DOI:10.19652/j.cnki.femt.2305514

基于多目标多智能体强化学习的低轨卫星切换策略

李 瑞[1] 杨巧丽[2] 张新澳[1]
(1.南京信息工程大学电子与信息工程学院 南京 210044;2.国防科技大学第六十三研究所 南京 210007)

摘 要:针对低轨卫星通信系统(LSM)中地面用户流量需求分布不均衡和用户并发切换过多等挑战,提出了一种基于多目标多智能体协同深度强化学习的低轨卫星切换策略,以地面小区用户流量需求满意度、切换时延、用户冲突为优化目标,采用多智能体协同深度学习算法对目标进行优化,其中每个智能体仅负责一个小区用户的卫星切换策略,智能体之间通过共享奖励实现协作,从而达到多目标优化的效果。仿真结果表明,所提的切换策略的平均用户流量满意度为73.1%,平均切换时延为343 ms,对比启发式算法能够更好满足地面小区用户的流量需求、平衡卫星网络的负载。

关键词:低轨卫星网络;多星切换;多目标优化;多智能体深度强化学习
中图分类号:TN927 **文献标识码**:A **国家标准学科分类代码**:510.5015

Low earth orbit satellite switching strategy based on multi-objective multi-agent reinforcement learning

Li Rui[1] Yang Qiaoli[2] Zhang Xin′ao[1]
(1. School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. The 63rd Research Institute of National University of Defense Technology, Nanjing 210007, China)

Abstract: To address the challenges of uneven traffic demand distribution and excessive concurrent handover among ground users in low earth orbit satellite communication systems, this paper proposes a low earth orbit satellite handover strategy based on multi-objective multi-agent collaborative deep reinforcement learning. The strategy aims to optimize the ground cell user traffic demand satisfaction, handover delay, and user conflict as the objectives, and adopts a multi-agent collaborative deep learning algorithm to optimize the objectives. Each agent is only responsible for the satellite handover strategy of one cell user, and the agents cooperate with each other by sharing rewards, thus achieving the effect of multi-objective optimization. Simulation results show that the average user traffic satisfaction of the proposed handover strategy is 73.1%, and the average handover delay is 343 ms. Compared with heuristic algorithms, the proposed strategy can better meet the traffic demand of ground cell users and balance the satellite network′s load.

Keywords: low earth orbit satellite network; multi-satellite handover; multi-objective optimization; multi-agent deep reinforcement learning

0 引 言

低轨卫星星座网络(low earth orbit satellite, LSC)是一种利用低轨道卫星提供全球通信和互联网服务的技术,具有时延低、带宽高和覆盖广等特点[1]。近年来,随着非地球静止轨道(non-geostationary orbit, NGSO)如 Oneweb、Starlink、Telesat 和国内的鸿雁[2]、虹云[3]等星座的快速发展,低轨卫星通信系统受到了广泛关注。与地球静止轨道卫星和中地球轨道卫星相比,低地球轨道卫星(low earth orbit, LEO)的运行轨道高度最低(≤2 000 km),但是这也导致 LEO 卫星移动速度较快,为了实现对地面的全覆盖,一个 LEO 系统通常包含数千颗卫星,形成一个庞大的星座。因此,在某一时刻,地面上的一个用户通常同时被多颗卫星覆盖,需要选择合适的卫星建立星地链路。同时由于卫星和地面用户之间巨大的速度差异导致地面用户视野范围内的接入点变化迅速,从而导致卫星对地链路的频繁切换。这种切换具有并发性[4],地面用户发出的大量并发切换会导致网络拥塞,因此,动态的卫星切换策

收稿日期:2023-09-06

略需要进行深入研究。

目前，关于卫星切换策略的研究主要分为单一目标切换策略和多目标切换策略两类。文献[5]以最大流量为目标提出了一种基于网络流算法的改进切换策略。文献[6]以最大化整体通信质量为目标提出了一种基于加权二分图的低地球轨道卫星网络中卫星与网关站之间链路的切换策略。但卫星切换过程中影响切换性能因素有很多只考虑单一目标的策略不能满足用户的需求。文献[7]则采用多目标切换策略，将卫星吞吐率、卫星负载均衡和用户切换成功率作为切换目标将用户进行分组提出了一种基于用户群组的低轨卫星网络多星切换策略。但是该策略采用启发式优化算法，例如粒子群算法[8]、蚁群算法[9]等，当面对高维度的状态空间和动作空间时，可能会出现维度灾难而导致优化算法无法得到全局最优解。

