

面向巨型星座的智能负载均衡算法

罗树欣¹, 张超¹, 肖勇², 刘建平²

(1. 西安交通大学电子与信息学部信息与通信工程学院, 陕西 西安 710049;

2. 西安卫星测控中心宇航动力学国家重点实验室, 陕西 西安 710043)

摘 要: 针对巨型星座中卫星数量众多容易引发局部拥塞的问题, 提出基于协作多智能体深度强化学习的巨型星座负载均衡算法。首先对巨型星座中的卫星进行分簇设计, 实现巨型星座的分布式管理, 降低网络管理开销。然后, 利用 Q -混合多智能体神经网络深度强化学习设计各卫星自主决策的路由规划方案, 实现多传输任务的簇内协同。此外, 提出基于自动编码器的簇状态压缩机制, 提高多智能体深度强化学习的效率。仿真结果表明, 所提算法相比于传统的单任务路由算法, 传输成功率可提升 40% 以上, 证明所提算法能够避免局部拥塞的发生, 提高巨型星座的传输效率。

关键词: 巨型星座; 多智能体深度强化学习; 负载均衡; 路由算法

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-8930.2023042

Intelligent Load Balancing Algorithm of Mega Constellation

LUO Shuxin¹, ZHANG Chao¹, XIAO Yong², LIU Jianping²

1. School of Information and Communication Engineering, Faculty of Electronic and Information Engineering,
Xi'an Jiaotong University, Xi'an 710049, China

2. State Key Laboratory of Astronautic Dynamics, Xi'an Satellite Control Center, Xi'an 710043, China

Abstract: To overcome the local traffic congestion caused by the huge number of satellites in mega-constellation, a load balancing algorithm based on multi-agent deep reinforcement learning was proposed. Firstly, the satellites in the mega constellation were divided into clusters to perform the distributed management of the mega constellation, which could reduce the overhead of whole network. Then, based on the coordinated multi-agent deep reinforcement learning model, routing planning, which could be individually operated by satellites in the mega constellation, was designed to achieve the intra-cluster coordination. Additionally, a cluster state compression mechanism with autoencoder was proposed to compress the state space and improve the efficiency of multi-agent deep reinforcement learning. Finally, simulation results showed that compared with the traditional single-task routing algorithm, the proposed algorithm could increase the transmission success rate by more than 40% and the proposed algorithm could efficiently avoid local traffic congestion.

Keywords: mega constellation, multi-agent deep reinforcement learning, load balancing, routing algorithm

0 引言

6G 通信系统旨在提供全球覆盖的低时延和大带宽服务, 以实现空天地海高速互联^[1]。低地球轨道 (Low Earth

Orbit, LEO) 卫星星座以其覆盖广、时延低等特点被认为是未来无线网络的重要组成部分。随着卫星发射成本和设备成本的降低, 低轨巨型星座得到了快速发展^[2]。低轨巨型星座最显著的特点就是卫星数量众多, 甚至可达上万颗。

收稿日期: 2023-07-10; 修回日期: 2023-11-21

通信作者: 张超, chaozhang@xjtu.edu.cn

基金项目: 国家重点研发计划资助项目 (No.2020YFB1806102); 陕西省重点研发计划资助项目 (No.2023-YBGY-251)

Foundation Items: National Key Research and Development Program of China (No.2020YFB1806102), Key Research and Development Program of Shaanxi (No.2023-YBGY-251)

以 SpaceX 公司的巨型星座 Starlink 为例,截至 2023 年 2 月,Starlink 已部署 3 600 多颗卫星,并计划最终部署 42 000 多颗卫星^[3]。星座规模的扩大提高了卫星网络的传输容量,扩大了通信覆盖范围,但也给卫星网络的系统设计和系统管理带来了巨大挑战^[4]。

在巨型星座网络中,由于卫星数量众多,传输任务的跳数较传统星座网络大幅增加,这对路由规划提出了很高要求,具有很大的挑战性^[5]。一方面,卫星数量的增加会导致传输转发次数的增加,也会导致路由决策的计算量急剧增加。这样,路由决策的开销,例如用于路由路径计算的计算资源和用于传输数据包的传输资源^[6],会大量消耗巨型星座网络的可用资源。另一方面,考虑到建设成本,巨型卫星星座一般由小卫星构成,因此卫星的成本和尺寸都是受限的,其计算资源和传输资源也并不充分^[7]。此外,巨型星座的传输负载较大并且分布不均匀,容易产生网络拥塞^[8-9]。

虽然星座规模的扩张带来了诸多问题,但是巨型星座也为路由传输任务提供了更多的路径选择和设计自由度^[10]。如果能协调多个传输任务,尽量避免不同传输任务同时经过同一颗卫星,则可完成巨型星座网络的负载均衡,从而可以避免网络拥塞和减少端到端传输时延^[11]。在小型卫星网络中,一般使用多路径路由规划完成卫星网络负载均衡,例如参考文献[12-13]。这种算法依据网络拓扑信息和负载信息完成多条传输路径选取及数据分流,因此需要全局拓扑信息和链路状态信息。然而在巨型星座网络中,网络规模巨大并且拓扑变化频繁^[13],卫星无法及时获取全局信息,很难完成全局路由计算,因此该类型算法无法有效应用于巨型星座。

目前对巨型星座的路由算法研究主要面向单个传输任务,一般利用巨型星座网络结构构建理想化拓扑模型完成路由规划。为提高传输任务的可靠性,参考文献[14]利用巨型星座网络中的星间链路来降低数据传输失败的风险,提出了容错低时延冗余多路径路由算法。在保证上述路由算法时延最优的前提下,为降低路由算法的复杂度,参考文献[15]提出了一种基于网络拓扑结构的次优路由算法。参考文献[16]使用时间网络模型来描述巨型星座卫星的时变拓扑,其将整个空间划分为若干小立方体,并通过网格坐标而非卫星坐标来定位卫星,从而降低了路由计算复杂度和网络复杂度。需要指出的是,这两种算法使用了集中式路由计算机制,对于节点的部署和信息收集有较高要求,不适用于大规模的星座系统。根据巨型星座网络拓扑的特征,参考文献[17]以最小开销原则设计了分布式路

径规划算法。该算法可用于较大规模的卫星星座网络,且在链路故障时实现了较低的端到端时延。然而,以上所述的路由算法都是面向单个传输任务的,未考虑不同传输任务冲突造成的局部传输拥塞问题,缺少负载均衡机制。

综上所述,巨型星座由于其卫星数量众多、网络节点稠密、卫星设备资源受限等特点,在面对大量传输任务时,容易发生局部传输拥塞,从而导致传输时延增加、效率降低。为解决该问题,根据巨型星座网络的特点,本文提出了基于协作多智能体深度强化学习的负载均衡策略,该算法能够协调各卫星传输任务,避免数据包被大量传输至同一区域,提高了巨型星座的传输效率并降低了传输时延。

1 巨型星座系统

1.1 网络模型

巨型星座由 $N \times M$ 颗卫星组成,其中 N 为卫星轨道条数, M 为每条卫星轨道上的卫星数。由于极地用户少,巨型星座一般采用 Walker-Delta 星座实现特定纬度带的传输覆盖。因此,本文针对此类倾斜轨道的巨型星座系统进行讨论。

