

基于改进深度Q学习的网络选择算法

马彬 陈海波* 张超

(重庆邮电大学重庆市计算机网络与通信技术重点实验室 重庆 400065)

(重庆邮电大学计算机科学与技术学院 重庆 400065)

摘要: 在引入休眠机制的超密集异构无线网络中, 针对网络动态性增强, 导致切换性能下降的问题, 该文提出一种基于改进深度Q学习的网络选择算法。首先, 根据网络的动态性分析, 构建深度Q学习选网模型; 其次, 将深度Q学习选网模型中线下训练模块的训练样本与权值, 通过迁移学习, 将其迁移到线上决策模块中; 最后, 利用迁移的训练样本及权值加速训练神经网络, 得到最佳选网策略。实验结果表明, 该文算法显著改善了因休眠机制导致的高动态性网络切换性能下降问题, 同时降低了传统深度Q学习算法在线选网过程中的时间复杂度。

关键词: 超密集异构无线网络; 改进深度Q学习; 网络选择

中图分类号: TN915

文献标识码: A

文章编号: 1009-5896(2022)01-0346-08

DOI: [10.11999/JEIT200930](https://doi.org/10.11999/JEIT200930)

Network Selection Algorithm Based on Improved Deep Q-Learning

MA Bin CHEN Haibo ZHANG Chao

(Chongqing Key Laboratory of Computer Network and Communication Technology,
Chongqing University of Post and Telecommunications, Chongqing 400065, China)

(Institute of Computer Science and Technology, Chongqing University of Post and
Telecommunications, Chongqing 400065, China)

Abstract: In ultra dense heterogeneous wireless network with sleep mechanism, in view of the problem that the network dynamic is enhanced and the handoff performance is reduced, a network selection algorithm based on improved deep Q-learning is proposed. Firstly, according to the dynamic analysis of the network, a deep Q-learning network selection model is constructed; Secondly, the training samples and weights of the offline training module in deep Q-learning network selection model, which are transferred to the online network decision-making module through the transfer learning; Finally, the training samples and weights of transfer are used to accelerate the process of training neural network, and the optimal network selection strategy is obtained. Experimental results demonstrate that the proposed algorithm improves significantly the performance degradation of high dynamic network handoff caused by sleep mechanism and the time complexity of traditional deep Q-learning algorithm for online network selection.

Key words: Ultra dense heterogeneous wireless network; Improved deep Q-learning; Network selection

收稿日期: 2020-10-30; 改回日期: 2021-05-26; 网络出版: 2021-08-24

*通信作者: 陈海波 860452738@qq.com

基金项目: 重庆市教委科学技术研究重大项目(KJZD-M201900602), 重庆市教委科学技术研究重点项目(KJZD-M201800603), 重庆市基础研究与前沿探索项目(CSTC2018jcyjAX0432), 重庆市研究生科研创新项目(CYS20256)

Foundation Items: The Major Project of Science and Technology Research of Chongqing Education Commission (KJZD-M201900602), The Key Project of Science and Technology Research of Chongqing Education Commission (KJZD-M201800603), The Foundation Research and Advanced Exploration Project of Chongqing (CSTC2018jcyjAX0432), The Project of Science Research Innovation of Chongqing Graduate Students (CYS20256)

1 引言

随着无线移动通信的发展，由5G异构蜂窝网络、无线局域网等多种接入技术构成的超密集异构无线网络，可为终端提供多种接入方式，支持终端的无缝移动。超密集组网会带来较高的能耗问题，引入休眠机制会一定程度降低能耗，同时，会进一步增强网络的动态性，导致终端服务质量和网络吞吐量等性能均下降。如何在高动态的超密集异构无线网络，保证终端所获得的吞吐量，并提升网络系统综合切换性能，成为当前研究需要解决的重要课题^[1]。

目前，国内外相关研究工作较多，从采取的研究方法来划分，网络选择算法大致可以分成两类：(1)基于参数阈值的网络选择算法^[2-5]，主要是以接收信号强度(Receive Signal Strength, RSS)等参数进行网络选择判决。文献[4]基于RSS的阈值，通过比较候选网络的RSS数值大小，从而进行网络选择。该类算法实现简单，选网参数易于获取，计算复杂度较低；但是，该类算法容易导致乒乓效应，无法完全体现接入网络的服务质量。(2)基于模糊逻辑或强化学习等人工智能方法的网络选择算法^[6-14]。文献[6]采用模糊逻辑算法，根据终端应用对QoS参数的需求，通过设计不同的隶属度函数，合理地选择网络。该类算法选网效率较高，但是需要事先建立相应的模糊推理规则库，在输入参数增加的情况下，模糊规则库的数量会激增，导致推理时间复杂度过大。文献[8]基于神经网络算法，根据不同业务类型，对分类后的参数进行训练，从而进行网络选择。该类算法具有强大的学习能力，能够根据环境自适应地进行调整。文献[11]提出一种基于体验质量(Quality of Experience, QoE)感知的网络选择方案，将QoS的网络参数映射成QoE参数，然后利用QoE参数构造回报函数，最后采用Q学习算法进行网络选择。该类算法能够通过不断的学习强化已有收益，从而选择高收益网络；但是，如果网络环境过于复杂，会导致网络控制模块学习效果下降，继而导致无法选择到最佳网络。

