第 46 卷 第 12 期 计 算 机 工 程 2020 年 12 月
Vol. 46 No. 12 Computer Engineering December 2020

·移动互联与通信技术· 文章编号：1000-3428(2020)12-0185-08 文献标志码：A 中图分类号：TN929.5

毫米波网络中基于 Q-Learning 的阻塞感知功率分配

施 钊，孙长印，江 帆

（西安邮电大学 通信与信息工程学院，西安 710121）

摘 要： 毫米波通信可在 5G 无线通信系统超密集网络场景中提供显著的系统容量增益，但毫米波通信场景中干扰复杂多变，并且小区边缘用户动态链路的高阻塞率会引起中断问题。为此，基于 Q-Learning 算法，提出一种考虑毫米波链路高间歇性概率的功率分配方案。基于泊松簇过程对随机部署的基站用户系统进行建模，分析链路阻断对有用信号和干扰信号带来的不同影响，并将利己利他策略引入 Q-Learning 算法的状态和回报函数设计中，通过机器学习策略得到功率分配最优解。仿真结果表明，与未考虑链路阻塞概率的 CDP-Q 方案相比，该方案由于根据链路动态链接状况进行最优功率分配，显著提升了系统总容量。

关键词： 毫米波通信；链路阻塞；Q-Learning 算法；功率分配；泊松簇过程；利己利他策略

开放科学（资源服务）标志码（OSID）：

中文引用格式： 施钊，孙长印，江帆. 毫米波网络中基于 Q-Learning 的阻塞感知功率分配[J]. 计算机工程，2020，46(12)：185-192.

英文引用格式： SHI Zhao，SUN Changyin，JIANG Fan. Block-aware power allocation based on Q-Learning in millimeter-wave network[J]. Computer Engineering，2020，46(12)：185-192.

Block-aware Power Allocation Based on Q-Learning in Millimeter-Wave Network

SHI Zhao，SUN Changyin，JIANG Fan

（School of Communication and Information Engineering，Xi'an University of Posts and Telecommunications，Xi'an 710121，China）

【Abstract】 millimeter-Wave（mm-Wave）communication is expected to provide significant capacity gains in ultra-dense network scenarios of the 5G wireless communication system. To address the complex interference in the mm-Wave communication scenario and the interruption caused by the high block rate of the dynamic links of the cell edge users，this paper proposes a power allocation strategy scheme based on Q-Learning algorithm considering the high intermission rate of mm-Wave communication. Poisson Cluster Process（PCP）is used in the modelling of randomly deployed base station user systems，and the different influences of link block on the useful signals and interference signals are analyzed. Then the egoistic and altruistic strategy is introduced in the design of state and reward function of the Q-Learning algorithm，and the machine learning strategy is used to get the optimal solution to power allocation. Simulation results show that，compared with the CDP-Q scheme that does not consider the link block rate，the proposed algorithm significantly improves the total capacity of the system due to the optimal power allocation based on the dynamic status of links.

【Key words】 millimeter-Wave（mm-Wave）communication；link block；Q-Learning algorithm；power allocation；Poisson Cluster Process（PCP）；egoistic and altruistic strategy

DOI： 10.19678/j.issn.1000-3428.0056421

0 概述

在移动通信领域，频谱资源是承载无线业务的基础，是推动产业发展的核心资源。目前低于 6 GHz 的频谱几乎已经被分配殆尽，而 6 GHz 以上的频谱资源非常丰富，由于其业务划分与使用相对简单，能够提供连续的大带宽频带，因此已成为一种具有前景的替代方案[1]。其中，高频段的毫米波（millimeter-Wave，

基金项目： 国家自然科学基金（61801382，61871321）；国家科技重大专项（2017ZX03001012-005）；陕西省自然科学基金重点项目（2019JZ-06）；陕西省重点研发计划“重点产业创新链（群）-工业领域”项目（2019ZDLGY07-06）。

作者简介： 施 钊（1993—），男，硕士研究生，主研方向为 5G 移动通信系统功率控制；孙长印，副教授；江 帆，教授。

收稿日期： 2019-10-28 **修回日期：** 2019-12-19 **E-mail：** m18789419040@163.com

mm-Wave)通信更被认为是一种有效解决无线电频谱稀缺问题的方法,将会为未来无线蜂窝网络提供显著的容量增益[2-4]。

目前,毫米波频段将用于 5G 移动通信网络已成为全球共识。在毫米波频率下可提供的大带宽有可能将网络吞吐量提高 10 倍[3]。但是其面临的信号衰耗大、覆盖距离短和通信链路对阻塞敏感等问题也不可忽视[5]。克服这些问题的一种有效方法是增加接入点的密度[6-7],但是随着接入点密度的增加,当所有毫米波基站重用相同的时频资源时,小区间干扰会越来越严重,网络管理的复杂度也越来越高[8-9],这将极大地限制毫米波小区的系统容量。对于毫米波信号衰耗大、覆盖距离短的问题,可以通过波束成形技术进行最优波束对准,提高系统和速率。对于毫米波通信链路高阻塞问题,现场测量结果[3]表明,由于各种因素(如基站与其服务用户之间的距离远近、不同障碍物阻挡等)引起的阻塞,毫米波链路的可用性可能是高度间歇性的,这进一步恶化了超密集网络由于复杂干扰导致的无线环境,对保证用户业务质量带来严重挑战。