随着人工智能技术的快速发展，深度强化学习(deep reinforcement learning，DRL)等人工智能技术在信息技术领域得到了广泛应用，也为卫星通信资源分配提供了新的方法[10-12]。文献[13]提出了一种综合考虑缓存容量、剩余服务时间和剩余空闲信道 3 个因素的多属性决策切换策略，采用基于深度强化学习的缓存感知智能切换策略，以使系统的长期效益最大化。文献[14]以平衡卫星负载情况作为优化目标提出了一种新的基于多智能体强化学习的卫星切换策略，文献[15]根据接收信号强度、速度、网络带宽利用率和切换成本的属性设计了一种基于强化学习的多属性星地切换策略。但是以上的切换策略，只针对少量用户或均匀用户的切换场景进行设计，难以适应 LSC 的快速发展。随着低轨卫星通信网络的用户量和业务量的急剧增长，其切换决策的复杂度也随之提高。尤其是在热点区域，当大量地面用户集中在一定区域时，会导致批量用户同时发起切换业务，这会产生大量的请求信令，造成网络拥塞，严重影响网络性能。

针对上述问题，本文提出了一种基于多目标多智能体强化学习的低轨卫星切换策略(MO-MACDQN)。该策略建立了可靠的卫星切换模型，能够较好地满足在有限资源下的非均匀地面用户的流量需求，同时能够降低卫星切换延时和卫星负载情况。通过仿真与传统的切换策略对比分析结果可知，该方法提高了系统的鲁棒性，并有效降低了模型的训练负担。

1 系统模型和问题建立

1.1 系统模型

如图 1 所示，LEO 卫星网络由 M 颗卫星组成，服务于特定时间段内的 N 个地面小区用户(本文将所服务的用户根据地理位置聚类为小区用户)，将该特定时间段划分为 t 个时隙。M 颗卫星由 $S=\{S_m \mid m=1,2,3,\cdots,M\}$ 表示，地面上静态小区用户由 $U=\{U_n \mid n=1,2,3,\cdots,N\}$ 表示。时隙由 $T=\{1,2,3,\cdots,t\}$ 表示。假设在 t 时刻内，卫星的拓扑结构和覆盖区域保持不变。卫星集、用户集和时隙集可以分别用 S、U 和 T 表示。在单个时隙中，每个地面小区用户只能选择一颗卫星进行访问，而一颗卫星可以为多个地面小区用户提供服务。本文考虑多波束的 LEO 低轨卫星即每个卫星包含 k 个波束，在同一时隙内单个卫星最多覆盖 k 个地面小区用户。

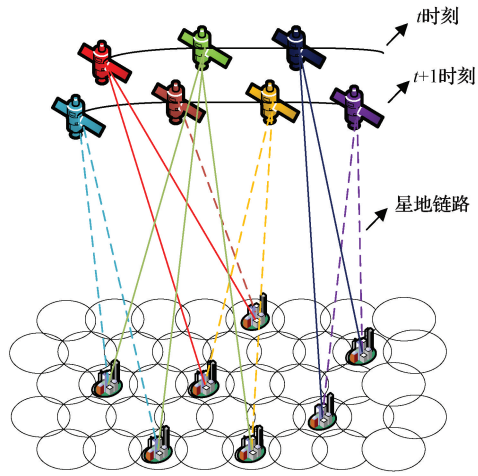

图 1 低轨卫星切换模型

为了满足地面小区用户的业务量需求，地面小区用户应该合理选择接入卫星。因此，将 t 时刻地面用户 U_n 与卫星 S_m 的接入策略表示为：

$$X_{U_n}=\begin{bmatrix} x^1_{U_n,1} & x^2_{U_n,1} & \cdots & x^t_{U_n,1} \\ x^1_{U_n,2} & x^2_{U_n,2} & \cdots & x^t_{U_n,2} \\ \vdots & \vdots & \ddots & \vdots \\ x^1_{U_n,S_m} & x^t_{U_n,2} & \cdots & x^t_{U_n,S_m} \end{bmatrix} \tag{1}$$

