

基于SAC的低轨卫星光网络负载均衡路由优化算法

唐可意, 李志刚

(中国电子科技集团公司第三十四研究所, 广西 桂林, 541004)

摘要:针对卫星动态性与复杂环境因素导致的网络负载不均衡及通信延迟问题,设计了基于深度强化学习(DRL)中Soft Actor-Critic(SAC)的低轨卫星光网络负载均衡路由优化算法。通过模拟卫星网络环境并动态调整决策策略,实现路由路径的动态优化,旨在合理分配网络资源并提升网络性能。仿真结果表明:该算法模型在优化数据流分配与提高系统可靠性方面具有显著优势;在50%和100% 2种流量强度下,链路平均负载率分别优化29.9%和42.0%,同时路由延迟亦得到有效改善。

关键词:负载均衡;深度强化学习;低轨卫星光网络;路由优化;卫星互联网

中图分类号: TN927.2 **文献标志码:** A **文章编号:** 1002-5561(2026)03-0036-05

DOI: 10.13921/j.cnki.issn1002-5561.2026.03.006

Load-balancing routing optimization algorithm for low earth orbit satellite optical networks based on SAC

TANG Keyi, LI Zhigang

(The 34th Research Institute of CETC, Guilin Guangxi, 541004, China)

Abstract: To address the issues of network load imbalance and communication delay caused by satellite dynamics and complex environmental factors, a load-balancing routing optimization algorithm for low earth orbit Satellite optical network based on Soft Actor-Critic(SAC) in deep reinforcement learning(DRL) is proposed. By simulating the satellite network environment and dynamically adjusting decision-making strategies, the algorithm achieves dynamic optimization of routing paths, aiming to rationally allocate network resources and improve network performance. Simulation results demonstrate that the proposed algorithm exhibits significant advantages in optimizing data flow distribution and enhancing system reliability. Under traffic intensities of 50% and 100%, the average link load rate is optimized by 29.9% and 42.0%, respectively, while routing delay is also effectively reduced.

Key words: load balancing, deep reinforcement learning, low earth orbit satellite optical network, routing optimization, satellite internet

0 引言

作为实现全球覆盖通信的主要手段,卫星通信一直以来都在不断发展^[1]。随着第五代移动通信技术(5G)和未来6G网络的逐步部署,低地球轨道(LEO)卫

星网络凭借其低时延、全球覆盖、传输损耗小、低成本等优点,已成为最具前景的卫星通信载体^[2],在工业、军事等领域展现出巨大的潜力和开发利用价值^[3],对推动全球互联网普及、提升网络性能也具有重要意义^[4]。在此背景下,卫星光通信凭借组网灵活、全天候工作及抗电磁干扰能力强等优势,同时针对星上资源有限的问题,收发终端需小型化以契合激光通信要求,使空天光网络成为高效、高速信息传输的理想方案^[5]。然而,受地理环境、资源竞争、需求差异以及LEO卫星动态运动与空间环境复杂性的共同影响,卫星光网络通信面临通信质量不稳定、星上资源短缺、负载不均衡等日益严峻的挑战,已成为当前研究的热点。

收稿日期:2025-10-14。

基金项目:中国电子科技集团公司第三十四研究所发展基金项目(K134002024SH02)资助。

作者简介:唐可意(2000—),男,硕士,助理工程师,现就职于中国电子科技集团公司第三十四研究所。主要从事空天地一体化光通信网络、卫星激光组网仿真以及机器学习等方面的研究工作,参与光网络仿真与数字孪生、天基光网络等多项卫星通信相关科研项目。



近年来,机器学习在网络路由算法优化中扮演着越来越重要的角色^[5]。机器学习通过其自适应、智能化的特性,在实时交互中不断更新策略、适应复杂多变环境的能力,能够优化传统的网络路由算法,使路由决策更加灵活、精准、高效^[6]。随着网络规模和复杂性的不断增加,将深度学习与强化学习相结合,所形成的深度强化学习(DRL)的应用有望为网络优化提供更多创新的解决方案,提高网络的整体性能和稳定性^[7]。