为完成组网,巨型星座中的卫星之间使用卫星星间链路 (Inter Satellite Link, ISL) 进行通信。依据卫星相对运动模型的不同,ISL 分为轨道内 ISL 和轨道间 ISL^[18]。轨道内 ISL 负责连接同一轨道上的卫星,由于轨道内卫星相对位置不发生变化,因此轨道内 ISL 可以持续连接。轨道间 ISL 负责连接不同轨道上的卫星,由于不同轨道间卫星的相对运动,通常需要使用天线追踪技术维持轨道间 ISL。需要注意的是,某些卫星在经过极区时由于链路状态的剧烈变化,通常会选择关闭轨道间 ISL^[19]。本文考虑使用 4 个固定的 ISL 进行卫星网络通信,即两个连接相同轨道卫星的轨道内 ISL 和两个连接相邻轨道卫星的轨道间 ISL。参考 Starlink 卫星星座一期工程的核心星座,图 1 给出本文采用的倾斜轨道巨型星座网络拓扑模型。

路由规划和计算需要知道网络拓扑状态信息和节点负载信息,前者保证数据信息的可靠传输,后者保证数据信息的低时延传输。一般而言,卫星网络路由策略往往由管理卫星或地面控制中心获取相关信息并完成路由计算,然后分发给各颗卫星^[20]。然而,直接将这种集中式路由由计算机应用于巨型星座网络会产生两个问题:第一,各卫星需要将自身的负载信息和链路状态信息传送给管理卫星或地面控制中心作为路由决策参考,但巨型星座卫星规模巨大,集中式的信息收集会造成巨量的传输资源消耗;

第二，新的路由策略分发至目标卫星时，其负载信息和链路状态信息可能已发生改变，原路由决策失去了时效性。为了适应巨型星座的以上特点，降低路由策略的开销，本文设计了一种巨型星座的分簇管理机制，实现了巨型星座的分布式管理和分布式路由决策。

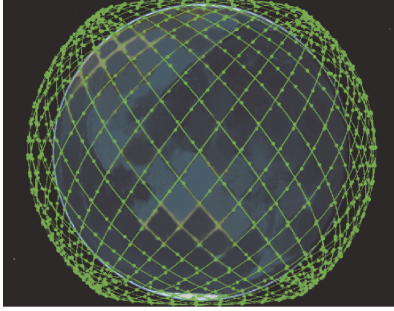


图1 倾斜轨道巨型星座网络拓扑模型

1.2 星座分簇机制

令集合 $C = \{C_1, C_2, \dots, C_K\}$ 表示分簇后的巨型星座，其中 K 为分簇数量。簇 C_i 定义为 $C_i = \{V_i, E_i\}$ ， V_i 为簇 C_i 内的卫星节点集合， E_i 为簇 C_i 内的星间链路集合。为了便于管理，每颗卫星只能属于一个簇。每个簇包含一个簇首和若干个簇成员，簇首负责收集簇成员的负载信息和链路信息，完成路由策略计算，并将路由策略分发至簇中的每个簇成员。分簇机制的开销包括簇首之间的信息洪泛、簇成员之间的信息洪泛及簇首到簇成员之间的信息洪泛。这3种洪泛分别完成全局信息的收集、簇首的重新选举和簇成员的信息收集。参考文献[21]运用博弈论讨论了小卫星集群管理问题，提出集群内卫星数量相同时，集群管理负载最低。同样，本文的巨型星座分簇机制也需要簇内的信息收集和簇间的信息交换。因此，本文也考虑每个簇中分配相同数量卫星以降低开销。

对一个由 $N \times M$ 颗卫星组成的巨型星座，分簇规模为 $X \times Y$ 时，其分簇数量为 $K = \frac{N}{X} \times \frac{M}{Y}$ 。对于簇 C_i ，其簇首 $C_{i,H}$ 的轨道编号定义为 $o_{i,H}$ ，轨道内编号定义为 $p_{i,H}$ ，则分布在此分簇内的卫星 v_j 的轨道编号 o_j 满足

$$o_{i,H} - \left\lceil \frac{X}{2} \right\rceil < o_j < o_{i,H} + \left\lceil \frac{X}{2} \right\rceil \quad (1)$$

式中， $\lceil \cdot \rceil$ 表示向上取整，分布在此分簇内的卫星 v_j 轨道内编号 p_j 需满足

$$p_{i,H} - \left\lceil \frac{Y}{2} \right\rceil < p_j < p_{i,H} + \left\lceil \frac{Y}{2} \right\rceil \quad (2)$$

图2给出巨型星座拓扑分簇模型。

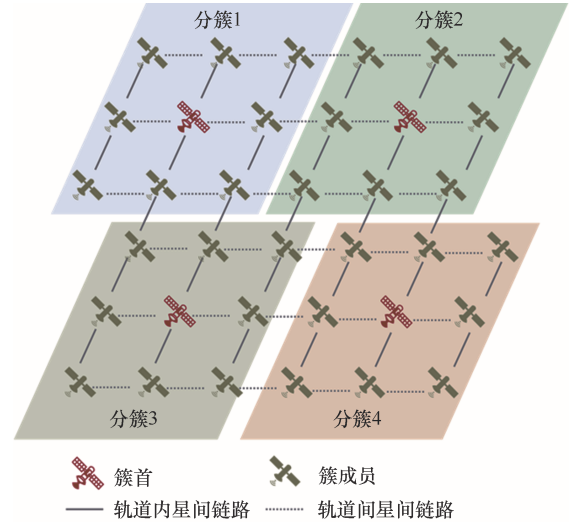


图2 巨型星座拓扑分簇模型

2 巨型星座智能负载均衡算法

2.1 负载均衡问题

在卫星网络中，如果传输路径上的某个节点发生拥塞，那么拥塞节点等待发送数据包の時延就会变长，最终导致数据包传输任务的时延变长。在卫星网络中，传输任务时延一般包括传播时延、传输时延和等待时延^[22]。对于从节点 i 到节点 j 的一跳数据包传输任务，令其路由由路径（Routing Path, RP）为 $rp_{i,j}$ 。那么，其传输时延 $d_{rp_{i,j}}$ 可以表示为

$$d_{rp_{i,j}} = \frac{d_{i,j}}{v_p} + \frac{l_p}{R_s} + n \frac{l_p}{R_s} \quad (3)$$

式中， $d_{i,j}$ 为节点 i 到节点 j 之间的距离， v_p 为信号传播速度， l_p 为数据包长度， n 为缓冲区数据包的数量， R_s 为发送速率。与传统的卫星网络不同，巨型星座中卫星数量众多，由星间距离差异引起的传输时延变化对传输任务总体时延影响不大^[23]，因此卫星网络拓扑可以建模为曼哈顿网络模型^[24]。一般使用虚拟距离 $d_{i,j}$ 来表示卫星之间的距离，为

$$d_{i,j} = \min \left\{ |o_i - o_j|, N - |o_i - o_j| \right\} + \min \left\{ |p_i - p_j|, M - |p_i - p_j| \right\} \quad (4)$$