其中，$x^t_{U_n,S_m}\in\{0,1\}$，当 $x^t_{U_n,S_m}=1$ 时表示 t 时刻用户 U_n 接入卫星 S_m。

本文考虑了功率损耗和传输损耗，忽略多径衰落和阴影效应的影响，功率损耗可以表示为[16]：

$$L=0.002\ 45\left(\frac{\theta D f}{c}\right) \tag{2}$$

式中：θ 表示天线主接收方向与入射信号主波束方向的夹角；D 为天线直径；c、f 分别为光速和传输信号频率。

传动损耗包括许多部分。在该模型中，本文只考虑自由空间损耗，自由空间损耗是传输损耗的主要部分。自由空间损失为：

$$FSL=32.4+20\lg l^t_{S_m,U_n}+20\lg f \tag{3}$$

式中：$l^t_{S_m,U_n}$ 是卫星 S_m 和用户 U_n 在 t 时刻的距离。

根据上式并假设卫星 S_m 的发射功率为 P_{S_m}，则用户 U_n 的接收功率可以表示为：

$$P_{r_{nm}}=P_{S_m}+G_t+G_r-L-FSL \tag{4}$$

式中：G_t 为卫星 S_m 的发射天线增益；G_r 为用户 U_n 的接收天线增益。

在多波束卫星的前向链路中，每颗卫星产生 k 个波束服务于来自不同的小区用户，即小区用户 U_n 与卫星之间

的信道向量为 $H_{U_n}=\{h_{U_n,1},h_{U_n,2},h_{U_n,3},\cdots,h_{U_n,S_m}\}$。假设小区用户补偿了卫星运动引起的多普勒频偏，且天空晴朗，不考虑雨衰，则第 n 个小区用户与第 m 个卫星之间的信道系数[17]，可进一步表示为：

$$h_{U_n,S_m}=\frac{\sqrt{G_tG_r}}{4\pi\dfrac{l^t_{S_m,U_n}}{\lambda}} \tag{5}$$

式中：G_t 为小区用户 U_n 到卫星 S_m 之间的发射天线增益；G_r 表示接收天线增益；λ 表示波长。

1.2 问题建立与优化模型

由于地面小区用户的流量需求不均匀性，定义 d^t_n 为 t 时刻小区用户的流量需求。本文假设地面小区用户在卫星服务周期内的 U_n 流量需求保持不变，即小区用户 U_n 的总需求量可以表示为 $D_n=\sum_{t=1}^{T}d^t_n$。

在 t 时刻，若小区用户 U_n 与卫星 S_m 建立通信链路，则为小区用户 U_n 提供的流量（用吞吐量表示）可以计算为：

$$r^t_n=B\log_2\left(1+\frac{|h_{U_n,S_m}|^2\cdot P_{r_{nm}}}{K_BT_{rx}B_d}\right) \tag{6}$$

式中：B_d 是共享频谱带宽；T_{rx} 是接收机噪声温度；K_B 为玻尔兹曼常数

小区用户 U_n 在整个卫星服务周期内所提供的流量表示为 $R_n=\sum_{t=1}^{T}r^t_n$。为了度量服务质量，定义每个用户小区的流量满意率由整个卫星服务周期内提供流量与需求流量的比表示即 $Z_n=\dfrac{R_n}{D_n}$。

本文卫星切换时延主要包括切换过程中的信号传播时延和卫星验证用户身份的时延。前者与用户和卫星之间的距离有关，后者与卫星的星载处理能力有关。根据文献[18]假设与卫星接入的用户平等地共享计算资源并且平均分配各卫星的计算资源和传输带宽，则卫星 S_m 为用户 U_n 分配的可用计算资源可以表示为：

$$\alpha_{U_nS_m}(t)=\frac{F_j}{N_m(t)} \tag{7}$$

式中：F_j 为常数，表示为卫星 S_m 所提供计算资源；$N_m(t)$ 为 t 时刻访问卫星 S_m 的用户个数。卫星 S_m 与用户 U_n 之间的身份验证时延可以表示为：

$$\Gamma_{U_nS_m}(t)=\frac{B}{\alpha_{U_nS_m}(t)} \tag{8}$$

式中：B 为切换过程中每个用户的计算量。本文假设在星间切换请求报告发送后，用户与目标卫星之间进行了 6 次信令交换。因此，用户 U_n 在 t 时刻切换到卫星 S_m 的切换时延为：

$$HT^t_{U_nS_m}=\Gamma_{U_nS_m}(t)+6\frac{l^t_{S_m,U_n}}{c} \tag{9}$$

在低轨卫星通信系统中，同一颗卫星的计算资源被同时接入的小区用户平均地共享。因此，当接入卫星的小区用户数量增加时，每个小区用户分配到的计算资源将会减少，同时会导致认证延迟、切换延迟等增加。另外，减少同一颗卫星的接入用户数量也可以降低卫星的负载，实现卫星负载的均衡。基于上述分析，可以推断，减少同时切换到同一颗卫星的小区用户数量（即本文所称的冲突）可以使这些用户获得更多的网络资源，从而提高网络性能。因此，将冲突数量最小化作为本文的优化目标之一。

本文假设用户冲突分为两种情况，一种是多个小区用户 U_i、U_j 同时接入或切换到同一卫星，另一种是小区用户 U_i 切换到已经服务于另一用户 U_j 的卫星。即小区用户的冲突情况可以表示为：

$$\phi^t_{U_i,U_j}=\begin{cases}1, & (t=1,X^t_{U_i}=X^t_{U_j})\\ & \text{或}(t\neq1,X^{t-1}_{U_i}=X^t_{U_j})\\ 0, & \text{其他}\end{cases} \tag{10}$$