在现有基于DRL的卫星网络路由研究中,已有学者进行了初步探索。文献[8]提出了一种基于Soft Actor-Critic(SAC)的LEO卫星网络路由优化算法。文献[9]采用双延迟深度确定性策略梯度(DDPG)算法实现路由优化。文献[10]研究了方向引导的蜂群优化算法。文献[11]基于深度Q网络(DQN)满足服务质量需求。文献[12]利用DDPG算法优化网络吞吐量。然而,上述工作在应对LEO光网络中负载不均衡、资源动态分配及多业务并发等问题时,仍存在策略探索效率低、奖励机制设计不全面等不足。

针对卫星光通信网络中拓扑结构快速变化及多类型业务并存的特点,本文在所构建的卫星光网络架构基础上,基于DRL中的SAC算法进行优化创新。面向LEO激光组网中通信质量不稳定、负载不均衡及资源短缺等问题,提出一种基于SAC的低轨卫星光网络负载均衡路由优化算法(下文简称SAC-RLRA),并围绕该算法展开系统研究。

1 低轨卫星光网络环境建模

本文选择极地轨道作为研究场景。极地轨道的轨道平面与地球极轴基本重合,卫星在地球极地区域上方通过,具有全球覆盖、监测范围广等特点,适合作为路由优化算法的研究对象。极轨道卫星存在2种星间链路:同轨链路(同一轨道内卫星之间)与异轨链路(不同轨道之间)。同轨链路永久存在且传播距离稳定;异轨链路因卫星间存在相对运动,稳定性较低。由于第一轨道与最后一轨道内卫星之间的相对速度过大,形成“裂缝”现象,两者之间不存在星间链路。异轨卫星间的动态变化相对规律,在一定时间片内可视为稳定连接。据此,可将卫星分布与链路连接信息转化为矩阵形式进行处理。具体地,卫星分布矩阵的元素表示对应轨道位置上是否存在卫星,星间链路分布矩阵的元素表示2颗卫星之间是否存在链路连接。

唐可意,李志刚:基于SAC的低轨卫星光网络负载均衡路由优化算法

在完成卫星节点与链路关系建模后,从源卫星节点到目的卫星节点的数据包路由路径 l_{path} 可表示为

$$l_{\text{path}} = [s_{\text{source}}, s_n, \dots, s_{\text{target}}] \quad (1)$$

式中: s_{source} 表示源卫星节点, s_{target} 表示目的卫星, s_n 表示传输时路由路径经过的中间节点。

2 SAC-RLRA

2.1 算法设计

为实现卫星互联网路由优化,本文设计了SAC-RLRA。该算法属于DRL路由方案中的“全网链路权重调整”类别,通过动态调整链路传输优先级来合理分配网络资源。在算法选型上,本文选择SAC算法作为核心创新基础,主要是因为它采用策略与价值函数联合训练、以最大化熵为目标函数,具备了优异的探索-利用平衡能力,能很好地适配低轨卫星光网络高维动态状态与连续动作空间等复杂场景。

SAC算法以DDPG为基础,在传统Actor-Critic框架中引入最大熵强化学习思想^[13]。该算法的核心是通过最大化期望奖励与策略熵共同优化智能体策略,并以最小化值函数与Q函数的误差提升策略精度。最大熵强化学习的优化目标为最大化带熵的累积奖励,可表示为

$$J(\pi) = \mathbb{E}_{(s_t, a_t) \sim \pi} \left[\sum_{t=0}^T r(s_t, a_t) + \alpha H(\pi(\cdot|s_t)) \right] \quad (2)$$

式中: $\mathbb{E}[\cdot]$ 表示数学期望, $(s_t, a_t) \sim \pi$ 表示智能体与环境进行的交互, $r(s_t, a_t)$ 表示在状态 s_t 时执行动作 a_t 所得到的奖励值; α 表示温度系数,确定了熵项对于奖励的相对重要性, $H(\pi(\cdot|s_t))$ 则表示策略 π 在状态 s_t 下的熵。而SAC算法通过策略梯度方法来优化目标函数,即