式中， (o_i, p_i) 为卫星 i 的轨道编号和轨道内编号。对于一个从源节点 s 到目的节点 d 的传输任务，其整体传输时延 $D(s, d)$ 可以表示为

$$D(s, d) = \sum_{rp_{i,j} \in \text{rp}} d_{rp_{i,j}} \quad (5)$$

对于卫星网络传输任务，其传输能量开销 $E(s, d)$ 由两

部分组成,即数据包传输能量开销 E_t 和路由计算开销 E_f 。

$$E(s, d) = \theta_1 E_t + \theta_2 E_f \quad (6)$$

式中, θ_1 和 θ_2 为能量权重因子。对于一个传输任务,传输能量开销可以表示为

$$E_t = \sum_{v_i \in \pi p} e_{t,i} \quad (7)$$

式中, v_i 为传输任务的传输路径上途径的卫星节点, $e_{t,i}$ 为卫星 v_i 在当前传输任务的能量消耗。 $e_{t,i}$ 由 3 部分组成,分别为信号发射功耗 $e_{i,s}$ 、信号接收功耗 $e_{i,r}$ 以及传输任务固定能量开销 $e_{c,i}$, 即

$$e_{t,i} = e_{i,s} R_i + e_{i,r} R_i + e_{c,i} \quad (8)$$

其中, R_i 为卫星节点 i 的传输速率。

路由计算的能耗 E_f 由 3 部分组成。第一部分为簇内每个成员上传经验值的能量开销 E_u , 第二部分为簇首下发路由策略的能量开销 E_d , 最后一部分是用于路由策略计算的固定开销 E_p 。因此, E_f 可表示为

$$E_f = \mu_1 N_{CM}^2 E_u + \mu_2 N_{CM} E_d + \mu_3 E_p \quad (9)$$

式中, μ_1 、 μ_2 和 μ_3 为权重因子。

负载均衡问题优化目标及约束可表示为

$$\begin{aligned} \arg \min_{\pi p} \sum_i \sum_k \{ & \omega_1 D_k(s_k, d_k) + \omega_2 E_k(s_k, d_k) \} \\ \text{s.t. 式(3)~式(9)} \\ & \sum_{\pi p_{i,j}} r_{\pi p_{i,j}} \leq R_{i,j} \\ & \pi p_{i,j}, v_i \in \mathcal{RP} \\ & s_k, d_k \in \mathcal{V} \end{aligned} \quad (10)$$

式中, ω_1 和 ω_2 为权重因子, D_k 为第 k 个传输任务的传输时延, E_k 为第 k 个传输任务的能量消耗, $R_{i,j}$ 为卫星节点 i 和卫星节点 j 之间的最大传输速率, s_k 和 d_k 为第 k 个传输任务的源节点和目标节点。

数据包发送到下一个节点后,其状态变化将不再受当前节点的影响,因此传输过程可以建模为马尔可夫过程,并通过深度强化学习解决相关问题。

2.2 多智能体深度强化学习模型

深度强化学习 (Deep Reinforcement Learning, DRL) 的智能体可以动态感知环境的变化,并通过与环境的互动交流获取学习经验^[25],在游戏、自动驾驶、机器人控制等领域已经取得了巨大的成功^[26]。同时, DRL 在网络路由领域也得到了广泛的运用,如动态智能路由路径选择^[27]。多智能体深度强化学习 (Multi-Agent Deep Reinforcement Learning, MADRL) 是深度强化学习的进一步发展^[28],其主要解决复杂环境中深度强化学习的多个智能体间的协

作问题,可以获取全局的最大收益。在更复杂的环境中,单个智能体往往无法感知环境的全局状态,只能观察其周围一定范围内的信息,因此无法做出最佳决策,而 MADRL 通过协调每个智能体的决策策略可以实现全局奖励的最大化。所以, MADRL 在处理大规模复杂网络的决策任务上具有比较大的优势。

在不同的环境中,不同智能体间可能存在协作或竞争关系。多智能体强化学习可以划分为完全协作型、完全竞争型及混合型^[29]。在完全协作型博弈模型中,每个智能体的奖励函数相同,可以通过协作以获取最大奖励。因此本文采用了完全协作型博弈,从而可以通过优化不同传输任务的传输路径以达到总时延最小的目标。

本文采用了完全协作型多智能体深度强化学习模型的 Q -混合多智能体神经网络^[30]。 Q -混合多智能体神经网络借鉴了演说家-评论家 (Actor-Critic) 模型^[31]的思想,解决了训练阶段和执行阶段信息不对等的问题,即训练时可以获取全局信息,但执行时智能体只获取部分观测值。 Q -混合多智能体神经网络由两部分组成:混合网络和智能体网络。混合网络只在训练时存在,通过全局信息指导智能体网络进行动作决策及更新;智能体网络则类似于演说家网络,负责输出动作决策,在训练阶段和执行阶段保持相同的输入。 Q -混合多智能体神经网络的结构如图 3 所示。各智能体网络接收观测输入 (o_i, u_{i-1}) 并输出行动价值函数 $Q_i(\tau, u_i)$, 行动价值函数 $Q_i(\tau, u_i)$ 代表智能体在状态 τ 采取动作 u_i 时的长期奖励值。混合网络接收所有智能体的行动价值函数 $Q_i(\tau, u_i)$, 并依据状态值 s_i 输出总体行动价值函数 $Q_{tot}(\tau, u)$, 据此完成环境中多智能体联合行为的价值评估。

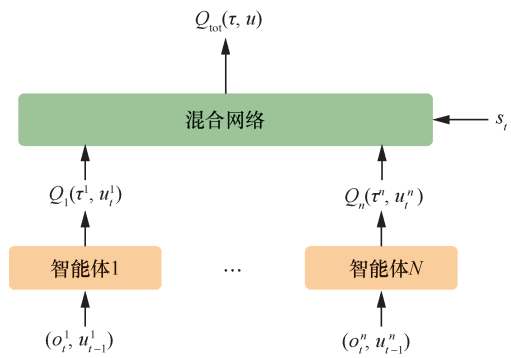


图 3 Q -混合多智能体神经网络结构

每个智能体可以使用不同的网络,也可以使用相同的网络。图 4 给出智能体网络结构,其包含 3 部分:输入层多层感知器 (Multi-Layer Perceptron, MLP)、中间层门控

循环单元 (Gate Recurrent Unit, GRU) 和输出层多层感知器。MLP 负责完成特征提取并输出行动价值函数。GRU 负责提取输入数据的时间特征, 解决智能体的部分观测问题。MLP 及 GRU 可以使用深度学习框架 Pytorch 实现, 通过 Tensor 函数可完成 MLP 的构建, 通过 GRUCell 函数可以完成 GRU 的构建。智能体网络以环境观测值 o_t 作为输入, 经 MLP 提取特征后作为 GRU 的输入。GRU 综合第一层 MLP 的输出及前置状态 h_{t-1} 作为输入, 提取数据时间特征后的输出值再经一层 MLP 提取特征后即可获取行动价值函数 $Q_n(\tau^n, u_t^n)$ 。 $Q_n(\tau^n, u_t^n)$ 选取动作时采用 ϵ -greedy 贪心策略^[32]。