对于引入休眠机制的超密集异构无线网络环境，随着该网络环境下基站数量的激增，同时引入休眠机制以节能，导致基站数量发生动态改变，继而导致网络动态性增强，网络拓扑结构时变性提高。同时，终端自身的移动性和基站之间由于密集布网所产生的干扰，导致网络动态性进一步加剧，而现有的相关文献，解决的是常规异构无线网络下的网络选择问题，并未考虑到如此高动态的网络环境，这样会使终端在通过现有选网算法切换到目标

网络之后，可能因目标网络突然休眠，导致所获得的吞吐量出现快速下滑，无法为终端提供持续稳定的吞吐量，基站之间产生的干扰又会严重影响到终端的服务满意度，最终发生系统切换性能严重降低的问题。因此，上述算法对于这一严峻问题，关注不足。

通过上述对引入休眠机制的超密集异构无线网络环境进行动态性分析，可以得出，该网络环境具有高动态特性，会使现有的网络选择算法出现切换性能严重下降的问题。因此，本文基于一种改进深度Q学习算法，以保证终端所获得的网络吞吐量，缓解系统因高动态性网络环境导致的切换性能严重降低的问题。同时，针对传统的深度Q学习算法在进行网络选择时，由于在线训练神经网络导致时延过大，出现算法失效的情况，本文利用迁移学习，加速训练神经网络，以降低在线选网的时间复杂度。综上，本文的主要贡献可概括为：

(1)针对由无线局域网和引入休眠机制的超密集蜂窝网络异构而成的超密集异构无线网络环境，进行动态性分析，以期缓解系统切换性能降低的问题。

(2)本文采用迁移学习对深度Q学习算法进行改良，提出一种基于改进深度Q学习的网络选择算法，降低了传统深度Q学习算法在线上选网过程中的时间复杂度。

2 算法流程

本文算法的流程图如图1所示，主要包括参数采样及初始化过程、深度Q学习选网模型、最优策略及网络选择3个阶段。第1阶段通过周期性采样网络参数的值来初始化深度Q学习选网模型，该模型由线下训练模块与线上决策模块构成，上述两个模块均采用深度Q网络构建；第2阶段利用迁移学习对线下训练模块和线上决策模块进行协同交互；第3阶段通过深度Q学习选网模型得到最优策略并进行网络选择。图1的历史信息数据库包含网络参数的采样值以及历史选网信息数据，作为深度Q学习选网模型的训练数据。

3 参数采样及初始化过程

3.1 参数采样

由于本文采用深度Q学习算法对网络选择行为进行建模，因此，在通过深度Q学习算法进行建模的过程中，需要周期性采样网络参数的值来构建深度Q学习的动作空间、状态空间和回报函数，以初始化深度Q学习选网模型，本文采样的参数如下。