针对上述问题,研究者提出较多解决方案。文献[8]通过使用 Q-Learning 算法,提出一种基于簇的分布式功率分配方案(Cluster based Distributed Power allocation using Q-Learning,CDP-Q)。该算法提升了系统容量,可满足所有用户所需的服务质量(Quality of Service,QoS),但未考虑毫米波通信的链路阻塞问题。文献[10]方法通过协作 Q-Learning 算法来最大化毫微微蜂窝总容量,同时保证宏蜂窝用户的容量水平。文献[11]提出一种基于 Q-Learning 的下行链路容量优化资源调度方案,在保持宏小区用户和毫微微小区用户之间公平性的同时,提高小区边缘用户的吞吐量。然而,在文献[10-11]中,毫微微用户的 QoS 均未被考虑。文献[12]考虑了毫米波多跳通信以克服链路阻塞,在直视路径被障碍物阻断时,通过中继器绕开障碍物来提高可靠性,但这需要足够高的节点密度以确保有合适的中继节点可用。文献[13]提出一种用于链路和中继选择的联合优化算法。考虑到包括反射路径的链接的阻塞概率,该算法选择的链接将预期的交付时间最小化,但并未考虑多路径的联合使用,并且在基站用户随机分布机制下,一条路径被阻塞时则需切换路径,这种切换机制会引入额外的等待时间。文献[14]建立一种基于波束的模型来评估毫米波覆盖概率,其不仅考虑视距传输,而且还考虑了一阶反射径的影响。在非视距情况下,反射径能够显著提高覆盖率,但该模型未考虑系统容量的影响。

本文针对毫米波通信链路高阻塞可能引起中断的问题,提出一种基于 Q-Learning 算法的功率分配方案。通过对链路阻塞问题进行分析,指出链路阻塞造成的影响有利有弊,因为链路阻塞有可能中断有用信号,同时也可能会中断干扰信号,其与基站、用户的随机分布以及周围环境等因素有关。在此基础上,将链路阻塞因素引入最优功率分配问题求解模型,借助利己利他策略[15]对利弊情况区别对待,并利用 Q-Learning 算法学习训练利己利他最佳策略,从而减小干扰,提升系统总容量,同时为用户提供所需 QoS。

1 单智能体的 Q-Learning 算法

本节通过简单案例介绍 Q-Learning 算法的基本概念[16]。Q-Learning 是一种无模型的强化学习方法,其主要解决的问题是:一个能够感知环境的智能体,通过与环境的交互反复学习,以状态为行,以行为为列构建一张 Q 表来存储 Q 值,根据 Q 值选取能够获得最大收益的动作。本文将此问题模型看作是马尔科夫决策过程(Markov Decision Process,MDP)[17-18]。将 MDP 记为(A,S,R_t,S_{t+1})元组,其包含以下元素:

1)智能体:执行学习行为并与环境交互的行为主体。

2)$A=\{a_1,a_2,\cdots,a_n\}$:一组智能体可能采取的有限的动作集合。

3)$S=\{s_1,s_2,\cdots,s_n\}$:一组有限的状态集合。基于 t 时刻的状态s_t,智能体选择动作$a_t\in A$。

4)$R_t=r(s_t,a_t)$:智能体在 t 时刻、状态 s_t 下执行特定动作a_t后的回报值函数,反映该动作的好坏,从而确定下一状态$S_{t+1}=\Phi(s_t,a_t)$。

5)S_{t+1}:奖励被反馈给智能体,从而确定下一状态S_{t+1},并重复该过程,直至 Q 表不再有更新或者达到设定的收敛条件。

首先假定智能体处在某一环境下,并且它可以感知周边的环境。其中,元素 $Q(s_t,a_t)$就是在 t 时刻、$s_t(s_t\in S)$状态下采取动作$a_t(a_t\in A)$所得到的最大累积回报值。

显然,策略的好坏不是由一次学习过程的回报值所决定的,而是由长时间累积的回报值来决定,因此,定义评估函数为:

$$V^{\pi}(s_t)=R_t+\gamma R_{t+1}+\gamma^2 R_{t+2}+\cdots=\sum_{i=0}^{\infty}\gamma^i R_{t+i} \tag{1}$$

式(1)表示的是从初始状态s_t不断通过执行策略 π 进行学习获得的累积回报值,其中,γ 是折扣因子,$\gamma\in[0,1]$,通过调节 γ,可以控制后续回报值对累积回报值的影响,γ 接近 0 表示智能体只在乎眼前的利益,做出的行为是为了最大化眼前的奖励,γ 接近 1 时表示智能体更看重长远的利益,目的是使$V^{\pi}(s_t)$最大化。

根据式(1)可以得出最佳策略 π^* 为:

$$\pi^*=\underset{\pi}{\operatorname{argmax}}\ V^{\pi}(s),\ \forall s\in S \tag{2}$$

