OEE | 光电工程 Opto-Electronic Engineering

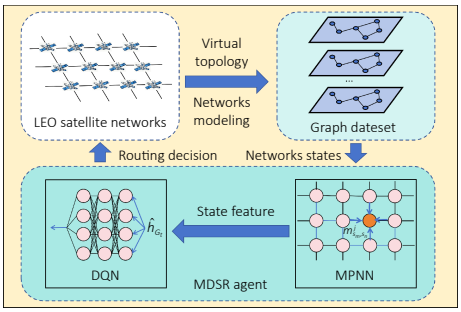
Article

2026 年 6 月, 第 53 卷, 第 6 期

面向大规模 LEO 卫星网络的深度图强化学习路由算法

黄思博*, 杨明川, 张祯昊

哈尔滨工业大学电子与信息工程学院, 黑龙江 哈尔滨 150001



摘要: 大规模低地球轨道 (LEO) 卫星网络路由技术面临着网络拓扑动态变化导致的路由传输性能不佳、算法模型可扩展性不足等关键挑战。为此, 利用改进的消息传递神经网络 (MPNN) 和深度 Q 网络 (DQN), 本文提出了一种基于 MPNN 和 DQN 的大规模 LEO 卫星网络路由算法 (MDSR)。其中, 改进的 MPNN 模块提取网络中卫星和星间链路的状态特征, DQN 模块基于这些特征实现路由决策。MDSR 算法在极轨道大规模 LEO 卫星网络上训练并测试。结果表明, 训练时 MDSR 模型的收敛速度和初始奖励均优于未改进 MPNN 模型。测试时 MDSR 的平均丢包率比其他强化学习算法降低约 48%, 比 Dijkstra 最短路径 (DSP) 算法降低约 35%; 平均端到端时延比 DSP、Q 学习 (Q-learning)、DQN 和深度确定性策略性梯度 (DDPG) 算法分别降低了约 20.5%、7.9%、2.7% 和 5.4%; 平均吞吐量比 DSP、Q-learning、DQN 和 DDPG 分别提升了约 21.7%、7.6%、5% 和 1.2%。同时, MDSR 无需额外训练即可适应不同星座网络构型, 具备较强的拓扑泛化能力。

关键词: 大规模低地球轨道卫星网络; 路由算法; 消息传递神经网络; 深度 Q 网络

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250386 | CSTR: 32245.14.oee.2026.250386 | 中图分类号: TN927.2; TP181 | 文献标识码: A

黄思博, 杨明川, 张祯昊. 面向大规模 LEO 卫星网络的深度图强化学习路由算法 [J]. 光电工程, 2026, **53**(6): 250386

Huang S B, Yang M C, Zhang Z H. Deep graph reinforcement learning routing algorithm for large-scale LEO satellite networks[J]. *Opto-Electron Eng*, 2026, **53**(6): 250386

1 引 言

随着 6G 标准的演进和天地一体化信息网络的发展, 卫星网络因其覆盖能力强、部署方式灵活和成本低廉等优点[1], 被视为未来 6G 通信网络架构中的关键技术之一。近年来, 卫星网络用户数量激增, 卫星业务数据日趋多样复杂, 对网络的规模和服务类型提出了更高要求。在此背景下, 随着卫星造价和发射成本的降低, 大规模低地球轨道 (low earth orbit, LEO) 卫星网络成为关键解决方案。大规模 LEO 卫星网络拥有 500 颗以上的网络节点和数以千计的通信链路资源, 具有网络规模大、网络拓扑结构动态复杂、传输速率高、多业务共存等特点, 能为用户提供高速率、低时延、高可靠性的通信服务[2]。其中, 路由技术是支撑高质量通信服务的重要基石[3]。通过多跳路由选择机制, 该技术不仅能实现用户间的长距离通信, 还能优化网络资源分配, 从而显著提升数据传输的可靠性。

大规模 LEO 卫星网络的动态拓扑为路由算法设计带来严峻挑战。为应对这一挑战, 路由算法需要具备实时感知网络状态变化、动态适配网络架构特征、智能调

收稿日期: 2025-12-12; 修回日期: 2026-03-02; 录用日期: 2026-03-12; 网络出版日期: 2026-06-25

基金项目: 国家自然科学基金面上项目 (62071146)

*通信作者: 黄思博, zerohsb2001@163.com

版权所有©2026 中国科学院光电技术研究所

配网络资源与传输约束以实现最优路由决策，并能灵活部署于不同星座配置这四大核心能力[4]。然而，现有算法大多仅基于小型 LEO 卫星网络展开设计，难以满足大规模网络业务需求，且其在多种构型网络场景中的适配性也未得到验证。因此，开发符合大规模 LEO 卫星网络特性的新型路由算法已成为当务之急。

近年来，人工智能 (artificial intelligence, AI) 技术蓬勃发展，以强化学习 (reinforcement learning, RL)、深度学习 (deep learning, DL) 为代表的技术被广泛运用于卫星网络，在路由技术、多维资源优化、波束成形等关键领域发挥重要作用[5]。其中，RL 通过智能体与网络环境的交互来获取信息并制定路由选择策略；DL 则通过强化神经网络对状态特征的认知，更有效地利用 LEO 卫星网络的资源。由此，二者结合而成的深度强化学习 (deep reinforcement learning, DRL) 兼具感知与决策功能，能够实时感知 LEO 卫星网络拓扑变化、负载变动以及服务质量，并据此做出最优决策，为提升卫星网络性能提供了可靠的方案。

与此同时，LEO 卫星网络的规模扩大、数据业务激增和网络拓扑动态变化导致路由表频繁更新，对基于 DRL 的路由算法设计提出了更加严格的要求[6]。现有 DRL 路由算法通常采用反向传播神经网络进行策略选择。这类算法虽能在一定程度上根据网络状态特征拟合学习路由决策，但其策略通常只能收集邻域卫星与链路信息，缺少全局视角；同时，此类算法难以准确地捕获网络状态在时序上的隐含关联，导致对网络状态的感知不够精准。图神经网络 (Graph neural networks, GNN) 的出现为上述问题提供了潜在解决方案[7]。图神经网络以节点、边和权重的方式储存网络信息。由于大规模 LEO 卫星网络可被建模为包括节点、边与权重的图结构，利用 GNN 能够更加精准地捕获网络的全局状态特征。进而，DRL 可基于 GNN 所感知的网络状态特征进行路由决策，输出最优路径。因此，采用 GNN 与 DRL 相结合的方法来设计大规模 LEO 卫星网络路由算法，具有重要的意义和价值。

本文对消息传递神经网络 (message process neural networks, MPNN) 进行改进，并将其与深度 Q 网络 (deep Q networks, DQN) 结合，提出了一种基于 MPNN 和 DQN 的大规模 LEO 卫星网络路由算法 (MPNN and DQN based large-scale LEO satellite networks routing algorithm, MDSR)。MDSR 旨在提升大规模 LEO 卫星网络路由的传输性能，并能够在不需额外训练成本的条件下，适配不同星座网络构型。本文主要贡献如下：

1) 本文设计了 MDSR 算法的基本框架。MDSR 算法框架包括改进的 MPNN 模块和 DQN 模块，分别对大规模 LEO 卫星网络精确感知和策略选择。同时，设计了 MDSR 智能体所需的状态空间、动作空间和奖励函数。

2) 本文改进了 MPNN 层。更新函数采用长短期记忆 (long short term memory, LSTM) 网络替代了初始 MPNN 模型中的循环神经网络 (recurrent neural networks, RNN) 更新节点隐藏状态，使其能够更好地处理时序拓扑图之间的隐含关联，有效缓解梯度消失与梯度爆炸问题。

3) MDSR 算法在不同大规模 LEO 卫星网络构型中进行测试，仿真结果表明在未经额外训练的情况下，MDSR 算法在多项路由传输性能指标上仍优于对比算法，展现了良好的拓扑泛化能力。

2 相关工作

2.1 早期 LEO 卫星网络路由算法

早期 LEO 卫星网络路由算法设计侧重于规避网络拓扑的动态性，保证算法性能的稳定。常见方法包括虚拟拓扑[8] 和虚拟节点方案[9]。进入 21 世纪，算法设计从拓扑结构转为网络性能指标的优化，包括负载均衡[10-11]、拥塞控制[12] 和服务质量 (quality of service, QoS) 保障[13-15] 等方面。然而，这些算法难以有效解决拓扑结构变化导致的链路不稳定问题和数据业务激增导致的网络拥塞引发传输超时与丢包问题。因此，这些路由算法难以适应如今的大规模 LEO 卫星网络的业务需求。近年来，随着 DRL、软件定义网络[16]、数字孪生[17] 等新兴技术的发展，大规模 LEO 卫星网络领域开始结合这些技术，为解决路由问题提供了新的思路。

2.2 基于 AI 技术的大规模 LEO 卫星网络路由算法

AI 技术目前已经被广泛应用于解决大规模 LEO 卫星网络路由问题。由于卫星网络拓扑具有周期性动态变化的特点，网络状态随时间发生转移。因此，路由决策过程可以被建模为 MDP 过程。此时，DRL 成为处理此类路由问题的一种理想方案。通过对端到端时延、吞吐量、丢包率和其他优化目标设置约束，DRL 不仅可以监控相邻卫星和链路的实时状态，还能动态调整路由策略，以寻求更优解决方案。许多学者已将 DRL 运用在卫星

路由算法的设计中，在增强抗拥塞能力[18]、提供 QoS 保障[19-20] 和设计安全路由协议[21] 等方面取得进展。然而，目前基于 DRL 的路由算法存在感知范围有限的问题，智能体只能够检测邻近卫星和链路状态，无法感知全局网络状态，这极易导致在路由选择的过程中陷入局部最优。同时部分模型与卫星规模高度耦合，可扩展性较弱，在其他星座网络构型下的测试效果往往显著下降。

为应对上述问题，近年来部分学者开始关注 GNN。GNN 通过分析图结构数据来捕捉元素之间的连通关系。其核心机制为节点通过获取相邻节点和边的特征信息，更新自身的隐藏状态，从而学习相关特征之间的关联[22]。同时，经过训练后的 GNN 模型可以迁移到不同的拓扑结构中，具有很强的泛化能力。因此，GNN 在流量预测、结构分类等任务中得到了广泛的运用[23-24]。近年来，GNN 与 DRL 的结合被应用于通信网络领域中的路径规划、流量预测，在光网络和地面通信网络中得到普及[25-26]。GNN 与 DRL 结合，能够更精准地捕捉网络的状态特征变化，实现网络感知和路由决策，提升模型的性能和可扩展性。在大规模 LEO 卫星网络中，GNN 和 DRL 的结合不仅能实现对全球网络状况的精确感知，还能基于所学习到的网络特性制定路由策略[27-29]。该方法在优化卫星网络资源利用率的同时，极大地改善了端到端时延和丢包率等关键路由传输性能指标。此外，它还提高了算法在不同星座构型下的泛化能力[30-33]。

现有研究大多将 GNN 与 DRL 相结合的方法应用于小规模 LEO 卫星网络路由场景中，针对大规模 LEO 卫星网络场景的应用有限。在不同星座网络构型中，路由算法的泛化性能缺乏充分验证。本文聚焦于此，旨在通过在不同星座构型下对路由传输性能指标进行测试，验证所提算法在大规模 LEO 卫星网络中的有效性与泛化能力。

3 本文方法

3.1 大规模 LEO 卫星网络路由模型

3.1.1 大规模 LEO 卫星网络路由场景

本文首先采用虚拟拓扑的方法，将大规模 LEO 卫星网络划分为等间隔的时间片，每个时间片代表一段时间内的网络拓扑图。网络拓扑图集合 G 可以表示为 $G=\{G_0,G_1,\cdots,G_t,\cdots,G_{k-1}\}$，$k$ 为网络拓扑图的总数。其中 $G_t=(N,L)$，N 为卫星集合，L 为星间链路 (inter satellite link, ISL) 集合。N_{sum} 为网络中卫星的总数，L_{sum} 为网络中 ISL 的总数。大规模 LEO 卫星网络路由场景如图 1 所示，每颗卫星最多拥有 4 条 ISL，包括 2 条同轨道 ISL 和 2 条异轨道 ISL。在大型 LEO 卫星网络中，数据在发送卫星中被分割打包，以数据包的形式在 ISL 中传输。数据包经过中继卫星的转发，到达目的卫星。其所经过的卫星集合构成一条传输路径。这就是路由在网络中的传输过程。