(1)接收信号强度表示为

$$RSS = \rho_{ij} - \eta \lg(d_{ij}) + u_1 \quad (1)$$

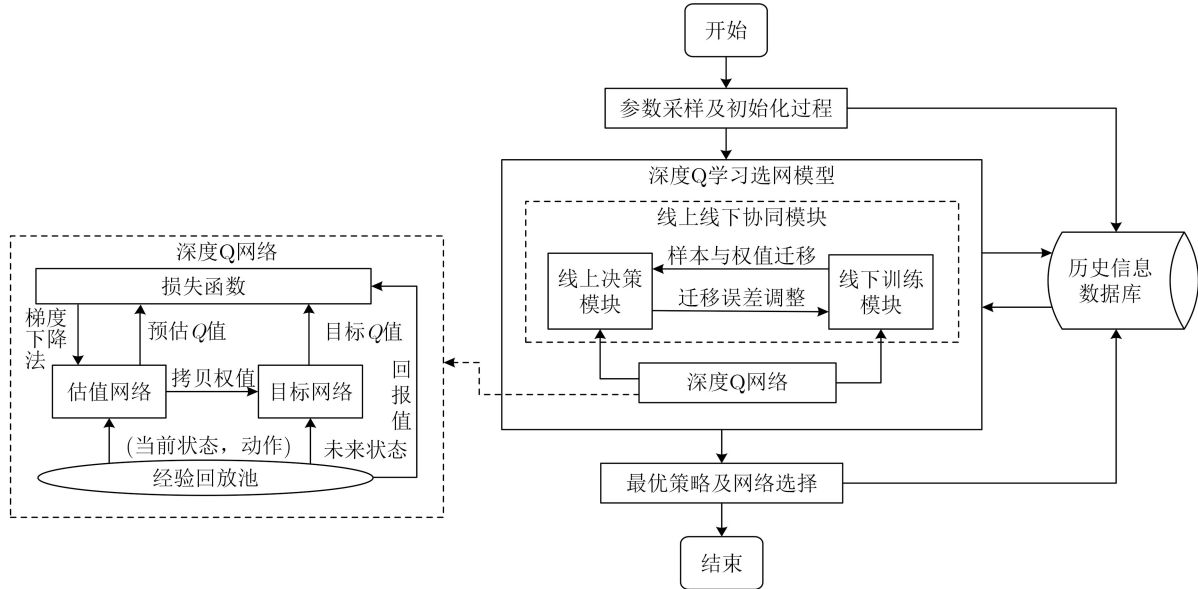


图1 本文算法流程图

其中, d_{ij} 为用户终端 i 到网络接入点 j 的距离, ρ_{ij} 为用户终端 i 接收到网络接入点 j 的发射功率, η 为路径损耗因子, u_1 为服从均值为0、标准差为 σ_1 的高斯白噪声。

(2)终端所获下行吞吐量可表示为

$$C = W \log_2(1 + \text{SINR}) \quad (2)$$

其中, W 为网络的可用带宽, SINR 为用户终端的信干噪比。

(3)根据文献[15]建立的干扰模型, 干扰影响因子可表示为

$$\psi = E \left[\exp \left(I \cdot \left(-\frac{\beta \cdot r_j^\alpha}{\rho_{ij}} \right) \right) \right] \quad (3)$$

其中, I 为干扰信号强度, β 为用户终端所能接受的最低信干噪比阈值, 本文设置为1 dB, r_j 为网络接入点 j 的半径, α 为路径衰落因子, $E(\cdot)$ 为期望函数。

(4)根据泊松点过程模型的特征, 休眠概率可表示为

$$P = P(\lambda_{\text{now}} < \lambda_{\text{min}}) = 1 - \sum_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} \frac{e^{-\lambda_{\text{now}}} \lambda_{\text{now}}^{\lambda_{\text{min}}}}{\lambda_{\text{min}}!} \quad (4)$$

其中, λ_{now} 为基站当前所接入的终端数, λ_{min} 为基站处于活跃状态的最低接入终端数, λ_{max} 为基站最大可容纳终端数, 当 $\lambda_{\text{now}} < \lambda_{\text{min}}$ 时, 基站进行休眠操作。

3.2 初始化过程

本文将超密集异构无线网络环境中终端可以接入的候选网络(基站和访问点)用集合 $N = \{n_1, n_2, \dots, n_i\}$ 表示; 其中终端在 t 时刻接入候选网络 n_i 表示为 $a_t(n_i)$, 则本文的动作空间可定义为 $A_t = \{a_t, a_t \in \{a_t(n_1), a_t(n_2), \dots, a_t(n_i)\}\}$ 。

本文将状态空间定义为 $S_t = (\text{rss}_t, c_t, \psi_t, p_t)$, 其中, rss_t 表示在 t 时刻各候选网络的接收信号强度集合, c_t 表示在 t 时刻各候选网络的吞吐量集合, ψ_t 表示在 t 时刻各候选网络的干扰影响因子集合, p_t 表示在 t 时刻各候选网络的休眠概率集合。

本文为最大化终端所获得的吞吐量, 通过考虑网络的吞吐量、干扰影响因子和休眠概率, 结合文献[16]中回报函数的构造方式, 将回报函数定义为

$$R(S_t, a_t) = \begin{cases} \frac{1 - e^{-C_t(n_i)}}{P_t(n_i)\psi_t(n_i)}, & a_t = a_t(n_i) \\ -1, & \text{其他} \end{cases} \quad (5)$$

其中, $C_t(n_i)$ 表示在 t 时刻终端接入候选网络 n_i 获得的吞吐量, $\psi_t(n_i)$ 表示在 t 时刻终端接入候选网络 n_i 后受到的干扰影响因子, $P_t(n_i)$ 表示在 t 时刻候选网络 n_i 的休眠概率。

4 深度Q学习选网模型

4.1 改进深度Q学习算法

本文将引入休眠机制的超密集异构无线网络环境下的选网问题, 基于深度Q学习算法进行建模求解, 该算法过程利用Q函数实现。Q函数表示在状态 S 下执行动作 a , 以及采取后续动作所产生累计回报值的期望, 定义为

$$Q(S, a) = \sum_{t=0}^{\infty} \gamma^t E(R(S_t, a_t)) \quad (6)$$

其中, $\gamma^t \in [0, 1]$ 为折扣因子, 用于调整对未来回报的重视程度, 随着时刻 t 的增加而呈现指数型下降趋势, $E(\cdot)$ 为期望函数。

由式(6)可得, $Q(S, a)$ 在 $t \rightarrow +\infty$ 时才能收敛到

最佳 Q 值，在实际的网络选择过程中很难实现。因此，深度 Q 学习算法利用神经网络构建 $Q(S, a; \theta)$ ，其中， θ 为权值，使得 $Q(S, a; \theta) \approx \max(Q(S, a))$ 来进行近似求解。

