基于深度强化学习的卫星信道动态分配算法

唐一强，杨霄鹏，肖　楠，朱圣铭
（空军工程大学信息与导航学院，西安，710077）

摘要　在卫星通信系统中，频率和信道是十分珍稀的资源，针对如何利用可靠且高效的方法来进行资源的开发这一亟需解决的难题，提出了一种基于 *Q*-learning 深度强化学习的动态卫星信道分配算法 DRL-DCA，该算法将卫星和环境交互建模为马尔科夫决策过程，通过环境的反馈提升卫星的决策能力，实现用户业务请求的高效应答，提升卫星通信的服务质量，降低通信阻塞发生概率。仿真分析表明该算法能够有效地提升通信的吞吐量，降低通信的阻塞率。

关键词　卫星通信；深度学习；*Q* 算法
DOI　10.3969/j.issn.1009-3516.2022.02.010
中图分类号　TN927　　**文献标志码**　A　　**文章编号**　1009-3516(2022)02-0061-07

A Dynamic Allocation Algorithm of Satellite Channels Based on Deep Reinforcement Learning

TANG Yiqiang, YANG Xiaopeng, XIAO Nan, ZHU Shengming
(Information and Navigation School, Air Force Engineering University, Xi'an 710077, China)

Abstract　In satellite communication systems, frequencies and channels are very rare resources. How to use reliable and efficient methods to develop resources has become a severe problem that needs to be solved urgently. This paper proposes a dynamic satellite channels allocation algorithm DRL-DCA. This algorithm is to model satellite and environment interaction on a Markov decision-making process, improving satellite decision-making ability through environmental feedback, realizing efficient response to user business requests, improving the service quality of satellite communication, and reducing the probability of communication blocking. The simulation analysis shows that the proposed algorithm can effectively improve the communication throughput and reduce the communication blocking rate.

Key words satellite communication; deep learning; *Q*-learning algorithm

卫星通信以其特殊的优势，已经成为移动通信领域中不可替代的重要部分，其主要的特点如下[1]：通信容量大，最高支持超过 300 Gbps 的容量；通信范围广，理论上只需 3 颗地球同步轨道卫星(GEO)就可以覆盖除南北极以外的全球区域；能够实现“动中通”，支持包括空中飞行器、陆地移动设备和海上移动设施的移动中不间断通信；业务种类丰富，涵盖了语音、图像、视频等多业务的移动通信。

与此同时，不断增长的用户业务对卫星通信的要求越来越高，传统的卫星通信技术已经不能满足

收稿日期：2021-07-05
基金项目：国家自然科学基金(61871474)
作者简介：唐一强(1997—)，男，四川绵阳人，硕士生，研究方向为低轨道卫星通信的信道资源分配优化。E-mail:2927247454@qq.com

引用格式：*唐一强，杨霄鹏，肖楠，等. 基于深度强化学习的卫星信道动态分配算法[J]. 空军工程大学学报(自然科学版)，2022，23(2)：61-67. TANG Yiqiang, YANG Xiaopeng, XIAO Nan, et al. A Dynamic Allocation Algorithm of Satellite Channels Based on Deep Reinforcement Learning[J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(2): 61-67.*

业务增长的需求。多波束天线[2]是解决这个问题的重要方法[2]。多波束天线使用卫星蜂窝通信,利用多个具有高增益的点波束实现目标区域的通信,从而提升系统的频带和容量资源。但是,地球上的终端用户往往分布是不均匀的,这会造成卫星各个波束之间的业务量差别很大,信道资源的需求不尽相同。因此,必须对卫星信道资源进行合理的调度,以提高卫星通信系统的性能。目前使用最广泛的信道资源管理主要包括固定信道分配(FCA)、混合信道分配(HCA)和动态信道分配(DCA)[3]。

在固定信道分配方式中,即使在波束中没有用户使用信道资源,其他的波束用户也不能使用该信道资源,这不仅造成信道资源的浪费,同时会增加网络拥塞,降低网络吞吐量。混合信道分配方案由固定信道分配和灵活信道分配两部分组成,但是固定信道分配占比很大,导致信道资源利用率往往较低。在动态信道分配方案中,会充分考虑用户的业务需求、信道增益和业务拥塞等影响因素,避免同频干扰和波束间共信道干扰(co-channel interference, CCI)的影响,允许各波束任意选择可用的信道,可以在保证用户业务质量的前提下最大化信道资源利用效率[4]。

