

基于深度强化学习的低轨卫星下行功率分配方案^{*}

张华明^{1,2}, 李强^{1†}

(1 中国科学院上海微系统与信息技术研究所, 上海 201800; 2 中国科学院大学, 北京 100049)
(2020 年 6 月 22 日收稿; 2020 年 9 月 2 日收修改稿)

Zhang H M, Li Q. Downlink power allocation scheme for LEO satellites based on deep reinforcement learning[J].
Journal of University of Chinese Academy of Sciences, 2022, 39(4): 543-550. DOI:10.7523/j.ucas.2020.0045.

摘 要 当前的卫星资源分配方案大多为同步轨道卫星设计, 针对低轨卫星的高动态特性, 以及存在频率和功率资源受限的问题, 提出一种基于深度强化学习的功率分配算法。首先对低轨卫星功率分配场景进行建模, 引入一种时隙划分方案来简化低轨卫星的动态特性模型, 进一步提出一种基于深度强化学习算法的功率分配策略, 该策略通过调节单颗低轨卫星各个波束中子载波的功率值, 降低同频干扰, 能达到提升低轨卫星频谱效率的目的。仿真结果表明, 所提算法能够在较短时间内收敛并达到稳定状态, 在总功率一定的条件下, 该方案能有效提升单颗低轨卫星的吞吐量, 其频谱效率明显高于注水算法和 Q 学习算法。

关键词 低轨卫星; 频谱效率; 功率分配; 深度强化学习

中图分类号: TN927 **文献标志码**: A **DOI**: 10.7523/j.ucas.2020.0045

Downlink power allocation scheme for LEO satellites based on deep reinforcement learning

ZHANG Huaming^{1,2}, LI Qiang¹
(1 Shanghai Institute of Microsystem & Information Technology, Chinese Academy of Sciences, Shanghai 201800, China;
2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract Most of the current satellite resource allocation schemes are designed for geosynchronous orbit satellites. In view of the highly dynamic characteristics and limitation of frequency and power resources in LEO satellites, a power allocation algorithm based on deep reinforcement learning is proposed. First of all, we model the LEO satellite power allocation scenario, and introduce a time slot division scheme to simplify the dynamic characteristics model of the LEO satellite. Then a power allocation policy is proposed based on deep reinforcement learning algorithm which can reduce the co-channel interference by adjusting the power value of the subcarriers in each beam of a single LEO satellite, thus improving the spectral efficiency of the LEO satellite. Simulation results illustrate that the proposed algorithm can converge and reach a stable state in a relatively short time. Under the condition of constant total power, this scheme can effectively improve the throughput of a single LEO satellite. The spectral efficiency based on deep reinforcement learning algorithm is significantly

^{*} 国家重点研发计划项目(2019YFB1803101)资助

[†] 通信作者, E-mail: qiang.li@mail.sim.ac.cn

higher than that of water-filling algorithm and Q-learning algorithm.

Keywords LEO satellite; spectrum efficiency; power allocation; deep reinforcement learning

近年来,随着卫星通信网络的快速发展,接入用户呈海量增长,低轨卫星(LEO 卫星)由于星地链路时延小,对天线功率和尺寸要求更低而广受青睐^[1]。目前卫星通信系统可用频率资源非常有限,包括 S 频段上下行链路的 2×15 MHz 和 Ka 频段上下行链路的 $2 \times 2\ 500$ MHz^[2]。此外,相较于地面蜂窝网络,低轨卫星上的功率资源有限,当卫星的太阳能电池板位于阴影处时,卫星只能使用电池能源。因此,合理调配资源,提高卫星频谱效率(spectrum efficiency, SE)和能效(energy efficiency, EE)是当前的研究热点。如何智能化、实时化的合理调度星上资源成为低轨卫星通信系统研究中的热点之一。

宽带通信卫星多采用多波束技术,通过波束间频率复用、功率分配等技术可以显著提高系统容量。对于卫星多波束功率分配,传统的固定分配方式已无法适应高效能的发展需求,目前许多研究都开始考虑卫星在空间、时间以及其传播路径的变化,功率分配方式逐渐向自适应动态分配方式发展。文献[3]根据流量需求和信道条件,使用拉格朗日乘数法优化了功率分配,但未考虑同频信道干扰。文献[4]考虑同频信道干扰,证明功率分配问题的 NP-hard 特性,并由此激发提出启发式算法,将问题转化成两阶段优化模型。文献[5]研究了多波束卫星通信系统载波和功率的联合分配算法,使系统在总发射功率限定下能满足更多用户需求。文献[6]将资源分配问题分解为功率分配、带宽分配和对偶变量更新 3 个问题,提出对应分配算法,结果能有效提升系统容量。但以上两篇文献的方法没有对信道进行建模,不适合低轨卫星时变信道环境。文献[7]利用动态规划方法讨论了卫星网络中的最优功率分配问题,然而文中每个用户的能量消耗是以已知的概率分布产生的,而实际上是随机事件,具有一定的局限性。