对于传统的深度 Q 学习算法，终端需要多次进行交互以获取足够样本，但是，这样会使神经网络的训练时间过长，导致算法因延迟过高而失效。综上，本文引入迁移学习^[17]，提出一种改进深度 Q 学习算法，来解决上述问题，该算法能够减少终端与网络环境实时交互的次数。同时，相比传统深度 Q 学习算法以随机赋权值的方法来训练神经网络，迁移线下训练权值之后，使得训练的起点更加接近神经网络收敛条件，从而加速了算法的收敛性。

4.2 线上线下同模块

4.2.1 样本与权重的生成及迁移

本文所提的深度 Q 学习选网模型由线下训练模块与线上决策模块构成，而上述两个模块均由深度 Q 网络构建。为了加速线上决策模块的神经网络训练过程，通过迁移线下训练模块的训练样本与权值，并对迁移后上述两个模块产生的训练误差进行校正，直到误差恒定，整个迁移学习过程结束。其中，训练样本与权值的生成及迁移过程如下：

神经网络的训练样本是由历史信息数据库中不同时刻的当前状态、动作、返回值以及未来状态所构成的，即 (S_t, a_t, R_t, S_{t+1}) ，其中， $t \in (0, +\infty)$ 。在深度 Q 网络中，为了高效地训练神经网络，通过设置经验回放池，以缓解训练过程中出现的迭代不稳定问题。因此，本文将线下训练模块的训练样本迁移到线上决策模块中，利用迁移的线下训练样本以及线上学习样本，构建线上决策模块的经验回放池，表示为

$$D_{\text{sum}} = D_{\text{on}} + \xi^t D_{\text{off}} \quad (7)$$

其中， D_{sum} 为经验回放池存放的样本总量， D_{on} 为线上学习样本的总量，初始值为0， D_{off} 为线下训练样本的总量， ξ^t 为样本迁移率， $\xi \in [0, 1]$ 为迁移率因子， ξ^t 随着时刻 t 的增加呈现指数型下降的趋势，即有 $t \rightarrow \infty$ ， $\xi^t \rightarrow 0$ 。

在线上决策模块的经验回放池构建完成后，将线下训练得到的神经网络权值 θ_{off} 迁移到线上决策模块中，作为神经网络训练的初始权值，即 $\theta_{\text{on}} = \theta_{\text{off}}$ 。

4.2.2 迁移误差调整

在神经网络进行迭代训练时，由于线下训练模块与线上决策模块之间在训练样本、权值存在差异的情况，可能出现训练样本、权值迁移过后，线上决策模块的神经网络训练效果不佳，从而导致神经网络的收敛速度未能达到预期效果。为了解决上述

问题，本文将线下训练与线上决策模块之间产生的训练误差定义为策略损失，为了将策略损失降至最低，采用策略模仿机制，通过线下训练模块中，预估 Q 值 $Q_{\text{off}}(S_t, a_t; \theta_{\text{off}})$ 的玻尔兹曼分布，将线下训练模块的估值网络转化为线下策略网络 $\pi_{\text{off}}(S_t, a_t; \theta_{\text{off}})$ ，表示为

$$\pi_{\text{off}}(S_t, a_t; \theta_{\text{off}}) = \frac{e^{\frac{1}{T} Q_{\text{off}}(S_t, a_t; \theta_{\text{off}})}}{\sum_{a_{t+1} \in A_{\text{off}}} e^{\frac{1}{T} Q_{\text{off}}(S_t, a_{t+1}; \theta_{\text{off}})}} \quad (8)$$

其中， T 为服从玻尔兹曼分布的参数， A_{off} 为线下训练时的动作空间。同样，利用线上决策模块的预估 Q 值 $Q_{\text{on}}(S_t, a_t; \theta_{\text{on}})$ ，将估值网络转化为线上策略网络 $\pi_{\text{on}}(S_t, a_t; \theta_{\text{on}})$ ，表示为

$$\pi_{\text{on}}(S_t, a_t; \theta_{\text{on}}) = \frac{e^{\frac{1}{T} Q_{\text{on}}(S_t, a_t; \theta_{\text{on}})}}{\sum_{a_{t+1} \in A_{\text{on}}} e^{\frac{1}{T} Q_{\text{on}}(S_t, a_{t+1}; \theta_{\text{on}})}} \quad (9)$$

线下训练与线上决策模块之间的策略损失采用交叉熵衡量，则策略模仿损失函数表示为

$$F(\theta_{\text{off}}) = \sum_{a_t \in A_{\text{off}}} \pi_{\text{off}}(S_t, a_t; \theta_{\text{off}}) \sum_{a_t \in A_{\text{on}}} \lg \pi_{\text{on}}(S_t, a_t; \theta_{\text{on}}) \quad (10)$$