式中：$\phi^t_{U_i,U_j}=1$ 代表用户 U_i、U_j 在 t 时刻有冲突。使用 $\varphi_{i,j}$ 记录冲突总数。

$$\varphi_{U_i,U_j}=\begin{cases}0, & U_i=U_j\\ \sum_{t=1}^{T}\phi^t_{U_i,U_j}, & \text{其他}\end{cases} \tag{11}$$

1.3 优化问题

为了能够在有限资源下的满足非均匀地面用户的流量需求，本文以最大化用户流量满意度、最小化卫星切换时间、最小化用户冲突为目标建立了可靠的卫星切换策略。其中用户流量满意度优化目标是使所有地面小区用户的业务满意率最小值最大化，即尽可能满足每个小区用户的业务需求，保证了低流量需求用户的服务质量，从而实现了用户间的流量分配公平性；最小化卫星切换时间的优化目标是使用户在切换卫星时尽可能减少由于信令切换和卫星与小区用户间距离所引起的延迟；最小化用户冲突的优化目标是通过对不同的卫星进行合理的用户分配来最小化卫星间的业务负载差距，避免同一个卫星的负载量过大导致信号传输中断的情况。所考虑的多目标优化数学模型如下：

$$\begin{aligned}
&P_1:\underset{X_{U_1},X_{U_2},\cdots X_{U_n}}{\text{maxmize}}\ \min_{n\in N}\{Z_n\}\\
&P_2:\min\frac{1}{N}\sum_{t=1}^{T}\sum_{n=1}^{N}\{HT^t_{U_n,S_m}\}\\
&P_3:\min\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_{U_i,U_j}\\
&\text{s.t. } C_1:x^t_{U_n,S_m}\in\{0,1\}\\
&\qquad C_2:\sum_{m=1}^{M}x^t_{U_n,S_m}=1
\end{aligned} \tag{12}$$

式中：C_1 表示待优化变量为一个二元变量；C_2 表示每个地面小区用户在同一时刻内最多只能接入一颗卫星。

该优化问题是由 3 个目标构成的复杂优化问题，其计算复杂度较高，且信道条件随机变化。如果采用传统算法

求解，一方面会增加计算难度，容易陷入局部最优解，效果不理想；另一方面，网络拓扑和环境的不断变化，会使算法不断重复计算，增加资源分配的计算消耗，降低资源调度的及时性和准确性。所以本文采用了 MO-CMADRL 算法，将上述问题转化为马尔可夫决策过程（MDP）。将以上 3 个目标转化为 MDP 中的即时奖励 R：

$$R=\begin{cases}\dfrac{1}{N}\sum_{n=1}^{N}Z_n\\ -\left\{\dfrac{1}{N}\sum_{t=1}^{T}\sum_{n=1}^{N}\{HT^{t}_{U_n,S_m}\}\right\}\\ -\left\{\dfrac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_{U_i,U_j}\right\}\end{cases}\tag{13}$$

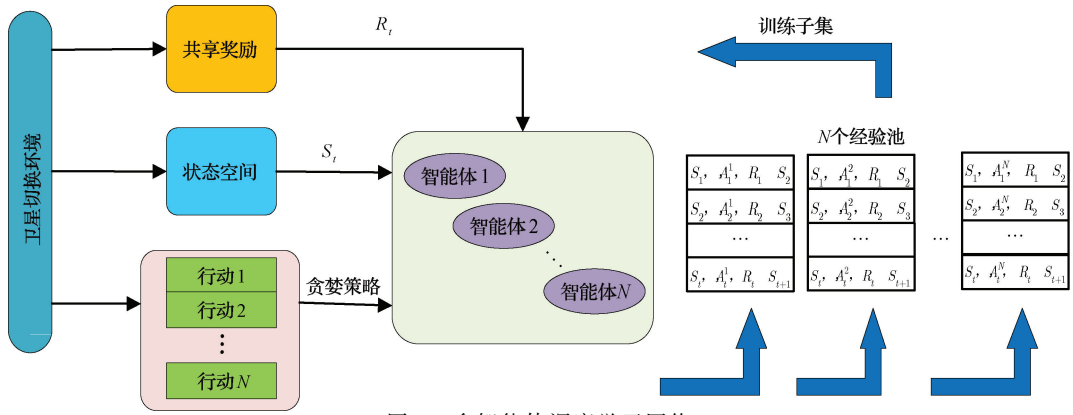

图 2　多智能体深度学习网络

1）状态空间 S_t，状态能抽象地表征环境。在 t 时刻，卫星切换环境可以观察整个流量需求、地面用户与卫星之间的位置信息作为全局状态，每个智能体都可以获得该状态，可以表示为：

$$S_t=\{l^{t}_{U_n,S_m},\theta^{t}_{U_n,S_m}\}\tag{14}$$

式中：$\theta^{t}_{U_n,S_m}$ 表示 t 时刻用户 U_n 与卫星 S_m 之间的夹角。

2）动作空间 A_t，智能体应该根据状态做出提高长期收益的决策。智能体在每个时隙中动态调整用户接入的卫星，可以表示为：

$$A_t=\{x^{t}_{U_n,1},x^{t}_{U_n,2},x^{t}_{U_n,3},\cdots,x^{t}_{U_n,S_m}\}\tag{15}$$