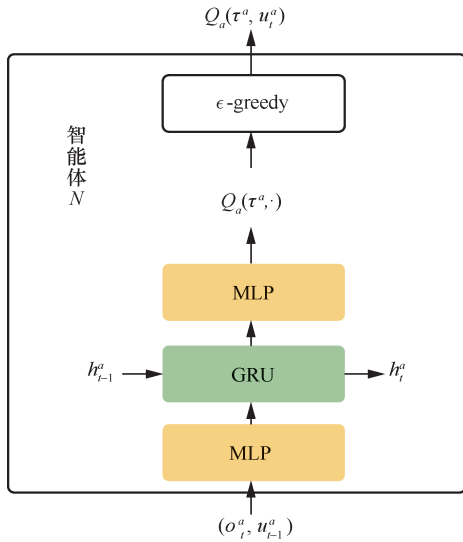


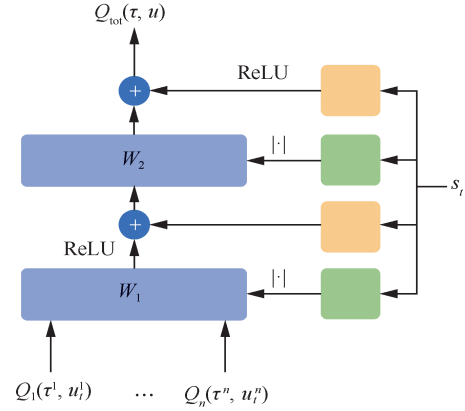
图4 智能体网络结构

图5 给出混合网络状态值 s_t 的组成单元及工作流程。混合网络首先获取环境状态值, 生成其自身的权重值 W_1 、 W_2 及偏置值。权重值生成时使用绝对值处理避免负值。权重值及偏置值参数生成完毕后, 混合网络以 n 个智能体的 Q 值作为输入, 经两层全连接单元及 ReLU 激活函数后输出总体行动价值函数 $Q_{\text{tot}}(\tau, u)$ 。

Q -混合多智能体神经网络实现智能体任务协作的关键有两点: 一是在训练过程中加入全局信息进行辅助; 二是采用混合网络对单智能体的奖励值进行合并而非简单的求和。 Q -混合多智能体神经网络求解的最佳动作是可以使各智能体奖励值最大化的动作组合, 即

$$\arg \max_u Q_{\text{tot}}(\tau, u) = \begin{pmatrix} \arg \max_{u^1} Q_1(\tau^1, u_t^1) \\ \vdots \\ \arg \max_{u^n} Q_n(\tau^n, u_t^n) \end{pmatrix} \quad (11)$$

式中, $Q_{\text{tot}}(\tau, u)$ 为环境中智能体的总奖励值, $Q_n(\tau^n, u_t^n)$ 为第 n 个智能体网络的奖励值。 Q -混合多智能体神经网络只要求总奖励值与各奖励值的单调性关系相同, 即总奖励值在环境中各智能体的奖励值最大的情况下达到最大值。

图5 混合网络结构状态值 s_t 的组成单元及工作流程

2.3 智能负载均衡算法设计

为解决巨型星座负载均衡问题, 本节给出基于 Q -混合多智能体神经网络的负载均衡算法。对巨型星座的每颗卫星而言, 如果进行路由决策时能够利用周围卫星的负载信息, 那么整个网络的拥塞避免是可能实现的。但协作路由决策是一个高状态维度和高选择维度的问题, 传统算法很难处理这类决策问题, 而多智能体深度强化学习在处理类似的问题时具有巨大的优势, 高状态维度以及高选择维度可以通过神经网络解决, 并且可以通过与环境的动态交互做出实时决策。

(1) 观测空间、动作空间、状态空间及奖励函数的设计
定义每颗卫星的观测空间 $o(t)$ 为

$$o(t) = \{o_s, p_s, o_d, p_d, b_u, b_d, b_l, b_r\} \quad (12)$$

式中, o_s 为当前任务源卫星节点的轨道编号, p_s 为当前任务源卫星节点在轨道上的卫星编号, o_d 为当前任务的目标卫星节点的轨道编号, p_d 为当前任务的目标卫星节点的轨道卫星编号, b_u 、 b_d 、 b_l 、 b_r 为其相邻4颗卫星的负载信息, 每颗卫星根据观测信息使用部署其上的智能体神经网络独立进行路由决策。

巨型星座中每颗卫星只与周围4颗卫星建立固定ISL。因此, 可选的下一跳节点只有4个, 分别为同轨道内两颗相邻卫星以及相邻轨道的两颗卫星。因此, 定义动作空间 $a(t)$ 为

$$a(t) = \{l, r, u, d\} \quad (13)$$

式中, l 和 r 为轨道间传输方向, u 和 d 为轨道内传输方向。

状态空间被定义为当前卫星及周围若干卫星的负载信息。传输任务协作的本质就是选取后续若干区域内负载较低的下一跳进行传输。因此，状态空间 $s(t)$ 定义为

$$s(t) = \{b_1, b_2, \dots, b_i\} \quad (14)$$

式中， b_i 为其分簇内第 i 颗卫星的负载信息。

路由策略的优化目标是能量开销最小前提下的数据包尽快到达，为兼顾传输时延和能量开销，奖励函数定义为

$$r = \begin{cases} \epsilon_1 e^{\frac{1-\text{dis}}{\text{DIS}}} + \epsilon_2 \frac{E_{\text{next}}}{E_{\text{max}}}, & \Delta_{\text{dis}} > 0 \text{ 且 } b_{\text{next}} < B_m \\ \epsilon_1 e^{\frac{1-\text{dis}}{\text{DIS}}} + \epsilon_2 \frac{E_{\text{next}}}{E_{\text{max}}} - Z, & \Delta_{\text{dis}} > 0 \text{ 且 } b_{\text{next}} > B_m \\ -\gamma, & \text{其他} \end{cases} \quad (15)$$

式中，DIS 为传输任务源节点至目标节点的初始距离， E_{next} 为下一跳目标卫星当前剩余能量， E_{max} 为卫星能量最大值， γ 为惩罚因子， b_{next} 为靠近目标卫星的下一颗卫星的负载， B_m 为设定的卫星负载阈值， Z 为折扣因子。dis 为传输任务所在节点至目标卫星的距离，其定义为

$$\text{dis} = \frac{1}{A_0 + A_p} \quad (16)$$

式中， A_0 为两颗卫星的轨道内距离， A_p 为两颗卫星轨道间距离。式(15)中引入非线性函数的目的是：当数据包与目标卫星的距离较远时，即 dis 远小于传输任务初始距离 DIS 时，奖励值变化不大，智能体可以探索各种协作方案；当数据包靠近目标卫星时，奖励值快速变化，保证数据包尽快到达，避免对其他数据包造成干扰，同时，尽可能选择剩余能量较多的下一跳卫星，以保证网络的可持续运行，其可以表示为下一跳目标卫星当前剩余能量 E_{next} 与最大值 E_{max} 之比。折扣因子 Z 保证下一跳传输方向为负载较低的方向，抑制不必要的前向传输。另外，当下一跳的传输方向远离目标卫星时，奖励值定义为惩罚因子 γ 。

(2) 训练及执行策略选择

多智能体深度强化学习分为训练阶段和执行阶段。训练阶段可以采用集中式或分布式训练策略。为避免每颗卫星均收集周围卫星状态信息的开销过大，不考虑在每颗卫星上完成智能体训练，而使用集中式智能体训练。