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{s \sim D, a \sim \pi_{\theta}} [\nabla_{\theta} \alpha \log \pi_{\theta}(a|s) - \nabla_{\theta} Q_{\phi}(s|a)] \quad (3)$$

式中: $s \sim D, a \sim \pi_{\theta}$ 表示从经验回放池 D 中采样状态 s ,以及从当前策略 π_{θ} 中采样动作 a ,用于控制策略的熵项。结合策略的熵项梯度和Q值函数的梯度,能够使得SAC算法能够更好地平衡探索和利用,并使得智能体的策略逐步优化。

SAC-RLRA的设计分为状态空间、动作空间和奖励机制3个部分,具体如下:

1) 状态空间。在低轨卫星光网络中共有 n 个节点,则 t 时刻卫星光网络的状态空间 $\mathbf{S}$ 可以用流量矩阵表示为

唐可意,李志刚:基于SAC的低轨卫星光网络负载均衡路由优化算法

$$S = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n,1} & s_{n,2} & \cdots & s_{n,n} \end{bmatrix} \quad (4)$$

式中: $s_{i,j}$ 表示节点*i*到节点*j*之间的数据流量。

2)动作空间。同样,含有*n*个节点的低轨卫星光网络在*t*时刻状态空间可以用*n*×*n*的流量权重分配矩阵*A*表示为

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \quad (5)$$

式中: $a_{i,j}$ 表示卫星节点*i*到卫星节点*j*之间的路由权重调整值。

3)奖励机制。本文选取传输延迟、链路负载均衡、通信阻塞率以及算法收敛性作为网络性能评价指标。对于智能体选定的路由路径*path*,路径总传播时延为路径所经各链路传播时延的累加和*D_{path}*。单条链路的链路负载可表示为

$$\eta = \frac{pP_{\text{SIZE}}}{B} \times 100\% \quad (6)$$

式中:*p*表示时间片内该链路中的数据包数量,*P_{SIZE}*表示单个数据包的平均大小,*B*表示链路带宽。此时间片内,路由传输路径*l_{path}*所经过的*m*条链路的平均链路负载率为

$$\eta_{\text{path}} = \frac{1}{m} \sum_{(s_i, s_j) \in l_{\text{path}}} \eta_{(s_i, s_j)} \quad (7)$$

式中:设路径*path*经过*m*条链路,*s_i*和*s_j*表示这个路径上经过的2个节点。

若在一个时间片内完成数据的传输,则对此次探索进行奖励计算。因此,综合考虑评价指标来进行奖励的设定和计算,根据智能体在一次探索中选取的路由传输路径*l_{path}*,计算出奖励*R*为

$$R = -\frac{\lambda_1 D_{\text{path}}}{a} + \frac{\lambda_2 \eta_{\text{path}}}{b} + \frac{\lambda_3 U_{\text{path}}}{c} + \lambda_4 \gamma_{\eta} \quad (8)$$

式中:阻塞率*U_{path}*定义为没有被网络处理的通信请求数与所有用户请求的数量的比值。 λ_1 、 λ_2 、 λ_3 、 λ_4 为指标参数权重控制因子,且 $\lambda_1 + \lambda_2 + \lambda_3 = 1$,*a*、*b*、*c*为奖励调整因子,负责将各状态计算出的数值结果调整至同一数量级。而 γ_{η} 为链路负载率优化奖励,其表达式为

$$\gamma_{\eta} = \begin{cases} 1, & \eta \geq \eta' \\ -1, & \eta < \eta' \end{cases} \quad (9)$$

