



XXXX

LD3QN-IR: 信息时效性与能量协同优化的卫星网络路由算法

杨晨¹, 王欣睿¹, 魏德宾¹, 潘成胜²

(1. 大连大学信息工程学院, 辽宁 大连 116622;

2. 南京邮电大学通信与网络技术国家工程研究中心, 江苏 南京 210023)

摘 要: 针对低轨卫星运动引起的网络拓扑和星间链路状态动态变化, 以及业务负载、链路资源和节点能量状态持续演化时, 现有卫星网络路由方法难以同时兼顾信息时效性、链路拥塞与节点能量消耗的问题, 提出信息时效性与能量协同优化的卫星网络路由算法 LD3QN-IR。该算法将网络邻接关系、链路剩余带宽、传输时延、端到端信息年龄及节点剩余能量比例联合构建为状态空间, 利用 LSTM 对历史网络状态进行时序编码, 并通过 CNN 进一步提取状态特征中的局部关联信息; 随后采用 Dueling Double DQN 对状态价值与动作优势进行分解, 并利用 Double Q 机制降低动作价值过估计。在此基础上, 通过综合考虑信息年龄、链路利用率、节点能量和路由前进性的奖励函数, 引导智能体在信息新鲜度、传输性能和能量消耗之间进行协同权衡。基于 STK、NS3 与 MATLAB 的联合动态拓扑仿真结果表明, LD3QN-IR 在不同网络负载条件下均获得最高的综合均衡得分, 表现出更优的多目标综合性能, 能够有效兼顾传输时延、信息新鲜度与节点能量消耗。

关键词: 低轨卫星网络; 信息年龄; 智能路由; 长短期记忆网络

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.

LD3QN-IR: An Information-Timeliness and Energy Co-Optimization Routing Algorithm for Satellite Networks

YANG Chen¹, WANG Xinrui¹, WEI Debin¹, PAN Chengsheng²

1. School of Information Engineering, Dalian University, Dalian 116622, Liaoning, China

2. National Engineering Research Center of Communications and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China

Abstract: In low-Earth-orbit (LEO) satellite networks, the network topology and inter-satellite link states were continuously altered by satellite motion, while continuous variations were observed in traffic loads, link resources, and node energy states. Consequently, information timeliness, link congestion, and node energy consumption could not be simultaneously balanced by existing satellite network routing methods. To address this problem, LD3QN-IR, a satel-

收稿日期: 2026-06-30; 修回日期: 2026-08-13

基金项目: 国家自然科学基金资助项目 (No.61931004)

Foundation Items: The National Natural Science Foundation of China(No.61931004)



lite network routing algorithm for the joint optimization of information timeliness and energy, was proposed. In the proposed algorithm, a state space was constructed by jointly incorporating network adjacency, residual link bandwidth, transmission delay, end-to-end age of information (AoI), and the residual energy ratio of each node. Historical network states were temporally encoded using a long short-term memory (LSTM) network, and local temporal correlations in the encoded state features were further extracted using a convolutional neural network (CNN). Subsequently, a Dueling Double deep Q-network (Dueling Double DQN) was adopted, through which the action-value function was decomposed into a state-value function and an action-advantage function, while action-value overestimation was reduced using the Double Q mechanism. On this basis, a reward function incorporating AoI, link utilization, residual node energy, and routing progress was designed, through which the agent was guided to balance information freshness, transmission performance, and energy consumption. Joint dynamic-topology simulations were conducted using STK, NS3, and MATLAB. The highest comprehensive balance score was achieved by LD3QN-IR under all tested network loads, and superior multi-objective performance was demonstrated, with end-to-end delay, information freshness, and node energy consumption effectively balanced.

Key words: Low-Earth-orbit satellite network, Age of information, Intelligent routing, Long short-term memory

1 引言

低轨卫星网络能够通过大规模星座实现全球广域覆盖,是天地一体化网络、远程监测、应急通信和时敏业务承载的重要基础设施^[1]。与地面网络相比,低轨卫星节点高速运动,星间链路可见性、传播距离和链路负载随时间持续变化;同时,星载设备受体积、载荷和能源补给方式限制,节点电池容量有限,且需要在光照区与卫星食区之间周期性充放电。在多跳转发场景中,如果路由策略长期选择少数中继节点,容易造成局部链路拥塞和节点放电深度累积,进而影响网络稳定性和星座服役寿命。

传统卫星网络路由研究多以时延、跳数、吞吐量或链路负载为优化目标,但对于状态更新类业务而言,仅降低平均传输时延并不一定能够保证接收端信息的新鲜度。信息年龄(Age of Information, AoI)能够刻画接收端所持有状态信息相对于当前时刻的陈旧程度,近年来已成为时敏通信系统的重要性能指标^[2]。已有研究从链路调度、吞吐约束和多跳路由等角度开展AoI优化^[3],但在低轨卫星多跳转发过程中,信息新鲜度、链路拥塞与节点能量消耗相互制约:过度追

求低时延或低AoI可能使业务集中于少数优质链路和中继节点,而过度强调能量均衡又可能增加传输路径长度和信息陈旧程度。

能量感知路由能够延长卫星网络持续运行时间^[4],但现有方法通常将能量约束作为独立指标处理,对AoI、端到端时延和能量消耗之间的动态权衡考虑不足。深度强化学习(Deep Reinforcement Learning, DRL)具备在动态环境中通过交互学习策略的能力,为低轨卫星网络智能路由提供了可行路径^[5]。然而,业务负载、链路剩余带宽、排队时延、信息年龄以及节点剩余能量等网络状态具有明显的时序相关性,当前时刻的传输性能和资源状态不仅与瞬时观测有关,还受到历史业务负载、资源占用、数据转发和能量消耗累积的影响。仅依据单一时刻的网络状态进行决策,难以充分利用历史状态中蕴含的演化信息。因此,需要构建能够同时刻画网络状态时序依赖和多目标约束的智能路由机制。

综上,现有卫星网络路由方法在动态拓扑条件下仍存在两方面不足:一是对信息新鲜度、链路拥塞与节点能量状态的协同考虑不足;二是基于瞬时网络状态的决策难以充分利用拓扑、业务负载及资源状态的历史演化信息。针对上述问

题, 本文提出信息时效性与能量协同优化的卫星网络路由算法 LD3QN-IR, 通过多维状态表征、时序特征提取和多目标奖励机制实现动态网络环境下的下一跳路由决策。

本文主要贡献如下:

(1) 针对低轨卫星运动引起的网络拓扑变化以及信息新鲜度、链路拥塞和节点能源状态难以协同优化的问题, 构建面向动态拓扑的 AoI—能量—拥塞多目标路由决策模型。将时变网络邻接关系、链路剩余带宽、传输时延、端到端信息年龄及节点剩余能量比例联合纳入状态空间, 并根据当前有效星间链路动态更新候选下一跳集合, 使智能体能够感知拓扑演化与网络资源状态变化。