式(2)表示对于所有的状态集合,使用策略 π^* 可使累积回报值$V^{\pi}(s)$达到最大。但在实际系统中,直接学习最佳策略 π^* 是不切实际的[19]。使用评估函数来判断动作的优劣更切合实际,评估函数定义如下:

$$Q(s,a)=r(s,a)+\gamma V^*(\Phi(s,a)) \tag{3}$$

由 $V^{\pi^*}(s)$的定义和式(2)、式(3)可以得出:

$$V^{\pi^*}(s)=\max_{a'} Q(s,a') \tag{4}$$

将式(4)代入式(3)可得:

$$Q(s,a)=r(s,a)+\gamma \max_{a'}(\Phi(s,a),a') \tag{5}$$

式(5)为一个递归函数,其为迭代逼近最优策略提供了基础。假定 $\hat{Q}$ 为智能体对实际 Q 函数的观测,对于每一个动作 $Q(s,a)=(1-\alpha)Q(s,a)+\alpha\{r(s,a)+\gamma \max_{a'} Q(\Phi(s,a),a')\}$,都有一个 $\hat{Q}(s,a)$ 与之对应。智能体多次观察当前状态 s 并选择动作 a,观察奖励 $r(s,a)$及下一状态 $\Phi(s,a)$,即:

$$\hat{Q}(s,a)\leftarrow r(s,a)+\gamma(\Phi(s,a),a') \tag{6}$$

为控制之前学习效果对整体的影响,引入学习因子 $\alpha(\alpha\in[0,1])$,其值接近 0 时表示几乎不再进行新的学习,而接近 1 时表示更看重当前的学习效果,而且学习因子会影响 Q-Learning 算法收敛的速度。因此,Q 函数更新为:

$$\begin{aligned}\hat{Q}(s,a)&=Q(s,a)+\alpha\{r(s,a)+\\&\gamma(\Phi(s,a),a')-Q(s,a)\}=\\&(1-\alpha)Q(s,a)+\alpha\{r(s,a)+\\&\gamma(\Phi(s,a),a')\}\end{aligned} \tag{7}$$

文献[16]证明了这种更新规则在某些条件下可以收敛到最优 Q 值,其中一个条件是每个状态-动作对必须进行无限次访问。如上所述,Q-Learning 在学习中获得奖励R_t,更新自己的 Q 值,并利用当前 Q 值来指导下一步的行动,在下一步的行动获得回馈值之后再更新 Q 值,不断重复迭代直至收敛。在此过程中,在已得到当前 Q 表的情况下,如何选择下一步的行为对完善当前 Q 表最有利,即如何对探索和利用进行折中最为重要。为此,本文引入一个随机因子 ε 来调节智能体进行学习的折中考虑,从而搜索到全局最优值并快速达到收敛。

由上述背景介绍可知,单智能体强化学习所在环境是稳定不变的,通常使用 MDP 来建模求解。然而,在多智能体系统中[20],每个智能体通过与环境进行交互获取奖励值来学习改善自己的决策,从而获得该环境下的最优策略。在多智能体强化学习中,环境是复杂的、动态的,这给学习过程带来很大困难。相比之下,对于单智能体面临的维度爆炸、目标奖励确定困难及不稳定性等问题,通常使用随机博弈来建模求解。本文考虑这两种方案的特点,针对网络模型随机分布即适合分布式解决方案的特性,选择单智能体解决方案。

2　系统模型与问题描述

2.1　系统模型

本文基于 5G 典型场景之一的密集室外城市场景,考虑密集部署的毫米波基站下行链路。在网络结构上,假设毫米波基站位置遵循基于密度λ_{BS}的泊松簇过程(Poisson Cluster Process,PCP)[21-22]。PCP 在实践中也被称为父子建模过程,其包含一个父过程和一个子过程,父过程形成簇的中心,子过程围绕父过程分布在簇中心一定的范围。本文假设共有 M 个簇,每个簇有 N 个毫米波基站,每个基站仅服务一个用户,用户以基站为中心、以 h 为半径随机抛洒,且每个基站天线数为 N_T、用户天线数为N_R。假设单个小区边缘用户与其相关联基站之间的距离大于设定界点距离 d 时有一定的概率会导致中断,此时用户的 QoS 会受影响,同时假定目标用户在整个操作过程中保持静止。

假设本文基站和用户之间的信道模型为毫米波信道,基站 i 和用户 k 之间的信道可以记为:$I_{i,k}(d)\times H_{i,k}$。其中,$I_{i,k}(d)$是 0-1 布尔变量,表示基站 i 和用户 k 之间的通信链路是否正常[23],d 为设定的毫米波动态链路高阻塞引起链路中断概率的界点距离。用 $P_1(x)$表示$I_{i,k}(d)$链路是否正常的概率。将毫米波基站与其相关联用户之间基于 2D 距离 x 的可视线概率记为$P_1(x)$,其由 3GPP 城市微街道峡谷模型[24]获得,如式(8)所示:

$$P_1(x)=\begin{cases}1, & x_{i,k}\leqslant d\\ \dfrac{\left(18+x_{i,k}\mathrm{e}^{-\frac{x_{i,k}}{36}}-18\mathrm{e}^{-\frac{x_{i,k}}{36}}\right)}{x_{i,k}}, & x_{i,k}>d\end{cases} \tag{8}$$

其中,$x_{i,k}$为当前毫米波基站 i 与其关联用户 k 之间的 2D 距离,当 $x\leqslant d$ 时,表示一定可视,链路无阻塞;当 $x>d$ 时,表示有概率不可视,链路有概率被阻塞。

$L_{i,k}$表示基站 i 和用户 k 之间的路径损耗,路径损耗$L_{i,k}$如式(9)所示:

$$L_{i,k}=\beta_1+10\beta_2\lg x_{i,k}+X_\zeta \tag{9}$$

其中,β_1 和β_2 是用于实现最佳拟合信道测量因子,$x_{i,k}$为当前毫米波基站 i 与其关联用户 k 之间的 2D 距离,$X_\zeta\sim N(0,\zeta^2)$表示对数阴影衰落因子。

