基于深度强化学习的多小区功率分配算法

惠庆琳

(1. 南京邮电大学通信与信息工程学院，江苏 南京 210003;
2. 南京邮电大学江苏省通信与网络技术工程研究中心，江苏 南京 210003)

摘　要: 在OFDM蜂窝网络下行链路中,功率控制与资源调度是决定系统性能的关键,对多小区功率分配和资源分配问题进行研究。首先,对多小区蜂窝网络资源分配和系统容量问题进行建模,控制基站的传输功率。其次,利用深度Q学习和卷积神经网络算法,最大限度地提高整个网络的总容量,提出一种基于深度Q网络(DQN,deep Q-network)的无线资源映射方法和适用于多小区功率分配的深度神经网络。通过仿真分析,与传统Q学习方法相比,提出的DQN可以获得更高的系统容量,并且在收敛速度和稳定性方面有显著提高。

关键词: 深度学习; 功率分配; 神经网络; 强化学习

Deep reinforcement learning based power allocation in multi-cell networks

HUI Qinglin

(1. College of Telecommunication & Information Engineering,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. Jiangsu Engineering Research Center of Communication and Network Technology,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: The OFDM based multi-cell downlink power and resource allocation issues is studied. Firstly, the problem of resource allocation and system capacity are modelled to adjust the transmission power of the base station. Secondly, deep Q learning and convolutional neural network algorithms are proposed to maximize the overall capacity of the entire network. Then a wireless resource mapping method based on deep Q network (DQN, deep Q-network) and a deep neural network for multi-cell power allocation are given. Finally, Simulation results show that DQN can achieve a higher total capacity, compared with the Q-learning method. Besides, DQN has significant improvement in convergence speed and stability.

Key words: Deep Learning; Power Allocation; Neural network; Reinforcement Learning
doi: 10.3969/j.issn.1006-8554.2020.10.003

0　引言

近10年来,随着LTE系统的普及和5G移动通信系统的出现,功率和无线资源分配问题一直是移动通信系统需要考虑的问题。无线通信技术的飞速发展扩大了通信网络规模。超密集网络对微基站进行密集排布,大大增加了系统的总容量和系统的覆盖范围,因此它成为了5G研究中的热门话题之一。但是,由于大量部署基站,特定区域中的用户数量会随着时间而变化,数据需求也会相应地变化。另外,为了满足大量的数据需求,过于高的部署密度也将导致用户间的干扰增加。因此,根据不同的环境信息进行基站功率的动态控制,在干扰管理、能效提升方面都具有十分重要且现实的意义。

与此同时,通信领域利用本身所拥有的大量数据分析和优化网络,这既现实又充满挑战。作为近年来的新兴技术,机器学习对数据预测和数据挖掘具有非常好的效果。

作为机器学习的重要部分,深度学习是一种深层的非线性网络结构,拟合数据和标签之间的非线性关系实现了复杂函数的逼近。DeepMind团队发布的端到端游戏引入了卷积神经网络(CNN)。由于CNN在诸如图像特征处理等大数据处理中具有明显的优势,因此CNN与强化学习的有机结合,即深度强化学习(Deep Reinforcement Learning, DRL)。此后,深度强化学习被广泛应用于各种问题,如AlphaGo、对抗网络架构、机器人控制等。

之后,Abtahi等人提出了将强化学习与深度置信网络结合,取代传统的效用价值函数逼近器,实现了车牌图像的字符分割[8]。廖晓闽等人构建了基于Q学习的误差函数,利用深度神经网络优化蜂窝系统的传输速率[13]。DeepMind团队结合DQN模型与目标Q网络模型,对原始的DQN模型进行改进,进一步提升其用于处理视觉感知任务的能力。白辰甲等人提出基于TD误差自适应校正的DQN主动采样方法,估计样本池中样本的真实优先级,提高了算法的学习速度[10]。