上述研究成果大多集中在同步轨道卫星的多波束功率分配,用户位置相对静态,信道状态相对固定,算法更新频率不需要很高。而低轨卫星运动速度快,过顶时间短,具有高动态性,用户与接入卫星的相对位置在实时变化,复杂度高的启发式算法收敛速度慢,效率低。

当前,机器学习在卫星无线资源分配的研究尚处于探索阶段,深度强化学习以其自主决策方面的优势在资源分配研究中取得显著成效,例如文献[8]采用深度 Q 学习方案联合优化星地网络中的缓存、计算和网络资源,提升了资源利用率。

本文提出一种基于深度强化学习的面向低轨卫星的功率分配方法,对卫星下行链路进行信道建模,考虑噪声和同频干扰的影响。同时考虑卫星的动态特性,并对其建模,将卫星在轨运行过程进行时间切片划分,每进入一个时间切片,更新信道信息。利用深度强化学习动态分配各个波束子载波的功率,在不了解网络环境动态特性的情况下,本文选择无模型(model-free)的强化学习方法,解决该序列决策问题,针对性地设计强化学习的状态、动作和奖励函数,从而有效提升系统总容量。为了提升算法速度,降低硬件存储消耗,对功率值离散化处理。该方案能充分利用强化学习的决策能力和深度学习的感知能力。

1 系统模型

1.1 多波束卫星覆盖模型

3GPP 在非地面网络与地面 5G 融合的场景下提出星地融合的 4 种网络架构^[9],本文采用无弯管、卫星承载 gNB(下一代 NodeB)并与用户直连的架构。卫星通过多波束技术可以进行频率复用,提高系统容量,同时能让能量更加集中。本文针对使用同一频段的多个波束,下行链路采用 OFDM 传输模式,同一波束内用户之间没有干扰,不同波束之间产生不同程度的同频干扰,这种简化降低了模型的复杂度,一般也不影响模型的有效性。用户在波束覆盖范围内随机分布。

假设卫星上搭载的信息控制中心可以收集所有的波束信息,包括链路的信号噪声干扰比(signal noise and interference ratio, SNIR)和传输功率。在时隙 t 接入波束 k 的用户数量表示为

$$u_t^k \in \{M_t^1, \dots, M_t^k, \dots, M_t^K\} \left(\sum_{k=1}^K M_t^k = M \right), (1)$$

其中 M_t^k 是时刻 t 接入波束 k 的用户数量, M 是总的用户接入数量, K 是卫星波束总个数。波束的地理分布如图 1 所示,圆表示波束的 3 dB 衰减边界。每个波束可以使用 N 个正交子载波。一旦

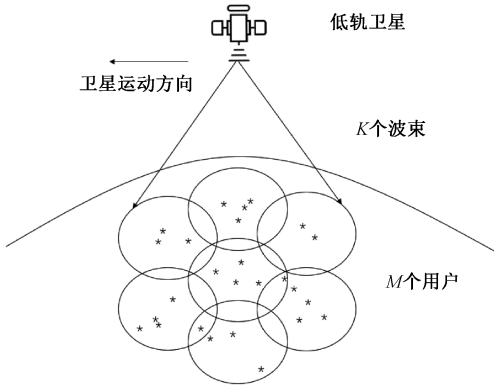


图1 低轨多波束卫星覆盖模型

Fig. 1 LEO multibeam satellite coverage model

有用户接入波束,该波束被激活并且传输功率为 p_t^n ,每个用户可以在时刻 t 连接到一个波束。每个子载波只能被分配给单个用户。信道模型采用自由空间损耗模型。

1.2 波束间同频干扰

在时刻 t ,自由空间传播损耗为 L , $L = 4\pi d/\lambda^2$,其中 d 为信号传输距离, λ 是信号波长^[10]。用 $p_t^{(k,n)}$ 和 $p_t^{(k',n)}$ 分别表示第 k 个波束和第 k' 个波束中第 n 个子载波对用户 m 的传输功率,则第 k 个波束中第 n 个子载波的有效信号功率 $E_t^{(k,n)}$ 和干扰信号功率 $I_t^{(k,n)}$ 可以表示成

$$E_t^{(k,n)} = p_t^{(k,n)} G_{k,m} \left(\frac{\lambda^2}{4\pi d} \right) = h_{k,m}^n p_t^{(k,n)}, \quad (2)$$

$$I_t^{(k,n)} = \sum_{k' \neq k} p_t^{(k',n)} G_{k',m} \left(\frac{\lambda^2}{4\pi d} \right) = \sum_{k' \neq k} h_{k',m}^n p_t^{(k',n)}. \quad (3)$$