此时第 k 个用户的信干噪比(Signal to Interference plus Noise Ratio,SINR)为:

$$\mathrm{SINR}_k=\frac{P_k\left|I_{k,k}(d)\times H_{k,k}\right|^2}{\sum\limits_{i\in Dk,i\neq k}P_i\left|I_{i,k}(d)\times H_{i,k}\right|^2+\sigma^2} \tag{10}$$

$$P_k=10^{(P_{k'}-L_{k,k})/10} \tag{11}$$

其中,P_k 为第 k 个基站发送到其服务用户 k 的实际功率,$P_{k'}$ 为 Q-Learning 算法中基站 k 选定某一状态所执行的行为(功率),$L_{k,k}$ 为基站 k 到用户 k 之间的路损,D_k 为干扰基站的集合,σ^2 表示加性高斯白噪声方差。因此,第 $P_k(P_k=P_{k'}-L_{k,k})$ 个用户的归一化容量为:

$$C_k=\mathrm{lb}(1+\mathrm{SINR}_k) \tag{12}$$

2.2 问题描述

本文旨在毫米波基站之间寻找最优功率分配,为此,首先建立最大化问题模型,然后通过问题求解使系统总容量最大化,同时满足所有用户的 QoS 和功率约束。本文优化问题(P1)可以表述为:

$$\underset{\mathrm{P1}}{\mathrm{maximize}}\sum_{m=1}^{M}\sum_{k=1}^{N}\mathrm{lb}(1+\mathrm{SINR}_k) \tag{13}$$

s. t.

$$P_k\leqslant P_{\max},k=1,2,\cdots,N \tag{14}$$

$$\mathrm{SINR}_k\geqslant q_k,k=1,2,\cdots,N \tag{15}$$

其中:目标函数(式(13))表示最大化网络总容量;M 为系统模型中划分簇的个数;N 表示划分的每个簇内的毫米波基站的数量;约束条件(式(14))指的是每个毫米波基站的功率限制,表示从每个毫米波基站分配给用户的功率不能超过最大功率 $P_{\max}$;式(15)中的 q_k 表示第 k 个用户所需的最小 SINR 值,称为阈值。P1 优化问题为:在使整个系统总容量最大的同时要满足每个用户所需的 QoS。

由于 P1 优化问题中 SINR 项的分母包含干扰项,而在毫米波通信网络中,干扰复杂不可忽略,因此,究其本质为一个耦合问题,此类优化问题是一个非凹函数,无法直接求解。传统的启发式方案所求得的仅为次优策略,与最优解误差较大。而本文所要解决的是一个耦合问题,再加上考虑到毫米波链路阻塞特性导致的 0-1 布尔问题,使得 P1 更加难以解决,因此,本文考虑使用 Q-Learning 算法来解决此问题。

本文设计的 P1 的解决方案还具有以下特征:

1)由于设定场景为密集室外城市场景,毫米波基站和用户数量众多,干扰复杂多样,没有行之有效的中心管理机构,因此本文以分布式结构来处理。

2)毫米波通信虽然带来了显著的容量增益,但其信号衰耗大、辐射范围有限,因此,合理假设只有距离接近的毫米波基站彼此干扰。可使用聚类机制将设定场景中的基站划分为多个集群,其中一个集群的干扰对其他集群的用户来说可以忽略不计。

3 PPCP-Q 分配方案

本节介绍 PCP 网络模型下基于 Q-Learning 算法的功率分配方案 PPCP-Q。与其他强化学习算法相比,Q-Learning 算法在复杂系统中具有非常好的学习性能。同时 PCP 还具有叠加性、稀释性和映象性[19]等特性,非常适应毫米波网络的复杂结构以及网络环境的动态变化,因此,本文采用 PPCP-Q 方案进行功率分配。PPCP-Q 算法分为 PCP 聚类和基于 Q-Learning 的分布式功率分配两个部分。

3.1 毫米波基站的泊松簇过程

本文考虑系统模型具有分布式特性以及毫米波通信信号损耗大和覆盖范围较小的特性,假定在应用场景中只有距离靠近的毫米波基站之间彼此干扰。本文基于 PCP 假设将毫米波基站划分为簇,PCP 过程是一种分布式聚类方法,并可生成非重叠簇。为具体说明其过程,给出以下定义:

1)簇头。基于 PCP 产生簇头,在本文中,被选定为簇头的毫米波基站与簇内其余毫米波基站之间没有优先级之分。

2)入簇(In Cluster,IC)和簇外(Out Cluster,OC)节点。在 PCP 中,将 IC 距离定义为 100 m,这是强干扰的表示。OC 距离定义为 200 m,其表示簇头周围簇的边缘覆盖范围。若 a 点距某一簇头在 100 m 范围之内则定义为 IC;若 a 点处于此簇的 OC 距离(即大于 100 m 小于 200 m 的范围),而此时又不属于任何其他簇的 IC 距离,则将 a 点作为 OC 节点加入此簇;若 a 点距离多个簇头的距离相同,此时随机选择一个簇加入即可。

3.2 基于 Q-Learning 的分布式功率分配

Q-Learning 算法是强化学习的典型方法之一,已被证明具有收敛性。在 PPCP-Q 算法中,毫米波基站被认为是 Q-Learning 算法的智能体。PPCP-Q 是一种分布式方法,其中多个智能体旨在通过反复与环境交互来发现最佳策略(功率)以最大化网络容量。