图 6 给出负载均衡策略工作流程。首先由每个卫星簇

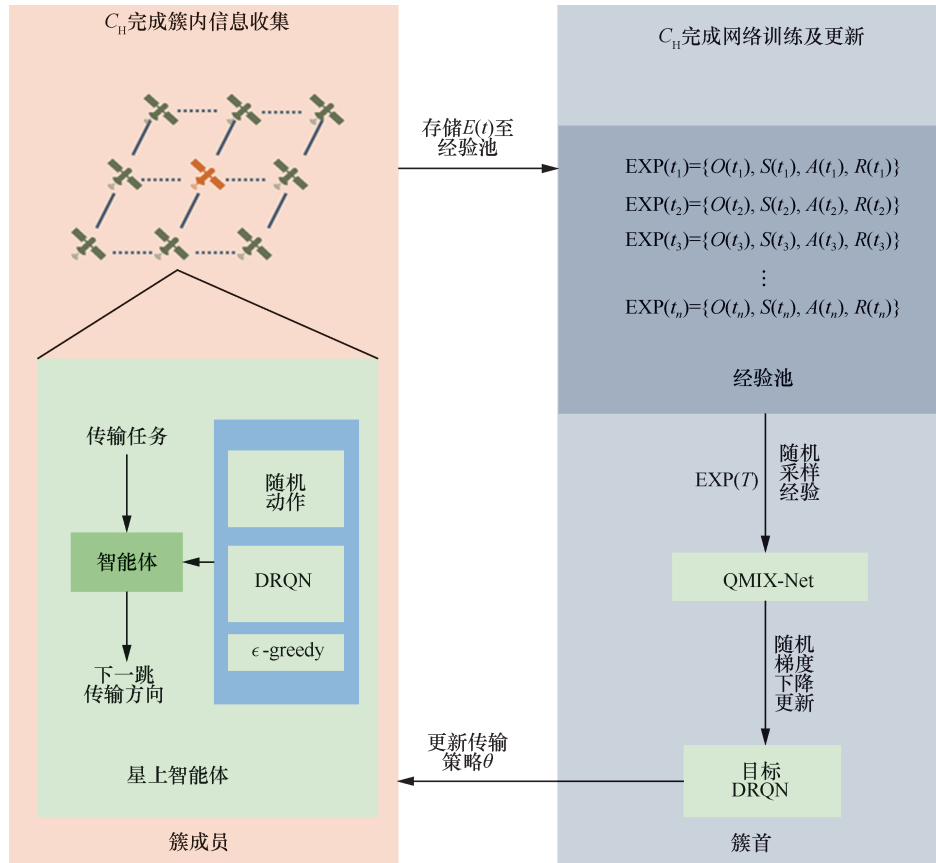


图 6 负载均衡策略工作流程

的簇首卫星收集簇内的负载信息以及每颗卫星的观测信息作为经验值，并存储至经验池，而后由部署于簇首的 Q -混合多智能体神经网络对经验池进行随机采样，依据采样数据进行智能体训练，同时定期更新目标深度递归 Q 网络 (Target-Deep Recurrent Q Network, Target-DRQN)。Target-DRQN 仅负责接收经验池数据并完成训练更新，不负责具体决策过程，以降低环境相关性对智能体训练的影响。训练完毕后，簇首卫星下发传输策略至簇内各卫星。执行阶段，每颗卫星依据簇首卫星下发的路由策略，完成传输任务观测，独立地执行路由决策，完成数据传输下一跳方向选取。因此，本文所提负载均衡算法采取了集中训练和分布式执行的策略。该策略在巨型星座中具有如下的优势：第一，星上设备只负责路由决策过程，计算开销很小，可以很好地延长网络工作寿命；第二，将智能体训练放置在每个簇的簇首卫星中，可以在较短的时间内完成簇内信息的收集，具有更好的环境适应性。

(3) 状态空间压缩

分簇规模对传输性能有重大影响。随着分簇规模的增大， Q -混合多智能体神经网络获得更多的状态信息，其混合网络可以协调的智能体数量增加，传输任务投递成功率可以得到有效提升。然而，分簇规模越大，状态空间的维数就会越大， Q -混合多智能体神经网络的计算复杂度就会越高。综上，需要在保留状态空间所包含的负载信息前提下，尽量压缩状态空间维度，兼顾计算效率和传输成功率。

本文采用自动编码器 (Auto Encoder, AE) 完成状态空间的压缩。自动编码器是一种无监督的神经网络模型，可用于数据特征挖掘和特征降维^[33]。图 7 给出簇内状态压缩机制具体执行流程。在执行簇内状态压缩时，首先对簇负载信息 B_i 进行矢量化作为自动编码器的输入，而后自动编码器依据输入 B_i 进行前向传播输出状态值 s_i ，经过压缩后的状态值 s_i 即可作为混合网络的输入。

巨型星座智能负载均衡的算法流程如算法 1 所示。

算法 1：巨型星座智能负载均衡算法

输入：当前卫星分簇集合 B ，当前卫星传输任务集合 M

1. 初始化经验池 D
2. 随机初始化智能体网络参数 θ
3. for episode = 1 to N do
4. for each 簇 C_i 的负载 B_i do
5. 压缩 B_i 至 B_{ic}
6. for each 簇 C_i 内的卫星 s do
7. 获取观测空间 $o(t) = \{p_s, o_s, p_d, o_d, b_u, b_d, b_l, b_r\}$

8. 选择动作 $a_t = \begin{cases} \text{随机动作, 以概率 } \epsilon \\ \arg \max_a Q_n(\theta_n), \text{ 以概率 } 1-\epsilon \end{cases}$
9. 依据动作 a_t 传输数据包
10. 获取奖励值 r_t
11. 更新观测值 $o(t)$ 至 $o(t+1)$
12. end
13. 存储经验值 $\text{EXP}(t) = \{O(t), S(t), A(t), R(t)\}$
14. if 经验池经验数量 > 缓冲区数量 then
15. 丢弃早期经验值
16. end
17. end
18. for epoch = 1 to T do
19. if 经验池大小 > 训练批数据大小 then
20. 簇首随机采样经验值 $\text{EXP}(t)$
21. 网络参数 θ 梯度求解
22. 网络参数更新: $\theta \leftarrow \theta + \Delta \theta$
23. end
24. if episode % $T_{\text{update}} = 0$ then
25. $\theta' \leftarrow \theta$
26. end
27. end

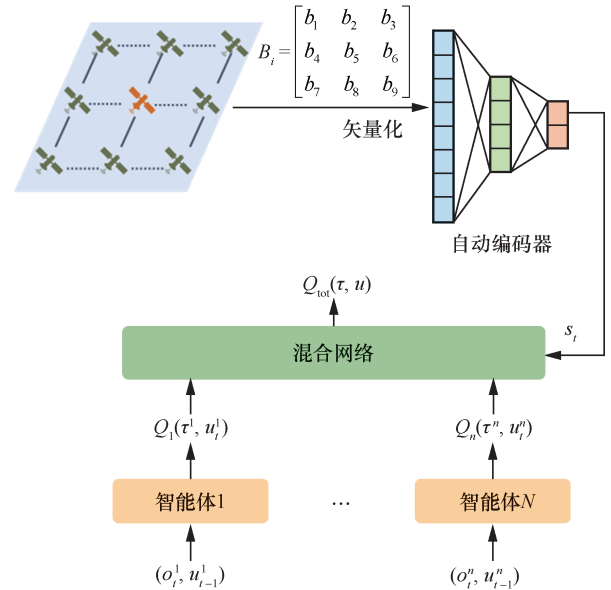


图 7 簇内状态压缩机制具体执行流程

3 算法仿真与结果分析

3.1 仿真环境说明

当前主流网络仿真软件，例如 OPNet、Qualnet 等，其仿真流程过于复杂，与路由过程不相关的流程较多，当

仿真网络规模巨大时仿真速度过于缓慢。为专注于路由过程仿真,本文使用 NetWorkX 库^[34]构建了巨型星座网络。仿真默认分簇规模为 6×6 。由于神经网络的超参数对神经网络性能有明显影响,根据目前的主流经验,中间层的神经元数量选择为 $\sqrt{N_{\text{PreviousLayer}} N_{\text{NextLayer}}}$, $N_{\text{PreviousLayer}}$ 和 $N_{\text{NextLayer}}$ 分别为中间层的前一层和后一层神经网络的神经元数量,其他神经网络关键参数见表 1。