因此,第 k 个波束中第 n 个子载波对用户 m 的信道增益计算式为

$$h_{k,m}^n = G_{k,m} \left(\frac{\lambda^2}{4\pi d} \right), \quad (4)$$

波束增益 $G_{k,m}$ 由下式计算:

$$G_{k,m} = G_{\max} \left(\frac{J_1(u)}{2u} + 36 \frac{J_3(u)}{u^3} \right)^2. \quad (5)$$

$u = 2.07123 \sin \theta_{k,m} / \sin \theta_{3dB}$, G_{\max} 是多波束卫星天线的最大增益, θ 是波束中心与用户终端之间的夹角。 $h_{k,m}^n$ 表示第 k 个波束中第 n 个子载波对用户 m 的信道增益, J_1 和 J_3 分别表示第一类一阶和三阶贝塞尔函数^[11]。

让 $\xi_t^{(k,n,m)}$ 表示时刻 t ,卫星的第 k 个波束服务于用户 m 的第 n 个子载波的 SINR,其中 $n \in \{1, \dots, N\}$,信干比 SINR 表示为

$$\xi_t^{(k,n,m)} = \eta_t^{(k,m)} \frac{h_{k,m}^{(n)} p_t^{(k,n)}}{\sum_{k' \neq k} h_{k',m}^{(n)} p_t^{(k',n)} + \sigma^2}. \quad (6)$$

由于采用全频率复用,干扰来自其他所有波束。 $\eta_t^{(k,m)}$ 表示用户 m 是否连接到第 k 个波束, $\eta_t^{(k,m)} \in [0,1]$, σ^2 是噪声功率。

1.3 问题表述

系统性能由总体容量衡量,单位 bps/Hz。波束在时刻 t ,在子载波 n 和用户连接的容量由下式给出

$$C_t^{k,n} = \frac{B}{N} \log_2 \left(1 + \sum_{m=1}^M \xi_t^{(k,n,m)} \right), \quad (7)$$

整个容量可以定义为

$$SC_t = \sum_{n=1}^N \sum_{k=1}^K C_t^{k,n}, \quad (8)$$

目标是通过调整波束在子载波上的功率提升整个网络的总体容量

$$\mathbf{p}_t^k = [p_t^{k,1}, \dots, p_t^{k,n}, \dots, p_t^{(k,N)}]. \quad (9)$$

优化问题可以表述为以下

$$\begin{cases} \text{argmax} SC_t \\ \text{s. t. } p_t^{(k,n)} \geq p_{\min}, \forall n, k, \\ \sum_{n,k} p_t^{(k,n)} \leq p_{\max}, \forall n, k \end{cases}, \quad (10)$$

其中 p_{\max} 是卫星的最大传输功率,各个波束总功率不能大于卫星最大功率, p_{\min} 是子载波的最小传输功率。

在网络初始化阶段,用户基于最大 SINR 准则完成和卫星波束的连接。用户和波束的链路干扰主要来自同频干扰,链路速率受附近波束到用户的信号强度的影响。本文的目标是调整波束中各个子载波的传输功率,从而提升整个低轨卫星的系统容量。考虑到用户的公平性,本文将下行带宽平均分配给用户,子载波的初始功率也由总功率平均分配。

上述问题是一个非凸优化问题,解决该问题的传统方法是启发式搜索,然而大多数这类算法运行时间长效率低,难以实时在线调整^[12],由于低轨卫星的动态特性,某一时刻分配的最优功率值在下一时刻不一定是最优的。为了解决该问题,引入深度强化学习算法。

2 基于深度强化学习的功率分配

2.1 信道统计信息的获取时间划分

为了适应低轨卫星的高速移动特性,需要高

频率的收集当前卫星的轨道参数和信道增益信息,以计算公式(4),再根据这些信息调整各个波束子载波功率值,而信息获取时间间隔的大小会影响功率分配方案的合理性,本文根据以下原则划分时间间隔。

星地信道主要受大尺度影响,即主要考虑路径损耗的影响,卫星和地面终端节点之间的路径损耗 $4\pi d/\lambda^2$ 写成 dB 形式为

$$L = 92.44 + \lg f + 20 \lg d, \quad (11)$$

其中: f 为卫星中心频率, d 为信号传输距离。将卫星运行过程划分成若干时隙,在每个时隙中,可以将信道视为不变。因此时隙的间隔只要充分小就可以假设路径损耗在该时隙内不发生改变。

时隙划分足够小需满足以下条件

$$[L(t + \Delta t) - L(t)]/L(t) \ll 1. \quad (12)$$

此外,卫星运动也会影响终端用户相对波束中心的位置,时隙的划分需满足以下条件

$$v \cdot \Delta t / 2r \ll 1, \quad (13)$$

其中: v 是卫星星下点的运动速度, r 是单个波束在地面的投影半径。

低轨卫星在轨道上持续运动,难以实现连续的子载波功率调整来适应其动态特性。将低轨卫星在轨运行的连续时间划分成若干离散时隙,通过上述时隙划分合理性的讨论,可以假设卫星相对地面终端的距离和信道状态信息不变,子载波在该时隙内维持固定的功率分配值。时隙改变后,卫星位置等状态信息发生改变,则改变功率分配值。该方案能够简化低轨卫星的动态特性。