受深度强化学习的成功启发,本文提出了一种基于DQN的多小区网络功率分配方法。针对无线网络的特点,将深度强化学习引入到无线网络的资源分配问题中。首先,考虑到无线网络的巨大状态空间,将发射功率离散化,从系统的收敛性和稳定性2个方面分析离散度的影响。其次,为了降低计算复杂度,提出了一种动态的状态添加策略。最后引入了深度神经网络和目标函数,实现了收敛速度的提高。仿真结果表明,该算法可获得更高的系统容量,并且在收敛速度和稳定性方面有显

著提高。

1 系统模型和问题建模

1.1 系统模型

本文采用下行正交频分复用的蜂窝网络，基站密集分布，同一小区用户不能重复使用同一频率，但其他小区可重复使用。在通信网络架构方面，需要考虑多个基站密集分布，因此，蜂窝网中所有基站对移动用户造成的干扰情况需要综合考虑。

假设有 M 个移动用户随机分布在小区内，N 个基站配备了单天线。1 个中央控制器可以收集整个网络的信息，包括信噪比和传输功率。每个用户将位置信息、干扰和传输速率通过导频信号传给中央控制器，由它制定频谱分配方案。网络模型如图 1 所示。

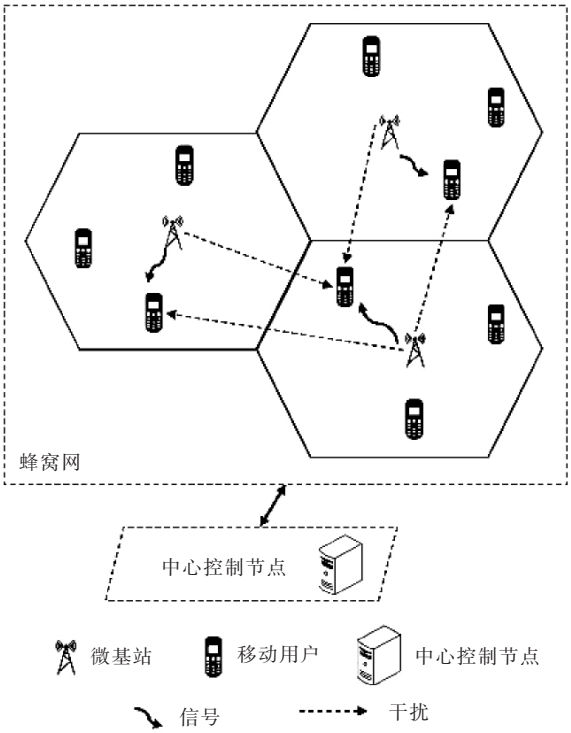

图 1 系统模型

1.2 问题建模

假设 $n=\{1,2,\cdots,N\}$ 表示基站的集合，$m=\{1,2,\cdots,M\}$ 表示移动用户的集合，$k=\{1,2,\cdots,K\}$ 表示可用载波频率的集合，系统工作方式采用随机游走模型。基站的分布遵循泊松点过程模型，复用 k 个正交子载波。每位用户连接基站时，他的发射功率为 p_t^n，在 t 时刻只能连接一个基站。每个子载波也只能分配给一位用户。系统采用六径衰落信道模型进行评估。

t 时刻，在第 k 个子载波上，第 n 个基站服务用户 m，接收的信噪比为

$$\eta_t^{(n,k,m)} = a_t^{(n,m)}\alpha_t^{(n,k,m)}\frac{g_{n,m}^{(k)}p_t^{(n,k)}}{\sum_{n'\neq n}g_{n',m}^{(k)}p_t^{(n',k)}+\sigma^2} \tag{1}$$

其中，$g_{n,m}^{(k)}$ 为信道增益。$p_t^{(n,k)}$ 为总发射功率。σ^2 为高斯白噪声。$\alpha_t^{(n,k,m)}$ 表示基站 n 是否分配载波 k 给用户 m，且 $\alpha_t^{(n,k,m)}\in[0,1]$。$a_t^{(n,m)}$ 表示用户 m 是否连接到基站 n：

$$a_t^{(n,m)}=\begin{cases}1 & m\in M_t\\ 0 & m\notin M_t\end{cases} \tag{2}$$

系统性能是根据测量的总容量(bps/Hz)进行分析。在 t 时刻，基站 BS_n^f 通过载波 k 关联用户获得的容量为：

$$C_t^{(n,k)} = \frac{B}{K}\log_2\left(1+\sum_{m=1}^{M}\eta_t^{(n,k,m)}\right) \tag{3}$$