3）奖励函数 R_t，将优化目标的增量表示为判断奖励的依据，本文有 3 个衡量智能体价值的指标，由小区用户流量满意度、卫星切换时间和用户冲突决定，奖励函数采用线性加权的方式计算。

$$R_t=\omega_1\frac{1}{N}\sum_{n=1}^{N}Z_n-\omega_2\frac{1}{N}\sum_{t=1}^{T}\sum_{n=1}^{N}\{HT^{t}_{U_n,S_m}\}-\omega_3\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_{U_i,U_j}\tag{16}$$

式中：ω_1、ω_2、ω_3 为线性加权的权值。

2.2　多智能体网络设计

由于式(15)定义的动作空间是离散的，本文使用性能良好的双深度 Q 学习（double deep Q-learning，DDQN）方法让智能体学习策略。同时，考虑到所有智能体都能观察全局环境状态，在多智能体框架中采用独立 Q 学习（independent Q-learning，IQL）[19]。即每个智能体都有一个独立的深度神经网络（deep neural networks，DNN）用作近似函数，生成与状态和动作相对应的 Q 值函数。通过训练 DNN 参数 ω，智能体可以学习最优策略。

智能体 n 的 Q 值函数可以表示为 $Q(S_t,A^{n}_{t};\omega_n)$，其中 ω_n 为智能体 n 的 Q 网络参数。本文采用全连接网络（FCN）作为函数逼近器。为了避免局部最优，充分挖掘动作空间，在动作智能体选择行动策略处采用 ε 贪婪策略，当智能体随机选择时概率为 ε，当动作智能体选择 Q 值最大时概率为 1－ε，ε 贪婪策略可以表示为：

$$A_t=\begin{cases}\text{随机选择}, & \text{概率为 }\varepsilon\\ \max Q(S_t,A^{n}_{t};\omega_n), & \text{概率为 }1-\varepsilon\end{cases}\tag{17}$$

采取行动后，智能体将获得全局奖励 R_t，卫星切换环境将进入一个新的状态 S_{t+1}。然后每个智能体将以四元组 $(S_t,A^{n}_{t},R_t,S_{t+1})$ 样本的形式存储到经验回放池中[20]。最后，采用重放记忆方法对网络进行训练，即通过对经验回放池中的样本进行提前训练从而缓解经验分布的非平稳性。

在 DDQN 体系结构中，每个智能体还拥有一个与 Q 网络结构相同的目标网络。目标网络可表示为 $\hat{Q}(S_t,A^{n}_{t};\omega^{-}_{n})$，其中 ω^{-}_{n} 为智能体 n 的目标网络参数。在训练期间，每个智能体从经验池中抽取 k 个过渡项 $\{S_j,A^{n}_{j},R_j,$

$S_{j+1}\}$，其中 $j=1,2,3,\cdots,k$ 计算出定义的目标值为：

$$y_j^n = r_j + \gamma \hat{Q}(S_{j+1}, \underset{A_j^n}{\text{argmax}} Q(S_{j+1}, A_j^n; \omega_n); \omega_n^-) \tag{18}$$

式中：$\gamma \in [0,1]$ 为折扣因子。

则均方误差损失计算为：

$$L(\omega_n) = \frac{1}{k}\sum_{j=1}^{k}(y_j^n - Q(S_t, A_t^n; \omega_n))^2 \tag{19}$$

最后，用 Adam 算法和目标网络参数 ω_n 更新 Q 网络参数[21]。

3 仿真环境及仿真性能分析

3.1 仿真环境设计及参数设置

本文对 Ka 波段前向低轨卫星通信链路的低轨卫星切换进行了仿真研究。在经度范围[115°－120°]，纬度范围[28°N－33°N]的区域内选取 11 个城市作为地面小区用户，小区用户的长期流量需求如图 3 所示。低轨卫星星座参考铱星(Iridium)系统[22]包含 66 颗低轨道卫星，假设每颗卫星的星载资源和工作频率相同且每颗卫星的最大波束个数都为 48，总带宽设置为 30 MHz。使用 STK 软件对铱星系统运行轨道场景进行模拟建立极轨道星座模型，得到起止时间内各卫星节点与地面站的位置数据，如图 4 所示。