式中: η' 为预设门限值。当链路的负载大于门限值即给予奖励,小于门限值给予一定惩罚。

2.2 算法流程

SAC-RLRA采用离线训练、在线决策模式,具体算法流程如算法1所示。

算法1 SAC-RLRA

输入 卫星数量*n*,链路数量*l*,数据包数量*p*

输出 路由路径*l_{path}*

- 1)初始化策略网络参数 ϕ
- 2)初始化2个Q网络*Q_{θ₁}*、*Q_{θ₂}*
- 3)初始化2个V网络*V_ψ*、*V_{ψ̄}*
- 4)初始化经验回放池子*D*
- 5)初始化温度系数 α
- 6) while 探索未完成 do
- 7) for 每个时间步*t* do
- 8) 根据当前状态*S*和策略网络选择动作
- 9) 智能体执行选择的动作,计算奖励值
- 10) 智能体获取下一个状态*S'*
- 11) 将状态数据存储至经验回放池*D*
- 12) if 经验回放池*D*中数据量足够
- 13) 随机采样一批数据
- 14) 更新*Q*值网络参数 θ_1 、 θ_2
- 15) 更新网络参数 ϕ
- 16) 最大化策略熵目标更新温度系数 α
- 17) 软更新目标网络参数 $\bar{\psi}$
- 18) end if
- 19) end for
- 20) end while

3 仿真验证

3.1 仿真设计

为模拟第2节设计的卫星光网络环境,本文基于Python开发了一个LEO光网络结构模型。该模型包含5个轨道,每个轨道部署12颗卫星,共计60颗卫星。同时,使用Python构建了SAC-RLRA模型,并分别在网络流量强度为50%和100%的场景下进行了仿真实验。研究旨在根据链路延迟、负载及拥塞状况的变化,为数据包选择传输路由路径,以优化网络负载均衡并合理分配资源,力求达到理想状态。

3.2 仿真结果分析

基于所设计的算法模型,在卫星节点数量为60的网络环境下对SAC-RLRA进行训练。在总计30 000步的训练过程中,每500步统计一次平均奖励值,结果如图1所示。图中阴影区域表示不同训练轮次下奖励

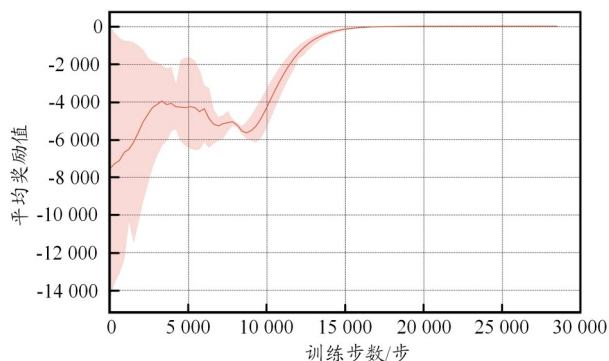
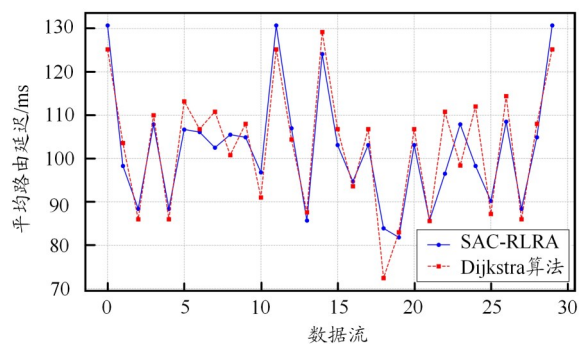


图1 SAC-RLRA 平均奖励值随训练次数的收敛情况

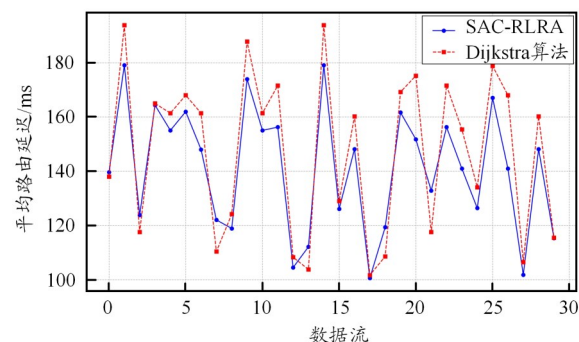
值的标准差。可以看出,在前7 000步时,SAC-RLRA模型处在快速探索学习,积累经验的阶段,奖励值为负且波动较大;在7 500步左右时,算法模型逐渐开始收敛,随着策略的逐步优化和收敛,平均奖励值逐渐趋于稳定,并在15 000步后达到稳定状态。