文献[5]提出了一种结合终端的地理位置信息为用户实时地分配资源的融合波束覆盖信道动态分配算法(fusion beam coverage-dynamic channel allocation algorithm ,FBC-DCA),结果表明该算法对于带宽利用率有一定的提升。文献[6]分析了以最大化容量为目标的资源自适应动态分配算法,结果显示该算法比较适合非对称的通信需求。文献[7]提出基于多终端和多业务优先级的动态信道分配算法,算法根据不同终端和不同业务级别分配信道,结果表明该算法能较好地提升满意度。文献[8]提出一种为多类型呼叫、多类型业务和多类型终端的动态信道预留策略,结合遗传与粒子群混合算法动态求解最佳的预留信道阈值分布,仿真表明算法能够较好地为高等级用户提供满意的服务质量。文献[9]结合用户的运动状态,对用户的预测轨迹上择取抽样点,把这些点的平均干扰选为动态信道分配的指标,仿真结果得出该策略能将用户的平均信噪比提升大约 0.5 dB。

通过对上述关于 DCA 算法研究现状的分析可知,目前 DCA 算法更多地关注单一独立时刻的信道分配,忽略了当前时刻的信道分配会对之后的信道资源产生影响,即信道分配具有时域相关性,这就造成了信道资源使用不够充分和阻塞率高的问题。深度强化学习能够有效解决具有时域相关性的序列决策问题[10],因此本文提出一种基于深度强化学习的动态卫星信道分配算法 DRL-DCA,将卫星和环境交互建模为马尔科夫决策过程,通过环境的反馈提升卫星的决策能力。

1 深度强化学习技术

强化学习(reinforcement learning, RL)是指一类从(与)环境交互中不断学习的问题以及解决这类问题的方法,包含智能体和环境 2 个可以交互的对象。

智能体(agent)可以感知外界环境的状态(state)和反馈的奖励(reward),并进行学习和决策。环境(environment)是智能体外部的所有事物,并受智能体动作的影响改变其状态,并反馈给智能体相应的奖励。如图 1 所示,智能体的决策功能是根据外界环境的状态来做出不同的动作(action),而学习功能是根据外界环境的奖励来调整策略。

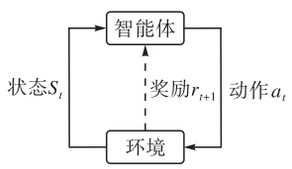

图 1 智能体与环境的交互

深度强化学习则是利用深度学习的强大感知能力来定义问题和优化目标,利用强化学习强大的决策能力来解决策略和值函数的建模问题[4]。

智能体每一个动作并不能直接得到监督信息,需要通过整个模型的最终监督信息(奖励)得到,智能体在当前状态下的工作不会立刻被评判,需要在下一个状态时获得奖励或惩罚[11]。这种算法是用不断试错的学习方式得到最优的策略,可使得决策持续获得收益。在本文研究的模型中,我们将深度强化学习建模为一组马尔科夫决策过程(Markov decision process,MDP)。

2 系统模型

本文将卫星建模为智能体,地面用户业务量建模为环境,卫星获得的收益值通过最优化目标函数对信道分配作出决策。具体而言,卫星在时刻 t 通过对用户和业务的观测获得状态 $S_t \in S$ 的抽象表达,接着卫星根据优化策略选择执行动作 $a_t \in A(s_t)$,以概率 $p_Q(s_{t+1} \mid s_t, a_t)$ 进入下一个状态,与此同时获得收益 r_{t+1},由奖励信号对动作策略进一步优化更新,使智能体的收益最大化。

在本文的卫星多波束间动态信道分配问题中，考虑卫星在地面的多波束为 $n\in\{n|n=1,2,\cdots,N\}$。卫星通信带宽 B_{tot} 被平均分成 M 个信道，每个信道的带宽为 $B_{\text{av}}=B_{\text{tot}}/M$，卫星的可用信道数目为 $m\in\{m|m=1,2,\cdots,M\}$，地面用户总数为 K 个，即时通信的用户数目 $k\in\{k|k=1,2,\cdots,K\}$。在这个系统中，每个用户终端 u 得到的信道资源分配用 $\boldsymbol{w}_u$ 表示，$\boldsymbol{w}_u=[w_{u,1},w_{u,2},\cdots,w_{u,m}]^{\text{T}}$，其中 $w_{u,m}$ 表示用户 u 在第 m 个信道上的增益。由此可已得到整个通信系统为所有即时用户分配信道的增益矩阵为 $\boldsymbol{W}=[w_1,w_2,\cdots,w_u]$，$\boldsymbol{W}\in\boldsymbol{R}^{M\times K}$。在下行链路上，卫星发送给用户 u 的卫星信号为：$\boldsymbol{s}_u=[s_{u,1},s_{u,2},\cdots,s_{u,m}]^{\text{T}}$，其中 $s_{u,m}$ 表示卫星在第 m 个信道上发送给用户 u 的信号。从卫星到用户终端的总衰减记为：$L=\{l_{i,j}|1\leqslant i,j\leqslant K\}$，$L$ 由自由空间路径损耗 $\boldsymbol{A}$、发射卫星天线增益 G_s 和用户端天线增益 G_u 构成。其中：