(2) 设计面向信息时效性、链路资源和节点能量的协同奖励机制。通过联合考虑 AoI、链路利用率、节点剩余能量及路由前进性, 对不同路由目标进行统一反馈, 引导智能体在信息新鲜度、传输性能与能量消耗之间实现动态权衡。

(3) 构建 LSTM-CNN-D3QN 路由决策网络。利用 LSTM 对连续历史拓扑和网络资源状态进行时序编码, 并通过 CNN 进一步提取局部时序关联特征, 结合 Dueling Double DQN 完成动态候选下一跳的动作价值评估, 提高复杂动态网络条件下的路由决策能力。

(4) 基于 STK、NS3 与 MATLAB 搭建动态低轨卫星网络联合仿真平台, 并与 LUR、Stackelberg 和 RT-ATT-MAD2QN 等方法进行对比。实验结果表明, LD3QN-IR 在不同网络负载条件下均取得最高综合均衡得分, 能够较好兼顾端到端时延、信息新鲜度和节点能耗。

2 相关工作

2.1 面向 AoI 的网络路由算法

AoI 用于描述状态更新分组从源端生成到接收端成功获得之间的时间跨度, 与传统平均时

延、吞吐量等指标不同, 它更关注接收端信息的新鲜程度。Gao 等^[3]面向低轨卫星-地面融合网络研究了平均 AoI 最小化问题, Lin 和 Liao^[6]进一步在 NOMA 低轨卫星物联网场景中考考虑能量约束下的 AoI 优化, Wang 等^[7]从天地融合 IoT 网络角度联合优化 AoI 与能效。Li 等^[8]针对低轨卫星星座中的多跳信息更新问题, 基于连续时变图研究面向 AoI 优化的多跳路由机制, 通过路径选择降低信息更新过程中的端到端 AoI。上述研究为卫星网络中的信息时效性建模提供了重要基础, 但现有方法更多关注 AoI 最小化或 AoI 与能效之间的优化关系, 对于多跳路由过程中信息新鲜度、链路资源与中继节点能源状态的协同考虑仍显不足。因此, 有必要在路由决策中进一步综合信息时效、链路状态与节点能源约束。

2.2 低轨卫星网络能量感知路由

在能量受限的卫星网络中, 路由策略不仅影响通信性能, 还会影响节点电池循环寿命。Yang 等^[4]从 QoS 需求与可用能量角度设计了 LEO 卫星网络路由算法, Jing 等^[9]面向低轨卫星激光通信建立了考虑电池寿命的能效路由模型, Liu 等^[10]进一步将深度强化学习用于巨型卫星星座中的能量感知路由。这类研究能够缓解节点能量消耗不均^[11]问题, 但通常更关注生命周期、时延边界或负载均衡, 对信息新鲜度指标的建模不足。因此, 如何在路由过程中兼顾节点能源状态与信息时效性仍有进一步研究空间。

2.3 基于深度强化学习路由

近年来, DRL 已被用于空间-空中-地面一体化网络负载均衡路由、低轨巨型星座服务功能约束路由等动态路由场景^[12]。DQN 及其改进结构^[13]已被用于链路选择、负载均衡和实时路由等任务。Wen 等^[14]提出了融合 LSTM 与 DQN 的卫星网络多业务路由算法, 将链路时延、带宽、丢包率、流量、网络拓扑和业务类型作为状态输入, 用于多业务 QoS 保障与负载均衡优化。



Double Q 学习通过分离动作选择与动作评估缓解 Q 值过估计问题^[15], Dueling 结构通过分解状态价值和动作优势提高动作价值估计效率^[16]。与已有 LSTM-DQN 类方法主要面向 QoS 保障和负载均衡不同, 本文关注信息新鲜度、链路拥塞与节点能源状态之间的协同路由问题, 在传统链路状态基础上进一步引入 AoI 和节点剩余能量信息, 并通过多目标奖励反馈实现联合决策。

3 系统模型与问题建模

3.1 卫星网络结构

本文以铱星类低轨卫星通信网络作为研究对象。与传统依赖地面中继的卫星通信系统不同, 铱星类星座通过空间段、用户段和地面段协同构成空天地一体化通信网络。空间段由多轨道面的低轨卫星组成, 卫星之间通过星间链路 (Inter-Satellite Link, ISL) 形成空间多跳转发网络; 用户段由固定终端、移动终端等接入设备组成, 负责业务数据的上行接入和下行接收; 地面段由网关机、网络控制中心和互联网接入设施组成, 承担数据汇聚、外部网络互联等功能。用户业务通常先由终端上传至可见接入卫星, 再经一条或多条星间链路转发至目标区域上空卫星, 最终通过下行链路或网关机送达目的端。本文主要对完成接入后的入口卫星与出口卫星之间的空间段多跳转发过程进行建模与优化, 用户段和地面段仅用于业务接入与数据接收, 不参与本文的路由决策优化。

3.2 问题建模

星间链路示意图如图 1 所示。

考虑卫星轨道运动引起的星间连接关系动态变化, 将低轨卫星网络表示为时变图 $G(t) = (V, E(t))$, 其中 $V = \{v_1, v_2, \dots, v_n\}$ 为卫星集合, 并且 $|V| = n$, $E(t)$ 为时刻 t 的有效星间链路集合。随着卫星沿轨道运动, 卫星间相对位置、可见关系及

表 1 主要符号说明

符号	含义
$A(t)$	加权邻接矩阵
$D_{i,j}(t)$	卫星节点 i 与 j 之间的距离
$\text{delay}_{i,j}(t)$	节点 i 与 j 之间链路传输时延
$\text{rbw}_{i,j}(t)$	链路 (i,j) 的剩余可用带宽
$C_{i,j}^{\max}$	链路 (i,j) 的最大链路容量
$\text{AoI}(i,j,d,t)$	选择节点 j 为下一跳时的预期端到端信息年龄
$\text{AoI}_{i,j}(t)$	归一化后的预期端到端信息年龄
$\text{RER}_j(t)$	节点 j 的剩余能量比例
$U_{i,j}(t)$	链路 (i,j) 的利用率
K_i, K_j	当前节点 i 和候选节点 j 到目的节点的最短剩余跳数
$B(t)$	节点间链路的剩余可用带宽矩阵
w_1, w_2, w_3, w_4	奖励函数各优化项的权重系数
$\chi_{ij}(t)$	链路可用状态
$U_{\max}$	链路利用率阈值
$\text{RER}_{\min}$	剩余电量比例阈值

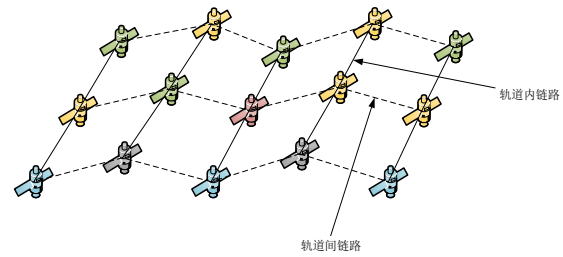


图 1 星间链路示意图

通信距离不断变化, 因此 $E(t)$ 随时间动态更新。记时刻 t 卫星 i 与卫星 j 之间的星间链路可用状态为 $\chi_{ij}(t)$, 其定义如式(1)所示:

$$\chi_{ij}(t) = \begin{cases} 1, & \text{节点 } i \text{ 与 } j \text{ 在时刻 } t \text{ 满足 ISL 连接条件,} \\ 0, & \text{其他.} \end{cases} \quad (1)$$