训练完成后,使用生成的模型处理LEO卫星星座网络的路由构建问题。分别在50%和100%流量强度场景下,模拟30次数据传输。为模拟地缘环境造成的不平衡现象,为边缘节点设置不同的选取概率;每次数据流传输根据边缘节点选取概率,随机生成1 000对源-目的节点的数据流量,并分别采用Dijkstra算法与SAC-RLRA处理数据流路由,对生成的路由路径进行数学分析比较,得到时延性能对比如图2所示。可以看出,流量强度为50%时,2种算法平均路由延迟接近;流量强度为100%时,SAC-RLRA延迟曲线显著低于Dijkstra算法。对30次数据流结果取算术平均,得到50%流量强度下Dijkstra算法与SAC-RLRA的平均延迟分别为99.68 ms和101.05 ms;100%流量强度下分别为153.0 ms和141.6 ms,SAC-RLRA较对比算法降低约7.45%。

不同流量强度下,2种算法的负载均衡情况如图3所示。可以看出,在流量强度为50%时,Dijkstra算法的负载率曲线整体处于较高水平,而SAC-RLRA则显著降低。对30次数据流的平均负载率取算术平均,得到Dijkstra算法的平均负载率为48.1%,SAC-RLRA为33.7%,负载率优化幅度为29.9%。当流量强度增至100%时,网络负载显著升高,2种算法的平均负载率均有所上升。其中,Dijkstra算法的负载率曲线大幅攀升,而SAC-RLRA仍保持在较低水平。对30次数据流结果取算术平均,Dijkstra算法的平均负载率为72.1%,SAC-RLRA为41.8%,负载率优化幅度达到42%。

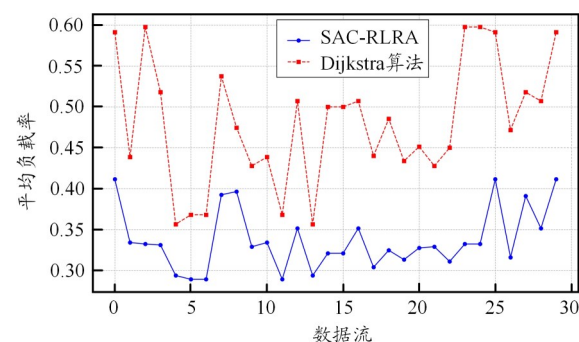


(a) 流量强度为50%

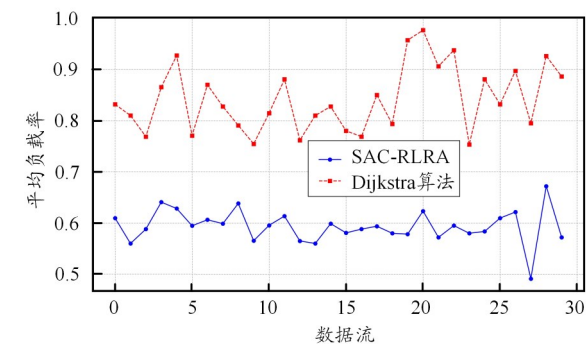


(b) 流量强度为100%

图2 不同流量强度下,2种算法的平均路由延迟对比



(a) 流量强度为50%



(b) 流量强度为100%

图3 不同流量强度下,2种算法的平均负载均衡情况对比

唐可意,李志刚:基于SAC的低轨卫星光网络负载均衡路由优化算法

通过对应的卫星光网络星座路由负载实验结果可知,与Dijkstra算法相比,SAC-RLRA在满足业务时延要求的情况下,能有效降低链路的平均负载率。由图3可知,随着网络流量强度不断增大,SAC-RLRA算法的负载均衡优化效果更加突出,在高负载场景下对卫星光网络流量分配的优化优势更为突出。