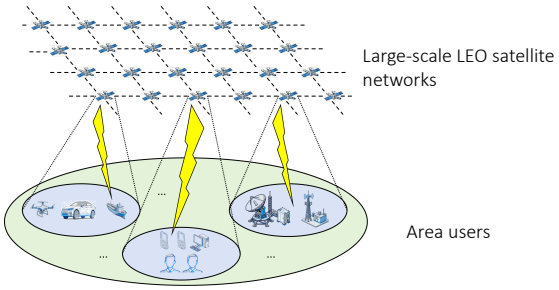


图 1 大规模 LEO 卫星网络路由场景

Fig. 1 Large-scale LEO satellite networks routing scenario

假设网络中存在一个数据包 p，从源卫星 s_0 向目的卫星 s_l 发送。从 s_0 到 s_l 可以构成一组路径集合 path_t。集合中的第 i 条路径 $path_t^i$ 可以表示为

$$path_t^i=\{<s_0,s_1>,\cdots,<s_m,s_n>,\cdots,<s_{l-1},s_l>\}, \quad (1)$$

其中 s_m 和 s_n 均为中继卫星。$<s_m,s_n>$ 代表 s_m 和 s_n 之间的 ISL。

在数据包传输过程中对路由选择影响最大的指标就是端到端时延 d_{ete}^i。d_{ete}^i 为传输时延 d_{trans}^i、传播时延 d_{prop}^i、处理时延 d_{proc}^i 和排队时延 d_{q}^i 之和[34]。

因此，d_{ete}^i 可以表示为

$$d_{\text{ete}}^i=\sum_{i=0}^{l-1}(d_{\text{trans}}^i+d_{\text{prop}}^i+d_{\text{q}}^i+d_{\text{proc}}^i), \quad (2)$$

d_{trans}^i 为卫星发送数据包所需要的时间。由数据包大小 p_{size} 和 ISL 的传输速率 R_{s_m,s_n}^i 决定。因此 d_{trans}^i 可以表示为

$$d_{\text{trans}}^i=\frac{p_{\text{size}}}{R_{s_m,s_n}^i}. \quad (3)$$

根据香农定理，R_{s_m,s_n}^i 可以表示为

$$R_{s_m,s_n}^i=B_{s_m,s_n}^i\log\left(1+\frac{P_{\text{rx}}}{N_0}\right), \quad (4)$$

其中，B_{s_m,s_n}^i 表示为 $<s_m,s_n>$ 的带宽，P_{rx} 为天线接收功率，N_0 为噪声功率。在卫星通信网络中，信号以电磁波的形式传播，仅在真空中经历自由空间损耗。因此，基于自由空间传播损耗模型，P_{rx} 可以表示为

$$P_{\mathrm{rx}} = P_{\mathrm{tx}} \cdot G_{\mathrm{tx}} \cdot G_{\mathrm{rx}} \cdot \left(\frac{\lambda}{4\pi e_{s_m,s_n}}\right)^2, \tag{5}$$

其中：P_{tx}为发射功率，λ为载波波长，e_{s_m,s_n}为$< s_m, s_n >$的长度，G_{tx}为发射天线增益，G_{rx}为接收天线增益。G_{tx}和G_{rx}均与卫星天线的参数有关，分别表示为

$$G_{\mathrm{tx}} = \left(\frac{2\pi D}{\lambda}\right)^2, \tag{6}$$

$$G_{\mathrm{rx}} = \eta\left(\frac{\pi D}{\lambda}\right)^2, \tag{7}$$

其中：D 为卫星天线的直径，η为接收卫星天线效率。

d^i_{prop}为数据包在 ISL 中传播的时间。它只与 ISL 的长度有关。因此d^i_{prop}可以表示为

$$d^i_{\mathrm{prop}} = \frac{e_{s_m,s_n}}{c}, \tag{8}$$

其中：c 为信号在电磁波传输的速度，$c = 3\times10^8$ m/s。

d^i_{proc}为卫星处理数据包的时间，包括对数据打包处理，查询路由表等。在本文中假设d^i_{proc}为固定值。

d^i_{q}指数据包在节点队列中等候的时间。在网络负载较低时，d^i_{q}可以忽略不计，但当网络负载严重，业务请求较多时，节点队列中数据包数量较多，会产生较大的d^i_{q}。p 进入到某一个中继卫星节点的队列中，p 之前有 x 个数据包等待发送，这时d^i_{q}可以表示为

$$d^i_{\mathrm{q}} = \sum_{x=1}^{x} d^i_{\mathrm{trans}s_x}, \tag{9}$$

其中，$d^i_{\mathrm{trans}s_x}$表示排在前面的第x个数据包发送所需要的传输时延。

假设s_0到s_l总共发送了N_p个数据包，且这些数据包均沿路径$path^i_t$传输，t 表示当前网络拓扑图的编号。则$path^i_t$下的总端到端时延d^t_{ete}可以表示为

$$d^t_{\mathrm{ete}} = \sum_{i=1}^{N_p} d^i_{\mathrm{ete}}. \tag{10}$$

路由模型选择路由策略不仅需要考虑端到端时延，需要考虑 ISL 的带宽使用情况。选择剩余带宽较大的 ISL 时，会更好地避免 ISL 中数据包的拥堵，降低网络拥塞发生的概率。假设$< s_m, s_n >$上有φ个数据包在传输，每个数据包的大小相等。此时$< s_m, s_n >$的剩余带宽$B^{r_i}_{s_m,s_n}$可以表示为

$$B^{r_i}_{s_m,s_n} = B^i_{s_m,s_n} - \varphi \cdot p_{\mathrm{size}}. \tag{11}$$

通过对大规模 LEO 卫星网络路由场景的描述，由此将复杂的网络路由问题转化为求解s_0到s_l的最短路径问题。

3.1.2 问题描述与优化目标

假设大规模 LEO 卫星网络运行周期为 T，按照之前的虚拟拓扑方法划分时间片，构造网络拓扑图集合 G。此时大规模 LEO 卫星网络时间划分间隔为δ，其作为网络的路由更新间隔。假设在 t 时刻，卫星的网络拓扑图刚好更新为G_t。此时s_0向s_l沿着路径集合path_t中的某条路径$path^i_t$发送数据包。则在$[t\cdot\delta,(t+1)\cdot\delta]$的时间范围内，数据包沿着路径$path^i_t$传输。因此，本文的优化变量为在当前时刻的网络拓扑图结构G_t中，为每个数据包p选择集合path_t中具体的传输路径$path^i_t$。

本文中，$path^i_t$需要在一定约束条件下，尽可能地满足最小化端到端时延、最大化利用剩余带宽以及最小化丢包率。因此，复杂的网络路由选择问题被转化为求解满足某个效用函数最优解的数学问题。当数据包p沿着路径$path^i_t$传输时，设计效用函数为

$$F(p, path^i_t) = -x\cdot U_\alpha(1-\mu_p) - y\cdot U_\beta\left(d^{\mathrm{ete}}_{p,path^i_t}\right) + z\cdot U_\gamma\left(B^r_{p,path^i_t}\right), \tag{12}$$

其中$U_\alpha(x)$为归一化函数，根据文献 [27] 中对$U_\alpha(x)$的描述，$U_\alpha(x)$表示方法如下：

$$U_\alpha(x) = \frac{x^{1-\alpha}}{1-\alpha}, \tag{13}$$

其中，μ_p代表是否发生丢包，当$\mu_p = 0$时代表数据包传输成功，$\mu_p = 1$则代表数据包发生丢失。α、β、γ、x、y、z分为$[0,1)$范围内的常数。为了统一归一化函数参数避免干扰，α、β、γ一般设置为一样的数值。x、y、z分别代表丢包率、时延和剩余链路带宽的权重，衡量三种约束指标的重要程度。$d^{\mathrm{ete}}_{p,path^i_t}$代表数据包$p$沿着路径$path^i_t$传输时产生的端到端时延。$B^{\mathrm{r}}_{p,path^i_t}$为路径的最大剩余带宽。同时，$d^{\mathrm{ete}}_{p,path^i_t}$需要小于网络所容许的端到端时延阈值$d^{\mathrm{ete}}_{\mathrm{th}}$，$B^{\mathrm{r}}_{p,path^i_t}$需要大于路径链路中设置的最小剩余带宽阈值$B^{\mathrm{r}}_{\mathrm{th}}$。因此，效用函数还需满足以下约束条件：

$$d^{\mathrm{ete}}_{p,path^i_t} \leqslant d^{\mathrm{ete}}_{\mathrm{th}}, \tag{14}$$

$$B^{\mathrm{r}}_{p,path^i_t} \geqslant B^{\mathrm{r}}_{\mathrm{th}}. \tag{15}$$

整个卫星周期内所有数据包的效用函数表示如下：

$$\max \sum_{t=0}^{k-1}\sum_{z=0}^{N_p-1} F(p_z, path^i_t), \tag{16}$$

其中，$path^i_t \in path_t$，为每个数据包p_z在 t 时刻选择的传输路径。通过所有数据包不同传输路径的选择来最大化效用函数。

因此，本文的优化目标为在式 (14)、(15) 的约束条

件下，最大化一个卫星周期 T 内的所有数据包效用函数之和。其对平均端到端时延、剩余带宽和丢包率的综合指标进行联合求解，来实现路由优化。

3.2 MDSR 算法设计

3.2.1 MDSR 算法架构

基于所构建的大规模 LEO 卫星网络路由模型与优化目标，结合 MPNN 与 DQN，本文提出了 MDSR 算法。MDSR 算法框架如图 2 所示。MDSR 框架首先采用虚拟拓扑的方法，将大规模 LEO 卫星星座网络的一个周期，以相等的路由更新间隔进行切片，并将切片的网络拓扑建模为时间序列图模型。MDSR 智能体以图模型作为输入，利用 MPNN 模型学习网络状态特征，随后将这些特征输入 DQN 的全连接层中处理，输出路由决策。通过训练，MDSR 智能体可以不断地与大规模 LEO 卫星网络进行交互，学习到最优的路由策略。在测试阶段，训练好的模型可以输出选择的路由策略，并能够输出路由的传输性能指标，便于算法分析。

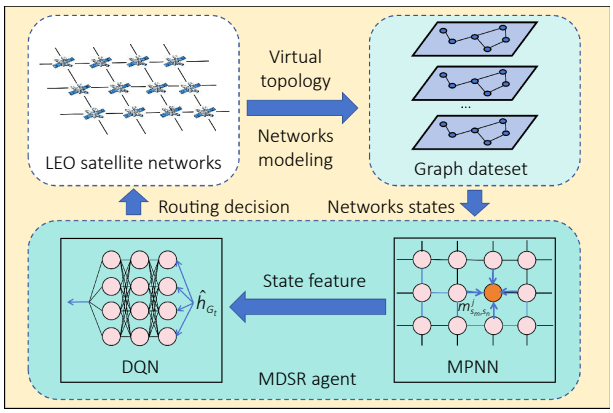


图 2 MDSR 算法框架

Fig. 2 MDSR algorithm framework

3.2.2 状态空间，动作空间和奖励函数设计

状态空间即卫星在时间片中所处环境状态的集合。在大规模 LEO 卫星网络模型中，状态空间包括卫星节点状态和星间链路状态。因此大规模 LEO 卫星网络状态集合 S 可以表示为

$$S=\{s_{G_0},s_{G_1},\cdots,s_{G_t},\cdots s_{G_{k-1}}\}, \tag{17}$$

其中网络拓扑图G_t的状态 s_{G_t}可以表示为 $s_{G_t}=\{\boldsymbol{X}_N^t,\boldsymbol{X}_L^t\}$，其中$\boldsymbol{X}_N^t$为卫星状态特征矩阵，$\boldsymbol{X}_L^t$为星间链路状态特征矩阵。卫星状态特征主要包括卫星的编号、卫星的队列拥塞程度和卫星此时剩余的能量。具体来讲，LEO 卫星 s_m的节点特征状态$\boldsymbol{X}_{s_m}^t$ 可以表示为

$$\boldsymbol{X}_{s_m}^t=[id_{s_m},co_{s_m},en_{s_m}]_{1\times 3}, \tag{18}$$

其中：id_{s_m}为s_m的编号；co_{s_m}为s_m的队列拥塞程度，在大规模 LEO 卫星网络中每个卫星内部存储一个数据包队列，负责存储、发送和接收数据包；en_{s_m}为s_m的剩余能量。因此，网络中所有卫星节点的状态矩阵集合可以表示为

$$\boldsymbol{X}_N^t=\{\boldsymbol{X}_{s_0}^t,\boldsymbol{X}_{s_1}^t,\cdots,\boldsymbol{X}_{s_m}^t,\cdots,\boldsymbol{X}_{s_{N_{\text{sum}}-1}}^t\}, \tag{19}$$