$\boldsymbol{A}=\text{diag}\{a_1,a_2,\cdots,a_k\}$；$G_s=\{g_{k,n}|1\leqslant k\leqslant K,1\leqslant n\leqslant N\}$；$G_u=\text{diag}\{g_1,g_2,\cdots,g_k\}$；$L=\boldsymbol{A}\cdot G_u\cdot G_s\cdot\boldsymbol{X}^{\text{T}}$。其中 $\boldsymbol{X}$ 表示用户 u 在最大信号接入准则的约束下选择波束接入的接入矩阵：

$$\boldsymbol{X}=\{x_{u,n}\mid x_{u,n}\in\{0,1\},\sum_{n=1}^{N}X_{u,n}=1 \tag{1}$$

式中：$x_{u,n}=1$ 表示用户 u 接入波束 n，否则表示用户没有接入波束 n。在用户终端 u 接收到的信号为：

$$y_k=\sum_{i=1}^{K}l_{u,i}w_i\otimes s_i+\sigma_k=l_{u,u}w_u\otimes s_k+\sum_{i=1,i\neq u}^{K}l_{u,i}w_i\otimes s_i+\sigma_k \tag{2}$$

式中：$\otimes$ 表示哈达玛积，最右式中第 1 项是用户的有用信号，第 2 项是共信道干扰，第 3 项是噪声。

下面计算用户信干噪比。首先，定义一个以信道为基的资源分配矩阵 $\boldsymbol{D}=\boldsymbol{W}^{\text{T}}$，$\boldsymbol{D}\in\boldsymbol{R}^{M\times K}$，记为 $\boldsymbol{D}_m=[d_{m,1},d_{m,2},\cdots,d_{m,k}]^{\text{T}}$，每一项表示卫星在信道 m 上的发射功率。由式(3)～(4)可知用户在各个信道上的有用信号功率 $\boldsymbol{P}_u$ 和共信道干扰功率 $\boldsymbol{I}_u$。

$$\boldsymbol{P}_u=|l_{u,u}|^2\text{diag}\{w_u\}[\text{diag}\{w_u\}]^{\text{H}} \tag{3}$$

$$\boldsymbol{I}_u=\text{diag}\{[g_u\cdot\boldsymbol{D}_m\cdot\boldsymbol{D}_m^{\text{H}}\cdot\boldsymbol{g}_u^{\text{H}}],m=1,2,\cdots,M\} \tag{4}$$

式中：$\boldsymbol{g}_u=[\boldsymbol{l}_{u,1},\boldsymbol{l}_{u,2},\cdots,\boldsymbol{l}_{u,k}]|(\boldsymbol{l}_{u,k}=0)$。由此可得干扰信号和噪声的和为：

$$\boldsymbol{P}_{nu}=\boldsymbol{I}_u+|\boldsymbol{\sigma}_k|^2\boldsymbol{E}_M \tag{5}$$

式中：$\boldsymbol{E}_M$ 表示 M 阶单位阵。由香农容量公式进一步推知用户 u 在资源分配下的理想可达速率为：

$$C_u=B_{\text{av}}\det[\log(\boldsymbol{E}_M+\frac{\boldsymbol{P}_u}{\boldsymbol{P}_{nu}})] \tag{6}$$

为了达到通信要求，可达速率不能低于某一个阈值，通常此阈值设为 C_{th}，只有当 $C_u\geqslant C_{\text{th}}$ 时，用户 u 才能正常通信，否则用户 u 将会掉话或阻塞。当用户有新的业务请求时，卫星系统查看目前是否存在可使用的信道资源，如果此时存在信道资源可供使用，卫星系统将会按照分配策略进行信道分配。卫星系统判断波束是否存在闲置的信道资源，主要从以下几个方面进行决策[12]：星上功率是否达到饱和状态、单个波束的功率是否达到饱和以及此次的信道分配是否会损害已分配用户的服务。为便于反映性能，定义一个性能指标 Ψ^t，以此来表示新的业务请求是否被阻塞。

$$\Psi^t=\begin{cases}0,\text{正常通信}\\1,\text{业务被阻塞}\end{cases} \tag{7}$$

在此，定义以波束为基的资源分配矩阵 $\boldsymbol{B}=[b_1,b_2,\cdots,b_n]$，每一项表示波束在对应的信道上的发射信号幅值大小。由此可推知卫星信道分配优化所有的约束条件为：

$$\sum_{n=1}^{N}\boldsymbol{b}_n^t\cdot(\boldsymbol{b}_n^t)^{\text{H}}\leqslant\boldsymbol{P}_{\text{all}},\forall t \tag{8}$$

$$\boldsymbol{b}_n^t\cdot(\boldsymbol{b}_n^t)^{\text{H}}\leqslant\boldsymbol{P}_b,\forall n,t \tag{9}$$

$$C_u\geqslant C_{\text{th}},\forall u\in U^t \tag{10}$$

$$\sum_{m=t}^{M}|w_{u^t,m}|^2\leqslant P_c,|w_{u^t,m}|^2\in\{0,P_c\},\forall m \tag{11}$$