由此, 有效星间链路集合可以表示为:

$$E(t) = \{(i,j) | \chi_{ij}(t) = 1\} \quad (2)$$

本文使用加权邻接矩阵 $A(t)$ 来表示卫星网络在时刻 t 的连接关系及链路权值, 如式 (3) - (4) 所示:

$$A(t) = [A_{ij}(t)]_{n \times n} \quad (3)$$

$$A_{ij}(t) = \begin{cases} D_{i,j}(t), & \chi_{ij}(t) = 1, \\ 0, & \chi_{ij}(t) = 0. \end{cases} \quad (4)$$

其中, $D_{i,j}(t)$ 表示卫星节点 i 与节点 j 的星间距离,

因此, $A(t)$ 不仅描述不同卫星之间的链路权值, 还反映时刻 t 的实际网络连接关系。当 $\chi_{ij}(t)=1$ 时, 采用卫星 i 与 j 的实时星间距离 $D_{ij}(t)$ 作为链路权值; 当链路不可用时, 对应邻接元素取 0。

由于卫星轨道运动会导致星间链路周期性建立、断开和切换, 因此不同卫星节点的可用邻居集合随时间动态变化。定义时刻 t 节点 i 的有效邻居集合为

$$\mathcal{N}_i(t) = \{j \in V | \chi_{ij}(t) = 1\} \quad (5)$$

其中, $\mathcal{N}_i(t)$ 表示时刻 t 与节点 i 存在有效星间链路的卫星节点集合。当拓扑状态发生更新时, 各节点的有效邻居集合根据最新链路状态同步更新, 为后续链路状态计算和下一跳路由选择提供候选节点集合。

卫星间距离可由卫星间的瞬时地心角与卫星的轨道高度计算得到, 如图2所示。

卫星间的瞬时地心角可由星下点经纬度计算得到。设 i 和 j 的星下点经纬度分别为 $(\lambda_1, \phi_1), (\lambda_2, \phi_2)$, α_{ij} 为两卫星对应的瞬时地心角。则瞬时地心角 $\alpha_{ij}(t)$ 为:

$$\alpha_{ij}(t) = \arccos[\sin \phi_1(t) \sin \phi_2(t) + \cos \phi_1(t) \cos \phi_2(t) \cos(\lambda_1(t) - \lambda_2(t))] \quad (6)$$

某一时刻卫星间距离的计算公式为:

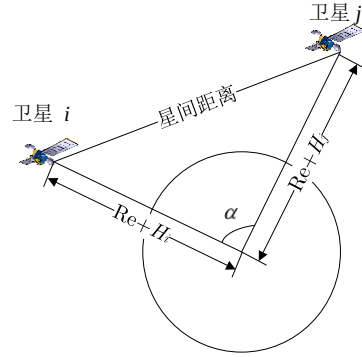


图2 卫星间距离示意图

$$D_{ij}(t) = \sqrt{(Re + H_i)^2 + (Re + H_j)^2 - 2 \cos \alpha_{ij}(t) \cdot (Re + H_i) \cdot (Re + H_j)} \quad (7)$$

其中, Re 为地球半径, H_i 是卫星 i 的高度, H_j 是卫星 j 的高度。STK 根据卫星轨道参数获得各卫星在时刻 t 的实时位置, 并输出对应的星下点经纬度, 由此计算瞬时地心角及星间距离。随着轨道运动, $D_{ij}(t)$ 连续变化, 并进一步影响链路传播时延和可用状态。

星间链路的传输时延由传播时延 $\text{delay}_{i,j}^p$ 与链路平均排队时延 $\text{delay}_{i,j}^{\text{queue}}$ 组成, 如式 (8) - (10):

$$\text{delay}_{i,j}^p(t) = \frac{D_{ij}(t)}{c} \quad (8)$$

$$\text{delay}_{i,j}^{\text{queue}}(t) = \frac{q(t) \times \text{Pkt}_{\text{avg}}}{C_{i,j}^{\text{max}}} \quad (9)$$

$$\text{delay}_{i,j}(t) = \text{delay}_{i,j}^p(t) + \text{delay}_{i,j}^{\text{queue}}(t) \quad (10)$$

其中, Pkt_{avg} 为队列中数据包的平均长度, $q(t)$ 表示在 t 时刻队列中的数据包数量, $C_{i,j}^{\text{max}}$ 为链路 (i,j) 的最大传输容量, c 表示光速, $\text{delay}_{i,j}$ 表示节点 i 到节点 j 的链路总传输时延, 由传播时延和排队

时延两部分组成。当 $\chi_{ij}(t)=0$ 时, 表示链路 (i,j) 在当前时刻不可用, 该链路不参与当前路由决策, 其对应传输时延不作为候选下一跳的有效状态信息。

设当前节点为 i , 候选下一跳为 j , 目的节点为 d 。则时刻 t 选择 j 时的预期端到端信息年龄可以表示为:

$$\text{AoI}(i,j,d,t) = \text{AoI}_{\text{current}}(t) + \text{delay}_{i,j}(t) + \text{AoI}_{\text{est}}(j,d,t) \quad (11)$$

其中, $\text{AoI}_{\text{current}}$ 为数据包本身的信息年龄, $\text{AoI}_{\text{est}}(j,d)$ 为从 j 到 d 的估计剩余信息年龄, 可以通过邻居交换信息获得。进一步, 为消除 AoI 数值尺度对路由决策的影响, 对当前节点 i 各候选下一跳对应的预期端到端信息年龄进行 Min-Max 归一化, 如式 (12) 所示:



$$\text{AoI}_{i,j}(t) = \frac{\text{AoI}(i,j,d,t) - \min_{i,j}(\text{AoI}(i,j,d,t))}{\max_{i,j}(\text{AoI}(i,j,d,t)) - \min_{i,j}(\text{AoI}(i,j,d,t))} \quad (12)$$

其中, $\min_{i,j}(\text{AoI}(i,j,d,t))$ 和 $\max_{i,j}(\text{AoI}(i,j,d,t))$ 分别表示当前候选下一跳集合中预期端到端信息年龄的最小值和最大值, $\text{AoI}_{i,j}(t) \in [0,1]$ 为归一化结果。