星间链路状态特征主要包括星间链路的长度，星间链路的带宽和星间链路的传输速率。具体来讲，假设卫星s_m和s_n之间存在一条 ISL$<s_m,s_n>$，为了方便表示，将其编号为l_i，则$<s_m,s_n>$的星间链路状态矩阵$\boldsymbol{X}_{l_i}^t$可以表示为

$$\boldsymbol{X}_{l_i}^t=[B_{s_m,s_n},e_{s_m,s_n},R_{s_m,s_n}]_{1\times 3}, \tag{20}$$

其中：B_{s_m,s_n}代表$<s_m,s_n>$的带宽；e_{s_m,s_n}代表$<s_m,s_n>$的长度，R_{s_m,s_n}代表$<s_m,s_n>$的传输速率。此时网络中所有 ISL 的星间链路特征集合可以表示为

$$\boldsymbol{X}_L^t=\{\boldsymbol{X}_{l_0}^t,\boldsymbol{X}_{l_1}^t,\cdots,\boldsymbol{X}_{l_i}^t,\cdots,\boldsymbol{X}_{l_{L\text{sum}-1}}^t\}. \tag{21}$$

动作空间为一个周期 T 发送卫星s_0到目的卫星s_l发送数据包时所选择的路径集合。在本文中，动作空间路径的初始选择以传播时延作为约束，采用 K 最短路径算法 (K-shortest path, KSP) 计算s_0到s_l之间的 K 条路径。假设当前时刻网络拓扑图为G_t时，采用 KSP 算法计算出的 K 条路径集合可以表示为

$$path_t=\{path_t^0,path_t^1,\cdots,path_t^{K-1}\}. \tag{22}$$

整个周期 T 的s_0到s_l的全部路径选择集合$Path$可以表示为

$$Path=\{path_{G_0},path_{G_1},\cdots,path_{G_t},\cdots,path_{G_{k-1}}\}, \tag{23}$$

$Path$即为动作选择集合。

奖励指环境根据 MDSR 智能体在当前状态下所选动作而给予的反馈[35]。本文的奖励函数的设计综合考虑数据包在路径传输时的状态，包括数据包到达目的卫星，数据包到达中继卫星和数据包丢弃这三种状态。当卫星节点队列的数据包达到队列的存储上限或星间链路的剩余带宽小于最低数据包传输带宽时，数据包则发生丢弃。假设 当前时刻为 t 且在拓扑图G_t中，数据包p沿着路径$path_t^i$传输，则下一时刻的奖励值可以表示为

$$r_{t+1}=\begin{cases}100+F(p,path_t^i)\ \text{arrived}\ s_l\\ 10+F(p,path_t^i)\ \text{arrived}\ s_m,s_n\\ -10+F(p,path_t^i)\ \text{dropped}\end{cases}. \tag{24}$$

在 DRL 中奖励主要统计累计奖励，因此累计奖励的计算根据文献 [36] 可以表示为

$$R=r_t+\xi r_{t+1}+\xi^2 r_{t+2}+\cdots+\sum_{k=0}^{\infty}\xi^k r_{t+1+k}, \tag{25}$$

其中：ξ为折扣因子，且$\xi\in[0,1]$。

3.2.3 改进 MPNN 层设计

本文采用 MPNN 模型作为 GNN 模型来获取大规模 LEO 卫星网络状态。为了使 MPNN 更加适应大规模 LEO 卫星网络拓扑图结构输入，本文对 MPNN 模型进行了改进，改进后的 MPNN 模块架构如图 3 所示。

改进的 MPNN 层包含消息传递阶段和读出阶段[37]。其中消息传递阶段由消息传递函数 (message process function, MPF)、信息聚合函数 (information aggregation function, IAF) 和隐藏状态更新函数 (hidden state update function, HSUF) 三部分构成。经过消息传递阶段的网络阶段状态，由读出阶段的读出函数 (readout function, RF) 读出网络的状态特征。

MPF 用于生成邻居卫星节点和 ISL 的消息。MPF 以当前卫星节点s_m，相邻卫星节点s_n和 ISL$<s_m,s_n>$的隐藏状态作为输入，利用ReLU函数作为激活函数提供非线性变化，加快模型收敛。在本文中，MPF 采用全连接层进行拟合，因此 MPF 可以表示为

$$m_{s_m,s_n}^j=ReLU(w_r^j(h_{s_m}^{j-1}\|h_{s_n}^{j-1}\|h_{<s_m,s_n>}^{j-1})), \tag{26}$$

其中，$h_{s_m}^{j-1}$、$h_{s_n}^{j-1}$和$h_{<s_m,s_n>}^{j-1}$分别代表输入到第 j 层 MPNN 网络上的s_m、s_n和$<s_m,s_n>$的隐藏状态，$\|$ 代表拼接操作，将节点隐藏状态与 ISL 隐藏状态进行拼接处理；w_r^j为第 j 层 MPNN 网络中大的全连接层参数权重，且$w_r^j\in\mathbb{R}^{C_r^j\times(2C_s^{j-1}+C_e^{j-1})}$，其中$C_r^j$为生成消息$m_{s_m,s_n}^j$的维度，根据状态空间的设计，可以得到：

$$h_{s_0}^0=\boldsymbol{X}_{s_0},h_{e_0}^0=\boldsymbol{X}_{e_0},C_s^0=3,C_e^0=3. \tag{27}$$

IAF 表征对所有邻居卫星节点的消息进行聚合，其输入是 MPF 的输出。IAF 不关注邻居卫星节点的顺序，即使获取邻居卫星节点消息的顺序不同，也不影响 IAF 输出的结果。本文中 IAF 采用均值聚合函数进行拟合，可以表示为

$$m_{s_m}^j=E_{s_n\in N(s_m)}(m_{s_m,s_n}^j). \tag{28}$$

HSUF 用于更新卫星节点的隐藏状态，其输入为 IAF 聚合后的隐藏状态和当前节点的隐藏状态。由于二者相互关联，因此可以将更新过程视为时间序列输入问题。原版 MPNN 使用了 RNN 来拟合更新函数，处理时间序列问题，但因 RNN 无法应对长序列依赖，且在训练时会发生梯度消失和梯度爆炸。为了克服这些限制，本文的 HSUF 选用 LSTM 进行拟合。LSTM 通过加入输入门、遗忘门和输出门等门控单位，更好地捕捉序列中的依赖关系，减缓训练时梯度累积造成的梯度消失和爆炸[36]。因此 HSUF 可以表示为

$$h_{s_m}^j=LSTM(h_{s_m}^{j-1},m_{s_m}^j). \tag{29}$$

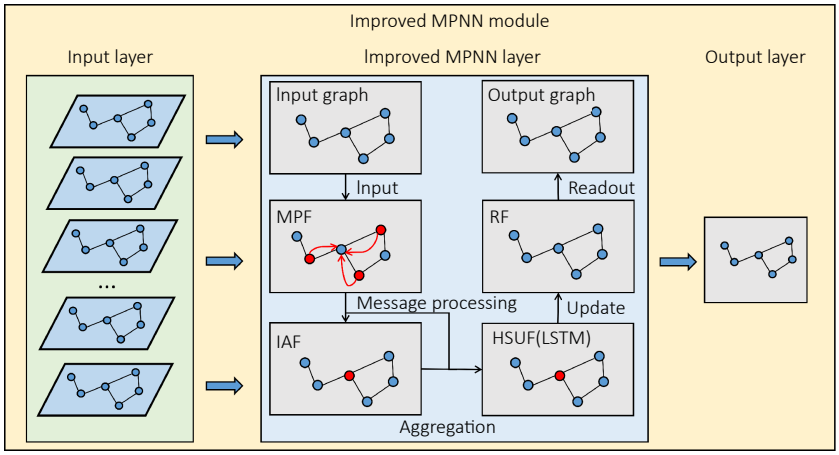


图 3 改进的 MPNN 模块

Fig. 3 Improved MPNN module

在一层 MPNN 网络中，所有的卫星节点均经历消息传递过程，因此经过 J 层 MPNN 网络后，可以得到 G_t 中的全部卫星节点的隐藏状态矩阵 $\boldsymbol{H}_s^J$ 和全部 ISL 的隐藏状态矩阵 $\boldsymbol{H}_e^J$。

RF 作用于读出阶段，获取 G_t 的大规模 LEO 卫星网络的特征。RF 函数可以分为读出卫星节点特征 $\hat{h}_s$ 和 ISL 特征 $\hat{h}_e$，$\hat{h}_s$ 和 $\hat{h}_e$ 分别表示如下：

$$\hat{h}_s = RF(\boldsymbol{H}_s^J), \tag{30}$$

$$\hat{h}_e = RF(\boldsymbol{H}_e^J). \tag{31}$$

将 $\hat{h}_s$ 和 $\hat{h}_e$ 进行拼接，即可得到拓扑图 G_t 的全部特征 $\hat{h}_{G_t}$：

$$\hat{h}_{G_t} = (\hat{h}_s \| \hat{h}_e). \tag{32}$$

3.2.4 DQN 模块

本文采用 DQN 方法对所获得的网络状态特征进行路由策略学习。如图 4 所示，DQN 模块采用神经网络来拟合 Q 值，通过获取的全局卫星网络状态特征 $\hat{h}_{G_t}$ 来选择动作并不断学习来更新网络参数、学习最优的路由选择决策。在本文中，Q 值根据贝尔曼方程计算[38]，同时 MPNN 读出的特征 $\hat{h}_{G_t}$ 作为输入 DQN 模块的状态 ψ_t，Q 值的更新可以表示为

$$Q(\psi_t, a_t; \theta_0, \omega_0) = r_{t+1} + \xi \max_{a_{t+1} \in A} Q(\psi_{t+1}, a_{t+1}; \theta_1, \omega_1), \tag{33}$$

其中：θ_0 和 ω_0 为主 Q 网络中 MPNN 层和全连接层的网络参数，θ_1 和 ω_1 为目标 Q 网络中 MPNN 层和全连接层的网络参数。

DQN 通过损失函数来计算主 Q 网络和目标 Q 网络之间的误差，主要采用均方误差函数 (mean square error, MSE) 来计算。在本文中，损失函数 $L(\theta_0, \omega_0; \theta_1, \omega_1)$ 的计算方式表示为

$$L(\theta_0, \omega_0; \theta_1, \omega_1) = E\left[\left(r_{t+1} + \xi \max_{a_{t+1}} Q(\psi_{t+1}, a_{t+1}; \theta_1, \omega_1) - Q(\psi_t, a_t; \theta_0, \omega_0)\right)^2\right]. \tag{34}$$

主 Q 网络参数采用梯度下降法更新，目标 Q 网络在经过一段时间后将主 Q 网络的网络参数复制到目标 Q 网络中。主 Q 网络参数更新和目标 Q 网络参数更新过程分别如下：

$$\nabla_{\theta_0, \omega_0} L(\theta_0, \omega_0) = E[(r_{t+1} + \xi \max_{a_{t+1}} Q(\psi_{t+1}, a_{t+1}; \theta_1, \omega_1) - Q(\psi_t, a_t; \theta_0, \omega_0)) \nabla_{\theta_0, \omega_0} Q(\psi, a; \theta_0, \omega_0)], \tag{35}$$

$$(\theta_1, \omega_1) = (\theta_0, \omega_0). \tag{36}$$

经过一段时间后，利用经验回放池的经验回放机制，将小批次样本送入网络中训练，来加快模型的收敛。

3.3 MDSR 算法训练流程

根据 MDSR 算法状态空间、动作空间、奖励函数和改进 MPNN 模块和 DQN 模块，对 MDSR 算法框架进行了设计。MDSR 算法包括训练阶段和测试阶段。训练阶段，算法模型根据大规模 LEO 卫星网络的动态变化的网络拓扑学习卫星网络特征，在 MDSR 智能体与大规模 LEO 卫星网络环境不断交互过程中学习最优的路由策略，其中，MDSR 算法训练流程如表 1 所示。在测试阶段，训练好的模型能够根据实时的大规模 LEO 卫星网络状态信息而输出最佳的路由策略。