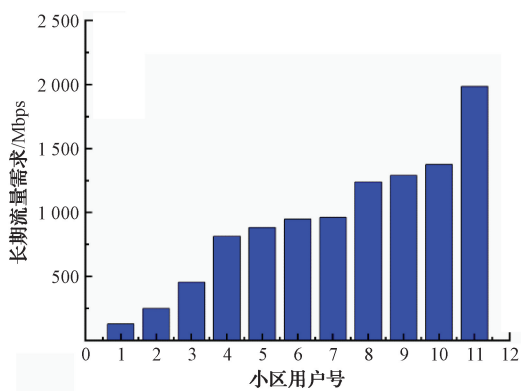

图 3 小区用户的长期流量需求

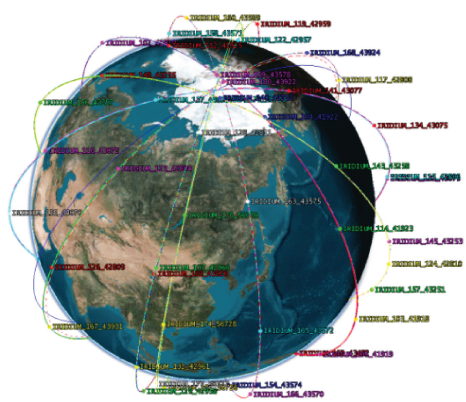

图 4 Iridium 星座轨道仿真

低轨卫星场景设置参数如表 1 所示，其中部分参数参考文献[23]。多目标多智能体协同深度强化学习算法参数如表 2 所示。

表 1 低轨卫星场景设置参数

卫星网络参数	取值
卫星轨道高度	785 km
卫星数量 M	66
地面用户小区数量 N	11
卫星的波束个数 K	48
天线增益 G_r	35.9 dBi
天线直径 D	3m
系统带宽 B_d	30 MHz
下行链路工作频率 f	20 GHz
接收天线噪声功率密度 T_{rx}	－174 dBm/Hz
星载功率 P_{s_m}	30 dBW
卫星提供的计算资源 F_j	2 200 MIPS
玻尔兹曼常数 K_B	1.36×10^{-23} J/K

表 2 多目标多智能体协同深度强化学习参数

MO-MACDQN 算法参数	取值
时隙 t	30 s
训练周期	500
每周期时隙数$\lvert T\rvert$	400
经验池容量	10 000
网络参数 ω_n^- 更新频率 C	100
激活函数	ReLU
折扣因子 γ	0.95
探索概率 ε	0.2

3.2 仿真性能分析

为了验证本文算法的性能，仿真环境选择在 win11，64 位操作系统，在 python 上进行仿真，使用 RTX 3070GPU 训练。选择文献[13]的 DQN 算法作为对比算法。两种算法在不同迭代次数下的总奖励变化趋势如图 5 所示。从图 5 可以看出，两种算法的加权目标总奖励在训练初期波动较大，这是因为在这一阶段算法处于探索阶段，在此期间算法经常探索错误的分配策略，因此受到大量的负奖励惩罚。MO-MACDQN 算法在训练次数约为 70 时累积的奖励曲线逐渐收敛，并趋于稳定，DQN 算法在训练次数达到 180 时才逐渐趋于稳定，并且抖动幅度大于本文算法，这是由于本文方法对单个智能体的学习内容进行了优化，将小区用户设置为多个智能体，智能体之间相互协作能够提高分配决策的效率。在智能体训练过程中，通过多次选择不同的小区用户组合，使每个智能体都能得到同等的学习机会，一定程度上避免了因智能体学习环境差异而产生的局部最优解。由此验证了 MO-MACDQN 在收敛性能上的优异性。

为了能够体现基于 MO-MACDQN 算法的低轨卫星

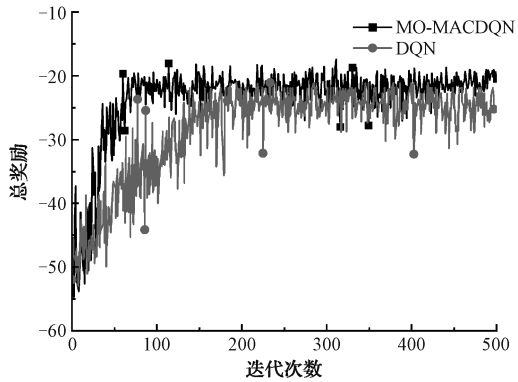

图5 不同方法迭代次数下的总奖励

切换策略对不同地面用户小区流量满意度、切换延迟和卫星负载上的性能提升，本文将所提算法与基于单目标遗传优化算法SOGA[24]和迁移深度强化学习算法TL-DQN[25]进行比较。在遗传算法中，以系统吞吐量最大化为目标，初始种群大小为200，选择算子采用精英保留策略和轮盘赌选择的方法。交叉方法采用单点交叉，当基本变异算子的变异概率为0.001时，交叉概率为0.1，迭代次数为200次。在迁移深度强化学习算法中，以最小化切换时延为目标，训练周期为600，经验池容量设置为5 000，折扣因子为0.9，学习率设置为0.001。