表 1 其他神经网络关键参数

仿真参数	值
经验池大小	$200 \times N \times M$
TTL	$2 \times N \times M$
折扣因子	0.99
更新周期	40
DRQN 隐藏层维度	4
超网络隐藏层维度	分簇大小
起始探索概率	1
中止探索概率	0.1
学习率	0.01

3.2 传输性能对比

为了说明所提算法的有效性和创新性,这里仿真了参考文献[17]所提的分布式生存路由算法 (Distributed Survivable Routing Algorithm, DSRA)。DSRA 也是针对大规模巨型星座而设计的,其通过一定范围之内的洪泛令每颗卫

星获取周围卫星链路的信息,从而可以确定质量最好的主要传输路径以及质量稍差的辅助传输路径,通过卫星编号确定传输方向并实现拥塞避免。此外,为了验证多智能体协同的效果,借鉴参考文献[34]所提单智能体深度强化学习路由算法 (Deep Q-Network Routing Algorithm, DQNRA),在巨型星座系统中也仿真该方案。DQNRA 在每颗卫星上部署智能体,定期交换周围邻居负载信息,以传输任务及周围 4 颗卫星的负载作为状态信息,进行下一跳动作选择。

图 8 给不同星座规模下路由传输成功率对比。为探究所提算法在不同星座规模下的有效性,分别在 6×6 、 12×12 、 24×24 以及 32×32 的星座规模下进行了仿真。图 8 (a) 中,根据稳定的传输成功率计算发现,所提算法的传输成功率达到了 69.2%,分别比 DQNRA 及 DSRA 提高 40.8%及 95.3%。这是因为 DQNRA 只能保证自身奖励最大化,而忽略了与其余智能体奖励的影响。DSRA 类似于 DQNRA,只保证自身的传输任务,缺乏协同机制。通过对比图 8 (a)、图 8 (b)、图 8 (c) 及图 8 (d),根据稳定的传输成功率,计算发现,随着星座规模的扩大,所提算法的传输成功率比 DQNRA 分别提升了 40.8%、48.2%、58.3%及 61.2%。这说明在规模更大的星座中,所提算法的性能优势更大。其主要因为更大规模的星座中传输跳数更多,可选路径就更多,进行传输任务协同的解空间也相对更大。以上证明了所提算法可以在巨型星座中获得更高的传输成功率。