式中：P_{all} 表示卫星的最大发射功率；P_b 表示单个波束的最大功率；U^t 表示当前时刻新请求的用户的信息。式(11)表示各个信道上发射的功率相同，并且单个用户只允许分配一个信道。

在上述的约束条件下，要达到的最优化目标可用式(12)表示：

$$P(U^t,W^t,u^t)=\min_{w_{u^t}}\sum_{t=1}^{T}\Psi^t \tag{12}$$

在此模型中我们用最小化阻塞率来衡量信道分配优劣[13-14]，阻塞概率计算公式为：

$$P_{\text{fail}}=U_{\text{block}}/U_{\text{arrival}} \tag{13}$$

式中：U_{block} 表示当前时刻被阻塞的用户数目，U_{arrival} 表示当前时刻达到卫星通信系统总的用户数目。

3　深度强化学习的信道分配算法分析

如图 2 所示，我们将卫星终端建模为智能体，把用户业务量和信道占用状态建模为环境，在智能体与环境交互的过程中使智能体的收益最大。智能体与环境的交互过程是一组马尔科夫决策过程：卫星根据对用户的观测获得当前的状态 S_t，接着卫星按照优化目标策略执行动作 a_t，在环境改变时以概率 $p(s_{t+1}|s_t,a_t)$ 转为状态 S_{t+1}，从环境中获得收益 r_{t+1}。

3.1 基于 *Q*-learning 算法的总体框架

Q-learning 是深度强化学习中一种非常经典的算法,该算法根据系统的状态-动作值函数 $Q(s,a)$ 进行不断地迭代更新,根据收益 r 评估选择接下来的动作的同时优化 Q 函数[15-16]。系统迭代公式为:

$$Q(s,a)\leftarrow Q(s,a)+\alpha[r+\gamma\max_{a'}Q(s',a')-Q(s,a)] \tag{14}$$

式中:α 为学习速率;γ 为折扣因子。可以看出 α 越大则 Q 值迭代后保留之前的效果越少;γ 越大,长期的回报对当前时刻的影响就越大。

如图 2 所示,在信道业务请求时刻,卫星根据当前的环境得到环境收益,经验池中的数目达到一定的量值后,每一次的训练过程都会从池中随机的选择一批数据,并与目标网络中的 Q' 一起对 Q 网络进行训练,改变 Q 值函数,完成卫星通信系统信道资源的灵活分配。

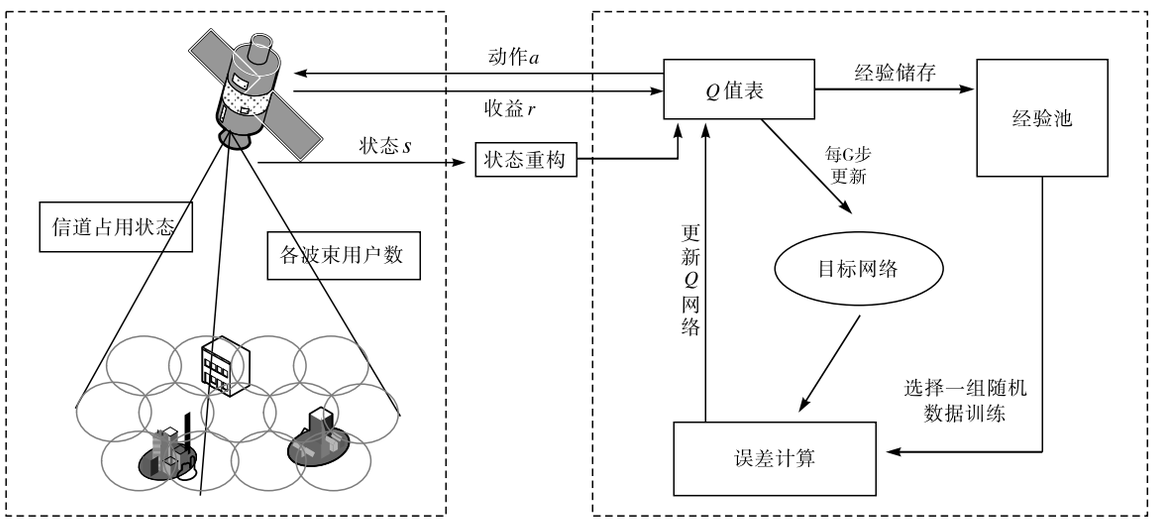

图 2 卫星系统信道分配模型

3.2 马尔科夫决策过程

为便于研究,将卫星与环境的交互看做是离散的时间序列。如图 3 所示,卫星从感知到的初始环境 S_0 开始,然后决定做出一个相应的动作 a_0,环境相应地发生改变到新的状态 S_1,并反馈给智能体一个即时的奖励 r_1,然后智能体又根据新的状态 S_1 作出下一个动作 a_1,环境改变为 S_2,反馈奖励 r_2,交互一直进行下去。