系统的总容量可以表示为：

$$R_t = \sum_{n=1}^{N}\sum_{k=1}^{K}C_t^{(n,k)} \tag{4}$$

为了在接近最优的子载波分配的基础上，通过调整子载波上基站的发射功率 $p_t^n=[p_t^{(n,1)}\cdots p_t^{(n,k)}\cdots p_t^{(n,K)}]$ 来提高整个网络的总容量。优化问题可表示为：

$$\begin{aligned}&\arg\max R_t\\ &\{\text{s. t. } C1: p_t^{(n,k)}\geqslant p_{\min},\forall n,k\\ &C2:\sum_{n,k}p_t^{(n,k)}\leqslant p_{\max},\forall n,k\end{aligned} \tag{5}$$

其中，$p_{\max}$ 为最大发射功率，$p_{\min}$ 为子载波上所需的最小发射功率。

本文通过调整基站的传输功率，实现整个系统的整体容量的提高。由于考虑到用户的公平性，本文将 1 个基站的下行子载波平均分配给所有承载用户。子载波间的初始功率采用注水算法进行分配。

上述问题是 1 个多目标非凸优化问题，传统的解决方法是启发式搜索算法。但是，这些算法大多非常低效，运行时间长，无法实时在线调整。将介绍一种深度强化学习算法来解决这个问题。

2 基于深度强化学习算法的求解方法

2.1 Q 学习算法

Q 学习参考每个状态的效用值函数，然后通过这些函数得到最优纳什策略。因此利用 Q 学习的方法能不断逼近动作 - 状态值函数。

在每个基站的学习过程中可以由 5 个部分组成$\{P,N,S,R(s,\vec{a})\}$。P 为概率传递函数。$N=\{1,2,\cdots,N_f\}$ 表示为代理(基站)的集合。$S=\{S_1,S_2,\cdots,S_m\}$ 为系统可能占据的状态集，其中 m 为可能的状态数。$R(s,\vec{a})$ 定义了效用值函数，在状态 s 采取行动 a 预计能够得到的累计收益，其中 $s\in S$，$A=\{a_1,a_2,\cdots,a_l\}$ 为可能的行为集合。

在 Q 学习算法中，每个基站通过连续的迭代学习来逼近自己的行为值函数。这个行为值函数一般由 $Q(s_m,a_l)$ 表示，其 $a_l\in A$，$s_m\in S$。因此，Q 值表的大小为 m×1。Q 值表 $Q(s_m,a_l)$ 表示在一段时间 s_m 范围内，行为 a_l 状态的累积回报的期望值：

$$Q_\pi(s,a) = E_\pi\left[\sum_{k=0}^{\infty}R_{t+1}+\gamma Q(S_{t+1},A_{t+1})\mid S_t=s,A_t=a\right] \tag{6}$$

其中，R_{t+1} 表示在状态 s 时即时状态值是在行为 a 之后获得的。γ 是衰减系数，且 $0\leqslant\gamma\leqslant1$。

Q 值表更新如下：

$$\begin{aligned}Q_n^{t+1}(s_m^n,a_l^n) &= (1-\alpha)Q_n^t(s_m^n,a_l^n)\\ &+\alpha(R_n^t+\gamma\max_{a'\in A}Q_n^t(s_{m'}^n,a'))\end{aligned} \tag{7}$$