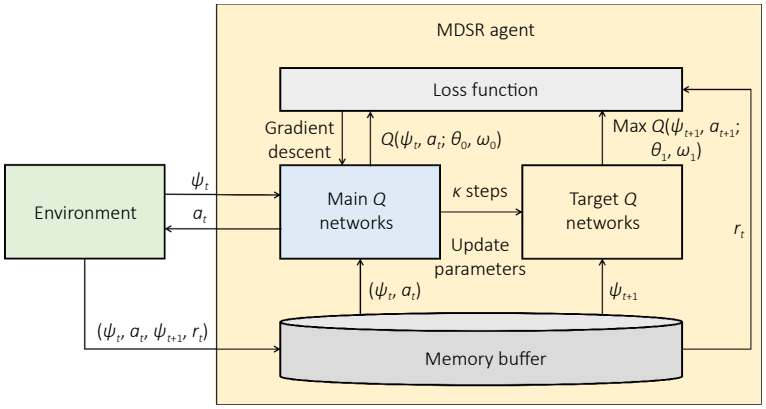


图 4 DQN 模块

Fig. 4 DQN module

表 1 MDSR 算法训练流程

Table 1 MDSR algorithm training process

算法：MDSR算法训练流程

初始化大规模LEO卫星网络路由场景，得到图结构数据集G；
初始化主Q网络参数θ_0,ω_0和目标Q网络参数θ_1,ω_1；
初始化经验回放池M, 训练轮次$epochs$, 训练步数$steps$, 经验回放采样步数τ,主Q网络更新步数κ；
for $epoch$ in range $(1,epochs)$：
  for $step$ in range $(1,steps)$：
    for G_t in G：
      for j in range $(1,J)$：
        for s_m in N：
          利用公式 (26)得到所有相邻卫星对s_m生成的消息；
          利用公式 (28)得到聚合后s_m的消息；
          利用公式 (29)更新s_m的隐藏状态；
        end for
      end for
      利用公式 (32)读出大规模LEO卫星网络状态特征$\hat{h}_{G_t}$；
    end for
    从所有$\hat{h}_{G_t}$中抽取状态ψ_t，输入到全连接神经网络中计算各个动作的Q值；
    根据$\varepsilon-$greedy策略选择动作a_t；
    将动作a_t输入到环境中执行；
    环境返回即时奖励r_t和下一时刻状态ψ_{t+1}；
    将样本$(\psi_t,a_t,\psi_{t+1},r_t)$储存入经验回放池$M$中;
    if $step>\tau$:
    从经验回放池M中随机取出批次样本进行训练;
    利用公式(34)计算损失函数, 利用公式(35)更新主Q网络参数θ_0,ω_0;
    if $step\%\kappa==0$:
    采用公式 (36)将主Q网络参数θ_0,ω_0复制给目标Q网络参数θ_1,ω_1;
  end for
end for

4 仿真结果与分析

4.1 仿真参数设置

本文采用 Python3.9 对大规模 LEO 卫星网络路由场景进行仿真，采用 PyTorch1.13.1 对神经网络模型进行构建，并调用 RTX 4050 加速模型训练。为了评估 MDSR 算法在不同大规模 LEO 卫星网络构型中的路由传输性能，本文设置了两种卫星星座：极轨道大规模 LEO 卫星星座和倾斜轨道大规模 LEO 卫星星座。极轨道大规模 LEO 卫星星座由 720 颗卫星组成，共有 18 个轨道，每个轨道上有 40 颗卫星。轨道高度为 720 km，轨道倾角为 86.4°。纬度大于 70°的地区设置为极区。大多数卫星配备 4 条 ISL，分别为 2 条轨道内 ISL 和 2 条轨道间 ISL；在极区内的卫星只有 2 条轨道内 ISL。极轨道星座存在反向缝，反向缝两侧的卫星有 2 条轨道内 ISL 和 1 条轨道间 ISL。倾斜轨道大规模 LEO 卫星星座由 1584 颗卫星组成，共有 72 个轨道，每个轨道有 22 颗卫星。轨道高度为 550 km，轨道倾角为 53°，所有卫星均配备 4 条 ISL。此外，为了判断不同路由更新间隔是否对 MDSR 算法性能产生影响，本文设置了 10 s、30 s、60 s 和 300 s 的不同路由更新间隔，来测试 MDSR 算法在不同路由更新间隔下是否稳定。

大规模 LEO 卫星网络的仿真参数如表 2 所示。MDSR 算法模型选择极轨道大规模 LEO 卫星网络路由场景进行训练，路由更新间隔为 30 s。每个卫星节点的队列的最大存储容量为 1000 个数据包。每条 ISL 的最大带宽为 100 MB，为了模拟卫星网络的链路带宽的拥塞情况，在初始设置的时候每条 ISL 的带宽随机设置为 $[10,100]$ MB。同时，卫星的最大能量为 100 J，在初始设置的时候也随机对每颗卫星的能量随机设置为 $[40,100]$ J。

表 2 大规模 LEO 卫星网络仿真参数

Table 2 Large-scale LEO satellite network simulation parameters

仿真参数	参数解释	参数设置
P_{rx}	卫星天线发射功率	10 W
D	天线直径	0.5 m
η	天线效率	0.65
λ	载波波长	0.013 m
d^{i}_{proc}	处理时延	1 ms
B_{s_m,s_n}	链路带宽范围	10~100 MB
N_p	发送数据包数目	1000
p_{size}	数据包大小	0.5 MB
co_{s_m}	卫星队列长度范围	0~1000
en_{s_m}	卫星能量范围	40~100 J
$d^{\mathrm{ete}}_{\mathrm{th}}$	端到端时延阈值	150 ms
$B^{\mathrm{r}}_{\mathrm{th}}$	剩余带宽阈值	10 MB
K	Ksp计算候选路由路径数	3

MDSR 智能体神经网络相关参数如表 3 所示。其中设置丢包率权重、端到端时延权重、剩余带宽权重分别为 0.1、0.6 和 0.3；令效用函数中的归一化参数相等，设置为 0.5。改进的 MPNN 模块采用两层消息传递网络进行拟合；全连接层采用 4 层全连接网络对 Q 值进行拟

合，利用 ReLU 激活函数提供非线性计算；模型优化器选择 Adam 优化器，同时损失函数采用了 MSE 函数；设置经验回放池大小为 10000，每一次进行计算的批次大小为 4，并随着训练次数的增加减少探索和权衡的比值。模型总共进行 300 轮训练，通过学习卫星网络场景的状态特征来进行路由决策。

表 3 MDSR 智能体仿真参数
Table 3 MDSR agent simulation parameters

仿真参数	参数设置
x, y, z	0.1, 0.3, 0.6
α, β, γ	0.5
MPNN层数	2
全连接层数	4
隐藏层维度	32
激活函数	ReLU
优化器	Adam
损失函数	MSE
批次大小	4
仿真轮数	300
ξ	0.99
$\varepsilon-\max$	1
$\varepsilon-\min$	0.05
学习率	0.001
经验回放池容量	10000

4.2 仿真对比算法设置

为了更好地评估 MDSR 算法的路由传输性能，本文选取以下几种算法进行对比。

1) Dijkstra 最短路径 (Dijkstra shortest path, DSP) 算法[39]：作为路由问题中最常用的路由算法，DSP 已被证明在大部分网络中能达到普遍最优解[40]。本章中，DSP 以端到端时延为约束，计算发射卫星与目的卫星的最短路径。

2) 基于 Q-learning 的大规模 LEO 卫星网络路由算法[21]：该算法采用 Q 表来代替传统路由表更新，将源-目的卫星节点对作为状态来降低 Q 表开销，并利用 Q 学习进行路由决策。

3) 基于 DQN 的大规模 LEO 卫星网络路由算法[41]：该算法未采用图神经网络进行泛化能力的表征，直接根据卫星网络状态输出路由决策，以该算法为对照说明加入图神经网络后的作用。

4) 基于确定性深度策略梯度 (deep deterministic policy gradient, DDPG) 网络的大规模 LEO 卫星网络路由算法[42]：该算法是以基于策略的 DRL 算法，其以演说家-批评家框架为学习核心，能够拟合连续动作空间。该算法训练模型的输入维度与卫星数量规模高度相关，因此其不能够适配其他构型的星座网络。

5) 采用 RNN 拟合 HSUF 函数的 MDSR 算法：该算法为 MPNN 模块中 HSUF 函数采用 RNN 拟合的 MPNN 模型，为了方便表示，将其命名为 HSUF-RNN，用于说明采用 LSTM 代替 RNN 拟合 HSUF 在算法训练和测试时所带来的影响。

4.3 仿真结果分析

4.3.1 MDSR 模型训练结果

根据 MDSR 算法仿真设置的大规模 LEO 卫星网络仿真参数和 MDSR 智能体仿真参数，对 MDSR 算法进行训练。训练轮次设置为 300 轮，每隔 5 轮记录一次累计奖励和模型的 MSE 损失。同时，为了验证改进的 MPNN 模块中用 LSTM 代替 RNN 拟合 HSUF 时对算法训练的影响，训练阶段采用了 HSUF-RNN 模型来进行对比。MDSR 模型和 HSUF-RNN 模型的奖励和损失随训练轮次变化曲线如图 5 和图 6 所示。

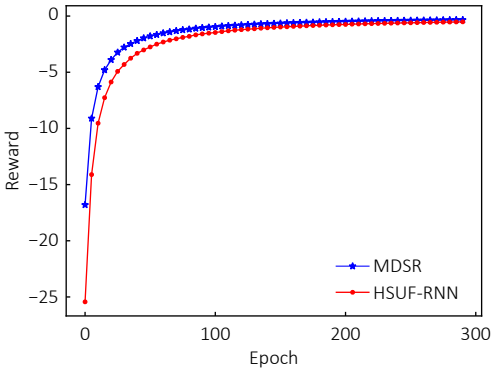


图 5 MDSR 和 HSUF-RNN 模型训练奖励曲线
Fig. 5 MDSR and HSUF-RNN model training rewards curve

仿真结果表明，MDSR 模型和 HSUF-RNN 模型训练过程中的奖励值均快速增加，损失值均快速下降，在 50 轮左右达到逐渐平和，在 150 轮左右基本收敛。其中，MDSR 模型所获得的初始奖励要高于 HSUF-RNN 模型，且初始损失也远小于 HSUF-RNN 模型。可以看出，在收敛速度方面，MDSR 模型要快于 HSUF-RNN 模型，

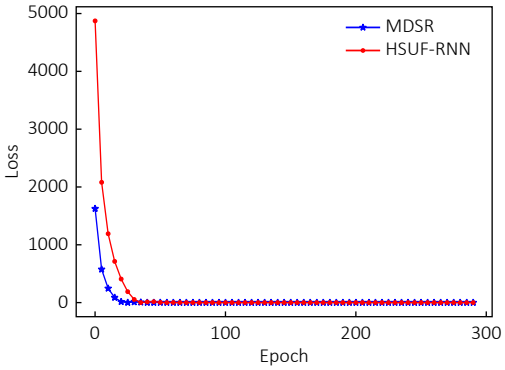


图 6 MDSR 和 HSUF-RNN 模型训练损失曲线

Fig. 6 MDSR and HSUF-RNN model training loss curve

这是因为二者的结构不同，MDSR 模型中通过 LSTM 来拟合 HSUF，通过多种门控单元使模型记住路由历史上下文信息，选择性遗忘噪声信息，长期记忆提升了路由选择策略质量，增加了模型获得的初始奖励，降低了初始损失。因此，通过对两者模型训练证明了采用 LSTM 拟合 HSUF 的 MDSR 模型在收敛速度和初始奖励方面要优于 HSUF-RNN 模型。

4.3.2 MDSR 算法在极轨道大规模 LEO 卫星网络中测试结果及分析

为了验证 MDSR 算法在极轨道大规模 LEO 卫星网络下的路由传输性能，本文在 10 s、30 s、60 s 和 300 s 路由更新间隔下进行了测试。测试时选择的数据发送卫星节点为 S_09_02，目的卫星节点 S_11_32，即第 9 个轨道 2 颗卫星到第 11 个轨道第 32 颗卫星。仿真时间设置为 6600 s，为极轨道大规模 LEO 卫星网络的一个运行周期。为排除其他干扰影响，对比算法中的 HSUF-RNN、DQN、Q-learning 和 DDPG 等强化学习模型训练场景与 MDSR 算法模型的训练环境相同。

图 7 展示了在极轨道大规模 LEO 卫星网络下六种算法的端到端时延随仿真时间变化曲线。图 7(a-d) 分别展示了在 10 s、30 s、60 s 和 300 s 的路由更新间隔下端到端时延的变化。结果表明，MDSR 算法和对比算法的端到端时延均随仿真时间的运行发生周期性变化。其中，DSP 算法的变化范围为 40~50 ms，而 MDSR、HSUF-RNN、Q-learning、DQN、DDPG 算法的波动范围在 30~40 ms。在 10 s、30 s、60 s 和 300 s 的路由更新间隔