低轨卫星主要依靠太阳能帆板充电和蓄电池储能, 受轨道阴影期和太阳光照条件的限制, 可用的电能总量存在严格上限。一旦节点电量耗尽, 不仅该卫星失效, 还可能造成网络拓扑断裂, 影响全局连通性。路由决策直接影响能耗分布。不同的转发路径会导致某些卫星承担过多的数据中继任务, 使其电池过快耗尽, 缩短整个星座的有效服务寿命。为此, 本文的路由算法考虑节点的剩余电量比例, 如式 (13)。

$$\text{RER}_j(t) = \frac{E_j^{\text{res}}(t)}{E_j^{\text{cap}}} \quad (13)$$

其中, $E_j^{\text{res}}(t)$ 是节点 j 当前剩余电量, E_j^{cap} 是电池的总容量, $\text{RER}_j \in [0,1]$ 为节点 j 的剩余能量比例。进一步, 路由算法的设计需要分散负载、避免拥塞, 从而提高网络的平均吞吐量, 为此, 计算链路利用率如下:

$$U_{i,j}(t) = 1 - \frac{\text{rbw}_{i,j}(t)}{C_{i,j}^{\text{max}}} \quad (14)$$

其中, $\text{rbw}_{i,j}(t)$ 表示链路 $e(i,j)$ 的剩余可用带宽, $C_{i,j}^{\text{max}}$ 表示当前链路的最大链路容量, $U_{i,j}(t) \in [0,1]$ 为链路利用率, $U_{i,j}(t)$ 越大, 表示当前链路负载越高。

根据上述提到的各类指标构建的具体目标函数如式 (15) 所示:

$$\begin{aligned} & \max_{\pi} \mathbb{E} [\alpha \cdot \text{RER}_j(t) - \beta \cdot U_{i,j}(t) - \gamma \cdot \overline{\text{AoI}}_{i,j}(t)] \\ & \text{s.t.} \quad U_{i,j}(t) \leq U_{\text{max}} \\ & \quad \text{RER}_j(t) \geq \text{RER}_{\text{min}} \end{aligned} \quad (15)$$

其中, α, β, γ 分别表示节点剩余能量、链路利用率和信息年龄在目标函数中的权重系数, 用于调节各优化目标的相对重要程度; U_{max} 为链路利用率阈值, 本文设为 0.8, RER_{min} 为剩余电量比例阈值, 设为 0.2。

4 LD3QN-IR 网络框架

在低轨卫星运行过程中, 卫星位置、星间链路可见性及网络邻接关系随时间变化; 与此同时, 业务负载、链路剩余带宽、排队时延、AoI 和节点剩余能量也持续演化。因此, 当前网络状态不仅与瞬时观测有关, 还受到历史拓扑变化、业务传输和资源消耗过程的影响。为提取上述网络状态的时序依赖特征, 本文引入 LSTM 对连续历史状态进行编码。

如图 3 所示, LD3QN-IR 由 LSTM、CNN 和 Dueling Double DQN 组成。首先, LSTM 对连续历史网络状态进行编码, 以提取拓扑、链路资源、AoI 及节点能量的时序演化特征; 随后, CNN 沿时间维进一步提取相邻状态之间的局部时序关联, 并通过自适应平均池化获得固定维度表示; 最后, Dueling 网络分别估计状态价值和动作优势, 结合 Double Q 机制完成候选下一跳的动作价值评估。

本文将卫星网络中的下一跳选择建模为基于多维网络状态的序贯决策过程。与仅依据时延、带宽或节点能量等单一指标进行路径选择的方法不同, LD3QN-IR 在每个路由决策时刻同时感知链路传输状态、信息时效状态和节点能源状态, 并依据协同奖励信号评价不同候选下一跳的综合决策效果。智能体通过与网络环境持续交互学习路由策略, 使所选择的下一跳不仅满足传输性能需求, 同时兼顾信息新鲜度、链路负载和节点能量消耗。

4.1 状态空间

状态空间 S_t 由网络连接与传输状态、信息时

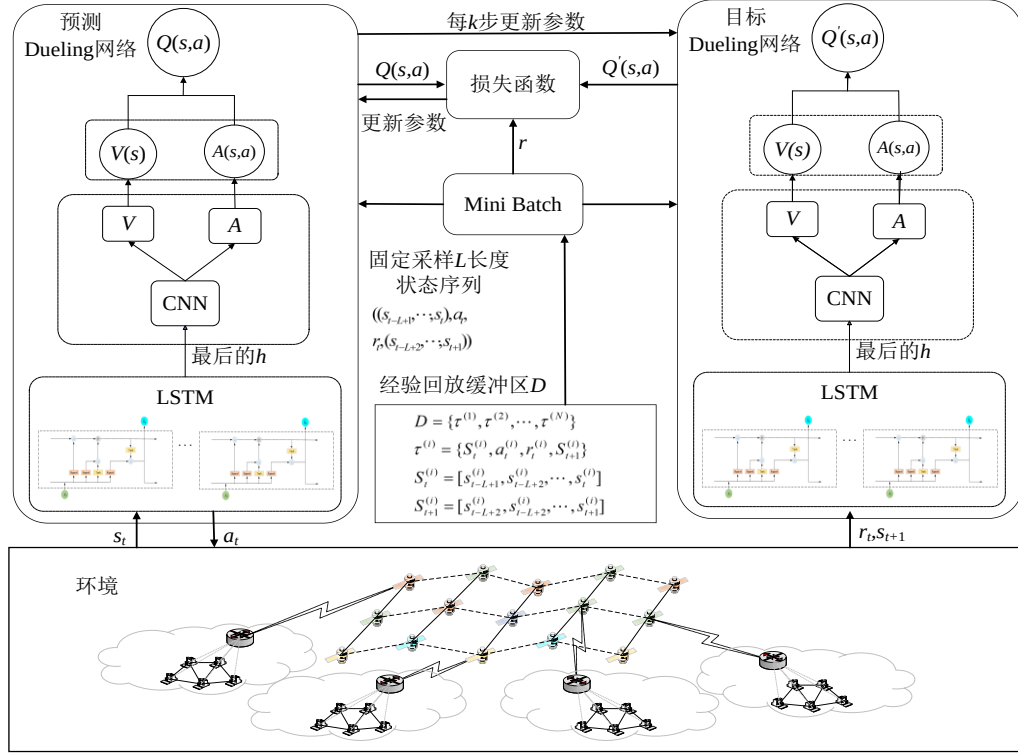


图3 LD3QN-IR网络框架

效状态和节点能源状态共同构成，定义为

$$S_t = \{A(t), B(t), \text{delay}(t), \text{RER}(t), \text{AoI}_{\text{est}}(t)\} \quad (16)$$

其中， $A(t)$ ， $B(t)$ ， $\text{delay}(t)$ ， $\text{RER}(t)$ 和 $\text{AoI}_{\text{est}}(t)$ 分别表示 t 时刻的加权邻接矩阵、链路剩余可用带宽矩阵、链路传输时延矩阵、节点剩余能量比例矩阵和端到端估计剩余信息年龄矩阵。