其中，α 表示学习速度，且 $0\leqslant\alpha\leqslant1$。

2.2 功率分配算法和 Q 学习映射

一种基于功率分配的 Q 学习算法是一种基于环境连续交互的多智能体算法。而在 Q 学习中的状态、行为、代理和状态值函数定义如下。

状态: $s_t^{n,k}=\{M_t^n,p_t^n\}$，其中 M_t^n 表示 t 时刻连接到基站 n 的用户数，p_t^n 表示第 n 个基站在 t 时刻的功率。为了降低算法的复杂度和网络的状态空间，对基站的能效进行离散化处理。离散公式如下。

$$\begin{aligned} p_t^n &= \tau \\ (P_{\max}^f - A_\tau) &\leqslant \sum_{k=0}^{K} p_t^{n,k} < (P_{\max}^f - A_{\tau+1}) \end{aligned} \tag{8}$$

其中，$\tau\in\{0,1,2,3,4,5\}$，$A_0=P_{\max}^f$，$A_6=0$，且 $A_1\sim A_5$ 是任意选择的阈值。

行为: $a_n=\{k_n,\vec{p}_t^{n,k}\}$，其中 k_n 表示第 n 个基站的第 k 个子载波。$\vec{p}_t^{n,k}$ 表示在基站 n 的第 k 个子载波上能效的调整值，且 $\vec{p}_t^{n,k}\in\{-|\vec{p}_t^{n,k}|,0,+|\vec{p}_t^{n,k}|\}$ 表示在采取行动后当前用户与用户 h 之间总吞吐量的变化 C_t^h。如果 C_t^h 增加，$\vec{p}_t^{n,k}=+|\vec{p}_t^{n,k}|$；反之亦然。

代理: 基站 n，$1\leqslant n\leqslant N_f$。

状态值函数: 第 n 个基站的第 k 个子载波的状态值函数定义为:

$$r_t^{n,k}=\begin{cases} 1-e^{-(C_t^{n,k})} & \sum_{k=1}^{K} p_t^{(n,k)} \leqslant P_{\max}^f \\ -1 & \text{Otherwise} \end{cases} \tag{9}$$

由于网络的动态性，网络的状态空间相对较大，各基站的 Q 值表大小也不相同。如果要确定 1 个大小固定的大尺寸 Q 值表，计算复杂度会高很多。因此，提出了一种动态的状态添加方法，即当存在新状态时，此状态将自动添加到状态集。动态方法的优点是不需要为每个代理定制 1 个 Q 值表，这样提高了 Q 值表的搜索效率和存储空间的利用率。但这也意味着 Q 值表需要很长时间才能收敛，因为当出现新状态时，网络需要再次收敛。因此，本文提出了一种 DRL 算法来加快 Q 值表的收敛速度。

2.3　深度 Q 网络(DQN)

本文使用 DQN 来解决资源分配问题。将值网络作为评判模块，遍历当前的各种动作。然后，采用 Q 学习算法训练值网络，得到最大价值的动作作为输出。

DQN 的网络结构为卷积层和全连接层，输出动作对应的概率，结构可以由图 2 表示。

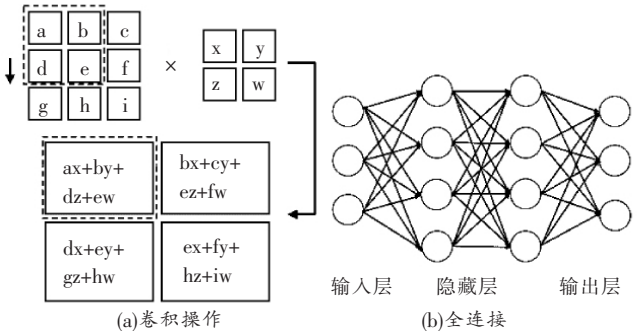

图 2　DQN 的网络结构

算法采用 Q 学习机制，主要根据如式(10)所示的迭代式来实现动作状态值函数的优化学习。

$$\begin{cases} Q_{d+1}(s_t,a_t) = Q_d(s_t,a_t) + \alpha_d E_d \\ E_d = r_t + \gamma \max\limits_{a'\in A} Q_d(s',a') - Q_d(s_t,a_t) \end{cases} \tag{10}$$