在策略损失存在的情况下，线上决策模块预估 Q 值 $Q_{\text{on}}(S_t, a_t; \theta_{\text{on}})$ 的梯度更新表示为

$$\Delta F(\theta_{\text{off}}) = [Q_{\pi}(S_t, a_t; \theta_{\text{on}}) - Q_{\text{on}}(S_t, a_t; \theta_{\text{on}})] \cdot \frac{\partial Q_{\text{on}}(S_t, a_t; \theta_{\text{on}})}{\partial F(\theta_{\text{off}})} \quad (11)$$

其中， $Q_{\pi}(S_t, a_t; \theta_{\text{on}})$ 表示在策略 π 下预估 Q 值的无偏差估计值。

当 $\Delta F(\theta_{\text{off}}) = 0$ 时，线下训练与线上决策模块之间的策略损失不再改变，即基于策略模仿机制的训练迭代过程实现收敛，则 $Q_{\pi}(S_t, a_t; \theta_{\text{on}}) \approx Q_{\text{on}}(S_t, a_t; \theta_{\text{on}})$ ，此时迁移学习过程结束。

5 最优策略及网络选择

终端在移动过程中，当其将要进入或者离开某个基站时，会出现网络选择决策时刻，此时终端需要进行网络选择，本文为了得到终端所要面临的网络选择决策时刻，根据网络的接收信号强度与终端的移动速度来进行预测，假设终端在基站覆盖范围内的移动模型如图2所示， O 点表示基站所在的位置， M 点为弦 AC 的中点， A 点表示终端在网络选择决策时刻 t_A 所处的位置， B 点表示终端自 A 点移动 Δl 后所处的位置，根据终端当前的运动趋势，预测在 C 点处将出现网络选择决策时刻 t_C ，由图2中 $\triangle OAM$ 和 $\triangle OBM$ 的关系可以得到

$$\left. \begin{aligned} r^2 &= l_{OM}^2 + (\Delta l + l_{BM})^2 \\ l_{OB}^2 &= l_{OM}^2 + l_{BM}^2 \end{aligned} \right\} \quad (12)$$

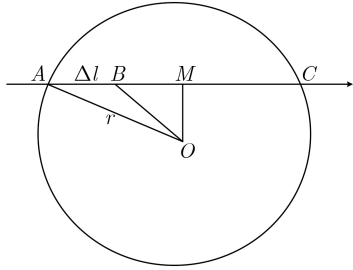


图2 终端移动模型图

其中,

$$l_{BM} = \frac{r^2 - l_{OB}^2 - (\Delta l)^2}{2\Delta l} \quad (13)$$

通过检测B点接收信号强度值, 可得到基站到B点的距离 l_{OB} , 终端在基站覆盖范围内的平均移动速度可表示为 \bar{V} , 则网络选择决策时刻 t_C 表示为

$$t_C = t_A + \frac{2(\Delta l + l_{BM})}{\bar{V}} \quad (14)$$

假设在网络决策时刻 t , 候选网络中最大 Q 值对应的网络为 n_m , 则终端在决策时刻 t 的最佳选网动作作为 $a_t(n_m)$, 以此类推, 终端在不同网络决策时刻所构成的最佳选网动作集合, 定义为最优策略

π^* , 该策略表示终端与候选网络在不同的网络决策时刻实现最佳匹配。

6 仿真结果与分析

6.1 系统模型和仿真参数设置

本文采用5G异构蜂窝网络和无线局域网(Wireless Local Area Network, WLAN)组成超密集异构无线网络环境, 无线接入网络均采用正交频分复用(Orthogonal Frequency Division Multiplexing, OFDM)技术。仿真场景如图3所示, 并在MATLAB平台进行仿真。仿真场景中, 有2个5G宏基站、4个5G微基站和3个WLAN接入点, 5G宏基站的半径均为800 m, 5G微基站的半径均为300 m, WLAN的半径均为80 m。假设用户随机分布在仿真区域内, 每隔一段时间随机改变运动方向。仿真过程中, 假设终端在网络选择决策时刻 k 获得的候选网络参数如表1所示。

仿真对比了本文算法与现有基于Q学习(Q-learning)的网络选择算法^[11]、基于深度Q学习网络(Deep Q-Network, DQN)的网络选择算法^[12]以及基于长短期记忆神经网络(Long Short-Term Memory, LSTM)的网络选择算法^[13]。