4.2 动作空间

在每个路由决策时刻，智能体从当前节点 i 的有效邻居集合中选择一个卫星作为下一跳。因此，时刻 t 节点 i 的可行动作空间记为 $\mathcal{A}_i(t)$ ，定义为：

$$\mathcal{A}_i(t) = \mathcal{N}_i(t) = \{j \in V | \chi_{ij}(t) = 1\} \quad (17)$$

动作 $a_t = j \in \mathcal{A}_i(t)$ 表示选择当前有效邻居节点 j 作为下一跳。受卫星轨道运动影响，不同时刻的有效邻居集合可能发生变化，因此动作空间随网络拓扑动态更新。本文每颗卫星最多维持4条有效星间链路，故有 $|\mathcal{A}_i(t)| \leq 4$

4.3 奖励函数

为协调信息时效性、链路拥塞、节点能源状态和路由前进性，本文构建多目标联合奖励函数。当数据包到达目的节点时给予奖励1；否则根据上述四项指标计算即时奖励：

$$r = \begin{cases} 1, & \text{Done} = 1 \\ w_1 \cdot [K_i - K_j] + w_2 \cdot \text{RER}_j - w_3 \cdot U_{i,j} - w_4 \cdot \text{AoI}_{i,j}, & \text{Done} = 0 \end{cases} \quad (18)$$

其中，Done表示当前路由任务是否到达目的节点，Done=1表示数据包已到达目标节点，否则Done=0， K_i ， K_j 分别表示当前节点 i 和候选节点 j 到目的节点的最短剩余跳数。引入 K_i 和 K_j 之差 $K_i - K_j$ 作为路由前进性评价项，当 $K_i - K_j < 0$

时，表明选择节点 j 后到目的节点的最短剩余跳数增加，该项产生负向反馈，从而抑制乒乓路由和不合理转发，反之则给予正向反馈。 w_1, w_2, w_3, w_4 分别表示路由前进性、节点剩余能量、链路利用率和信息年龄对应的权重系数，且



满足 $w_1 + w_2 + w_3 + w_4 = 1$ 。

4.4 算法伪代码

4.5 算法复杂度分析

为深入分析 LD3QN-IR 算法在业务负载与网络资源状态动态变化条件下的性能表现, 本文从收敛性、鲁棒性以及计算复杂度三个方面进行分析。

从收敛性角度来看, LD3QN-IR 在传统深度 Q 网络结构基础上融合了双 Q 学习机制与对决网络结构, 并引入长短期记忆网络以刻画网络状态的时序依赖特征。双 Q 学习通过将动作选择与动作评估过程分离, 有效缓解了 Q 值高估问题, 从而提高价值函数估计的准确性并改善训练过程的稳定性。与此同时, 对决网络结构将状态价值函数与动作优势函数进行解耦, 使网络能够更加精确地评估在不同网络状态下各路由动作的相对优劣, 提高策略学习效率。此外, 在状态空间设计中综合引入网络加权邻接矩阵、链路剩余可用带宽、节点间的传输时延、端到端剩余信息年龄、节点的剩余能量比例等状态特征, 并通过归一化处理降低不同指标量纲差异对训练过程的影响。在奖励函数构建过程中, 算法通过联合优化信息年龄、链路利用率、节点剩余能量比例和路由前进性等指标, 引导智能体在多目标约束条件下学习多目标权衡下的路由策略, 从而有助于提高策略学习过程的稳定性。

在鲁棒性方面, 低轨卫星网络中的业务负载、链路资源占用和节点能量状态会随着数据转发过程持续变化, 传统基于固定规则的路由策略难以根据网络运行状态及时调整^[17]。LD3QN-IR 在状态空间中联合引入链路剩余带宽、传输时延、信息年龄及节点剩余能量等信息, 使智能体能够感知业务负载和资源状态的变化, 并据此动态调整下一跳决策。同时, LSTM 通过对历史状态序列进行编码, 提取业务负载、链路资源及节点能源状态的时序相关信息, 从而增强模型在负

算法1 LD3QN-IR 伪代码

输入: 训练回合数 M , 每回合最大决策步数 N , 历史状态序列长度 L , 折扣因子 γ , 探索概率 ϵ , 学习率 ρ , 目标网络更新间隔 k 及软更新系数 τ 。

输出: 训练后的 LD3QN-IR 在线网络参数 θ 及对应的路由策略。

```

1. 随机初始化主网络参数  $\theta$ , 其中包括 LSTM 参数  $\theta^L$ , CNN 参数  $\theta^C$ ; 初始化目标网络参数  $\theta^- = \theta$ ; 经验回放池  $D$ ; 隐藏状态  $h_0$ ; 计数器  $t_d$ ;
2. for episode = 1 to  $M$  do
3.   初始化环境, 获得初始状态  $s_0$ ;
4.   构造初始序列  $S_0 = \{s_0, s_0, \dots, s_0\}$ , 长度为  $L$ ;
5.   for  $t = 1$  to  $N$  do
6.     根据当前拓扑  $G(t)$  构造有效邻居集合  $\mathcal{N}_i(t)$  和动作集合  $\mathcal{A}_i(t)$ 
7.     LSTM 时序特征提取:  $h_t = f_{\text{LSTM}}(S_t; \theta^L)$ 
8.     CNN 特征提取:  $z_t = f_{\text{CNN}}(h_t; \theta^C)$ 
9.     Dueling 结构计算 Q 值:  $Q(s_t, a) = V(z_t) + A(z_t, a) - \frac{1}{|\mathcal{A}|} \sum_{a'} A(z_t, a')$ 
10.    采用  $\epsilon$ -greedy 选择动作:  $a_t = \begin{cases} \text{从 } \mathcal{A}_i(t) \text{ 中随机选取动作, 依概率 } \epsilon \\ \arg \max_{a \in \mathcal{A}_i(t)} Q(s_t, a), \text{ 依概率 } 1 - \epsilon \end{cases}$ ,  $a_t$  为下一跳路由节点。
11.    执行  $a_t$ , 观察奖励  $r_t$  与下一状态  $s_{t+1}$ ;
12.    构造下一时刻序列:  $S_{t+1} = \{s_{t-L+2}, \dots, s_{t+1}\}$ 
13.    存储五元组到经验回放池:  $(S_t, a_t, r_t, S_{t+1}, done_t) \rightarrow D$ 
14.     $t_d = t_d + 1$ 
15.    if full( $D$ )
16.      随机采样小批量样本:  $(S_i, a_i, r_i, S'_i, done_i)$ 
17.      主网络选择下一动作:  $a_i^* = \arg \max_{a \in \mathcal{A}_i(t+1)} Q(S'_i, a | \theta)$ 
18.      目标网络计算目标 Q 值:  $y_i = r_i + \gamma(1 - done_i)Q(S'_i, a_i^* | \theta^-)$ 
19.      计算预测的当前 Q 值:  $Q(S_i, a_i | \theta)$ 
20.      定义损失函数:  $L_\theta = \frac{1}{N_b} \sum_i (y_i - Q(S_i, a_i | \theta))^2$ 
21.      梯度下降更新在线网络参数:  $\theta = \theta - \rho \nabla_\theta L$ 
22.      if  $t_d \bmod k = 0$ 
23.        更新目标网络参数:  $\theta^- = \tau \theta + (1 - \tau) \theta^-$ 
24.      end if
25.    end if
26.  end for
27. end for