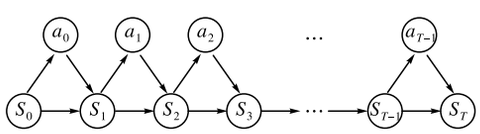

图 3 马尔科夫决策过程

马尔科夫决策过程在马尔科夫过程中加了一个额外的变量:动作 a,下一个时刻的状态 S_{t+1} 取决于当前的状态 S_t 和动作 a_t。

$$p(s_{t+1}\ s_t,a_t,s_{t+2}\ s_t,a_t,s_0,a_0)=p(s_{t+1}\mid s_t,a_t) \tag{15}$$

式中:$p(s_{t+1}\mid s_t,a_t)$ 为状态转移概率。

当到达终止状态时,交互过程就结束了,这一轮过程被称作一个回合(episode)或实验(trial)。为了应对环境中没有终止状态的情况发生,引入一个折扣回报来降低远期回报的权重[17],其定义为:

$$G(\tau)=\sum_{t=0}^{T-1}\gamma^t r_{t+1} \tag{16}$$

式中:$\gamma\in[0,1]$ 是折扣率。当 γ 接近于 0 时,智能体更关注短期的回报;当 γ 接近于 1 时,长期回报成为智能体考虑的重点。

3.2.1 状态空间设计

所有即时用户分配信道的增益矩阵 $\boldsymbol{W}^t=[w_1,w_2,\cdots,w_u]$,即时通信的用户数目 $k^t\in\{k\mid k=1,2,\cdots,K\}$,当前时刻新请求的用户的信息 U^t。文中的状态空间表示为:

$$s_t=F(W^t,K^t,U^t) \tag{17}$$

式中:F 表示函数映射关系。为避免和抑制同频干扰,相邻的小区不能使用同一信道。波束 n 对信道的占用情况 h_{up} 可以表示为:

$$h_{\text{up}}=\begin{cases}-1,\text{信道不可使用}\\0,\text{信道可使用}\\1,\text{信道可使用但被占用}\end{cases} \tag{18}$$

当前时刻的波束无可用信道资源时或者所有用户均有信道资源时,整个系统达到终止状态。系统没有达到终止状态时,卫星将继续根据当前时刻的可用信道资源进行动态的信道选取分配。

3.2.2 动作空间设计

卫星根据所处的环境和 Q 网络,动作依照概率

ε 选择最大状态-动作值 Q 函数去执行。首先确定状态 S_t 下的可执行动作集合 $A(S_t)$，在本文的场景下，$A(S_t) \subset M$，即可执行动作是可用信道的子集。由此可得出动作 a_t 的表达式为：

$$a_t=\{(n,m)\,|\,n\in N,\ m\in A(S_t)\} \tag{19}$$

动作 a_t 是在波束 n 和状态 S_t 下可用的信道集合之中，为其分配信道资源 m。当前时刻不存在可用信道时，可用信道为空，表示为 $A(S_t)=\varnothing$，这个时候的业务将会被阻塞无法正常进行通信。

3.2.3　收益空间设计

收益是卫星与环境交互过程中的回馈，一方面是对执行动作后的评价，另一方面也是信道资源分配的性能优劣的评估。在最优信道分配中，我们的目的在于降低阻塞发生的次数，提高服务的效率。所以，在文中设计收益与阻塞率呈负相关，收益为 0 时表示完全不能通信，所有的业务请求均被阻塞。用公式表示为：

$$r=R_{\max}(1-U_{\text{block}}/U_{\text{arrival}}) \tag{20}$$

式中：$R_{\max}$ 表示最大的奖赏值，$R_{\max}>0$。从式中我们可以看到，卫星通信系统的阻塞率越低，获得的收益就越大，通信系统的总体性能就越好。