2.2 功率分配的 Q 学习方案

传统的功率分配方法通常利用先验知识进行决策,在复杂环境下效果不佳。由于每个时隙的功率分配操作只取决于卫星波束覆盖的当前状态而不是历史状态,因此低轨多波束卫星的功率分配可以建模成马尔科夫决策过程。由于信道条件是时变的,采用无模型强化学习算法——Q 学习算法,在转移概率未知情况下, Q 学习算法能够提高决策能力。一个典型的强化学习框架由智能体和环境组成,两者通过智能体的动作、环境的状态和奖励相互作用。在低轨卫星功率分配场景下,每个波束作为一个智能体,通过连续迭代使 Q 值收敛。状态、动作和奖励函数定义如下:

1) 状态空间: $s^{k,n} = \{M^k, p_t^k\}$, 其中 M^k 表示接入波束的用户数量, p_t^k 表示波束的功率等级。为了降低算法的复杂度和网络的状态空间,将功率

值划分成若干离散等级。离散规则如下:

$p_t^k = \phi$, 当 $(P_{\max} - p_\phi) \leq \sum_{n=0}^N p_t^{k,n} < (P_{\max} - p_{\phi+1})$ 时, 其中 $\phi \in \{0, 1, \dots, U\}$, $p_0 = p_{\max}$, $p_U = 0$ 。功率离散度 U 由卫星具体运算能力决定,显然 U 越大,功率等级划分约细粒度,计算越精确,但是占用的存储空间越大。

2) 动作: 在马尔科夫决策过程中,智能体根据系统状态 $s^{k,n}$ 进行决策,动作决策应该满足约束条件,动作设计由如下二元组构成

$$\mathbf{A} = \{a_1, a_2, \dots, a_{3N} \mid a_k = (n_k, \Delta p_t^{k,n})\}, \quad (14)$$

其中: n_k 表示第 k 个波束的第 n 个子载波, $\Delta p_t^{k,n}$ 表示子载波的功率调整量。由于子载波有 N 个,功率调整量有 3 种选择(包括增加、减少和不变),故动作共有 $3N$ 种选择。

$$\Delta p_t^{k,n} \in \{-|\Delta p_t^{k,n}|, 0, +|\Delta p_t^{k,n}|\}. \quad (15)$$

对于子载波所服务用户附近的 h 个用户,若增加功率后这 h 个用户的总体容量 C_t^h 增加,则采取动作为 $\Delta p_t^{k,n} = +|\Delta p_t^{k,n}|$, 反之,则采用动作 $\Delta p_t^{k,n} = -|\Delta p_t^{k,n}|$, 若总体容量不变,选择 $\Delta p_t^{k,n} = 0$ 。此外,如果增加功率导致总功率超过卫星最大发射功率,或者减少功率导致子载波功率为负值,则功率调整量为“不变”,即 $\Delta p_t^{k,n} = 0$ 。Q 学习动作更新采用 ε -greedy 方式。

3) 奖励函数: Q 学习过程中,智能体需要通过决策获得最大的累计回报,奖励值在智能体采取动作后通过奖励函数计算获得,奖励函数设计越接近优化目标,越能达到更好的性能。该场景下,智能体采取动作后,应使吞吐量增加,同时波束总功率小于卫星最大功率,基于该原则第 k 个波束在第 n 个子载波上的奖励定义如下所示:

$$r_{ct}^k = \begin{cases} \sum_{n=1}^N 1 - e^{-(C_t^{(k,n)})}, & \sum_{n=1}^N p_t^{k,n} < P_{\max} \\ -1, & \text{Otherwise} \end{cases} \quad (16)$$

此外,为了保证用户 QoS,可引入服务阻塞率。用户速率若小于设置的传输速率阈值 C_{th} , 则判定为阻塞,服务阻塞率定义为用户阻塞个数 N_{bk} 与总用户 N_{us} 之比,即 $p_{bk} = N_{bk}/N_{us}$ 。

引入阻塞率后的奖励函数定义为

$$r_t^k = \beta_1 r_{ct}^k + \beta_2 r_{bu}^k, \quad (17)$$

其中: $r_{bu}^k = e^{-p_{bk}}$, 阻塞率越低,奖励值越高; β_1 和 β_2 是权重因子。

4) Q 学习迭代过程: 单个波束通过连续的迭

代学习更新行为价值函数。行为价值函数由一个 Q 值表 $Q(s,a)$ 表示,其中 a 属于动作集合 A , s 属于状态集合。 $Q(s,a)$ 表示在状态 s 执行动作 a 时,从一个无限时间的角度计算累计奖励值的期望值, Q 值更新方式具体如下

$$Q(s_t, a_t) = (1 - \alpha) Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a)], \quad (18)$$