在多智能体的学习中,智能体可以合作学习(Cooperative Learning,CL)或独立学习(Independent learning,IL)。在 CL 中,当前智能体与其他合作智能体共享其 Q 表。在 IL 中,每个智能体独立于其他智能体学习(即将其他智能体视为环境的一部分,忽略其行为),虽然这可能导致算法收敛时间变长,但与 CL 相比,智能体之间没有通信开销,因此,本文选择 IL。PPCP-Q 算法描述如下:

算法 1 PPCP-Q 算法

输入 状态集合,动作集合,学习因子 α,折扣因子 γ

输出 Q 表

1. 初始化:$t=0,\varepsilon=0.9,\forall s_t^k\in S,\forall a_t^k\in A,Q_k(s_t^k,a_t^k)=0$
2. 基于 3.1 节泊松簇过程对系统模型进行分簇
3. for 所有簇
4. for 每一个簇内智能体(毫米波基站)
5. for 每一个智能体的训练次数 episode
6. 设定状态,并随机选择一个状态 s_t^k
7. 生成一个 0 ~ 1 之间的随机数 sigma
8. if sigma < ε

9. $\varepsilon = \varepsilon \times 0.99$

10. 随机选择动作a_t^k

11. else

12. 根据$a_t^k = \text{argmax}\ Q_k(s_t^k, :)$选择 Q 表中该状态下最大 Q 值对应动作$a_t^k$

13. end

14. 执行动作a_t^k,得到对应的奖励R_t^k

15. 确定下一状态s_{t+1}^k

16. 根据式(7)更新 Q 表

17. 更新状态$s_t^k = s_{t+1}^k$

18. end for

19. end for

20. end for

Q-Learning 算法的输出为分配的功率,被表示为 Q 函数,智能体的 Q 函数被称为 Q 表,其中行是状态,列是行为(功率)。在 Q-Learning 算法中,行为、状态和回报函数定义如下:

1)行为。每个智能体可执行的动作是簇内毫米波基站可以使用的一组可能的功率。在仿真中,行为(分配的功率)集合 A 定义为 $A = \{a_1, a_2, \cdots, a_n\}$,它均匀地覆盖了最小功率($a_1 = P_{\min}$)和最大功率($a_n = P_{\max}$)之间的范围,步长为 1 dBm。

2)状态 $S = \{s_t^1, s_t^2, \cdots, s_t^k, \cdots, s_t^n\}$。$s_t^k$ 表示在时刻 t 智能体 k 所处的状态,$s_t^k = \{d_t^{i,k}, p_t^k\}$,$d_t^{i,k} \in \{0,1\}$,定义为在时刻 t 根据基站 i 与用户 k 之间的 2D 距离$x_{i,k}$是否会导致毫米波动态链路中断,表达式如下:

$$d_t^{i,k} = \begin{cases} 0, & x_{i,k} \leqslant d \\ 1, & x_{i,k} > d \end{cases} \tag{16}$$

其中,$i,k = 1,2,\cdots,N$。$x_{i,k}$小于等于设定的界点距离 d 表示链路正常通信,此时为状态 0;$x_{i,k}$大于 d 表示有概率导致链路中断,此时为状态 1。

选择一个簇,即A_1、A_2,利己利他策略应用示意图如图 1 所示。

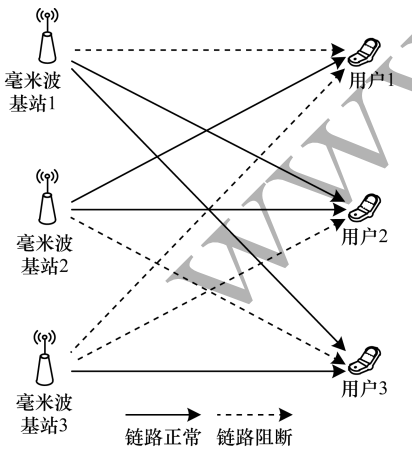

图 1　利己利他策略示意图

Fig. 1　Schematic diagram of egoistic and altruistic strategy

根据基站 i 与用户 k 之间的 2D 距离 $x_{i,k}$,智能体在时刻 t 的状态可分为以下 3 种情况:

情况 1　基站 1 与服务用户 1 链路阻塞,基站 1 与被干扰用户 2、用户 3 链路正常,即有用信号发生阻塞,用户 1 考虑利他功率分配策略,此时应最小化发射功率,避免对其他用户干扰。

情况 2　基站 2 及其服务用户 2 链路正常,基站 2 与用户 1 链路正常,与用户 3 链路中断,即部分被干扰用户链路阻塞,此时应考虑利己利他功率分配策略,即按功率等级适量发射功率,均衡干扰同时提升系统容量。

情况 3　基站 3 及其服务用户 3 链路正常,基站 3 与用户 1、用户 2 链路中断,即用户 3 所产生干扰信号完全阻塞,此时应考虑利己功率分配策略,最大化发射功率,提升系统容量。

上述 3 种情况分别对应状态中p_t^k 在时刻 t 智能体 k 发射功率的量化等级为:

$$p_t^k = \begin{cases} 0, & P_{\min} < p_t^k \leqslant P_{\max} - A_1 \\ 1, & P_{\max} - A_1 < p_t^k \leqslant P_{\max} - A_2 \\ 2, & P_{\max} - A_2 < p_t^k \leqslant P_{\max} \end{cases} \tag{17}$$