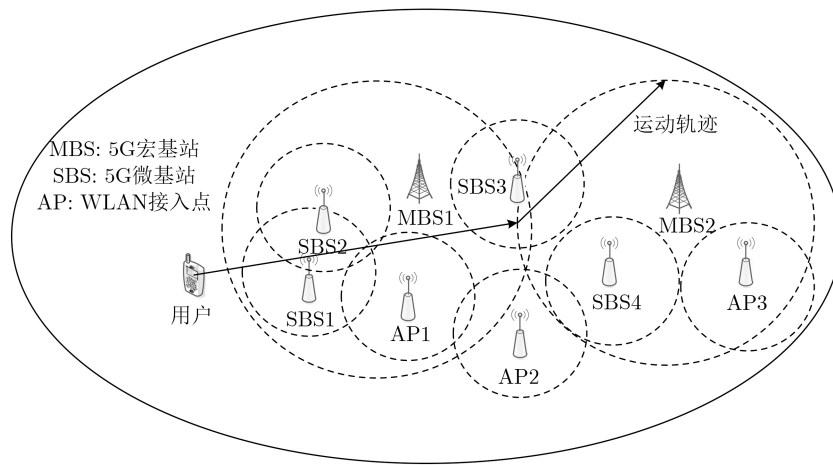


图3 超密集异构无线网络仿真场景图

表1 候选网络的参数值

网络	接收信号强度(dBm)	路径损失(dB)	噪声偏差(dBm)	吞吐量(kbps)	负载量(个)
MBS1	-85	48	6	1100	68
MBS2	-70	51	9	900	52
SBS1	-78	47	8	2700	16
SBS2	-72	53	8	2600	23
SBS3	-86	50	7	2900	25
SBS4	-95	49	6	3100	20
AP1	-60	45	9	4800	12
AP2	-75	43	6	6400	8
AP3	-71	47	7	5500	10

6.2 时间复杂度分析

时间复杂度是网络选择算法的一个重要指标，本文算法与另外3种算法的时间开销对比如图4所示，随着迭代次数的增加，4种算法所消耗的时间都在增加；但是，本文所采用的算法时间增加的幅度明显慢于另外3种算法。这是因为本文算法采用迁移学习对传统深度Q学习算法进行改进，极大地减少了神经网络训练时间，从而使整个算法的时间消耗降低。对于Q学习算法，在状态与动作空间迅速增大的时候，计算能力持续下降，耗时逐渐加大，与本文算法的时间差距逐渐拉开。而DQN算法和LSTM算法，则直接采用深度神经网络进行迭代运算，在迭代次数巨大的情况下，它们与本文算法的时间消耗差距将更为显著。

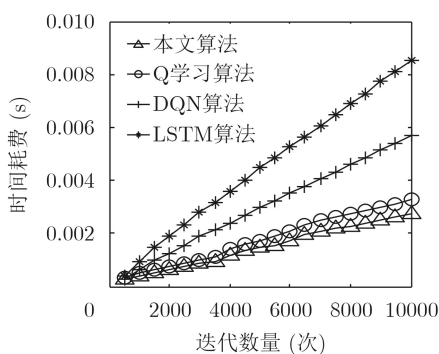


图4 算法时间开销

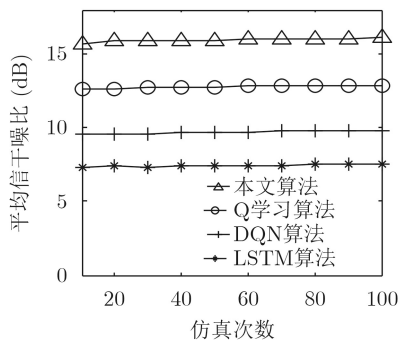


图5 平均信噪比

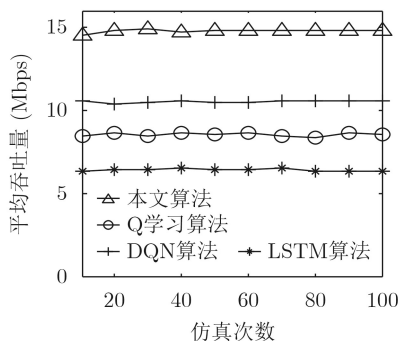


图6 平均吞吐量

6.3 信干噪比分析

图5给出了随着仿真次数的增加，终端在4种算法下受到的平均信干噪比数值情况。通过对比可以看到，采用本文算法所得到的平均信干噪比，高于另外3种算法。这是因为本文考虑到终端所受到的干扰，将干扰因素考虑进来，成功预测了基站未来对终端所造成的干扰影响情况，最大限度地降低了因基站所产生的干扰而对终端造成损失；同时，通过引用干扰影响因子参数来构建深度Q学习算法的回报函数，有效缓解了干扰对终端产生的影响，从而为用户带来更高的信干噪比。对于其他3种算法，由于没有专门考虑终端受到干扰的情况，为用户设计合适的回报函数来降低干扰影响，导致了平均信干噪比没有本文算法高。

6.4 网络吞吐量分析

图6给出了随着仿真次数的增加，终端在4种算法下平均吞吐量变化情况。通过对比可以得出，采用本文算法所得到的平均吞吐量，远高于另外3种算法。这是因为本文采用深度Q学习算法成功预测了基站未来因休眠机制所导致的状态变化情况，使得终端可以根据网络未来发生的动态性变化合理地选择网络，极大地降低了由于基站干扰和休眠情况造成的吞吐量损失；同时，本文根据用户获得的吞吐量定义深度Q学习算法的回报函数，更加符合用户的实际需求。对于3种算法，由于它们既没有充分考虑在未来网络环境下基站的状态，也没有为用户设计合适的回报函数来增加网络吞吐量，从而导致了吞吐量没有本文算法高。