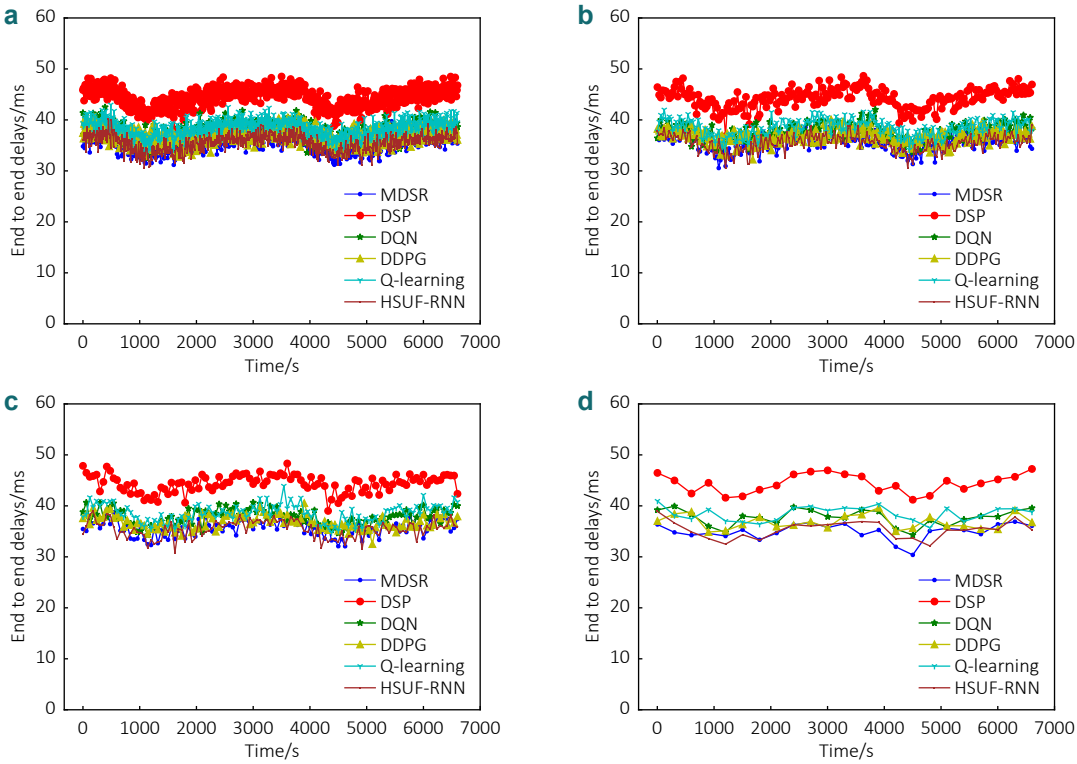


图 7 极轨道大规模 LEO 卫星网络下不同算法的端到端时延随仿真时间变化曲线。路由更新间隔分别为 (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

Fig. 7 End-to-end delay curves of different algorithms in polar orbit large-scale LEO satellite networks with simulation time. Routing update interval (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

下，不同算法的变化趋势不大，这也证明了算法在该仿真场景下的稳定性。对比来看，MDSR 算法的端到端时延曲线变化趋势与 HSUF-RNN、Q-learning、DQN 和 DDPG 相差不大，且均优于 DSP算法。

图 8 展示了在极轨道大规模 LEO 卫星网络下六种算法吞吐量随仿真时间变化曲线。图 8(a-d) 分别展示了在 10 s、30 s、60 s 和 300 s 路由更新间隔下的吞吐量变化。仿真结果表明，MDSR 算法和对比算法的吞吐量均随仿真时间变化而发生规律周期性变化。MDSR 算法与 HSUF-RNN 的性能表现几乎一致，优于其他对比算法，吞吐量保持在 0.4 Mbit/s 以上。相对而言，强化学习策略的算法的吞吐量均在 0.4 Mbit/s 附近波动，低于 MDSR。

DSP 算法的吞吐量保持在约 0.3 Mbit/s，低于其他算法。所有算法在整个仿真过程中都具备稳定性。

图 9 展示了在极轨道大规模 LEO 卫星网络下六种算法的丢包率随仿真时间变化曲线。图 9(a-d) 分别展示了在 10 s、30 s、60 s 和 300 s 路由更新间隔下的丢包率变化。仿真结果表明，MDSR 算法和对比算法的丢包率曲线趋势变化稳定，不受路由更新间隔的影响。MDSR 算法和 HSUF-RNN 的丢包率最低，约为 12 %，而 DSP 的丢包率约为 20 %。DQN、Q-learning和 DDPG 算法丢包率性能表现相似，均在 25 % 左右。因此在极轨道大规模 LEO 卫星网络路由仿真场景中，MDSR 明显在丢包率指标上优于 DQN 等未采用图神经网络方式的对比算法。

为了更直观地描述 MDSR 与几种对比算法的路由传输性能指标的影响，本文对不同路由更新间隔下不同算法的端到端时延、吞吐量和丢包率进行平均计算，做成柱状图进行对比，以便更精确直观地判断几种算法路由传输性能指标的差异。

图 10 展示了极轨道大规模 LEO 卫星网络下几种算法在不同路由更新间隔下的路由传输性能指标对比。图 10(a-c) 分别展示了平均端到端时延、平均吞吐量和平均丢包率三种路由传输性能指标对比。仿真结果表明，MDSR 算法在平均丢包率指标下优势明显，比 DSP 丢

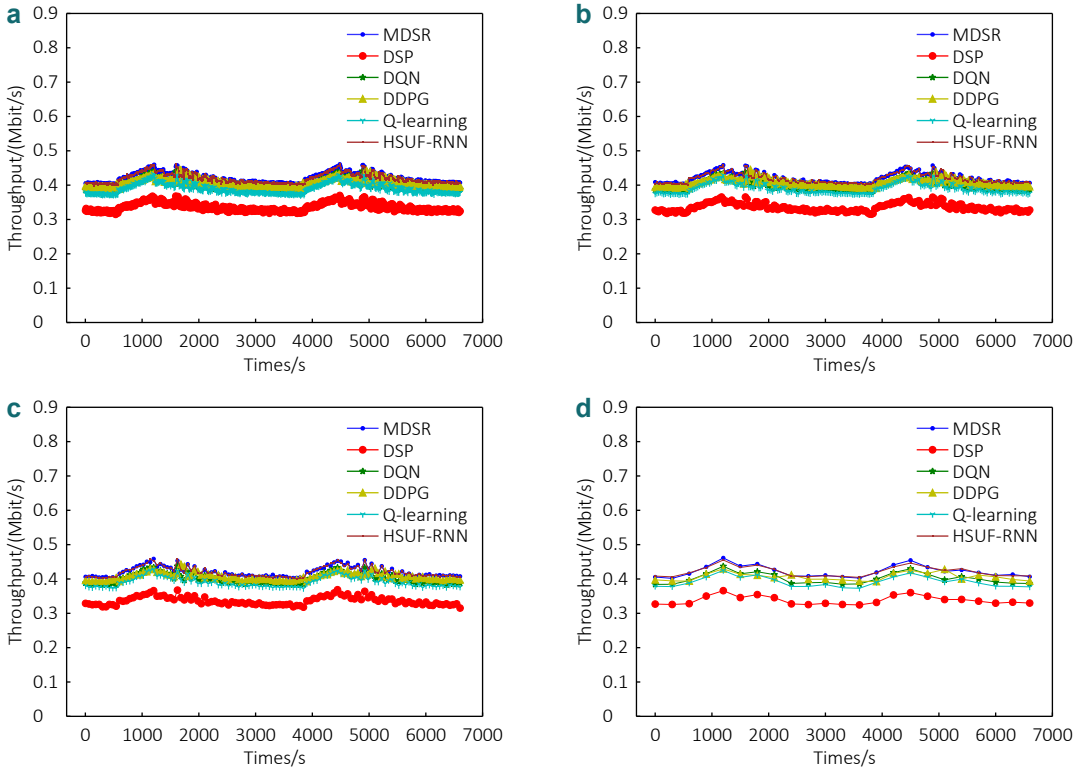


图 8 极轨道大规模 LEO 卫星网络下不同算法的吞吐量随仿真时间变化曲线。路由更新间隔为 (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

Fig. 8 Throughput curves of different algorithms in polar orbit large-scale LEO satellite networks with simulation time. Routing update interval (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

包率降低了约 35 %，比 DQN、Q-learning 和 DDPG 算法的丢包率降低了约 48 %。在平均端到端时延和平均吞吐量上，MDSR 算法要明显优于 DSP，虽然与 DQN、Q-learning 和 DDPG 等 RL 算法较为相近，但也占据一定的优势。其中，在平均端到端时延上，MDSR 相比于 DSP、Q-learning、DQN 和 DDPG 分别降低了约 20.5 %、7.9 %、2.7 % 和 5.4 %；在平均吞吐量上，MDSR 相比于 DSP、Q-learning、DQN 和 DDPG 分别提升了约 21.7 %、7.6 %、5 % 和 1.2 %。这是由于 MDSR 算法选择了图神经网络来表征大规模 LEO 卫星网络的网络状态，同时改进的 LSTM 能够使 MPNN 更好处理时序长序列依赖，精准适配了大规模 LEO 卫星网络的拓扑动态周期变化特性，能够更精准地令强化学习神经网络选择路径决策基于全图网络状态，降低了误选质量低链路的概率。因

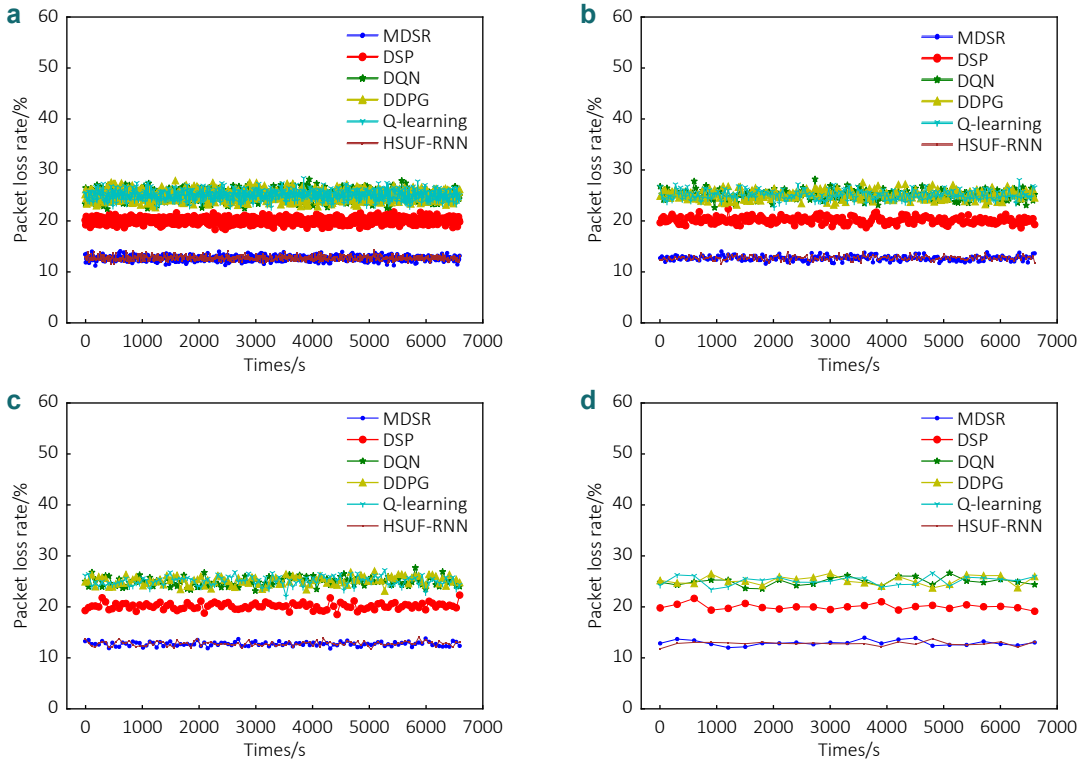


图 9 极轨道大规模 LEO 卫星网络下不同算法的丢包率随仿真时间变化曲线。路由更新间隔为 (a) 10 s；(b) 30 s；(c) 60 s；(d) 300 s

Fig. 9 Packet loss rate curves of different algorithms in polar orbit large-scale LEO satellite networks with simulation time. Routing update interval (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