其中,$k = 1,2,\cdots,N$,A_1 和 A_2 分别为 10 dBm 和 5 dBm。

3)回报函数。本文优化问题 P1 旨在最大化系统总容量,同时满足用户所需的 QoS。回报函数的设计基于优化目标,反映环境对智能体选择动作的满意程度。回报函数定义如下:

$$R_t^k = \begin{cases} 1 - e^{-\frac{C_t - G\text{lb}(1+q_k)}{C_t}}, & g_t \geqslant Gq_k \\ -1, & g_t < Gq_k \end{cases} \tag{18}$$

$$C_t = \sum_{g \in G, g \neq k}^{G} \text{lb}(1 + \text{SINR}_g) \tag{19}$$

其中:$k = 1,2,\cdots,N$;R_t^k 表示智能体 k 在时刻 t 得到的即时奖励;q_k 为多次实验观测所取的阈值常数,所有用户均考虑此值;回报函数R_t^k 取决于在 t 时刻除智能体 k 外环境基站得到的和容量C_t 以及用户所需最低 SINR 的q_k 值;G 表示该簇内除当前智能体外其余环境中毫米波基站的数量;g_t 表示当前簇内智能体 k 在时刻 t 得到一个行为(功率)对该簇内其余成员基站造成影响后,其余成员在链路未阻塞状态下得到的最小和速率,体现了利己策略。

回报函数的基本原理如下:

1)只有当环境中毫米波基站最小和速率g_t 大于等于 G 倍的阈值时,才能得到一个正的回报值,否则回报值为负值。

2)因为$(C_t - G\,\text{lb}(1+q_k))/C_t < 1$,并且$C_t$ 越大,该值越大,所以系统吞吐量越大,对应回报值也越大。

此外,在学习过程中,需要设置一个随机因子 ε 来调节智能体进行随机学习的比例,本文假设随机

因子 ε 的初值为 0.9,且智能体每进行一次随机学习,该值就更新为原值的 0.99 倍。基于此设置,智能体在初始学习时会频繁地进行随机学习,随着学习次数的增加和学习经验的累积,随机因子的值会逐渐趋近于 0,此时智能体会进行经验学习,选择每一步的最优策略并快速达到收敛。

4 仿真与结果分析

本节对构建的系统模型进行系统级仿真,仿真程序在 Matlab 环境下实现,以证明本文方案的有效性。

4.1 仿真环境与参数设置

考虑密集部署的毫米波基站的下行链路网络,在 1 km^2 的区域内分布 2 个 ~16 个簇。簇头基于泊松过程生成,每个簇内基于半径 R 范围随机分布 $N(N=5)$ 个成员基站,依据成员基站与簇头距离再划分成员基站不重叠簇的归属。簇头与簇内成员之间没有优先级之分,且簇与簇之间相互独立。簇内每个基站基于半径 h 随机抛洒一个用户,即每个基站支持一个用户。用户的 QoS 被定义为支持用户服务所需的最低 SINR,所有用户均考虑阈值 q_k = 10 dB。在 Q-Learning 算法中,学习率为 α,折扣因子为 γ,最大迭代次数被设置为 50 000 次。其他仿真参数如表 1 所示。

表 1 仿真参数

Table 1 Simulation parameters

参数	参数值
折扣因子 γ	0.9
学习因子 α	0.5
随机因子 ε	0.9
泊松簇过程密度 λ_{BS}	14
毫米波基站服务用户范围 h/m	25
发射天线数 N_T	1
接收天线数 N_R	1
界点距离 d/m	18
最佳拟合信道测量因子 β_1	34.6
最佳拟合信道测量因子 β_2	37.6
对数阴影参数 ζ/dBm	8
噪声功率谱密度 σ^2/(dBm · Hz^{-1})	−172
最小功率 P_{min}/dBm	0
最大功率 P_{max}/dBm	15
簇内分布半径 R/m	100

4.2 结果分析

本文基于 PCP 设定的系统模型基站和用户的分布情况如图 2 所示,图中以不同形状分别表示簇头、簇内毫米波基站和相关联的用户。

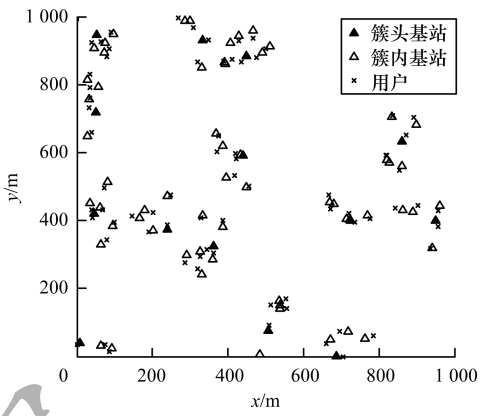

图 2 PCP 模型基站与用户分布

Fig. 2 BSs and users distribution of PCP model

图 3 所示为本文方案某个智能体在学习动作上的收敛情况。由于 PPCP-Q 方案考虑阻塞概率,使系统更加动态化,因此引入随机因子 ε 对探索和利用进行折中考量,调节智能体进行随机学习的比例,选择集体最优而非单次最优的一串最优动作。随着 ε 的值逐渐趋近于 0,智能体会进行经验学习,加快收敛速度。从图 3 中可以看出,在前3 000 次学习中,智能体进行了大量的随机动作选择,但是随着迭代次数的增加,设定的随机因子在逐渐减小,在 3 000 次 ~9 000 次学习过程中随机的动作次数逐渐减小,在 9 000 次学习之后,智能体在学习过程中动作的选择逐渐达到收敛状态。