地面小区用户在不同算法下所获得的用户流量满意度如图6所示。将地面小区用户按照流量需求从小到大排序，可以看出3种算法在低流量需求下小区用户的满意度能够保持较高值，随着用户的流量需求增加由于卫星的星载资源有限导致用户流量满意度也处于下降趋势，本文所提策略在不同流量需求下的用户满意度均高于SOGA算法和TL-DQN算法，能够更好的满足用户流量需求。3种算法的平均用户流量满意度分别为73.1%、59.6%、63.8%。

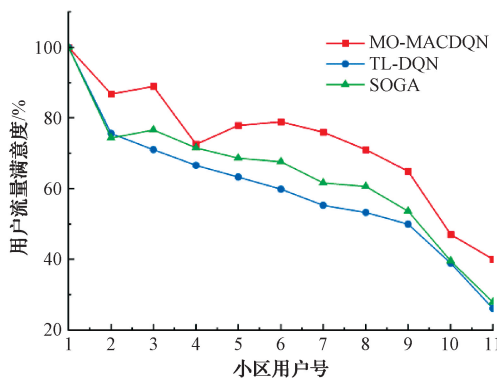

图6 小区用户平均流量满意度对比

图7和8所示为不同算法下小区用户的累计卫星切换时延和每个时隙内小区用户的冲突数量。从图7可以观察到，SOGA算法和TL-DQN算法在低流量需求下与本文算法的切换时延相近，但在高流量需求下切换时延显著增加。这是因为高业务负载下可动态分配的卫星资源减少，导致SOGA和TL-DQN算法的计算复杂度较高。相比之下，本文算法在计算复杂度上具有较大优势，累计切换时延比另外两种算法分别降低了11.2%、29.7%。从图8可以看出，本文算法在每个时隙下用户的冲突数量都低于另外两种算法，有效地减少了用户冲突，实现了卫星负载的平衡。

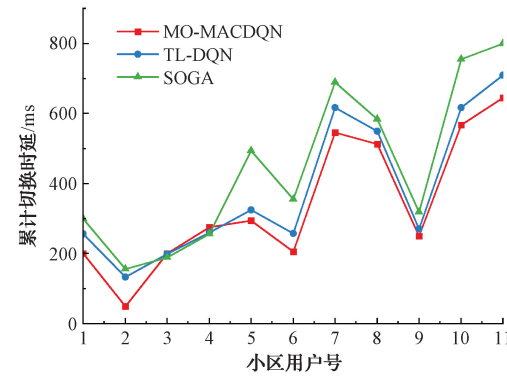

图7 小区用户累计切换时延对比

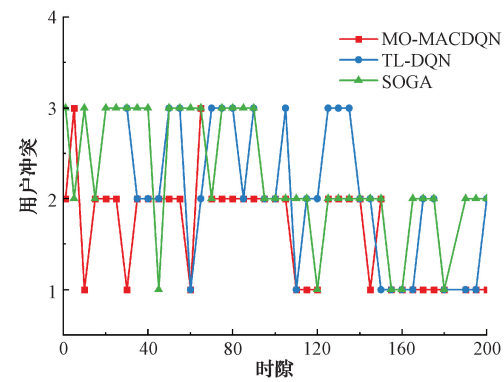

图8 每个时隙的用户冲突数量对比

4 结 论

本文针对低轨卫星通信系统中地面用户流量需求分布不均衡、切换时延过长和用户并发切换过多等问题，提出了一种多目标多智能体深度强化学习算法MO-MACDQN对低轨卫星的切换策略进行了研究。以地面小区用户流量需求满意度、用户切换冲突、卫星切换时延为优化目标，每个智能体负责一个小区用户的卫星切换策略，智能体之间共享相同的奖励协作实现多个目标的优化。仿真结果表明MO-MACDQN算法在有限的卫星资源下能够尽可能满足地面用户的流量需求，平衡卫星负载，并有效降低切换带来的延迟。该算法可以广泛应用于低轨卫星通信系统，但是其性能可能会受到网络规模的影响。随着网络规模的增大，智能体的数量也会增加，这可能会增加算法的复杂性和计算负担，未来能够继续研究该算法在多波束卫星资源分配和跳频资源分配上的应用。

参考文献

[1] YUN J, AN T, JO H, et al. Dynamic downlink interference management in LEO satellite networks without direct communications[J]. IEEE Access, 2023, 11:24137-24148.