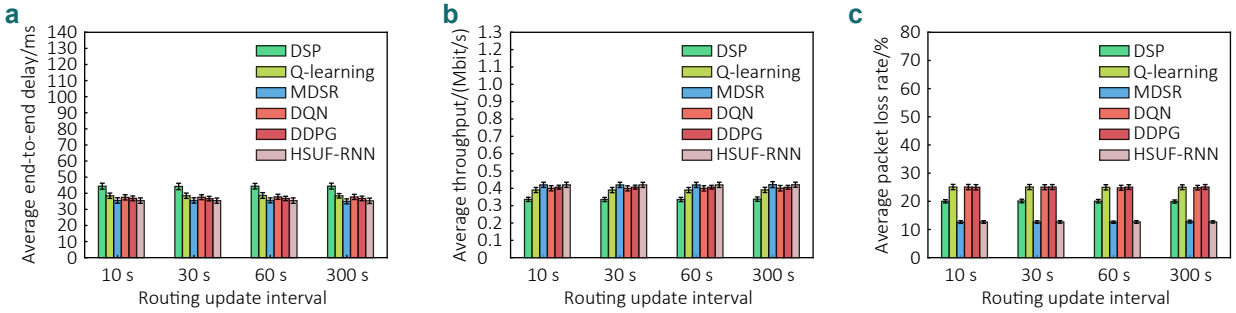


图 10 极轨道大规模 LEO 卫星网络下不同算法的路由传输性能指标在不同路由更新间隔下的对比。(a) 平均端到端时延；(b) 平均吞吐量；(c) 平均丢包率

Fig. 10 Comparison of the routing performance indicators of different algorithms in the polar orbit large-scale LEO satellite networks at different route update intervals. (a) Average end-to-end delay; (b) Average throughput; (c) Average packet loss rate

此能够优先选择质量稳定、带宽充足的链路，同时弥补 DSP 仅依靠传播时延进行决策和基于 DRL 算法感知精度不足的问题。在路由传输性能指标，尤其是丢包率上 MDSR 具有较大优势。与 HSUF-RNN 相比，MDSR 在路由传输性能上基本相似，这是由于两者的模型均为 GNN+DRL 的框架范式，训练后学习到了相似的路由选择策略，都可以利用 MPNN 进行网络状态特征决策，所以在测试结果上，路由传输性能表现相似。综上所述，综合对比下 MDSR 算法在路由传输性能指标上要优于其他 RL 对比算法和 DSP 算法。

4.3.3 MDSR 算法在倾斜轨道大规模 LEO 卫星网络中测试结果及分析

为了测试 MDSR 算法是否在不同的星座构型上仍然具备优势，以及在不需要额外的训练情况下的路由传输性能，验证其可扩展性，本文采用了倾斜轨道大规模 LEO 卫星网络，在路由更新间隔为 10 s，30 s，60 s 和 300 s 条件下进行测试。测试时选择的数据发送卫星为 S_09_12，目的卫星为 S_11_21，即数据发送卫星为第 9 轨道第 12 颗卫星，目的卫星为第 11 轨道第 21 颗卫星。在倾斜轨道大规模 LEO 卫星网络中，对比算法选择分别为 HSUF-RNN、基于 DSP、Q-learning 和 DQN 的大规模 LEO 卫星网络路由算法进行对比。由于基于 DDPG 的大规模 LEO 卫星网络路由算法模型的输入纬度与卫星网络规模高度耦合，其可扩展性有限，因此不将 DDPG 算法在倾斜轨道网络中对比。仿真时间设置为 6000 s，为倾斜轨道大规模 LEO 卫星网络的运行周期。

图 11 展示了在倾斜轨道大规模 LEO 卫星网络下五种算法端到端时延随仿真时间变化曲线。图 11(a-d) 分别展示了在 10 s、30 s、60 s 和 300 s 路由更新间隔下的端到端时延变化。仿真结果表明，在倾斜轨道大规模 LEO 卫星网络下，各种算法的端到端时延均随着仿真时间而发生周期性变化。同时在不同路由间更新间隔下，所有算法的端到端时延均没有明显的变化，这说明不同时间更新间隔对于算法的稳定性的影响不显著。其中，DSP 算法的端到端时延在 30~40 ms 内波动。MDSR、

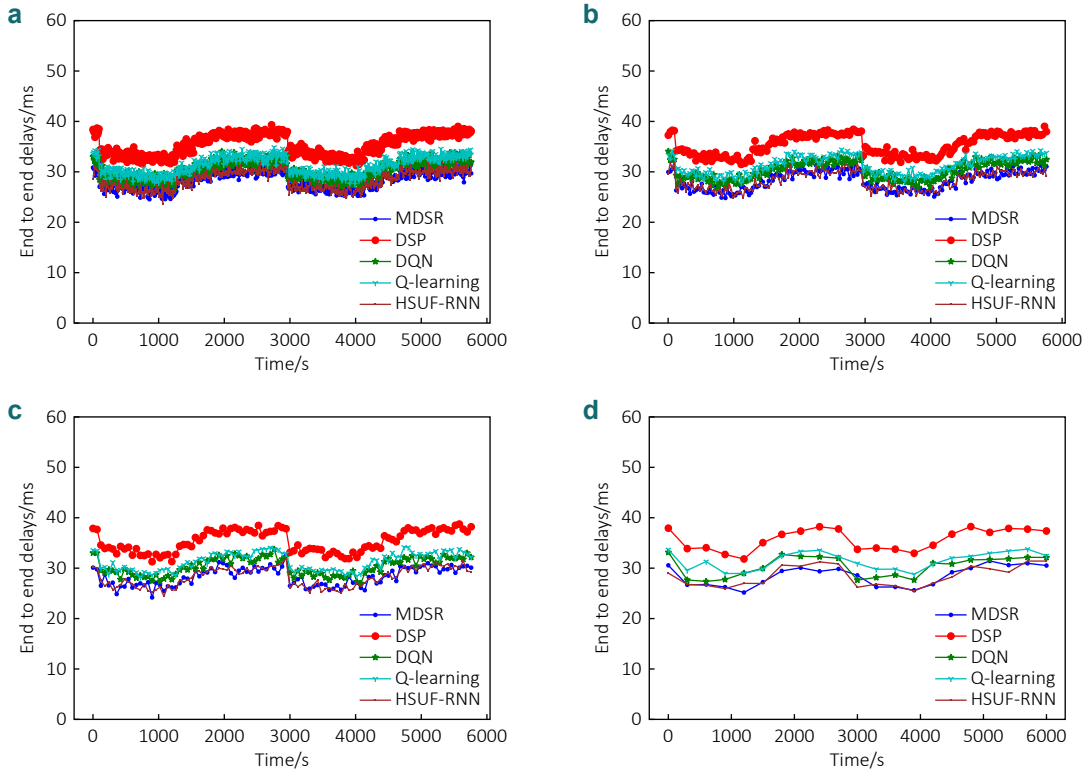


图 11 倾斜轨道大规模 LEO 卫星网络下不同算法的端到端时延随仿真时间变化曲线。路由更新间隔为 (a) 10 s； (b) 30 s； (c) 60 s； (d) 300 s

Fig. 11 End-to-end delay curves of different algorithms in inclined orbit large-scale LEO satellite networks with simulation time. Routing update interval (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

HSUF-RNN、DQN、Q-learning算法的平均端到端时延在 25~35 ms 内波动。可以看出，MDSR 算法在端到端时延指标上要明显优于 DSP 算法，与 HSUF-RNN、DQN、Q-learning 这些基于 RL 范式的路由算法性能差距不大。但 MDSR 和 HSUF-RNN 相比于基于 RL 的路由算法仍然具有一定的优势。

图 12 展示了在倾斜轨道大规模 LEO 卫星网络下五种算法吞吐量随仿真时间变化曲线。图 12(a-d) 分别展示了在 10 s、30 s、60 s 和 300 s 路由更新间隔下的吞吐量变化。仿真结果表明，各算法的吞吐量均随着仿真时间而发生周期性变化。同时在不同路由更新间隔下，所有算法的吞吐量均没有明显的变化，这也说明不同时间更新间隔对于算法的稳定性并无太多影响。其中，DSP 算法的吞吐量最低，在 0.3 Mbit/s 上下浮动，MDSR、HSUF-RNN、DQN 和 Q-learning 算法均在 0.3~0.4 Mbit/s 之间浮动，可以看出 MDSR 算法在吞吐量指标上要明显优于 DSP 算法，虽然与 HSUF-RNN、DQN、Q-learning 等强化学习算法的吞吐量值趋近，但 MDSR 算法要比另外两种对比算法在吞吐量上有轻微的优势。

图 13 展示了在倾斜轨道大规模 LEO 卫星网络下五种算法丢包率随仿真时间变化曲线。图 13(a-d) 分别展示了在 10 s、30 s、60 s 和 300 s 路由更新间隔下的丢包率变化。仿真结果表明，在倾斜轨道大规模 LEO 卫星网络下，各种算法的丢包率与极轨道大规模 LEO 卫星网络的丢包率基本趋于一致，更换了星座构型网络后，基本对丢包率影响不大。MDSR 算法的丢包率要明显优于所有对比算法。同时可以看出，不同路由更新间隔对算法的丢包率没有明显的影响，这也侧面证明了算法的稳定性。

同样，与极轨道大规模 LEO 卫星网络相同，为了更加直观地展示 MDSR 算法与其他对比算法在倾斜轨道上的路由传输性能指标对比，这里也对端到端时延、吞吐量和丢包率进行了平均值计算，来更直观和精确地对路由传输性能指标进行对比分析。

图 14 展示了倾斜轨道大规模 LEO 卫星网络下几种算法在不同路由更新间隔下的路由传输性能指标对

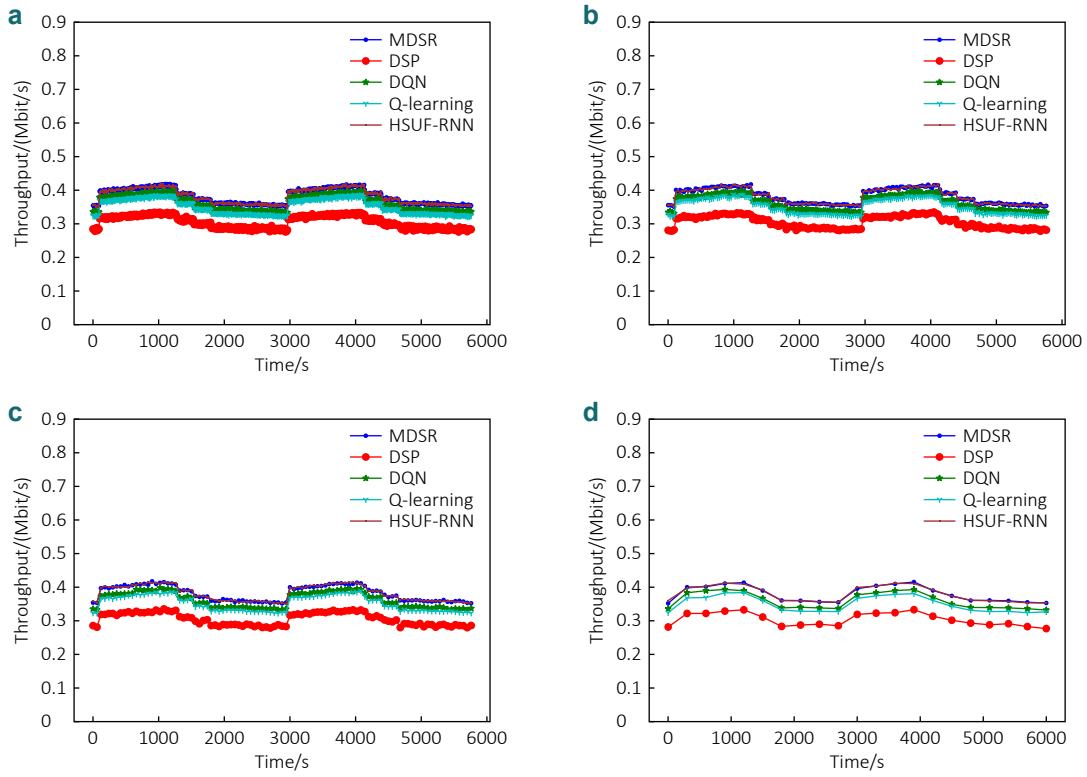


图 12 倾斜轨道大规模 LEO 卫星网络下不同算法的吞吐量随仿真时间变化曲线。路由更新间隔为 (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

Fig. 12 Throughput curves of different algorithms in inclined orbit large-scale LEO satellite networks with simulation time. Routing update interval (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