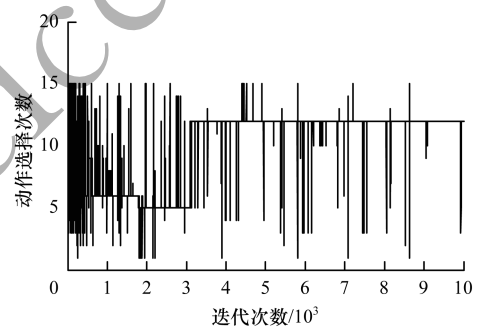

图 3 Q-Learning 算法动作收敛情况

Fig. 3 Action convergence of Q-Learning algorithm

将本文 PPCP-Q 方案与文献[8]的 CDP-Q 方案进行比较,如图 4 所示。可以看出,对于多种可能簇的大小,2 种方案在系统总容量上的取值均超过阈值,均满足用户所需的 QoS。图 5 为 PPCP-Q 与 CDP-Q 两种方案系统容量的 CDF 曲线对比。可以看出,PPCP-Q 方案相较于 CDP-Q 方案提供了较为明显的系统容量增益。增益机理分析如下:PPCP-Q 方案考虑到毫米波链路可用性是高间歇性的,同时在 Q-Learning 算法状态和回报函数设计中加入利己利他策略来求解目标函数,对有利有害情况区别对待并加以利用,在减小干扰的同时合理分配功率以最大化系统容量,同时满足用户的 QoS。而文献[8]的

CDP-Q 方案基于 Q-Learning 算法来训练智能体分配功率,其采用利己分配策略,但未根据毫米波通信链路阻塞特性施加不同功率分配策略,所以在系统性能角度上,其增益受到限制。总体而言,本文方案显著提升了系统容量,而且随着簇的个数增多,性能优势更为明显。

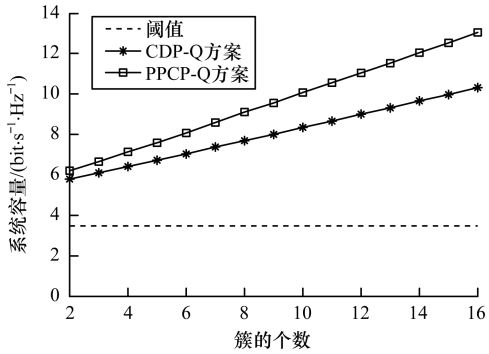

图 4　多簇情况下两种方案的系统总容量对比

Fig. 4　Comparison of the total capacity of two schemes in the case of multiple clusters

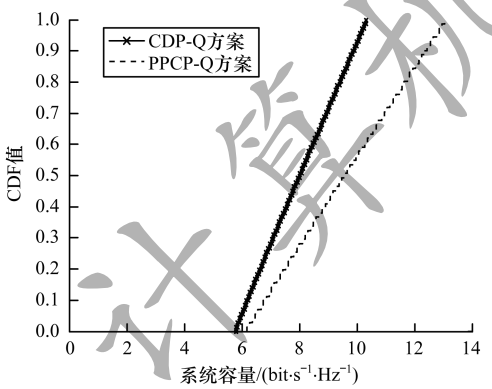

图 5　两种方案的系统容量 CDF 曲线对比

Fig. 5　Comparison of system capacity CDF curve of two schemes

5　结束语

随着多媒体应用的不断发展和移动流量的爆炸式增长,5G 网络中基站部署日趋密集,使毫米波通信在解决频谱资源短缺和提升系统性能的同时也面临高阻塞、大衰落和干扰复杂多变等问题,影响了毫米波链路的可用性,而且基站用户随机分布,无规律可循,使得功率分配问题更为复杂。对此,本文提出一种基于 Q-Learning 的功率分配方案。以毫米波基站为智能体,在状态和回报函数设计中加入利己利他策略,考虑多种可能的链路阻塞情况,充分利用功率资源以最大化系统容量,同时保证用户的 QoS。仿真结果表明,本文方案能够提升系统性能,实现优化目标。由于Q-Learning 算法在状态和行为设计上维度有限,因此下一步将考虑利用深度神经网络改进该方案,并通过添加更多指标,同时实现多个系统优化目标,提升系统的整体性能。

参考文献

[1] CHEN Liang, YANG Qi. Analysis of spectrum occupation and perspectives of 5G network [J]. Study on Optical Communications, 2016, 42(6): 68-69. (in Chinese)
陈亮,杨奇. 5G 网络中无线频谱资源分配的进展分析[J]. 光通信研究,2016,42(6):68-69.

[2] BOCCARDI F, HEATH R W, LOZANO A, et al. Five disruptive technology directions for 5G [J]. IEEE Communications Magazine, 2014, 52(2): 74-78.

[3] RANGAN S, RAPPAPORT T S, ERKIP E. Millimeter-wave cellular wireless networks: potentials and challenges [J]. Proceedings of the IEEE, 2014, 102(3): 366-373.

[4] GHOSH A, THOMAS T A, CUDAK M C, et al. Millimeter-wave enhanced local area systems: a high-data-rate approach for future wireless networks [J]. IEEE Journal on Selected Areas in Communications, 2014, 32(6): 1152-1163.

[5] AGRAWAL S K, SHARMA K. 5G millimeter Wave (mmWave) communications [C]//Proceedings of the 3rd International Conference on Computing for Sustainable Global Development. Washington D.C., USA: IEEE Press, 2016: 3630-3634.