```

载波动和资源状态变化条件下的决策稳定性。此外,联合奖励机制综合考虑信息年龄、链路利用率和节点剩余能量,使路由策略在关注传输性能的同时降低局部链路持续高负载和节点能源过度消耗的风险。

在计算复杂度方面,LD3QN-IR的主要计算开销来源于LSTM时序建模、CNN局部时序特征提取以及Dueling网络的动作价值估计。设历史状态序列长度为 L ,LSTM隐藏层维度为 h ,则LSTM的计算复杂度主要与序列长度、输入特征维度及隐藏层维度有关。本文在LSTM后采用一维卷积进一步提取局部时序关联特征,设卷积核大小为 k ,输入和输出通道数分别为 C_{in} 和 C_{out} ,则该卷积层的计算复杂度约为 $O(LC_{in}C_{out}k)$ 。本文设置 $L=4$ 、 $C_{in}=C_{out}=96$ 、 $k=3$,卷积后通过自适应平均池化得到固定维度特征,并输入Dueling网络完成动作价值估计。由于每个卫星节点仅需在有限的相邻节点中进行下一跳选择,因此在线决策阶段的计算开销总体可控。

在空间复杂度方面,训练阶段需要维护主Q网络与目标Q网络两组模型参数,并存储经验回放池中的交互样本。设网络层数为 M ,平均每层参数规模为 P ,则模型参数存储复杂度约为 $O(MP)$ 。此外,经验回放池需要存储 N 条历史交互样本,若单条样本维度为 d_s ,则其空间复杂度约为 $O(Nd_s)$ 。在实际部署阶段,由于不再需要经验回放机制且仅保留训练完成后的主Q网络进行推理,因此系统整体空间开销将显著降低,仅需存储单个网络模型参数即可完成实时路由决策。

5 实验结果与分析

5.1 仿真环境

为了评估LD3QN-IR路由算法在动态低轨卫星网络中的性能,本文构建了融合卫星轨道仿真、网络传输模拟与智能决策训练的联合实验平

表2 LD3QN-IR超参数设置

超参数	LD3QN-IR
折扣因子 γ	0.9
经验回放缓冲区大小 $ D $	$1e+4$
训练批次大小	128
软更新率 $\square$	0.001
奖励函数权重 ω_i	$\omega_1, \omega_2, \omega_3, \omega_4 = 0.35, 0.25, 0.2, 0.2$

台。仿真系统由Satellite Tool Kit (STK)、NS3网络模拟器和MATLAB深度学习框架协同组成。其中,STK用于构建卫星星座场景、传播卫星轨道,并周期性计算卫星位置、星间距离及链路可见性;NS3根据STK输出的拓扑快照动态建立或断开星间链路,模拟数据包产生、排队、转发和丢失过程,同时采集链路负载、剩余带宽、传输时延和节点能耗等网络状态;MATLAB负责LD3QN-IR模型的训练与策略推理,并将下一跳决策返回NS3执行。通过STK、NS3和MATLAB之间的周期性交互,实现卫星拓扑更新、网络状态采集、路由决策和数据转发的闭环仿真。

本文采用由66颗卫星组成的铱星星座,其中包含6个轨道面,每个轨道面部署11颗卫星,所有卫星轨道高度780 km,倾角为 86.4° ,极区边界南北纬 60° ,轨道接缝不建立链路。仿真过程中,每颗卫星最多维持4条ISL。在同一轨道面内,每颗卫星始终与前后相邻卫星建立链路;跨轨道面链路根据极区边界和反向缝位置周期性断开、重建或切换,从而形成随时间演化的网络拓扑。在网络结构参数设置方面,历史状态序列长度 L 设为4,LSTM采用单层结构,隐藏层维度为96。LSTM输出随后输入一维卷积层(Conv1D),其输入通道数和输出通道数均为96,卷积核大小为3,步长和填充均设为1,以保持卷积前后序列长度一致。

地面节点与卫星之间的接入链路带宽设置为1 Mbps,卫星节点之间的星间链路带宽设置为25 Mbps,数据包大小为1024字节。单次仿真持



续时间设置为 6000 s，拓扑更新时间间隔设置为 10 s，星间链路的最大通信距离设为 5000 km。每组实验独立重复 5 次，并取平均值作为最终结果。为了研究不同网络负载条件下路由算法的适应能力，本文随机生成多组源目的节点对，规模分别为 10、50、100 以及 200 个。随着源节点数量的增加，网络中的数据流密度逐渐提高，从而形成不同程度的链路竞争和路径拥塞。

5.2 仿真结果及分析

为了验证所提出的 LD3QN-IR 算法在低轨卫星网络中的综合性能优势，本文选取 LUR^[8] 算法，基于 Stackelberg 博弈的 DRL 方法^[18]（以下简称 Stackelberg）以及 RT-ATT-MAD2QN^[19] 算法进行了对比分析。上述方法分别代表 AoI 感知路由、博弈强化学习和多智能体强化学习等不同技术路线。

从三个关键维度进行量化评估：平均端到端时延、平均 AoI、平均电池能耗。相关实验结果图 4~图 6 所示，通过多指标对比可反映各算法在复杂动态环境下的性能差异。

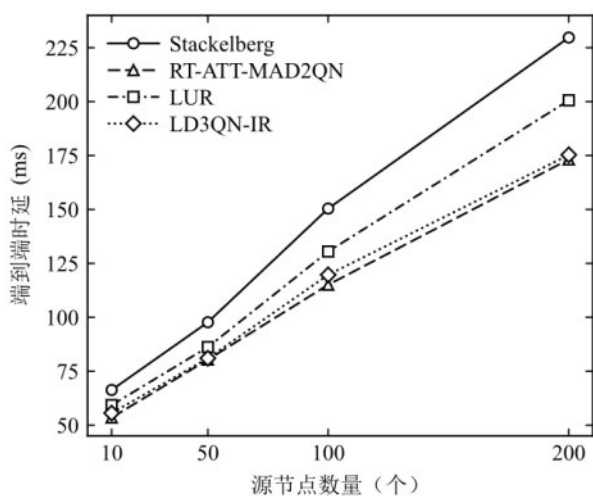


图 4 各算法的端到端时延

由图 4 可知，随着源节点数量由 $N=10$ 增加至 $N=200$ ，各算法的平均端到端时延均呈明显上升趋势，说明随着业务负载增加，链路竞争与排

队转发压力逐渐加剧。RT-ATT-MAD2QN 在各负载条件下均取得最低平均时延，体现了其在实时路由和传输时延优化方面的优势。LD3QN-IR 的时延性能始终处于次优水平，并优于 LUR 和 Stackelberg。以 $N=200$ 为例，LD3QN-IR 的平均端到端时延为 175.3 ms，仅较 RT-ATT-MAD2QN 高约 1.21%，但相比 LUR 和 Stackelberg 分别降低约 12.61% 和 23.68%。上述结果表明，LD3QN-IR 虽然同时考虑信息时效和节点能源状态，但仍能够保持接近时延优化方法的传输性能。