6.5 掉话率分析

图7为4种算法的掉话率与用户数之间的比较。由图可以看出，虽然4种算法的掉话率都在缓慢增加，但是，在用户数增加到40以后，本文算法的掉话率增加幅度最小，另外3种算法增加幅度明显高于本文算法。这是因为本文算法相比较其他3种算法，在网络动态性持续增加的情况下，能够预测到未来网络的变化情况，继而为用户提供较高质量的

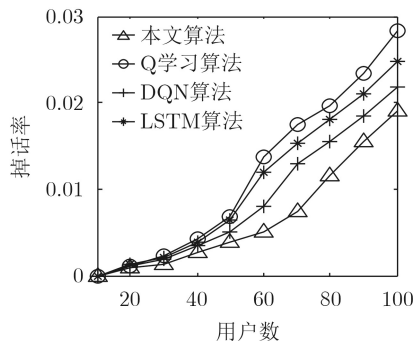


图7 网络掉话率

网络进行选择,有效地降低了切换失败的可能性。对于Q学习算法,由于不能够准确地对网络状态进行预测,在用户数增加的情况下,掉话率急剧增加。同样,对于DQN和LSTM算法,由于在训练深度神经网络的过程中,会造成选网时延较高的结果,使得掉话率上升明显。

6.6 乒乓效应分析

图8为用户采用4种算法后产生的总切换次数。由图可以看出,在用户数不断增加的情况下,采用本文算法产生的网络总切换次数,在所述4种算法中始终处于最低。这是因为本文考虑了因网络环境动态性增强,导致算法切换失效率增加,从而发生频繁切换的情况。本文算法成功预测了用户在进行网络选择之后的网络状态变化情况,从而使发生切换的次数大大降低。而另外3种算法,由于均未妥善解决基站因休眠机制带来的网络高动态性影响,导致网络的切换频发,乒乓效应加剧;因此,本文算法能够有效地降低无谓的网络切换。

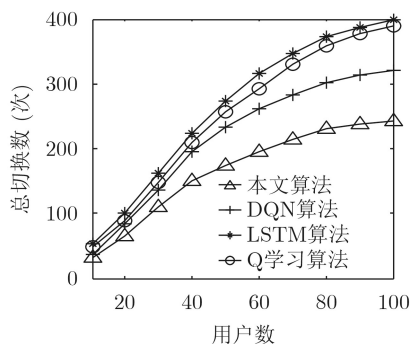


图8 网络总切换次数

7 结束语

本文提出一种基于改进深度Q学习的网络选择算法,缓解了引入休眠机制的超密集异构无线网络中,由于网络动态性的明显提升而引发系统切换性能降低的问题。通过利用网络参数来构建回报函数,从而尽可能地降低引入休眠机制后,网络高动态性所造成的影响;同时,通过迁移学习对传统的深度Q学习算法进行优化,使得算法的时间复杂度大大降低。实验结果表明,该算法在提升网络吞吐量的同时,降低了时间复杂度,减少了网络切换次数。