[2] 徐菁."鸿雁"星座闪亮亮相 移动通信或将全球无缝覆盖[J].中国航天,2018(11):35-36.

[3] 孙喆."长征"十一号火箭成功发射"虹云"工程首颗卫星[J].中国航天,2019(1):42.

[4] HUANG C M, CHIANG M S, DAO D T. A group based handover control scheme for mobile internet using the partially distributed mobility management (GP-DMM) protcol[C]. International Symposium on Pervasive Systems, Algorithms and Networks, 2017: 148-155.

[5] ZHANG S, LIU A, HAN C, et al. A network-flows-based satellite handover strategy for LEO satellite networks[J]. IEEE Wireless Communications Letters, 2021, 10(12): 2669-2673.

[6] FENG L, LIU Y, WU L, et al. A satellite handover strategy based on MIMO technology in LEO satellite networks[J]. IEEE Communications Letters, 2020, 24(7): 1505-1509.

[7] 朱洪涛,郭庆.基于用户群组的低轨卫星网络星切换策略[J].电信科学,2022,38(4):39-48.

[8] 邢燕好,于昊,张佳,等.基于粒子群参数优化的O-VMD数据处理方法研究[J].仪器仪表学报,2023,44(4):304-313.

[9] 杨立炜,付丽霞,王倩,等.多层优化蚁群算法的移动机器人路径规划研究[J].电子测量与仪器学报,2021,35(9):10-18.

[10] CAO L, ZHIMIN. An overview of deep reinforcement learning[C]. 2019 4th International Conference, 2019: 1-9.

[11] XU J, AI B. Deep reinforcement learning for handover-aware MPTCP congestion control in space-ground integrated network of railways[J]. IEEE Wireless Communications, 2021, 28(6): 200-207.

[12] XU H, LI D, LIU M, et al. QoE-driven intelligent handover for usercentric mobile satellite networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 10127-10139.

[13] LENG T, XU Y, CUI G, et al. Caching aware intelligent handover strategy for LEO satellite networks[J]. Remote Sensing, 2021, 13(11): 2230.

[14] HE S, WANG T, WANG S. Load-aware satelite handover strategy based on multiagent reinforcement learning[C]. IEEE GLOBECOM 2020, 2020: 1-6.

[15] JIA X, ZHOU D, SHENG M, et al. Reinforcement learning-based handover strategy for space-ground integration network with large-scale constellations[J]. Journal of Communications and Information Networks, 2022, 7(4): 421-432.

[16] YANG B, WU Y, CHU X, et al. Seamless handover in software-defined satellite networking[J]. IEEE Communications Letters, 2016, 20(9): 1768-1771.

[17] CHEN L, LAGUNAS E, CHATZINOTAS S, et al. Satellite broadband capacity on demand: Dynamic beam illumination with selective precoding[C]. 2021 29th European Signal Processing Conference (EUSIPCO), 2021: 900-904.

[18] DING X, ZHANG Z, LIU D. Low-delay secure handover for space air ground integrated networks[C]. 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 2020.

[19] ELLEUCH I, POURRANJBAR A, KADDOUM G. A novel distributed multi-agent reinforcement learning algorithm against jamming attacks[J]. IEEE Communications Letters, 2021, 25(10): 3204-3208.

[20] 刘子怡,李君,李正权.多用户蜂窝网络中基于深度强化学习的功率分配[J].国外电子测量技术,2023,42(3):30-35.

[21] LIN D, CHEN C H. PSO-Adam a case study of mobile positioning[C]. IEEE, 2022: 685-686.

[22] 李新桐,张亚生.一种适用于低轨卫星的SDN网络人工智能路由方法[J].电子测量技术,2020,43(22):109-114.

[23] ZHAO B, LIU J, WEI Z, et al. A deep reinforcement learning based approach for energy-efficient channel allocation in satellite internet of things[J]. IEEE Access, 2020, 8: 62197-62206.

[24] WANG L, HU X, MA S, et al. Dynamic beam hopping of multi-beam satellite based on genetic algorithm[C]. 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 2020: 1364-1370.

[25] 陈前斌,麻世庆,段瑞吉,等.基于迁移深度强化学习的低轨卫星跳波束资源分配方案[J].电子与信息学报,2023,45(2):407-417.

作者简介

李瑞，硕士研究生，主要研究方向为卫星通信、深度学习等。
E-mail：202212490445@nuist.cn

杨巧丽，副研究员，主要研究方向为卫星通信。
E-mail：1523013007@qq.com

张新澳，硕士研究生，主要研究方向为卫星通信。
E-mail：3216786657@qq.com