图 5 给出了不同业务负载下各算法的平均 AoI 变化情况。随着源节点数量增加，各算法的平均 AoI 均明显上升，其中 LUR 在各负载条件下均取得最低 AoI，说明面向信息新鲜度的专用路由方法能够有效降低状态更新过程中的信息陈旧程度。LD3QN-IR 的 AoI 性能始终处于第二位，并明显优于 RT-ATT-MAD2QN 和 Stackelberg。以 $N=200$ 为例，LUR 和 LD3QN-IR 的平均 AoI 分别为 993.6 ms 和 1014.5 ms，LD3QN-IR 仅较 LUR 高约 2.10%；与此同时，相比 RT-ATT-MAD2QN 和 Stackelberg，LD3QN-IR 的平均 AoI 分别降低约 21.75% 和 27.34%。

该结果说明，单纯以时延为目标并不能保证信息新鲜度的最优性，而 LD3QN-IR 在兼顾链路拥塞与更新频率调控的前提下，有效抑制了 AoI 的过快增长，体现出对时效性指标的显式建模与约束能力。

由图 6 可知，各算法的单位能耗均随源节点数量增加而上升。Stackelberg 在不同业务负载下始终具有最低能耗，体现了其在能源控制方面的优势。LD3QN-IR 的能耗始终处于第二位，并明显低于 LUR 和 RT-ATT-MAD2QN。以 $N=200$ 为例，LD3QN-IR 的单位能耗为 142.2 J/bit，仅较 Stackelberg 高约 4.41%，但相比 LUR 和 RT-ATT-MAD2QN 分别降低约 16.11% 和 21.57%。上述结果表明，通过在状态空间和奖励函数中显式考虑

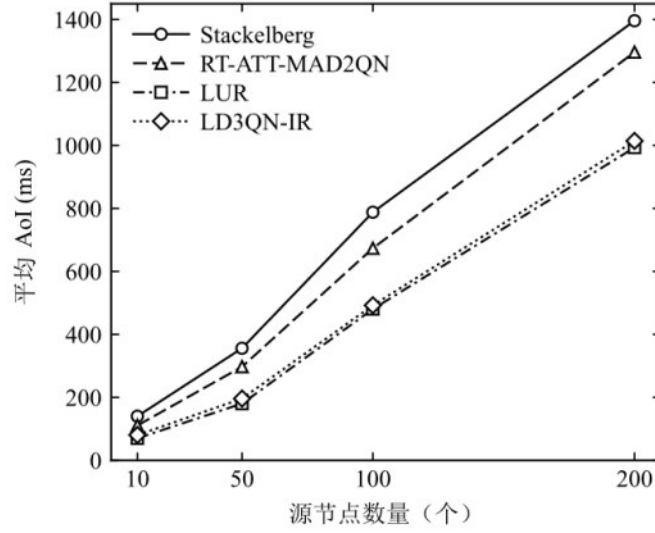


图5 各算法的平均信息年龄

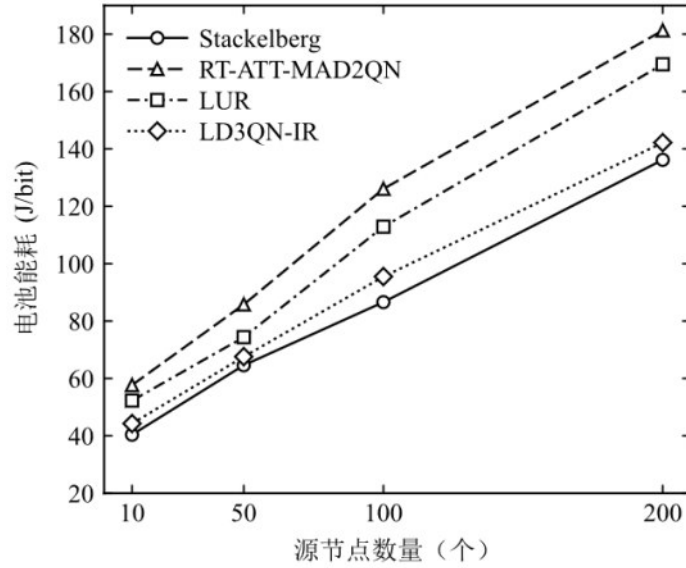


图6 各算法的电池能耗

节点剩余能量，LD3QN-IR 能够在维持较好传输性能和信息时效性的同时有效控制节点能源消耗。

为综合评价不同算法在端到端时延、AoI 和电池能耗方面的整体性能，本文首先以相同网络负载下各指标的最优值为基准，将三项指标转换为相对最优值接近度，并采用等权几何平均计算综合均衡得分，由于上述三项指标均为成本型指标，即数值越小表示性能越好，因此算法*i*在源

节点数量为*N*、指标*k*下的相对最优值接近度 $q_{i,k}(N)$ 定义为：

$$q_{i,k}(N) = 100 \times \frac{x_k^{\min}(N)}{x_{i,k}(N)} \quad (19)$$

$$x_k^{\min}(N) = \min_{j \in \Omega} x_{j,k}(N) \quad (20)$$

式中， $x_{i,k}(N)$ 表示算法*i*在源节点数量为*N*时指标*k*的原始值； $k \in \{D, A, E\}$ 分别表示端到端时延、平均 AoI 和单位能耗； Ω 为参与比较的算法集合。 $q_{i,k}(N)$ 的取值范围为 0-100，当其等于 100 时，表



示该算法在对应网络负载和指标上达到所有对比算法中的最优水平。

在此基础上,采用三项指标的等权几何平均计算综合均衡得分:

$$S_i(N) = [q_{i,D}(N)q_{i,A}(N)q_{i,E}(N)]^{\frac{1}{3}} \quad (21)$$

其中, $S_i(N)$ 表示算法 i 在源节点数量为 N 时的综合均衡得分, 得分越高, 表明该算法在传输时延、信息新鲜度和能量消耗之间具有越好的综合权衡能力。采用几何平均能够对某一指标明显偏低的情况施加更强惩罚, 从而避免单项性能优势完全掩盖其他指标的不足。

