.2878948

[21] ARAKI S, MAKINO S, HINAMOTO Y, et al. Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures[J]. EURASIP Journal on Applied Signal Processing, 2003, 2003(11); 1162. DOI: 10.1155/s1110865703305074

(编辑 苗秀芝)