图 9 给出不同星座规模下的不同算法的传输时延对比。

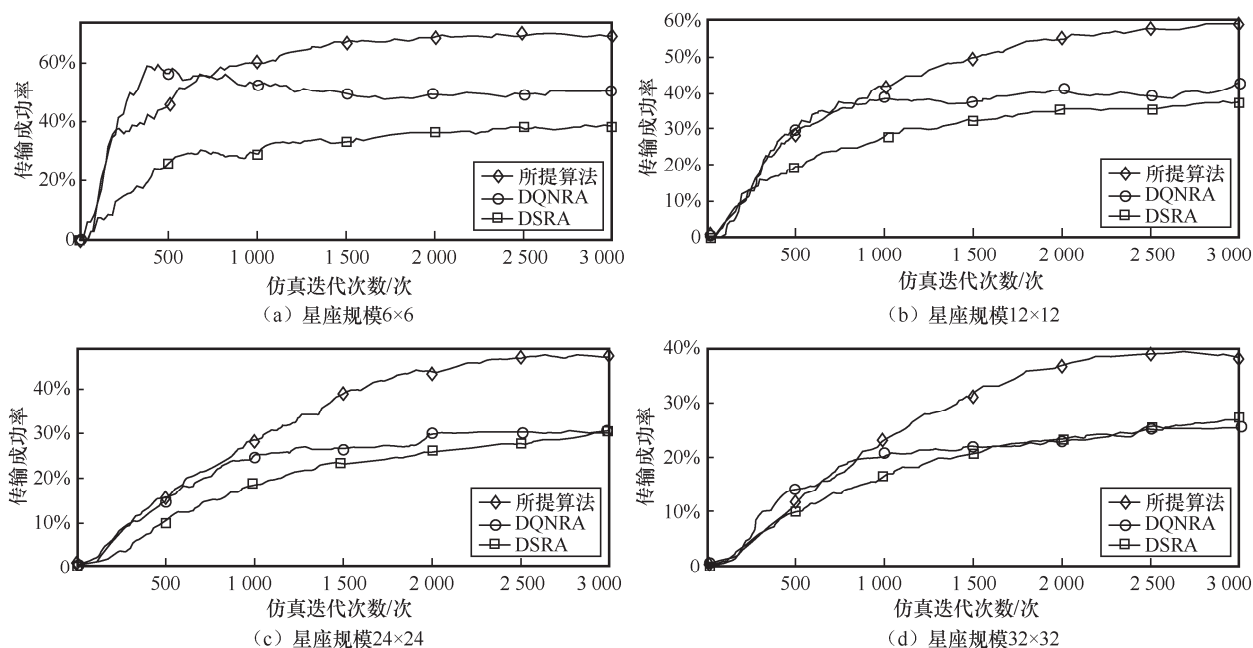


图 8 不同星座规模下路由传输成功率对比

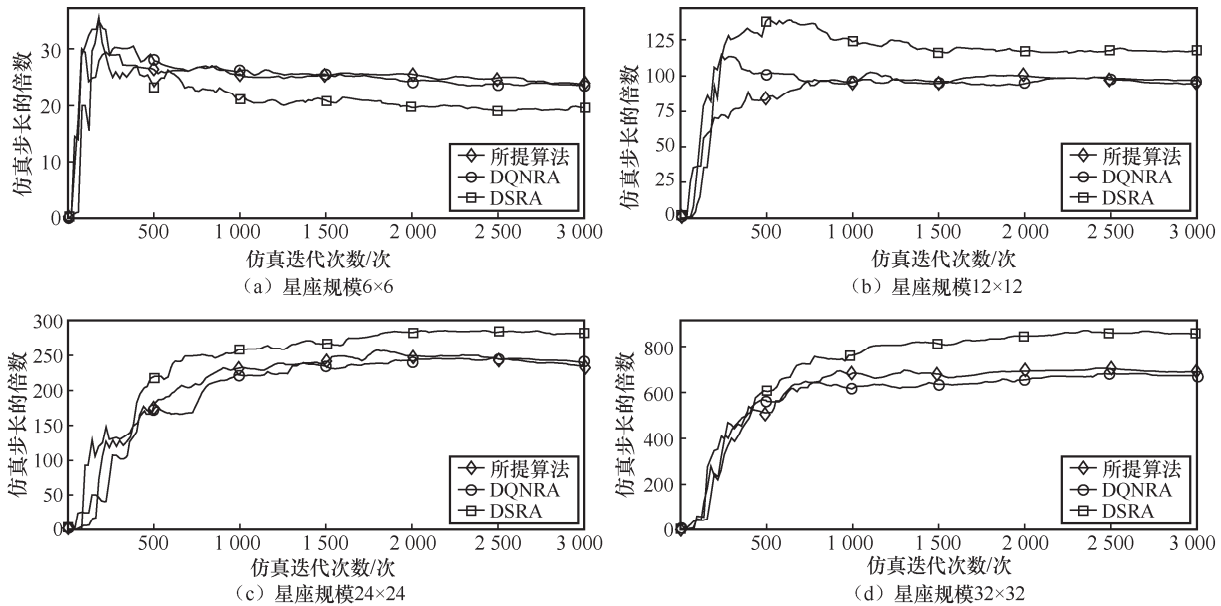


图9 各星座规模下不同算法传输时延对比

在前3种星座规模下,所提算法的传输时延和DQNRA几乎相同。这说明所提算法相较于DQNRA,在不增加传输时延的基础上可有效提升巨型星座的传输成功率。当星座规模为 6×6 时,DSRA的传输时延低于所提算法和DQNRA。而当星座规模为 12×12 、 24×24 和 32×32 时,所提算法和DQNRA的传输时延又显著低于DSRA。这是因为在星座规模比较小的情况下拥塞概率较低,未发生拥塞的传输任务可以迅速传输至目标卫星。星座规模增加时,传输任务的传输跳数增加,局部拥塞概率增加,DSRA的传输时延相应增加。因此,通过图8和图9,可以说明所提算法在星座规模较大时,传输成功率和传输时延两方面性能均优于DSRA和DQNRA两种算法。同时也证明了,所提算法在巨型星座中的有效性。

另外,在图8~图9中,仿真共迭代了3000次。可以看到,在 6×6 的星座下,所有算法大约1000次迭代后完成了收敛,而 12×12 、 24×24 、 32×32 的星座规模则分别需要迭代大约2200次、2500次和2650次。随着星座规模的扩大,每个传输任务可选的路径逐渐增多,相当于每个传输任务的动作空间变大,因此完成智能体之间的协同需要更长的训练时间。一旦收敛后,网络进入稳定的工作状态,随着时间的增长,训练时间可以忽略不计。

表2给出了不同算法下卫星最小剩余能量百分比。设所有卫星初始能量均为满电状态(即100%)。这里最小剩余能量百分比就是巨型星座全部卫星中剩余能量的最低

值与初始值的比。当星座规模为 12×12 时,所提算法相比DQNRA和DSRA提高2.63%和7.3%。类似的,星座规模为 6×6 、 24×24 和 32×32 时所提算法亦取得了最好的能效表现。其原因是DQNRA和DSRA无法完成负载均衡,大量的传输任务可能会集中在某些卫星上,从而导致该卫星能量消耗较大。通过最小剩余能量百分比,可以揭示出本文所提算法能够实现巨型星座的负载均衡。

表2 不同算法下卫星最小剩余能量百分比

算法	星座规模			
	6×6	12×12	24×24	32×32
所提算法	85.32%	72.84%	56.49%	39.19%
DQNRA	82.95%	70.21%	52.91%	29.33%
DSRA	77.18%	65.54%	44.68%	25.74%

3.3 分簇规模对性能的影响

图10是星座规模为 12×12 时不同分簇规模的性能对比。可以看出,随着分簇规模的逐渐增大,传输成功率逐渐提高,而平均传输时延逐渐降低。这是因为当分簇规模较大时,Q-混合多智能体神经网络可以尽可能完成智能体协调以避免拥塞,从而可以提高传输成功率和降低平均传输时延。图10还给出了不同分簇规模对计算时间的影响。当分簇规模为 6×6 时,Q-混合多智能体神经网络的混合网络和参数生成网络的神经元数量最多,因此计算时间最长。当分簇规模下降到 4×4 时,计算时间比分簇规模为

6×6 时减少了 50.1%。然而,随着分簇规模进一步降低,计算时间不再有明显变化,这是因为中间层的神经元数量已经减少到很低的水平,不会再产生明显变化。

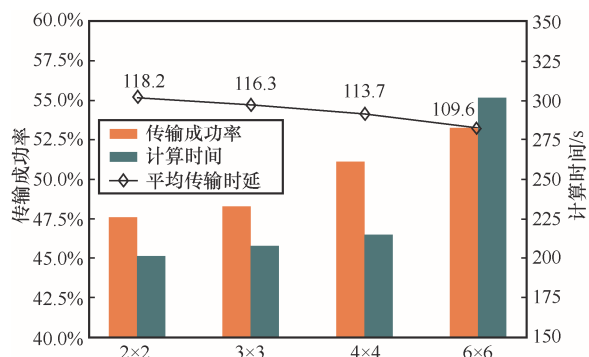


图 10 星座规模为 12 × 12 时不同分簇规模的性能对比

由上述仿真结果可以看出,分簇规模对传输性能及计算时间有显著的影响。首先,随着分簇规模的增加, Q -混合多智能体神经网络获取的状态信息更多, Q -混合多智能体神经网络的混合网络可以协调的智能体数量增加,可以尽可能地提高全局奖励值。因此,随着分簇规模的增大,传输成功率可以得到提升。另外,分簇规模对计算时间有较大的影响。 Q -混合多智能体神经网络中,状态值 s_t 作为参数生成网络的输入,输出值为参数生成网络的权重值。一般而言,中间层神经元数量取决于输入层和输出层。状态空间的维度取决于分簇规模,分簇规模的扩大引起了 Q -混合多智能体神经网络中参数生成网络的神经元数量增加。在实际应用时,由于训练时间只占很小的时间比重,可以在分簇规模引起的传输性能和计算时间变化间做适当权衡。

4 结束语

本文提出了基于 Q -混合多智能体神经网络的低轨巨型星座智能负载均衡算法。通过 Q -混合多智能体神经网络协调各卫星传输任务,指导部署于星上的智能体的决策行为,避免数据包被大量传输至同一区域造成拥塞,完成了巨型星座的智能负载均衡。仿真结果表明,本文所提算法在传输成功率和传输时延两方面的性能均优于目前主流的巨型星座路由算法。

参考文献:

[1] SHENG M, ZHOU D, BAI W G, et al. 6G service coverage with mega satellite constellations[J]. China Communications, 2022,

19(1): 64-76.

- [2] 阮永井, 胡敏, 云朝明. 低轨巨型星座构型设计与控制研究进展与展望[J]. 中国空间科学技术, 2022, 42(1): 1-15.
RUAN Y J, HU M, YUN C M. Advances and prospects of the configuration design and control research of the LEO mega-constellations[J]. Chinese Space Science and Technology, 2022, 42(1): 1-15.
- [3] CHAUDHRY A U, YANIKOMEROGLU H. Laser intersatellite links in a starlink constellation: a classification and analysis[J]. IEEE Vehicular Technology Magazine, 2021, 16(2): 48-56.
- [4] RADHAKRISHNAN R, EDMONSON W W, AFGHAH F, et al. Survey of inter-satellite communication for small satellite systems: physical layer to network layer view[J]. IEEE Communications Surveys & Tutorials, 2016, 18(4): 2442-2473.
- [5] YANG X. Low earth orbit (LEO) mega constellations: satellite and terrestrial integrated communication networks[D]. Guildford: University of Surrey, 2019.
- [6] 陈全, 杨磊, 郭剑鸣, 等. 低轨巨型星座网络: 组网技术与研究现状[J]. 通信学报, 2022, 43(5): 177-189.
CHEN Q, YANG L, GUO J M, et al. LEO mega-constellation network: networking technologies and state of the art[J]. Journal on Communications, 2022, 43(5): 177-189.
- [7] 周笛, 盛敏, 郝琪, 等. 巨型星座系统的网络运维与资源管控技术[J]. 天地一体化信息网络, 2020, 1(1): 26-35.
ZHOU D, SHENG M, HAO Q, et al. Network operation, maintenance and resource management in mega constellation system[J]. Space-Integrated-Ground Information Networks, 2020, 1(1): 26-35.
- [8] CHEN Q, GIAMBENE G, YANG L, et al. Analysis of inter-satellite link paths for LEO mega-constellation networks[J]. IEEE Transactions on Vehicular Technology, 2021, 70(3): 2743-2755.
- [9] QU Z C, ZHANG G X, CAO H T, et al. LEO satellite constellation for Internet of Things[J]. IEEE Access, 2017, 5: 18391-18401.
- [10] LEE Y, CHOI J P. Connectivity analysis of mega-constellation satellite networks with optical intersatellite links[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(6): 4213-4226.
- [11] XIE H R, ZHAN Y F, ZENG G M, et al. LEO mega-constellations for 6G global coverage: challenges and opportunities[J]. IEEE Access, 2021, 9: 164223-164244.
- [12] TANG F L, ZHANG H T, YANG L T. Multipath cooperative routing with efficient acknowledgement for LEO satellite networks[J]. IEEE Transactions on Mobile Computing, 2019, 18(1): 179-192.