定义阻塞率大于20%的链路为阻塞链路。本文对30次数据传输结果进行统计,结果如表1所示。可以看出,SAC-RLRA通过优化网络负载分布,显著减少了阻塞链路数量,有效降低了阻塞率,相比Dijkstra算法具有明显优势。

表1 阻塞链路统计结果

算法	平均阻塞链路数	
	流量强度 50%	流量强度 100%
Dijkstra	3	5
SAC-RLRA	0	1

4 结束语

本文提出了一种SAC-RLRA。该算法以卫星网络流量矩阵为状态空间,以链路路由权重为动作空间,综合时延、链路负载与阻塞率设计奖励函数,通过策略网络与价值网络的联合训练实现路由策略的动态优化。仿真结果表明:与传统Dijkstra算法相比,SAC-RLRA在流量强度为50%时可将链路平均负载率优化29.9%,在流量强度为100%时优化幅度达42%;在高负载场景下平均路由延迟降低7.45%,同时显著降低网络阻塞率并减少阻塞链路数量。所提算法在满足业务时延要求的前提下,能够实现更均衡的流量分配与更高效的资源利用,且在高负载场景下优势更为突出。该研究可为大型复杂低轨卫星光网络的路由决策与资源分配提供可靠优化方案,并对未来空地一体化光通信网络的性能提升具有重要参考价值。

参考文献:

[1] Wang Niwei, Xu Xiaofan, Zhang Chi, et al. Assumption and key tech-

nologies of next generation satellite network[C]//IEEE. Proceedings of 2023 15th International Conference on Communication Software and Networks(ICCSN). Shenyang:IEEE,2023:414-418.

[2] Al-Homssi B, Al-Hourani A, Wang Ke, et al. Next generation mega satellite networks for access equality: opportunities, challenges, and performance[J]. IEEE Communications Magazine, 2022, 60(4): 18-24.

[3] Liu Peng, Liang Jun, Qian Xinghua, et al. Research on link load-balancing between marine LEO satellite constellations based on SDN[C]//IEEE. Proceedings of 2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT). Sanya: IEEE, 2021: 977-982.

[4] Jin Liang, Wang Liting, Jin Xinghu, et al. Research on the application of LEO satellite in IOT[C]//IEEE. Proceedings of 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information(ICETCI). Changchun:IEEE,2022:739-741.

[5] 王天枢,林鹏,董芳,等.空间激光通信技术发展现状及展望[J].中国工程科学,2020,22(3):92-99.

[6] Luong N C, Hoang D T, Gong S M, et al. Applications of deep reinforcement learning in communications and networking: a survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.

[7] Raikar A M, Manjaiah D H, Elumalai R. Comparison of tcp congestion control algorithms: harnessing the power of traditional hybrid and machine learning fusion[C]//IEEE. Proceedings of 2024 international conference on electronics, computing, communication and control technology (ICECCC). Bengaluru:IEEE,2024:1-6.

[8] Zhu Jiacheng, Sun Xuebin, Chen Dianjun. Deep reinforcement learning for congestion control and routing optimization in LEO satellite networks[C]//IEEE. Proceedings of 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA). Dalian: IEEE, 2023: 739-742.

[9] 魏琳慧,刘国文,刘雨,等.基于深度强化学习的卫星互联网路由优化研究[J].天地一体化信息网络,2022,3(3):65-71.

[10] 杨迈.卫星光网络路由和资源管理方法研究[D].北京:北京邮电大学,2024.

[11] 董飞虎.低轨星座计算卸载与路由选择技术研究[D].北京:北京邮电大学,2024.

[12] Harsha Vardhan B K, Sharma D K, Prakash C B. Enhancing network performance in leo satellite networks through drl-based routing optimization[C]//IEEE. Proceedings of 2023 Global Conference on Information Technologies and Communications(GCITC). Bangalore:IEEE,2023:1-6.

[13] Hlavatý M, Kozáková A. Development of advanced control strategy based on soft actor-critic algorithm[C]//IEEE. Proceedings of 2023 International Scientific Conference on Computer Science (COMSCI). Sozopol:IEEE,2023:1-5.