3.3　算法实现

基于 3.2 节中所述的空间设计，本文基于深度强化学习的卫星信道分配算法的实现过程如下：

输入：状态空间 S，动作空间 $A(S)$，γ，学习速率 α，更新间隔 G，初始探索概率 $\varepsilon_{\text{init}}$；

初始化经验池和相关参数，$Q(S,a)$，$W=\varnothing$，$B=\varnothing$，$N_{\text{block}}=0$，$N_{\text{arrival}}=0$；
repeat
　$t=1$，T 个业务请求时刻；
　更新业务到达参数 $N_{\text{arrival}}=N_{\text{arrival}}+1$ 和探索概率
　$\varepsilon=\max(\varepsilon-\varepsilon_{\text{gap}},\varepsilon_f)$
　观测环境，得到即时奖励 r，依据 MDP 过程状态定义构建状态 S_t；
　计算当前的新业务波束的动作 $A(S_t)$；
　若无信道可用，即 $A(S_t)=\varnothing$，则：
　更新阻塞业务参数 $N_{\text{block}}=N_{\text{block}}+1$；
　由式(20)获得立即收益值 r；
　若有信道可以使用，则：
　由式(20)获得立即收益值 r；
　将 S_t，$A(S_t)$，r，到经验池中；
　以概率 ε 随机选择动作 $a_t\in A(S)$；
　否则，选择最大 Q 函数值的信道，即 $a_t=\arg\max_{a\in A_{(S_t)}}Q(S,a)$；
　进行信道分配，更新参数 W，B；
　Q 网络训练；
　在经验池中随机选择一批数据；
　根据式(14)对网络进行优化训练；
　时间每经过 G 步对 Q 网络复制到目标网络；
until $\varepsilon_f\leftarrow\varepsilon$
根据得到的 N_{block} 和 N_{arrival}，计算得到 $P_{\text{fail}}=N_{\text{block}}/N_{\text{arrival}}$

输出：最终策略 $\pi(S)=\text{argmax}_{a\in A_{(S_t)}}Q(S,a)$ 和最终的信道分配结果 $\boldsymbol{W}^t$；

算法结束

在上述算法中 ε_{gap} 表示衰减因子，ε_f 表示算法最终的探索概率。为了对探索和利用进行折中，本算法采用的是 ε 贪婪策略(ε-greedy)，随机地以 ε 概率进行动作。探索是指抛弃已经获得的信息，尝试一种新的方法，避免陷入到局部最优化，尽量实现全局最优；利用则是指按照获得的信息进行决策，充分开发历史经验信息的潜力[19]。算法中线性下降，逐渐减小，最终达到 ε_f，算法结束。

4　仿真结果与分析

本文的仿真基于 Matlab2019b 实验平台，分别选取了不同业务量分布、不同业务到达率作为仿真场景，并与固定信道分配(FCA)、混合信道分配(HCA)和融合波束覆盖信道动态分配算法(FBC—DCA)进行对比。实验结果表明，本文所提出的基于深度学习的卫星信道动态分配算法在多种场景下具有很好的性能，所有的场景下均有较低的阻塞率。

4.1　仿真参数设置

该算法仿真中，业务达到是参数为 λ 的泊松分布数据流(单位：次/业务时刻)，业务服务时长是 μ 的负指数分布(单位：业务时刻)。假定波束为 37 个，业务传输阈值 $C_{\text{th}}=500$ kbps，业务到达率 λ 次/业务时刻，业务时长随机变化，服务率 μ 固定为 20 个/业务时刻。在神经网络中为了减少 Q 网络训练过程中出现较大的波动，我们设定较小的学习率 $\alpha=0.01$。DRL-DCA 算法的主要仿真参数如表 1 所示。

表 1　DRL-DCA 仿真参数表

仿真参数	仿真数值
波束个数 N	37
信道个数 M	25
学习速度 α	0.01
激活函数	sigmoid
业务阻塞收益值	0
业务满意收益值	1
折扣因子 γ	0.99
初始探索概率 ε	1.0
最终探索概率 ε_f	0.01
业务传输阈值 C_{th}/kbps	500

定义以下 2 个性能指标,以便更好地对比算法的差别[20-21]:

吞吐量:卫星通信系统中单位时间内成功地传输的数据的数量,与算法的性能密切相关。

阻塞率:卫星通信系统中,信道处于繁忙状态的概率。

4.2 性能分析

如图 4 所示,系统吞吐量随着业务到达率的增加而增大,达到一定数量值后趋于稳定。这是因为在通信信道的数目足量时,单位时间内到达的业务量越多,系统的吞吐量自然愈多。到达率超过某一数值后,由于系统信道数目的限制,吞吐量不再随着到达率的增加而改变,总体上趋于一个定值。在稳定的状态下,DRL-DCA 的稳定吞吐量约为 2.4×10^7 bit,FBC-DCA 算法的吞吐量约为 2.2×10^7 bit,HCA 的吞吐量约为 2×10^7 bit,FCA 的吞吐量约为 1.8×10^7 bit。DRL-DCA 的吞吐量较 FBC-DCA 算法高出 0.2×10^7 bit,较 FCA 高出 0.6×10^7 bit,较 HCA 高出 0.4×10^7 bit。在本文的仿真条件及仿真稳定状态下,DRL-DCA 的吞吐量大约是 FBC-DCA 的 1.1 倍,是 HCA 的 1.2 倍,是 FCA 的 1.3 倍。仿真结果表明,本文所提出的 DRL-DCA 算法能够有效地提升系统吞吐量,改善通信质量。