比。图 14(a-c) 分别展示了平均端到端时延。平均吞吐量和平均丢包率三种路由传输性能指标对比。仿真结果表明，MDSR 算法路由传输性能指标表现上要优于其他对比算法。其中平均丢包率指标上要远远优于未采用图神经网络范式的三种对比算法。同时，MDSR 与 HSUF-RNN 在测试时候路由传输性能指标仍然相近，由于二者本身都采用了 GNN+DRL 的范式优化路由选择，均能够利用 MPNN 感知网络的全局状态来进行路由决策，可以获得更好的路由传输性能。这也侧面证明，即使在倾斜轨道大规模 LEO 卫星网络上，采用极轨道训练的模型仍具备较强的传输性能。MDSR 算法无需额外训练即可适应不同轨道构型，展现了出色的拓扑泛化能力、

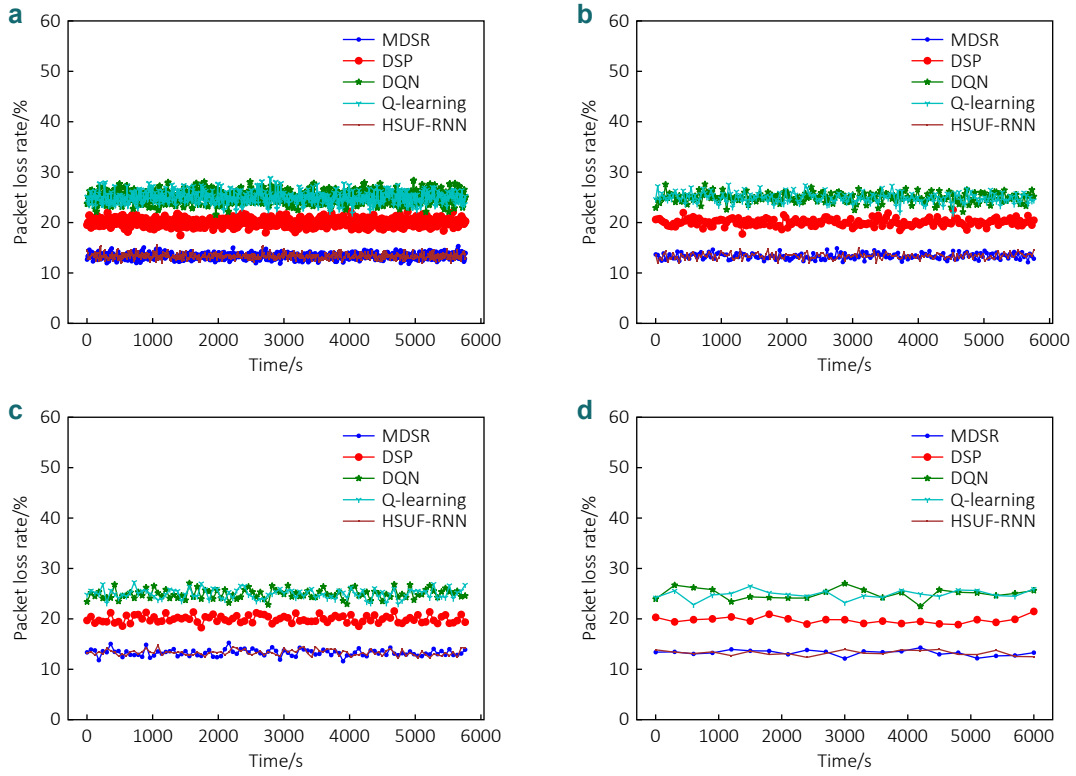


图 13 倾斜轨道大规模 LEO 卫星网络下不同算法的丢包率随仿真时间变化曲线。路由更新间隔为 (a) 10 s；(b) 30 s；(c) 60 s；(d) 300 s

Fig. 13 Packet loss rate curves of different algorithms in inclined orbit large-scale LEO satellite networks with simulation time. Routing update interval (a) 10 s; (b) 30 s; (c) 60 s; (d) 300 s

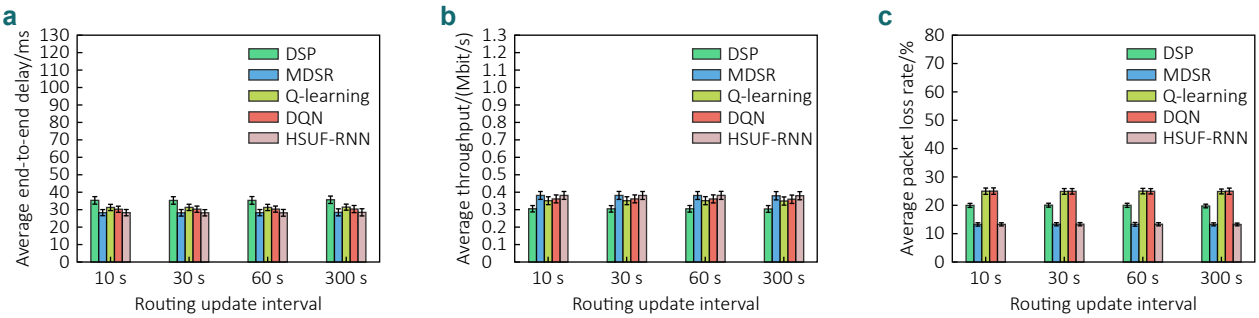


图 14 倾斜轨道大规模 LEO 卫星网络下不同算法的路由传输性能指标在不同路由更新间隔下的对比。(a) 平均端到端时延；(b) 平均吞吐量；(c) 平均丢包率

Fig. 14 Comparison of the routing performance indicators of different algorithms in the inclined orbit large-scale LEO satellite networks at different route update intervals. (a) Average end-to-end delay; (b) Average throughput; (c) Average packet loss rate

可扩展性与迁移能力，有效降低了训练成本。

综上所述，本章所提出的 MDSR 算法不仅在训练的极轨道星座网络的路由传输性能上要优于对比算法，并且不受路由更新间隔的影响。同时在倾斜轨道这种其他构型的大规模 LEO 卫星网络上仍然具有较好的路由传输性能指标，侧面证明了所提算法具备一定的拓扑泛化能力和迁移能力，具备良好的可扩展性。因此 MDSR 算法在大规模 LEO 卫星网络的路由问题中得到了验证，具有一定的优势。

5 结　论

本文主要针对大规模 LEO 卫星网络因动态拓扑变化而导致的传输性能下降和路由算法拓扑泛化能力不足的问题，提出了 MDSR 算法。MDSR 采用虚拟拓扑方法将大规模 LEO 卫星网络按照等时间间隔进行划分，构建图结构数据作为 MDSR 算法的输入。利用改进的 MPNN 网络来感知网络状态特征，并通过 DQN 算法学习最佳路由策略。本文在极轨道和倾斜轨道两种大规模 LEO 卫星网络中验证了 MDSR 的性能，并与 DSP、基于 RL 的大规模 LEO 卫星网络算法进行对比。仿真结果表明，MDSR 算法在训练时的收敛速度和初始奖励要高于 HSUF-RNN 模型，且在传输性能指标和拓扑泛化能力方面均优于其他对比算法。其中，在极轨道大规模 LEO 卫星网络中，MDSR 的平均丢包率比其他 RL 算法降低约 48%，比 DSP 降低约 35%；平均端到端时延比 DSP、Q-learning、DQN 和 DDPG 分别降低了约 20.5%、7.9%、2.7% 和 5.4%；平均吞吐量比 DSP、Q-learning、DQN 和 DDPG 分别提升了约 21.7%、7.6%、5% 和 1.2%。此外，MDSR 在不同的星座网络中表现出强大的泛化能力，无需额外训练即可保持较高性能，具备一定的可扩展性。未来，研究将继续探索随机链路连接和流量分区等优化方案，以进一步增强 MDSR 的鲁棒性，实现更均衡的网络负载分配。

作者贡献

黄思博：算法的设计、实验仿真与分析，论文写作；杨明川：指导论文选题、论文修改；张祯昊：资料收集和部分仿真结果分析。

利益冲突

所有作者声明无利益冲突

参考文献

1. Zhu X M, Jiang C X. Integrated satellite-terrestrial networks toward 6G: architectures, applications, and challenges[J]. *IEEE Internet Things J*, 2022, **9** (1): 437−461.
2. Liu Z Y, Wang S, Zhang R, et al. Routing in hierarchical hybrid satellite networks: a survey[J]. *IEEE Trans Netw Sci Eng*, 2026, **13**: 4883−4911.
3. Wang H, Wang J X, Qi W, et al. Survey on dynamic routing networking methods in integrated satellite-terrestrial network[C]// *Proceedings of 2023 IEEE/CIC International Conference on Communications in China*, Dalian, 2023: 1–5. https://doi.org/10.1109/ICCCWorkshops57813.2023.10233772.
4. Zhang J X, Chang Z Y, Zhang Y L, et al. Survey on routing technology of mega constellation[J]. *Space-Integr-Ground Inf Netw*, 2024, **5** (1): 2−13.
张佳鑫, 常朝阳, 张易隆, 等. 巨型星座路由技术综述[J]. 天地一体化信息网络, 2024, **5** (1): 2−13.
5. Bhattacharyya A, Nambiar S M, Ojha R, et al. Machine learning and deep learning powered satellite communications: enabling technologies, applications, open challenges, and future research directions[J]. *Int J Satell Commun Network*, 2023, **41** (6): 539−588.
6. Lee K L ,Lin P F,Chen C, et al. A survey on constellation routing methods in 6G non-terrestrial networks[C]// Proceedings of 2024 International Computer Symposium (ICS). IEEE,2024,5(4):57-74. https://doi.org/10.1109/ICS64339.2024.00027
7. Moorthy S K, Jagannath J. Survey of graph neural network for internet of things and NextG networks[J]. *IEEE Commun Surv Tutorials*, 2026, **28**: 1803−1844.
8. Werner M. A dynamic routing concept for ATM-based satellite personal communication networks[J]. *IEEE J Sel Areas Commun*, 1997, **15** (8): 1636−1648.
9. Ekici E, Akyildiz I F, Bender M D. A distributed routing algorithm for datagram traffic in LEO satellite networks[J]. *IEEE/ACM Trans Network*, 2001, **9** (2): 137−147.
10. Li X, Tang F L, Chen L, et al. A state-aware and load-balanced routing model for LEO satellite networks[C]//*Proceedings of 2017 IEEE Global Communications Conference*, Singapore, 2017: 1–6. https://doi.org/10.1109/GLOCOM.2017.8254443.
11. Liu Z L, Li J S, Wang Y R, et al. HGL: a hybrid global-local load balancing routing scheme for the internet of things through satellite networks[J]. *Int J Distrib Sens Netw*, 2017, **13** (3): 1–16. https://doi.org/10.1177/1550147717692586.
12. Dai S Y, Rui L L, Chen S Y, et al. A distributed congestion control routing protocol based on traffic classification in LEO satellite networks[C]//*Proceedings of 2021 IFIP/IEEE International Symposium on Integrated Network Management*, Bordeaux, 2021: 523–529.
13. Zhang T, Li H Y, Zhang S, et al. STAG-based QoS support routing strategy for multiple missions over the satellite networks[J]. *IEEE Trans Commun*, 2019, **67** (10): 6912−6924.
14. Kumar P, Bhushan S, Halder D, et al. fybrrLink: efficient QoS-aware routing in SDN enabled future satellite networks[J]. *IEEE Trans Netw Serv Manage*, 2022, **19** (3): 2107−2118.
15. Xie X, Huang L Y, Tang C W, et al. Multi-objective routing algorithms for low-earth orbit satellite network[J]. *Int J Satell*

Commun Network, 2023, **41** (5): 427−440.