其中: R_{t+1} 表示在状态 S_t 采取动作 a_t 后的即时奖励值。

由于低轨卫星的动态特性,尽管对功率值做了离散化处理,波束覆盖网络的状态空间仍然很大,并且每个智能体的 Q 值表不同。因此,采用动态状态添加方法,只有当新状态出现,该状态才会被自动添加到状态表。

Q 学习算法根据过去出现过的状态,统计和迭代 Q 值。一方面 Q 学习适用的状态和动作空间非常小;另一方面 Q 学习泛化能力差。

2.3 功率分配的深度强化学习(DQN)方案

Q 学习算法的值函数是通过 Q 值表进行更新的,而深度 Q 学习(deep Q-network, DQN)通过神经网络的参数更新来进行值函数更新^[13]。更新方式采用了梯度下降算法,值函数更新如下

$$w_{t+1} = w_t + \alpha \frac{1}{U_{mb}} \sum_{i=1}^{U_{mb}} [r + \gamma \max_a Q^-(s, a; w^-) - Q(s_i, a_i; w)] \nabla Q(s_i, a_i; w), \quad (19)$$

其中: $r + \gamma \max_a Q^-(s, a; w^-)$ 是时序差分目标, $Q^-(s, a; w^-)$ 是通过值函数近似的网络目标; U_{mb} 是批量更新的大小。

引入深度神经网络拟合 Q 值表,该网络包含 4 层神经层,输入层包含接入第 k 个波束的用户 m 所占用的子载波 n 的信道状态信息。中间的两个隐藏层主要增加网络优化的非线性,提高网络的适应能力。输出层包含所有动作对应的 Q 值。神经网络基本结构如图 2 所示。

利用目标网络 Q^- 计算标签值,作为学习的目标,标签值 y_t 计算式为

$$y_t = \begin{cases} r_t, & a_{t+1} = \emptyset \\ r_t + \gamma \max_a Q^-(s, a; w^-), & \text{others} \end{cases} \quad (20)$$

损失函数计算采用以下方式:

$$L(w) = E(y_t - Q(w))^2 + c \|w\|_2, \quad (21)$$

其中: $Q(s,a)$ 是 Q 学习计算得出的功率分配策略,由公式 (17) 迭代训练获得。损失函数第一部

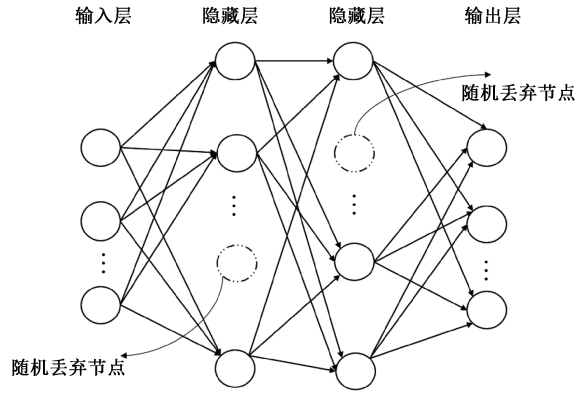


图 2 DQN 神经网络的基本结构

Fig. 2 Basic structure of DQN neural network

分是为了使神经网络尽可能拟合 Q 值表,第二部分是正则化项,约束参数 w 避免网络过拟合, c 是惩罚因子, $c \in (0,1)$ 。利用随机梯度下降法 Adam 优化算法进行训练,来最小化损失函数 $L(w)$ 。

除了对参数正则化处理,网络隐藏层以固定概率随机丢弃一部分节点,提高网络结构的多样性,进一步防止网络过拟合^[14]。

基于深度强化学习的低轨卫星功率分配方案的架构如图 3 所示。

本文提出的基于 DQN 的功率分配算法具体实现如下:

1) 初始化阶段

①初始化通信场景,建立低轨卫星多波束网络覆盖模型。初始化信道参数、时隙长度、用户和卫星位置更新参数。

②初始化 DQN 方案中的相关参数。低轨卫星上搭载的信息控制中心设定学习速率 α ,折扣因子 γ ,初始探索概率 ε 。设置训练网络的周期数 N_{epochs} 、经验池大小 U_{max} 、训练数量 U_{st} 、采样大小 U_{mb} 、时隙计数器 N_c 、目标网络更新时隙间隔 D 。

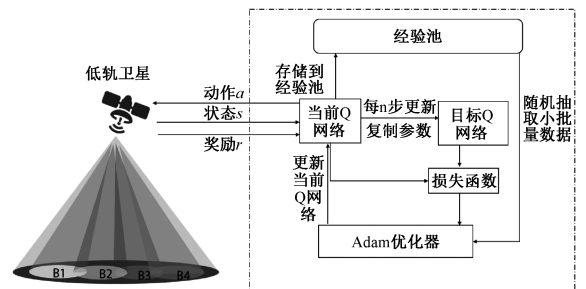


图 3 低轨卫星功率分配的 DQN 方案架构

Fig. 3 DQN scheme architecture for LEO satellite power allocation

③平均分配初始化功率。

2) 训练和运行阶段

①设置循环周期 $\text{epoch} = 1$ 。每执行一遍循环过程(循环过程包括下面步骤(2)至(5)), epoch 计数加 1,直到循环 N_{epochs} 次停止。