[6] BALDEMAIR R, IRNICH T, BALACHANDRAN K, et al. Ultra-dense networks in millimeter-wave frequencies [J]. IEEE Communications Magazine, 2015, 53(1): 202-205.

[7] BAI T, HEATH R W. Coverage in dense millimeter wave cellular networks [C]//Proceedings of Asilomar Conference on Signals, Systems and Computers. Washington D. C., USA: IEEE Press, 2013: 2062-2066.

[8] AMIRI R, MEHRPOUYAN H. Self-organizing mm-wave networks: a power allocation scheme based on machine learning [C]//Proceedings of the 11th Global Symposium on Millimeter Waves. Boulder, USA: GSMM Press, 2018: 1-4.

[9] CHEN Ming, HUA Ye, GU Xinyu, et al. A self-organizing resource allocation strategy based on Q-learning approach in ultra-dense networks [C]//Proceedings of IEEE International Conference on Network Infrastructure and Digital Content. Washington D.C., USA: IEEE Press, 2016: 155-160.

[10] SAAD H, MOHAMED A, ELBATT T. Distributed cooperative Q-learning for power allocation in cognitive femtocell networks [C]//Proceedings of IEEE Vehicular Technology Conference. Washington D. C., USA: IEEE Press, 2012: 1-5.

[11] WEN Bin, GAO Zhibin, HUANG Lianfen, et al. A Q-learning-based downlink resource scheduling method for capacity optimization in LTE femtocells [C]// Proceedings of the 9th International Conference on Computer Science and Education. Washington D. C., USA: IEEE Press, 2014: 625-628.

[12] KWON S, WIDMER J. Relay selection for mmWave communications [C]//Proceedings of the 28th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications. Washington D. C., USA: IEEE Press, 2017: 1-6.

[13] HE Zhifeng, MAO Shiwen. A decomposition principle for link and relay selection in dual-hop 60 GHz networks [C]// Proceedings of the 35th Annual IEEE International Conference on Computer Communications. Washington D. C., USA: IEEE Press, 2016: 1-9.

[14] TATINO C, MALANCHINI I, AZIZ D, et al. Beam based stochastic model of the coverage probability in 5G millimeter wave systems[C]//Proceedings of the 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks. Washington D.C., USA: IEEE Press, 2017: 1-6.

[15] SUN Changyin, LEI Yuan, JIANG Fan, et al. Beamforming for virtual cell with balanced interference[J]. Journal of Xi'an University of Posts and Telecommunications, 2017, 22(1): 18-21. (in Chinese)
孙长印, 雷远, 江帆, 等. 虚拟小区干扰均衡的波束形成算法[J]. 西安邮电大学学报, 2017, 22(1): 18-21.

[16] WATKINS C J C H, DAYAN P. Technical note: Q-learning[J]. Journal of Machine Learning, 1992, 8(3): 279-292.

[17] GAO Zhenhai, SUN Tianjun, HE Lei. Causal reasoning decision-making for vehicle longitudinal automatic driving[J]. Journal of Jilin University (Engineering and Technology Edition), 2019, 49(5): 1392-1404. (in Chinese)
高振海, 孙天骏, 何磊. 汽车纵向自动驾驶的因果推理型决策[J]. 吉林大学学报(工学版), 2019, 49(5): 1392-1404.

[18] WANG Xiaolei, CHEN Yunjie, WANG Chen, et al. Scheduling method of virtual network function based on Q-learning[J]. Computer Engineering, 2019, 45(2): 66-67. (in Chinese)
王晓雷, 陈云杰, 王琛, 等. 基于 Q-learning 的虚拟网络功能调度方法[J]. 计算机工程, 2019, 45(2): 66-67.

[19] SHANG Sihan. Research of interference management techniques in ultra dense networks[D]. Chengdu: University of Electronic Science and Technology of China, 2018. (in Chinese)
尚思翰. 超密集网络中的干扰管理技术研究[D]. 成都: 电子科技大学, 2018.

[20] ASL Z D, DERHAMI V, YAZDIAN-DEHKORDI M. A new approach on multi-agent multi-objective reinforcement learning based on agents' preferences[C]//Proceedings of Artificial Intelligence and Signal Processing Conference. Washington D.C., USA: IEEE Press, 2017: 75-77.

[21] CHUN Y J, HASNA M O. Analysis of heterogeneous cellular networks interference with biased cell association using Poisson cluster processes[C]//Proceedings of International Conference on Information and Communication Technology Convergence. Washington D.C., USA: IEEE Press, 2014: 319-322.

[22] MA Zhonggui, LIU Liyu, YAN Wenbo, et al. Deployment model of three-layer heterogeneous cellular networks based on Poisson clustered process[J]. Chinese Journal of Engineering, 2017, 39(2): 309-311. (in Chinese)
马忠贵, 刘立宇, 闫文博, 等. 基于泊松簇过程的三层异构蜂窝网络部署模型[J]. 工程科学学报, 2017, 39(2): 309-311.

[23] ALONZO M, BUZZI S. Cell-free and user-centric massive MIMO at millimeter wave frequencies[C]//Proceedings of the 28th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications. Washington D.C., USA: IEEE Press, 2017: 1-2.

[24] GAPEYENKO M, PETROV V, MOLTCHANOV D, et al. On the degree of multi-connectivity in 5G millimeter-wave cellular urban deployments[J]. IEEE Transactions on Vehicular Technology, 2019, 68(2): 1973-1974.

编辑 金胡考