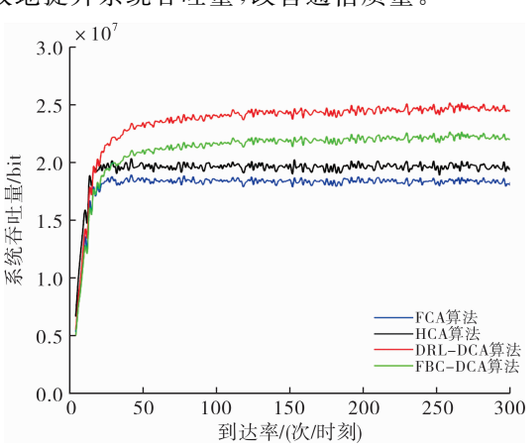

图 4 3 种算法的吞吐量比较

下面讨论本文提出的 DRL-DCA 算法对于不同业务分布的通信性能。图 5 展示了用户具有不同的业务量的阻塞率,图 6 展示了用户不同业务量需求条件下的平均到达率下的阻塞率。可以看出,通信的阻塞率随着到达率的增加而增大。按照排队论知识,在到达率小于服务率 μ(20 次/时刻)时,卫星信道够及时处理到达的业务。但当到达率超过系统的服务率 μ 时,业务就需要排队等待处理,系统阻塞率就随之增加。从图 5 与图 6 的对比中可以发现,用户具有相同业务量的通信阻塞率要低于不同业务量的阻塞率,这也证实了目前不同业务量条件下通信的严峻形势。在较大到达率时,本文提出的 DRL-DCA 算法在阻塞率性能上均低于另外 3 种算法的阻塞率。

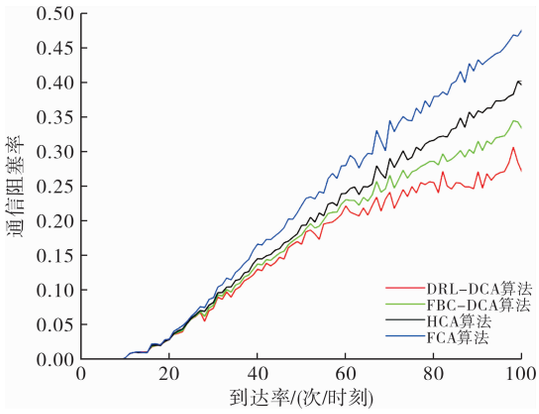

图 5 用户不同业务量的阻塞率

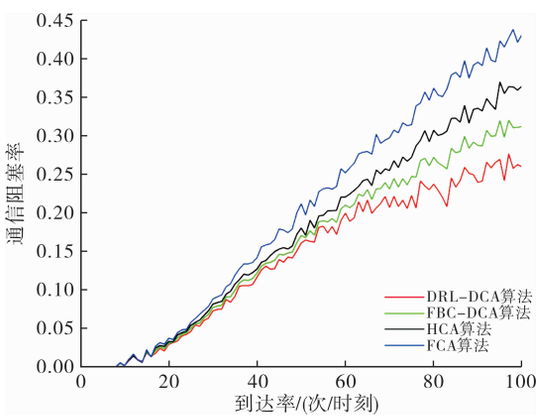

图 6 用户平均业务量的阻塞率

图 7 展示了 DRL-DCA 算法的收敛性变化趋势。以 $\lambda=100$ 次/时刻作为算法的收敛性分析的仿真场景。从图中可以看出,在前 2 500 步内算法的性能没有明显的改善,这主要是因为经验池中必须满足一定的经验数量时,才会按照信道分配的经验对网络进行训练,在训练中,网络迅速优化,大约在 3 000 步时算法性能趋于稳定。在多波束卫星的实际运行中,系统将会产生众多的经验条目,这些条目会有助于算法的训练[21],系统也将能够在短时间内收敛。

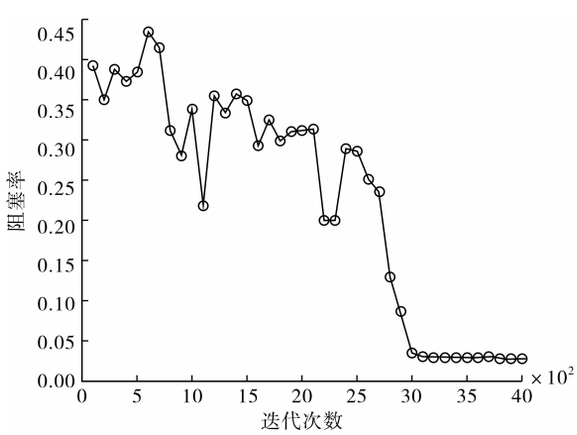

图 7 DRL-DCA 算法的收敛性分析

5　结语

本文在强化学习的基础上，以 Q-learning 算法为思路进行卫星信道的动态分配，详细地分析了算法的实现流程，并在仿真参数的设定下对本文所提出的算法进行了实验仿真，最后通过 Matlab2019b 仿真得出结果，对本文的算法进行佐证。