②信息控制中心观测状态 s_t , 即此时连接到各波束的用户数量和各波束的功率等级。波束以 ε 的概率随机选取动作 a_t , 即对任意子载波调节功率;以 $1-\varepsilon$ 的概率选取最优动作 a_t , 选取子载波调整功率。获得新状态 s_{t+1} , 依据式(16)或式(17)计算状态的即时奖励,采用式(16)的算法记为最优 DQN 算法,采用式(17)的算法记为公平 DQN 算法。将 (s_t, a_t, r_t, s_{t+1}) 4 元组存储到经验池中,若经验池的容量超过 U_{max} ,则丢弃最早的 4 元组。

③当循环周期 epoch 满足条件: $\text{epoch} > U_{st}$ 时,从经验池中随机抽取 U_{mb} 个样本,由式(20)计算标签值,根据式(21)计算损失值 $L(w)$,用 Adam 优化算法训练当前 Q 网络并更新权值 w 。

④当 $\text{mod}(t, D) = 0$ 时(即每经过 D 个时隙),更新目标网络 Q^- 参数 $w^- = w$ 。

⑤当前时隙 t 的计数增 1,进入下一时隙,若 $t > N_c$,停止本轮循环,执行下一步骤,否则继续转步骤②执行。

⑥更新循环周期 epoch ,增加 1,若 $\text{epoch} > N_{\text{epochs}}$,停止循环,输出已经训练好的当前 Q 网络进行动作决策,否则继续转步骤②执行。

3 仿真分析

3.1 仿真场景和参数设置

仿真场景由若干卫星波束和随机分布在卫星波束覆盖区域的用户组成。在波束重叠区域每个用户根据最大 SINR 连接到卫星。假设噪声功率为 $\sigma^2 = 10^{-7}$ W。考虑到地面用户终端的移动速度远远小于卫星的移动速度,例如卫星在 780 km 轨道高度时,运行速度 7.46 km/s,因此可以忽略用户的移动性。卫星波束在地面的覆盖区域朝一定方向以固定速度移动,因此仿真可以将卫星位置的移动过程转换为用户位置在卫星波束覆盖区域中的移动。用户朝一定方向以和卫星的相对速度运动,移出卫星覆盖范围的用户,在反方向随机新增相同用户数,位置在每个时隙更新一次。仿真参数设置参考现有低轨卫星通信技术^[15-17],具体参数设置如表 1 所示。仿真运算依托 Python 语

表 1 参数配置

Table 1 Parameter configuration

参数/单位	数值
轨道高度/km	780
子载波数/个	32
卫星覆盖半径/km	2 200
卫星最大发射功率/dBW	30
子载波最小发射功率/W	0
工作带宽/MHz	50
天线增益/dBi	30
单颗卫星波束个数/个	16
用户数量/个	150
卫星中心频率/GHz	30
波束偏轴角/rad	0.175
传输速率阈值/(kbit/s)	500
功率离散度	9
DQN 学习率 α	0.4
DQN 贪婪因子 ε	0.2
DQN 折扣因子 γ	0.9
目标网络更新间隔	20

言和 Pycharm 编译环境运行,使用 Keras 库对模型进行构建和训练。

时隙长度为 4 s,根据卫星轨道参数计算结果如下,满足时隙足够小的条件:

$$\begin{aligned} [L(t + \Delta t) - L(t)]/L(t) &\approx 0, \\ v \cdot \Delta t/2r &= 0.096. \end{aligned}$$

根据波束个数和子载波个数 DQN 输入参数个数为 5 328,设置隐藏层 1 参数 8 000,参数丢弃率 0.5,隐藏层 2 参数 4 000,参数丢弃率 0.4,输出层参数 1 536 个。

3.2 仿真结果及分析

为了体现所提功率分配方案的性能,多次计算最优 DQN 算法的收敛曲线并做平均。图 4 显示模型收敛过程的频谱效率。对于每一次迭代,每过 4 s 更新时隙,用户和卫星位置改变,信道环境随之更新,结果中虚线表示最优 DQN 算法单次收敛过程,黑色实线表示 10 次 DQN 算法收敛过程的平均结果。由图像可见深度强化学习算法通过多轮迭代在收敛之后能够达到较高的频谱效率。并且收敛速度较快,从 10 次算法的运行结果统计来看,卫星经过约 60 次迭代就能使算法收敛,可达到稳定高效的频谱效率,平均约 4.12 (bit/(s·Hz))。在少数情况下,如用户初始位置不同,经过约 50 次迭代就能使算法接近收敛。公平 DQN 算法权重因子 $\beta_1 = 0.7, \beta_2 = 0.3$,相比最优 DQN 算法其收敛速度更慢,曲线波动更大。