- [13] 巢孟愿. 卫星网络多路径路由算法与切换策略研究[D]. 杭州: 浙江大学, 2014.
- CHAO M Y. Research on multipath routing algorithm and handover strategy in satellite networks[D]. Hangzhou: Zhejiang University, 2014.
- [14] ZENG G M, ZHAN Y F, PAN X H. Failure-tolerant and low-latency telecomm and in mega-constellations: the redundant multi-path routing[J]. IEEE Access, 2021, 9: 34975-34985.
- [15] LIU Y L, ZHU L D. A suboptimal routing algorithm for massive LEO satellite networks[C]//Proceedings of 2018 International Symposium on Networks, Computers and Communications (ISNCC). Piscataway: IEEE Press, 2018: 1-5.
- [16] LI J, LU H C, XUE K P, et al. Temporal netgrid model-based dynamic routing in large-scale small satellite networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(6): 6009-6021.
- [17] QI X X, ZHANG B, QIU Z L. A distributed survivable routing algorithm for mega-constellations with inclined orbits[J]. IEEE Access, 2020, 8: 219199-219213.
- [18] 倪少杰, 岳洋, 左勇, 等. 卫星网络路由技术现状及展望[J]. 电子与信息学报, 2023, 45(2): 383-395.
- NI S J, YUE Y, ZUO Y, et al. The status quo and prospect of satellite network routing technology[J]. Journal of Electronics & Information Technology, 2023, 45(2): 383-395.
- [19] 郑爽, 张兴, 王文博. 低轨卫星通信网络路由技术综述[J]. 天地一体化信息网络, 2022, 3(3): 97-105.
- ZHENG S, ZHANG X, WANG W B. Survey of low earth orbit satellite communication network routing technology[J]. Space-Integrated-Ground Information Networks, 2022, 3(3): 97-105.
- [20] 石晓东, 李勇军, 赵尚弘, 等. 基于SDN的卫星网络多QoS目标优化路由算法[J]. 系统工程与电子技术, 2020, 42(6): 1395-1401.
- SHI X D, LI Y J, ZHAO S H, et al. Multi-QoS objective optimization routing algorithm of satellite network based on SDN[J]. Systems Engineering and Electronics, 2020, 42(6): 1395-1401.
- [21] LIU J, ZHANG X Y, ZHANG R, et al. Reliable and low-overhead clustering in LEO small satellite networks[J]. IEEE Internet of Things Journal, 2022, 9(16): 14844-14856.
- [22] TANG F X, HOFNER H, KATO N, et al. A deep reinforcement learning-based dynamic traffic offloading in space-air-ground integrated networks (SAGIN)[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(1): 276-289.
- [23] QI X G, MA J L, WU D, et al. A survey of routing techniques for satellite networks[J]. Journal of Communications and Information Networks, 2016, 1(4): 66-85.
- [24] 沙鹏翔. 基于分簇的巨型低轨星座高效管理方法研究[D]. 南京: 南京大学, 2021.
- SHA P X. Research on efficient management method of giant LEO constellation based on clustering[D]. Nanjing: Nanjing University, 2021.
- [25] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2018, 41(1): 1-27.
- LIU Q, ZHAI J W, ZHANG Z C, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.
- [26] 裴培, 何绍溟, 王江, 等. 一种深度强化学习制导控制一体化算法[J]. 宇航学报, 2021, 42(10): 1293-1304.
- PEI P, HE S M, WANG J, et al. Integrated guidance and control for missile using deep reinforcement learning[J]. Journal of Astronautics, 2021, 42(10): 1293-1304.
- [27] NAZARI M, OROOJLOOY A, TAKÁČ M, et al. Reinforcement learning for solving the vehicle routing problem[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 9861-9871.
- [28] 孙戡, 曹雷, 陈希亮, 等. 多智能体深度强化学习研究综述[J]. 计算机工程与应用, 2020, 56(5): 13-24.
- SUN Y, CAO L, CHEN X L, et al. Overview of multi-agent deep reinforcement learning[J]. Computer Engineering and Applications, 2020, 56(5): 13-24.
- [29] ZHANG K, YANG Z, BAŞAR T. Multi-agent reinforcement learning: A selective overview of theories and algorithms[J]. Handbook of reinforcement learning and control, 2021: 321-384.
- [30] RASHID T, SAMVELYAN M, DE WITT C S, et al. QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning[EB]. 2018.
- [31] 陈亮, 梁宸, 张景异, 等. Actor-Critic 框架下一种基于改进DDPG的多智能体强化学习算法[J]. 控制与决策, 2021, 36(1): 75-82.
- CHEN L, LIANG C, ZHANG J Y, et al. A multi-agent reinforcement learning algorithm based on improved DDPG in Actor-Critic framework[J]. Control and Decision, 2021, 36(1): 75-82.
- [32] 刘建伟, 高峰, 罗雄麟. 基于值函数和策略梯度的深度强化学习综述[J]. 计算机学报, 2019, 42(6): 1406-1438.

LIU J W, GAO F, LUO X L. Survey of deep reinforcement learning based on value function and policy gradient[J]. Chinese Journal of Computers, 2019, 42(6): 1406-1438.

- [33] 韩冲, 王俊丽, 吴雨茜, 等. 基于神经进化的深度学习模型研究综述[J]. 电子学报, 2021, 49(2): 372-379.

HAN C, WANG J L, WU Y X, et al. A review of deep learning models based on neuroevolution[J]. Acta Electronica Sinica, 2021, 49(2): 372-379.

- [34] LIU J H, ZHAO B K, XIN Q, et al. DRL-ER: an intelligent energy-aware routing protocol with guaranteed delay bounds in satellite mega-constellations[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(4): 2872-2884.

[作者简介]



罗树欣(1998-),男,西安交通大学硕士生,主要研究方向为巨星座路由及仿真开发。



张超(1982-),男,博士,西安交通大学教授,主要研究方向为卫星互联网、6G技术等。



肖勇(1984-),男,博士,西安卫星测控中心高级工程师,主要研究方向为航天器飞行任务规划与调度。



刘建平(1975-),男,博士,西安卫星测控中心教授级高级工程师,主要研究方向为航天器飞行任务规划与调度。