其中，α_d 是学习速率，$\gamma\in(0,1)$ 为折扣因子，s'为执行动作后 α_t 获得的观测值，a'为动作集合 A 中第 d 次迭代下的动作状态值函数可执行的动作。从式(10)可以看出，要实现动作状态值函数的逼近，即

$$Q_{d+1}(s_t,a_t)\approx Q_d(s_t,a_t) \tag{11}$$

则　$r_t+\gamma\max\limits_{a'\in A}Q_d(s',a')-Q_d(s_t,a_t)\to 0$

对于每一次迭代 d，最小化目标函数，进行参数更新，如式(12)。

$$\min E=\min\left(r_t+\gamma\max_{a'\in A}Q_d(s',a')-Q_d(s_t,a_t)\right) \tag{12}$$

因此，将公式(12)作为误差函数，采用梯度下降法，得到动作－状态值函数的最优解。

3　仿真结果

功率控制仿真参数设置如表 1 所示，假设每个用户连接基站时的信噪比最大，其中网络环境中有 N_f 个基站，所有基站共享频谱带宽。假设噪声功率 $\sigma^2=10^{-7}$，功率调节每个基站的子载波振幅如表 2 所示，β 更新如表 3 所示。在每个步骤中，由于节点的可移动性，网络更新拓扑一次。

表 1　仿真所需参数与取值

参数	物理意义	取值
B	带宽/MHz	10
BS_S	基站数量/个	32
UE_S	用户数量/个	200
r	半径/m	15
p	初始功率/dbm	0－5
$p_{\max}$	基站的最大发射功率/dbm	30
$p_{\min}$	载波的最小发射功率/dbm	－100
$v_{\max}$	传输速率/($m\cdot s^{-1}$)	1
α	学习速率	0.5
γ	衰减系数	0.9

表 2　载波振幅

参数	取值
A_1	25
A_2	20
A_3	15
A_4	10
A_5	5

表 3　β 更新

迭代步骤	取值
0～30	0.8
30～60	0.6
60～100	0.3
100～	0.1

由图 3 表示，模型收敛后两种模式的频谱效率。对于每次迭代，用户进行移动并更新网络拓扑。从图 3 可以看出，DQN 的频谱效率最高，比 Q 学习算法更稳定。

网络的收敛性如图 4 所示，可知 Q 学习的波动较大。Q 学习在网络拓扑结构发生变化时需重新计算和收敛。但在动态场景中，虽然 DQN 也会出现波动，与 Q 学习相比还是比较稳定

的。同时，随着深度神经网络策略的加强，DQN 大大提高了频谱效率。

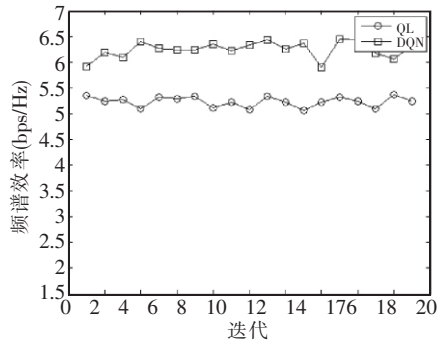

图 3 频谱效率比较

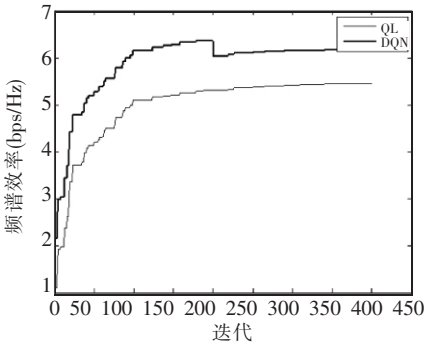

图 4 收敛速度比较

图 5 为功率离散化三层收敛对比图。结果表明，在不同的离散粒度下，整个网络的收敛速度很小。但可以明显看出，当离散粒度为 6 时，整个网络更稳定，频谱效率更高。如果功率离散为 15，则环境状态的数目太大。模型收敛后，系统稳定，波动小。当离散度为 3 时，模型收敛后出现波动很大程度上是因为离散的粒度太粗，不能代表环境的真实变化。功率离散过高或者过低，系统不稳定，频谱效率变低。