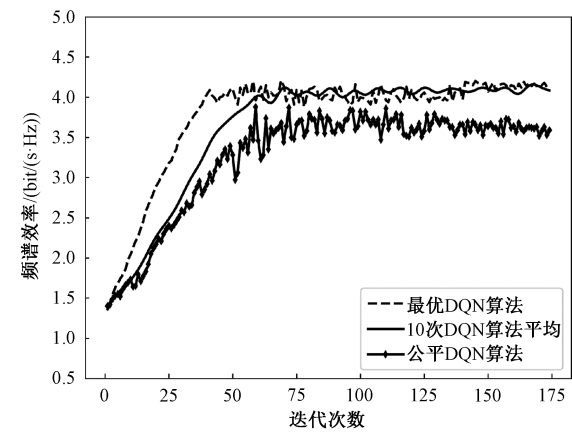


图 4 算法收敛图像

Fig. 4 Algorithm convergence image

为了分析所提功率分配算法的有效性,对本文所提的两种 DQN 算法和 Q 学习算法、注水功率分配算法进行对比^[18],如图 5 所示。在每个时隙计算一次系统的频谱效率。注水算法的基本思想是对信道条件好的多分配功率,对信道条件差的少分配功率,没有考虑同频信道干扰,也忽略了用户的位置分布。可以看出最优 DQN 算法的瞬时频谱效率明显更高,也更加稳定。

仿真实验还统计了各个算法对应的用户阻塞率。其中,公平 DQN 算法通过奖励函数的设计尽量保证每个用户的速率不低于一个阈值。实验结果如表 2 所示。可以看出,公平 DQN 算法在奖励函数设计中考虑用户阻塞率后,相比最优 DQN 算法整体频谱效率有所降低,但是阻塞率变得更小,保证了一定的公平性。并且公平 DQN 算法的频谱效率仍然高于注水算法和 Q 学习算法。注水功率分配则将大部分功率分配给了信道增益较大的用户。

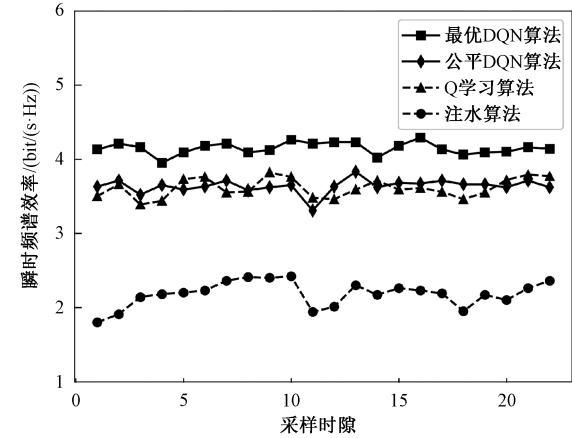


图 5 4 种方案的频谱效率对比

Fig. 5 Comparison of spectrum efficiency of four solutions

表 2 用户阻塞率对比		
Table 2 The comparison of user blocking probability		
功率分配方法	用户阻塞率	频谱效率/(bit/(s·Hz))
最优 DQN 算法	0.43	4.12
公平 DQN 算法	0.37	3.61
Q 学习算法	0.41	3.57
注水功率分配	0.59	2.24

最后,比较了各个算法在不同噪声功率环境下的频谱效率,如图 6 所示。其中,公平 DQN 算法和注水算法的频谱效率随着噪声功率的升高下降更快,注水功率算法在噪声较高时频谱效率会受到严重的影响。

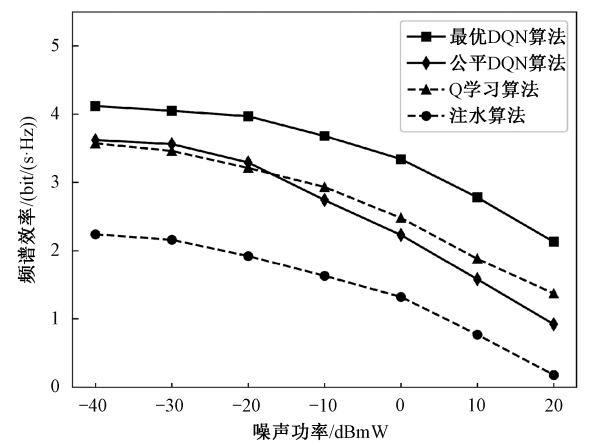


图 6 不同噪声功率下的频谱效率

Fig. 6 Spectrum efficiency under different noise power values

4 结论

本文针对低轨卫星的功率分配问题,提出一种基于深度强化学习的解决方案,将卫星运动过程进行时隙划分,以单个时隙作为强化学习训练过程中的单步,提取该时隙中的用户位置、信道增益信息,并对时隙大小进行了合理划分。进一步针对性地设计强化学习的状态、动作和奖励函数,引入深度神经网络加速收敛。为了减少状态空间,降低算法复杂度,对功率值进行了离散化处理。