参考文献

- [1] YAN Xiaohuan, ŞEKERCIOĞLU A, and NARAYANAN S. A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks[J]. *Computer Networks*, 2010, 54(11): 1848–1863. doi: [10.1016/j.comnet.2010.02.006](https://doi.org/10.1016/j.comnet.2010.02.006).
- [2] XIE Shengdong and WU Meng. Adaptive variable threshold vertical handoff algorithm[C]. 2008 International Conference on Neural Networks and Signal Processing, Nanjing, China, 2008: 366–369. doi: [10.1109/ICNNSP.2008.4590373](https://doi.org/10.1109/ICNNSP.2008.4590373).
- [3] LIU Min, LI Zhongcheng, and GUO Xiaobing. An efficient handoff decision algorithm for vertical handoff between WWAN and WLAN[J]. *Journal of Computer Science and Technology*, 2007, 22(1): 114–120. doi: [10.1007/s11390-007-9016-8](https://doi.org/10.1007/s11390-007-9016-8).
- [4] ZAHARAN A H, LIANG Ben, and SALEH A. Signal threshold adaptation for vertical handoff in heterogeneous wireless networks[J]. *Mobile Networks and Applications*, 2006, 11(4): 625–640. doi: [10.1007/s11036-006-7326-7](https://doi.org/10.1007/s11036-006-7326-7).
- [5] HAIDER A, GONDAL I, and KAMRUZZAMAN J. Dynamic dwell timer for hybrid vertical handover in 4G coupled networks[C]. 2011 IEEE 73rd Vehicular Technology Conference (VTC Spring), Budapest, Hungary, 2011: 1–5. doi: [10.1109/VETECS.2011.5956636](https://doi.org/10.1109/VETECS.2011.5956636).
- [6] 马彬, 张文静, 谢显中. 面向终端个性化服务的模糊垂直切换算法[J]. *电子与信息学报*, 2017, 39(6): 1284–1290. doi: [10.11999/JEIT160839](https://doi.org/10.11999/JEIT160839).
- [7] ALSAMHI S H and RAJPUT N S. An intelligent hand-off algorithm to enhance quality of service in high altitude platforms using neural network[J]. *Wireless Personal Communications*, 2015, 82(4): 2059–2073. doi: [10.1007/s11277-015-2333-2](https://doi.org/10.1007/s11277-015-2333-2).
- [8] 马彬, 李尚儒, 谢显中. 异构无线网络中基于人工神经网络的自适应垂直切换算法[J]. *电子与信息学报*, 2019, 41(5): 1210–1216. doi: [10.11999/JEIT180534](https://doi.org/10.11999/JEIT180534).
- [9] MA Bin, LI Shangru, and XIE Xianzhong. An adaptive vertical handover algorithm based on artificial neural network in heterogeneous wireless networks[J]. *Journal of Electronics & Information Technology*, 2019, 41(5): 1210–1216. doi: [10.11999/JEIT180534](https://doi.org/10.11999/JEIT180534).
- [10] NURJAHAN, RAHMAN S, SHARMA T, et al. PSO-NF based vertical handoff decision for ubiquitous heterogeneous wireless network (UHWN)[C]. 2016 International Workshop on Computational Intelligence (IWCI), Dhaka, Bangladesh, 2016: 153–158. doi: [10.1109/IWCI.2016.7860357](https://doi.org/10.1109/IWCI.2016.7860357).
- [11] YANG Bingtao, WANG Xue, and QIAN Zhihong. A multi-armed bandit model-based vertical handoff algorithm for heterogeneous wireless networks[J]. *IEEE Communications Letters*, 2018, 22(10): 2116–2119. doi: [10.1109/LCOMM.2018.2861731](https://doi.org/10.1109/LCOMM.2018.2861731).

- [11] CHEN Jiamei, WANG Yao, LI Yufeng, *et al.* QoE-aware intelligent vertical handoff scheme over heterogeneous wireless access networks[J]. *IEEE Access*, 2018, 6: 38285–38293. doi: [10.1109/ACCESS.2018.2853730](https://doi.org/10.1109/ACCESS.2018.2853730).
- [12] HAN Zijun, LEI Tao, LU Zhaoming, *et al.* Artificial intelligence-based handoff management for dense WLANs: A deep reinforcement learning approach[J]. *IEEE Access*, 2019, 7: 31688–31701. doi: [10.1109/access.2019.2900445](https://doi.org/10.1109/access.2019.2900445).
- [13] ALJERI N and BOUKERCHE A. A two-tier machine learning-based handover management scheme for intelligent vehicular networks[J]. *Ad Hoc Networks*, 2019, 94: 101930. doi: [10.1016/j.adhoc.2019.101930](https://doi.org/10.1016/j.adhoc.2019.101930).
- [14] 马彬, 王梦雪, 谢显中. 超密集异构无线网络中基于位置预测的切换算法[J]. 电子与信息学报, 2020, 42(12): 2899–2907. doi: [10.11999/JEIT190751](https://doi.org/10.11999/JEIT190751).
MA Bin, WANG Mengxue, and XIE Xianzhong. Handoff algorithm based on location prediction in ultra-dense heterogeneous wireless network[J]. *Journal of Electronics & Information Technology*, 2020, 42(12): 2899–2907. doi: [10.11999/JEIT190751](https://doi.org/10.11999/JEIT190751).
- [15] ZHANG Xuefei, XIE Yuxuan, CUI Yushan, *et al.* Multi-slot coverage probability and SINR-based handover rate analysis for mobile user in HetNet[J]. *IEEE Access*, 2018, 6: 17868–17879. doi: [10.1109/ACCESS.2018.2821761](https://doi.org/10.1109/ACCESS.2018.2821761).
- [16] SAAD H, MOHAMED A, and ELBATT T. A cooperative Q-learning approach for distributed resource allocation in multi-user femtocell networks[C]. 2014 IEEE Wireless Communications and Networking Conference (WCNC), Istanbul, Turkey, 2014: 1490–1495. doi: [10.1109/WCNC.2014.6952410](https://doi.org/10.1109/WCNC.2014.6952410).
- [17] SUN Yaohua, PENG Mugen, and MAO Shiwen. Deep reinforcement learning-based mode selection and resource management for green fog radio access networks[J]. *IEEE Internet of Things Journal*, 2019, 6(2): 1960–1971. doi: [10.1109/JIOT.2018.2871020](https://doi.org/10.1109/JIOT.2018.2871020).

马彬：男，1978年生，教授，博士生导师，主要研究方向为异构无线网络、认知无线电网络等。

陈海波：男，1994年生，硕士生，研究方向为异构无线网络。

张超：男，1994年生，硕士生，研究方向为异构无线网络。

责任编辑：马秀强