仿真结果表明，本文提出的 DRL-DCA 算法能够有效地提升卫星通信系统的吞吐量，能够在较大业务量的情况下，提升系统的吞吐量，降低系统的阻塞率，提升用户的使用体验。通过对比 4 种算法，我们发现不同的业务量将会对系统的通信质量产生重大影响，不同的算法表现出较大的性能差异。总体上来说，本文提出的 DRL-DCA 算法能够达到较低的阻塞概率和较高的吞吐量，多波束卫星通信的性能有了实质性的提高。

参考文献

[1] 杨廷卿，林善亮. 卫星移动通信系统组成及应用的探讨[J]. 通讯世界，2020，27(1)：120-121.

[2] 杨宇龙. 多波束卫星通信系统的资源分配算法研究[D]. 哈尔滨：哈尔滨工程大学，2020.

[3] 谢奕钊，易爱. 低轨卫星通信系统信道分配策略分析[J]. 电子测试，2019(13)：94-95，123.

[4] ZOU Q，ZHU L. Dynamic Channel Allocation Strategy of Satellite Communication Systems Based on Grey Prediction[C]//2019 International Symposium on Networks，Computers and Communications (ISNCC). Istanbul，Turkey：IEEE，2019：1-5.

[5] 丁亚南，庞文镇，张艳君，等. GMR-1 卫星移动通信系统中融合波束覆盖的动态信道分配算法[J]. 移动通信，2020，44(9)：43-57.

[6] 胡圆圆，宋高俊. Ka 频段下多波束卫星通信的资源分配[J]. 通讯技术，2013，46(10)：22-25.

[7] 别玉霞，卜瑞杰，刘海燕. 多优先级的卫星网络信道分配算法[J]. 计算机科学，2017，44(3)：132-136，144.

[8] 郭佳妮. 适应高速终端的卫星移动通信系统信道分配策略研究[D]. 北京：北京邮电大学，2017.

[9] 李航，赵明，王京. 阴影衰落信道下多波束卫星移动通信系统的动态信道分配策略[J]. 电讯技术，2016，56(6)：618-623.

[10] 刘召，许珂. 多波束卫星动态信道资源分配算法[J]. 移动通信，2019，43(5)：27-32.

[11] XU X，WANG C，JIN Z. Perturbed ISL Analysis in LEO Satellite Constellation[C]//2020 IEEE 3rd International Conference on Electronics and Communication Engineering（ICECE）. Xi'an，China：IEEE，2020：12-17.

[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] LIU S，HU X，WANG W. Deep Reinforcement Learning Based Dynamic Channel Allocation Algorithm in Multibeam Satellite Systems[J]. IEEE Acsess，2018，6：15733-15742.

[14] YAN X，AN K，LIANG T，et al. Effect of Imperfect Channel Estimation on the Performance of Cognitive Satellite Terrestrial Network[J] IEEE Access，2019，7：126293-126304.

[15] HU X，LIU S，CHEN R，et al. A Deep Reinforcement Learning Based Framework for Dynamic Resource Allocation in Multibeam Satellite Systems[J]. IEEE Communications Letters，2018，22(8)：1612-1615.

[16] XU X，WANG C，JIN Z. Perturbed ISL Analysis in LEO Satellite Constellation[C]//2020 IEEE 3rd International Conference on Electronics and Communication Engineering（ICECE）. Xi'an，China：IEEE，2020：12-17.

[17] BANKEY V，UPADHYAYP K，DA COSTAD B，et al. Performance Analysis of Multi-Antenna Multiuser Hybrid Satellite-Terrestrial Relay Systems for Mobile Services Delivery[J]. IEEE Access，2018，6：24729-24745.

[18] 王磊，郑军，贺川，等. 高通量多波束通信卫星系统资源分配方法[J]. 中国空间科学技术，2021，8(54)：1-10.

[19] 包文倩. 多波束卫星通信系统资源分配研究[D]. 北京：北京邮电大学，2018.

[20] 蔡睿妍. 卫星网络综合加权信道分配策略[J]. 大连大学学报，2020，41(3)：5-7.

[21] 钟旭东，何元智，任保全，等. 基于合作博弈的认知卫星网络信道分配与上行功率控制算法[J]. 计算机科学，2020，47(1)：252-257.

（编辑：徐楠楠）