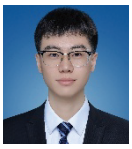
16. Zhang G Y, Wang K, Lai C Q, et al. Deep reinforcement learning-based routing optimization for software-defined satellite networks[C]//*Proceedings of 2024 International Conference on Ubiquitous Computing and Communication*, Chengdu, 2024: 436–441. https://doi.org/10.1109/IUCC65928.2024.00082.
17. Jiang X Y, Zhang T, Liu L. Research on satellite QoS routing algorithm based on digital twin[C]//*Proceedings of 2023 11th International Conference on Intelligent Computing and Wireless Optical Communications*, Chongqing, 2023: 118–122. https://doi.org/10.1109/ICWOC57905.2023.10200404.
18. Liu X Y, Ma T, Qin X H, et al. A DRL empowered multipath cooperative routing for ultra-dense LEO satellite networks[C]//*Proceedings of 2023 IEEE Global Communications Conference*, Kuala Lumpur, 2023: 5961–5966. https://doi.org/10.1109/GLOBECOM54140.2023.10436959.
19. Wang C, Wang H W, Wang W D. A two-hops state-aware routing strategy based on deep reinforcement learning for LEO satellite networks[J]. *Electronics*, 2019, **8** (9): 920.
20. Song J X, Ju Y, Liu L, et al. Trustworthy and load-balancing routing scheme for satellite services with multi-agent DRL[C]//*Proceedings of the IEEE INFOCM 2023 - IEEE Conference on Computer Communications Workshops*, Hoboken, 2023: 1–6. https://doi.org/10.1109/INFOCOMWKSHPS57453.2023.10226127.
21. Huang Y X, Wu S F, Kang Z Y, et al. Reinforcement learning based dynamic distributed routing scheme for mega LEO satellite networks[J]. *Chinese Journal of Aeronautics*, 2023, **36** (2): 284−291.
22. Nie M S, Chen D M, Wang D Q. Reinforcement learning on graphs: a survey[J]. *IEEE Trans Emerg Top Comput Intell*, 2023, **7** (4): 1065−1082.
23. Xue L X, Yin K J, Wang R G, et al. Interactive instance proposal network for HOI detection[J]. *Opto-Electron Eng*, 2022, **49** (7): 210429.
薛丽霞, 尹凯建, 汪荣贵, 等. 基于交互实例推荐网络的人-物交互检测方法研究[J]. 光电工程, 2022, **49** (7): 210429.
24. Ye S J, Wang Y X. Graph neural network-based WSI cancer survival prediction method[J]. *Opto-Electron Eng*, 2024, **51**(4): 240011.
叶世杰, 王永雄. 基于图神经网络的 WSI 癌症生存预测方法[J]. 光电工程, 2024, **51**(4): 240011.
25. Almasan P, Suárez-Varela J, Rusek K, et al. Deep reinforcement learning meets graph neural networks: exploring a routing optimization use case[J]. *Comput Commun*, 2022, **196**: 184−194.
26. Sun W H, Li Z M, Sun H M. Multi-agent reinforcement learning value function factorization approach based on graph neural network[J]. *Comput Eng*, 2024, **50** (5): 62−70.
孙文洁, 李宗民, 孙浩淼. 基于图神经网络的多智能体强化学习值函数分解方法[J]. 计算机工程, 2024, **50** (5): 62−70.
27. Wang H, Ran Y Y, Zhao L, et al. Dynamic routing algorithm for LEO satellite networks based on deep graph reinforcement learning[J]. *J Chongqing Univ Posts Telecommun (Nat Sci Ed)*, 2023, **35** (4): 596−605.
汪昊, 冉泳屹, 赵雷, 等. 基于深度图强化学习的低轨卫星网络动态路由算法[J]. 重庆邮电大学学报 (自然科学版), 2023, **35** (4): 596−605.
28. Huang Y X, Yang D, Feng B H, et al. A GNN-enabled multipath routing algorithm for spatial-temporal varying LEO satellite networks[J]. *IEEE Trans Veh Technol*, 2024, **73** (4): 5454−5468.
29. Xu P, Feng M J, Zhou J X, et al. Inter-satellite routing for LEO satellite networks: a GNN and DRL integrated approach[C]//*Proceedings of 2024 IEEE/CIC International Conference on Communications in China*, Hangzhou, 2024: 1346–1351. https://doi.org/10.1109/ICCC62479.2024.10681995.
30. Gao Y, Huang K H, Wang K, et al. Intelligent hierarchical routing strategy based on regional division for large-scale LEO satellite constellations[C]//*Proceedings of 2025 IEEE/CIC International Conference on Communications in China*, Shanghai, 2025: 1–6. https://doi.org/10.1109/ICCCWorkshops67136.2025.11148142.
31. Yu G, Kwak M, Park S, et al. GNN-aided efficient multipath routing in satellite constellation networks[C]//*Proceedings of 2025 International Conference on Mobile, Military, Maritime IT Convergence*, Cebu, 2025: 55–58. https://doi.org/10.1109/ICMIC66299.2025.11257771.
32. Zhang S B, Liu A J, Han C, et al. GRLR: routing with graph neural network and reinforcement learning for mega LEO satellite constellations[J]. *IEEE Trans Veh Technol*, 2025, **74** (2): 3225−3237.
33. Ran Y Y, Ding Y J, Chen S W, et al. Fully-distributed dynamic packet routing for LEO satellite networks: a GNN-enhanced multi-agent reinforcement learning approach[J]. *IEEE Trans Veh Technol*, 2025, **74** (3): 5229−5234.
34. Deng X, Chang L, Zeng S Y, et al. Distance-based back-pressure routing for load-balancing LEO satellite networks[J]. *IEEE Trans Veh Technol*, 2023, **72** (1): 1240−1253.
35. Hsu Y H, Lee J I, Xu F M. A deep reinforcement learning based routing scheme for LEO satellite networks in 6G[C]//*Proceedings of 2023 IEEE Wireless Communications and Networking Conference*, Glasgow, 2023: 1–6. https://doi.org/10.1109/WCNC55385.2023.10118680.
36. Hamrouni A, Ghazzai H, Setti G, et al. LSTM-based dynamic routing with non-ISL LEO satellite constellations for remote IoT connectivity[C]//*Proceedings of 2024 IEEE Global Communications Conference*, Cape Town, 2024: 1473–1478. https://doi.org/10.1109/GLOBECOM52923.2024.10901596.
37. Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry[C]//*Proceedings of the 34th International Conference on Machine Learning*, Sydney, 2017: 1263–1272.
38. Wang X, Wang S, Liang X X, et al. Deep reinforcement learning: a survey[J]. *IEEE Trans Neural Netw Learn Syst*, 2024, **35** (4): 5064−5078.
39. Zhang X, Yang Y, Xu M W, et al. ASER: scalable distributed routing protocol for LEO satellite networks[C]//*Proceedings of 2021 IEEE 46th Conference on Local Computer Networks*, Edmonton, 2021: 65–72. https://doi.org/10.1109/LCN52139.2021.9524989.
40. Haeupler B, Hladík R, Rozhoň V, et al. Universal optimality of Dijkstra via beyond-worst-case heaps[C]//*Proceedings of the 2024 IEEE 65th Annual Symposium on Foundations of Computer Science*, Chicago, 2024. https://doi.org/10.1109/FOCS61266.2024.00125.
41. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, **518** (7540): 529−533.
42. Tan H N. Reinforcement learning with deep deterministic policy gradient[C]//*Proceedings of 2021 International Conference on Artificial Intelligence, Big Data and Algorithms*, Xi'an, 2021: 82–85. https://doi.org/10.1109/CAIBDA53561.2021.00025.

作者简介



【通信作者】黄思博 (2001-)，男，硕士研究生，主要研究方向为深度强化学习和卫星网络路由算法。

E-mail：zerohsb2001@163.com



张祯昊 (2001-)，男，硕士研究生，主要研究方向为卫星通信波束资源调度算法、深度强化学习。

E-mail：zhzhang2022@163.com



杨明川 (1977-)，男，博士，教授，主要研究方向为卫星通信新技术、空间信息网络、星地一体化网络高效传输技术和复杂环境下信号快速识别技术。

E-mail：mcyang@hit.edu.cn

Deep graph reinforcement learning routing algorithm for large-scale LEO satellite networks

Huang Sibo*, Yang Mingchuan, Zhang Zhenhao

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150001 China

Abstract

Objective Large-scale low earth orbit (LEO) satellite networks can provide strong coverage, high data transmission rates and low latency. This makes them an important part of future 6G communication systems. Routing technology is a critical component of large-scale LEO satellite networks, providing high-quality and low-latency services for network communications through packet transmission mechanisms. However, routing technology for large-scale LEO satellite networks faces critical challenges, including poor routing transmission performance due to dynamic changes in network topology and insufficient scalability of algorithmic models. In recent years, the development of artificial intelligence (AI) has provided new avenues for addressing these challenges. To address these challenges, this study utilizes AI technology, combining graph neural networks (GNNs) and deep reinforcement learning (DRL) to propose a routing algorithm that enhances transmission performance across diverse network topologies while demonstrating robust topological generalization capabilities.

Methods This study utilized message passing neural networks (MPNNs) from GNN models and deep Q-networks (DQNs) from DRL models, and proposed a large-scale LEO satellite network routing algorithm based on MPNN and DQN (MDSR). Firstly, the MDSR algorithm employed a virtual topology approach to partition large-scale LEO satellite networks into fixed time intervals and construct them as graph-structured data for input into MDSR. Secondly, we designed the MDSR framework. The MDSR framework included two parts: an improved MPNN module and a DQN module. The improved MPNN module included a message processing stage and a readout stage. The message processing stage consisted of a message processing function (MPF), an information aggregation function (IAF), and a hidden state update function (HSUF). The readout stage consisted of a readout function (RF). Specifically, we improved the HSUF by replacing the original recurrent neural networks (RNNs) with long short-term memory (LSTM) networks. This aimed to enhance the model's ability to learn long-range dependencies and mitigate vanishing and exploding gradients. The improved MPNN module extracted the satellites and inter satellite links state features and their interconnections from the networks. The DQN module then used these features to make routing decisions. Finally, in order to compare the routing and transmission performance of different algorithms in different satellite network configurations, this study selected the HSUF as the RNN for the MPNN + DQN (HSUF-RNN) based algorithms, the Dijkstra shortest path (DSP) algorithm, the DQN based algorithm, the Q-learning-based algorithm, and the deep deterministic policy gradient (DDPG)-based algorithm for comparison with MDSR. MDSR was trained alongside other reinforcement learning algorithms used for comparison in a routing scenario involving a large-scale LEO satellite network in a polar orbit with a routing update interval of 30 s. Simultaneously, MDSR and other comparison algorithms underwent routing transmission performance testing under 10 s, 30 s, 60 s and 300 s routing

Foundation item: National Natural Science Foundation of China (62071146)

*Correspondence: Huang S B, E-mail: zerohsb2001@163.com

update intervals for large-scale LEO satellite networks in polar and inclined orbits. The experimental environment included the Windows 11 operating system, an NVIDIA GeForce RTX 4050 GPU, the PyTorch 1.13.1 + CUDA 11.6 deep learning framework, and the Python 3.9 programming language.

Results and Discussions The results demonstrate that, during training, MDSR achieved faster convergence and a higher initial reward than HSUF-RNN. During testing, in polar large-scale LEO satellite networks, MDSR achieved an average packet loss rate approximately 48% lower than those of the other RL algorithms and 35% lower than that of DSP. The average end-to-end delay of MDSR was approximately 20.5%, 7.9%, 2.7%, and 5.4% lower than that of DSP, Q-learning, DQN, and DDPG, respectively. The average throughput of MDSR was approximately 21.7%, 7.6%, 5%, and 1.2% higher than that of DSP, Q-learning, DQN, and DDPG, respectively. In inclined large-scale LEO satellite networks, MDSR adapted to different constellation network configurations without requiring additional training, demonstrating strong topological generalization capabilities. MDSR significantly outperformed static routing algorithms like DSP in terms of routing transmission performance. Compared to routing algorithms based on non-GNN+DRL paradigms, MDSR can accurately perceive the global network state to make routing decisions, thereby improving key routing metrics such as end-to-end delay, throughput, and packet loss rates. Therefore, in our experiments, MDSR enhanced routing and transmission performance. Without additional training, the MDSR algorithm demonstrated stable routing performance across different constellation network configurations, exhibiting strong topological generalization capabilities.

Conclusions This study proposes the MDSR algorithm, which has been developed to address the issues of degraded transmission performance and insufficient topological generalization capability in routing algorithms. MDSR utilizes the improved MPNN module to extract network state features, and the DQN module learns the best routing strategies through these features. MDSR is subjected to experimental evaluation in both polar and inclined orbit large-scale LEO satellite networks. It is also compared with the HSUF-RNN, DSP, DQN, Q-learning and DDPG algorithms. The results show that, during training, MDSR exhibits faster convergence and higher initial rewards compared to the HSUF-RNN model. During testing, MDSR outperforms other comparison algorithms in both routing transmission performance metrics and topological generalization capabilities. In future research, we will explore optimization schemes such as random link connections and traffic partitioning to further enhance the robustness of MDSR while also achieving more balanced network load distribution.

Keywords: large-scale low earth orbit satellite networks; routing algorithm; message passing neural networks; deep Q networks

Huang S B, Yang M C, Zhang Z H. Deep graph reinforcement learning routing algorithm for large-scale LEO satellite networks[J]. *Opto-Electron Eng*, 2026, **53**(6): 250386; DOI: 10.12086/oee.2026.250386

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文