在具体仿真场景下验证算法的性能,结果显示,相对功率注水算法,在带宽固定情况下,本方案能明显提升系统容量,即提升了频谱效率,可以满足更大的流量需求,适用于数据业务。在引入服务阻塞率后,所提方案能在一定程度上降低阻塞率,且频谱效率依然较高,该方案为低轨卫星通信系统资源分配提供了一种思路。

参考文献

[1] 汪春霆, 李宁, 翟立君, 等. 卫星通信与地面 5G 的融合初探(一)[J]. 卫星与网络, 2018(9): 14-21. DOI: 10.3969/j. issn. 1672-965X. 2018. 09. 004.

[2] 汪春霆, 李宁, 翟立君, 等. 卫星通信与地面 5G 的融合初探(二)[J]. 卫星与网络, 2018(11): 22-26, 28. DOI: 10.3969/j. issn. 1672-965X. 2018. 11. 005.

[3] Choi J P, Chan V W S. Optimum power and beam allocation based on traffic demands and channel conditions over satellite downlinks [J]. IEEE Transactions on Wireless Communications, 2005, 4(6): 2983-2993. DOI: 10.1109/TWC. 2005. 858365.

[4] Alexis I, Shankar B, Arapoglou P, et al. Power allocation in multibeam satellite systems: A two-stage multi-objective optimization [J]. IEEE Transactions on Wireless Communications, 2015, 14(6): 3171-3182.

[5] Nakahira K, Kobayashi K and Ueba M. Capacity and quality enhancement using an adaptive resource allocation for multi-beam mobile satellite communication systems [C] // IEEE Wireless Communications and Networking Conference WCNC 2006. April 3-6, 2006, Las Vegas, NV, USA: IEEE, 2006: 153-158. DOI: 10.1109/WCNC. 2006. 1683456.

[6] 史煜, 张邦宁, 郭道省, 等. 考虑波束间干扰的多波束卫星功率带宽联合分配算法[J]. 计算机工程, 2018, 44(2): 103-106, 113. DOI: 10.3969/j. issn. 1000-3428. 2018. 02. 018.

[7] Fu A C, Modiano E, and Tsitsiklis J. Optimal energy allocation and admission control for communications satellites [J]. IEEE/ACM Transactions on Networking, 2003, 11(3): 488-500. DOI: 10.1109/TNET. 2003. 813041.

[8] Qiu C, Yao H, Yu F R, et al. Deep Q-learning aided networking, caching and computing resources allocation in software-defined satellite-terrestrial networks [J]. IEEE Transactions on Vehicular Technology, 2019, 68(6): 5871-5883. DOI: 10.1109/TVT. 2019. 2907682.

[9] 3GPP. Study on New Radio (NR) to support non terrestrial networks (Release 15): 3GPP TR 38.811 [S/OL]. (2018-8-10) [2020-6-20]. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3234>.

[10] Ippolito L J, Joseph L. Satellite communications systems engineering: atmospheric effects, satellite link design and system performance [M]. Chichester, UK: John Wiley & Sons, 2017.

[11] Christopoulos D, Chatzinotas S, Zheng G, et al. Linear and nonlinear techniques for multibeam joint processing in satellite communications [J]. EURASIP Journal on Wireless Communications and Networking, 2012, 162(2012): 1-13. DOI: 10.1186/1687-1499-2012-162.

[12] 刘帅军. 卫星通信系统中动态资源管理技术研究[D]. 北京: 北京邮电大学, 2018.

[13] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533. DOI: 10.1038/nature14236.

[14] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[15] 翟继强, 李雄飞. OneWeb 卫星系统及国内低轨互联网卫星系统发展思考[J]. 空间电子技术, 2017, 14(6): 1-7. DOI: 10.3969/j. issn. 1674-7135. 2017. 06. 001.

[16] Pratt S R, Raines R A, Fossa C E, et al. An operational and performance overview of the IRIDIUM low earth orbit satellite system [J]. IEEE Communications Surveys, 1999, 2(2): 2-10. DOI: 10.1109/COMST. 1999. 5340513.

[17] 赵星惟, 吕源, 刘会杰, 等. LEO 通信卫星多波束天线构型方案设计[J]. 中国科学院研究生院学报, 2011, 28(5): 636-641. DOI: 10.7523/j. issn. 2095-6134. 2011. 5. 011.

[18] 张冬梅, 徐友云, 蔡跃明. OFDMA 系统中线性注水功率分配算法[J]. 电子与信息学报, 2007, 29(6): 1286-1289.