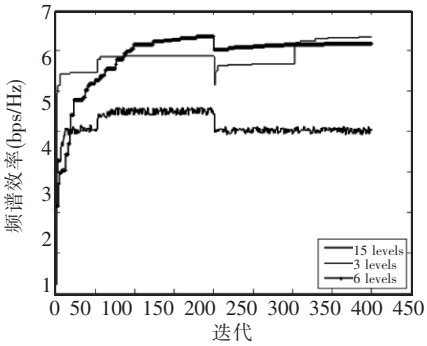

图 5 不同离散粒度的收敛性比较

4 结语

根据深度强化学习和卷积神经网络算法，对资源分配和系统容量的问题建模。为了最大限度地提高整个网络的总容量，提出一种基于深度 Q 网络的无线资源映射方法和用于多小区功率分配的深度神经网络。仿真结果表明，与传统 Q 学习方法相比，DQN 可以实现更高的总容量，获得了较高的系统容量增益。并且，仿真结果还表明 DQN 在收敛速度和稳定性方面有显著提高，对未来的多小区无线网络优化的研究具有指导意义。

参考文献:

[1] Ge X, Tu S, Mao G, *et al*. 5G Ultra-Dense Cellular Networks [J]. IEEE Wireless Communications, 2016, 23(1): 72 - 79.

[2] Hinton G E. Reducing the Dimensionality of Data with Neural Networks [J]. Science, 2006, 313(5786): 504 - 507.

[3] Deng L, Yu D. Deep Learning: Methods and Applications [J]. Foundations and Trends? in Signal Processing, 2013, 7(3).

[4] Silver D, Huang A, Maddison C J, *et al*. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529(7587): 484 - 489.

[5] Wang Z, Schaul T, Hessel M, *et al*. Dueling Network Architectures for Deep Reinforcement Learning [DB/OL]. arXiv: 1511.06581.

[6] Levine S, Finn C, Darrell T, *et al*. End-to-End Training of Deep Visuomotor Policies [J]. Journal of Machine Learning Research, 2015, 17(1): 1334 - 1373.

[7] Sergey Levine, Peter Pastor, Alex Krizhevsky, *et al*. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. 2018, 37(4 - 5): 421 - 436.

[8] Abtahi F, Zhu Z, Burry A M. A deep reinforcement learning approach to character segmentation of license plate images [C]//Iapr International Conference on Machine Vision Applications. IEEE, 2015.

[9] Mnih V, Kavukcuoglu K, Silver D, *et al*. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529 - 533.

[10] 白辰甲，刘鹏，赵巍，等. 基于 TD-error 自适应校正的深度 Q 学习主动采样方法 [J]. 计算机研究与发展，2019，56(2): 38 - 56.

[11] Bai Chenjia, Liu Peng, Zhao Wei, *et al*. Active sampling method for deep Q learning based on TD-error adaptive correction [J]. Journal of Computer Research and Development, 2019, 56(2): 38 - 56.

[12] Jongwon Y, Arslan M Y, Karthikeyan S, *et al*. Characterization of Interference in OFDMA-based Small-cell Networks [J]. IEEE Transactions on Vehicular Technology, 2018.

[13] 廖晓闽，严少虎，石嘉，等. 基于深度强化学习的蜂窝网资源分配算法 [J]. 通信学报，2019，40(2): 15 - 22.

[14] Liao Xiaomin, Yan Shaohu, Shi Jia, *et al*. Cellular network resource allocation algorithm based on deep reinforcement learning [J]. Journal of Communications, 2019, 40(2): 15 - 22.

[15] Shen Y, Zhao N, Xia M, *et al*. A Deep Q-Learning Network for Ship Stowage Planning Problem [J]. Nephron Clinical Practice, 2017, 24(S3): 102 - 109.

作者简介:

惠庆琳(1996 -)，女，江苏无锡人，硕士研究生在读，研究方向：深度学习、强化学习、无线网络优化研究。