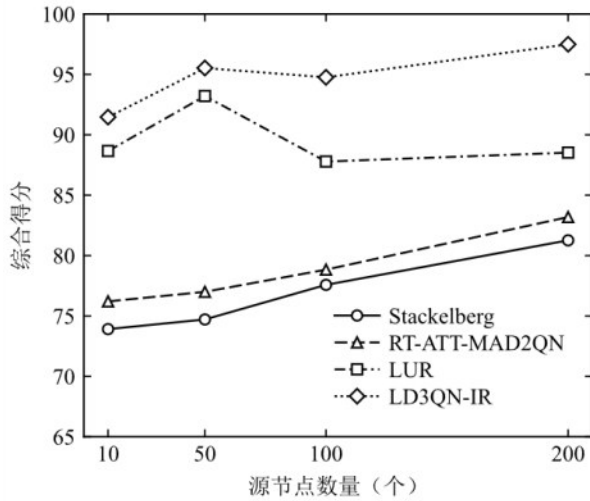


图7 基于时延、AoI和能耗等权几何平均得到的综合均衡得分

如图7所示, LD3QN-IR在 $N=10$ 、50、100、200时的得分分别为91.46、95.53、94.76和97.50, 均为所有算法最高; 相比次优LUR分别提高2.80、2.32、6.97和8.97分, 表明LD3QN-IR具有更优的多目标综合均衡性能。

5.3 消融实验

为验证LSTM时序建模及完整LD3QN-IR结构对路由性能的贡献, 在与主实验相同的实验设置下, 设置D3QN、LSTM-D3QN和LD3QN-IR三组模型进行消融对比。如表3所示。

由表3可知, 在D3QN基础上引入LSTM后,

模型在不同网络负载条件下的平均端到端时延、平均AoI和单位能耗均有所降低。在轻载和中等负载场景下, LSTM对时延和AoI的改善相对有限, 但对节点能耗具有较为明显的优化作用。例如, 在 $N=10$ 时, LSTM-D3QN相较D3QN的平均端到端时延和平均AoI分别降低0.84%和0.48%, 单位能耗降低15.27%; 当负载增加至 $N=200$ 时, 三项指标分别降低14.21%、11.00%和11.73%。这表明, 在网络负载较低时, 当前网络状态已能够提供较充分的传输决策信息, 而随着业务负载增加以及拓扑和资源状态变化更加复杂, 历史状态中包含的时序演化信息能够为下一跳决策提供更有效的辅助。

表3 不同网络负载下的消融实验结果

模型	指标	$N=10$	$N=50$	$N=100$	$N=200$
D3QN	Delay/ms	56.82	83.70	122.51	227.68
	AoI/ms	81.42	200.24	499.85	1229.93
	Energy/(J/bit)	55.41	74.52	108.32	178.80
LSTM-D3QN	Delay/ms	56.34	82.01	120.97	195.32
	AoI/ms	81.03	197.72	495.83	1094.67
	Energy/(J/bit)	46.95	69.26	97.46	157.83
LD3QN-IR	Delay/ms	55.6	81.0	119.7	175.3
	AoI/ms	80.4	196.3	492.6	1014.5
	Energy/(J/bit)	44.30	67.60	95.50	142.20

在LSTM-D3QN基础上进一步引入CNN后, 完整LD3QN-IR在三项指标上均获得进一步改善。当 $N=200$ 时, LD3QN-IR相较LSTM-D3QN的平均端到端时延、平均AoI和单位能耗分别降低10.25%、7.32%和9.90%; 相较基础D3QN模型则分别降低23.01%、17.52%和20.47%。结果表明, LSTM能够有效利用连续历史网络状态中的时序依赖信息, 而CNN通过沿时间维进一步提取相邻状态之间的局部时序特征, 可增强复杂动态网络状态下的特征表示能力。两者协同作用使LD3QN-IR在高负载及动态拓扑条件下获得更好的路由决策性能, 并在传输时延、信息新鲜度

和节点能耗之间取得更优的综合表现。

6 结束语

针对动态低轨卫星网络中信息时效性、链路拥塞与节点能源状态难以协同优化的问题, 本文提出 LD3QN-IR 路由算法。该算法通过联合表征动态拓扑、链路资源、AoI 和节点剩余能量, 并结合 LSTM - CNN 时序特征提取与 Dueling Double DQN, 实现面向动态候选下一跳的协同路由决策。基于 STK、NS3 与 MATLAB 的联合仿真结果表明, 在本文采用的等权综合评价指标下, LD3QN-IR 在不同网络负载下均取得最高综合均衡得分; 消融实验进一步表明, LSTM 时序建模和 CNN 局部时序特征提取有助于改善高负载条件下的路由性能。

本文实验主要基于铱星类轨道配置和固定拓扑更新间隔, 尚未系统考察不同轨道参数及拓扑更新粒度对算法性能的影响。后续将进一步评估算法在不同星座配置、拓扑更新间隔和业务模式下的泛化能力, 并探索基于多智能体强化学习的分布式协同路由机制, 以及队列控制、功率分配与链路调度的联合优化。

参考文献:

- [1] SU Y, LIU Y, ZHOU Y, et al. Broadband LEO satellite communications: architectures and key technologies[J]. IEEE Wireless Communications, 2019, 26(2): 55-61.
- [2] KAHRAMAN I, KOSE A, KOCA M, et al. Age of information in Internet of Things: a survey[J]. IEEE Internet of Things Journal, 2023, 11(6): 9896-9914.
- [3] GAO Z, LIU A, HAN C, et al. Non-orthogonal multiple access-based average age of information minimization in LEO satellite-terrestrial integrated networks[J]. IEEE Transactions on Green Communications and Networking, 2022, 6(3): 1793-1805.
- [4] YANG L, ZHANG H, QI Y, et al. Available energy routing algorithm considering QoS requirements for LEO satellite network[J]. Computer Communications, 2024, 217: 87-96.
- [5] TSAI K C, FAN L, WANG L C, et al. Multi-commodity flow routing for large-scale LEO satellite networks using deep reinforcement learning[C]//Proceedings of 2022 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2022: 626-631.
- [6] LIN C Y, LIAO W. Energy-aware age of information minimization for Internet of Things in NOMA-based LEO satellite networks[C]//Proceedings of 2024 IEEE 99th Vehicular Technology Conference. Piscataway: IEEE Press, 2024: 1-5.
- [7] WANG Q, LIANG X, ZHANG H, et al. AoI-aware energy efficiency resource allocation for integrated satellite-terrestrial IoT networks[J]. IEEE Transactions on Green Communications and Networking, 2025, 9(1): 125-139.
- [8] LI Y, XU Y, ZHANG Q, et al. Age-optimized multihop information update mechanism on the LEO satellite constellation via continuous time-varying graphs[J]. IEEE Internet of Things Journal, 2023, 10(8): 7189-7203.
- [9] JING Y, YANG Z, ZHAO Y, et al. Energy-efficient routing based on a genetic algorithm for satellite laser communication [J]. Optics Express, 2023, 31(5): 8682-8695.
- [10] LIU J, ZHAO B, 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, 2020, 8(4): 2872-2884.
- [11] DONG F, SONG J, ZHANG Y, et al. DRL-based load-balancing routing scheme for 6G space-air-ground integrated networks[J]. Remote Sensing, 2023, 15(11): 2801.
- [12] CHEN Y, CAO H, WANG L, et al. Deep reinforcement learning-based routing method for low earth orbit mega-constellation satellite networks with service function constraints [J]. Sensors, 2025, 25(4): 1232.
- [13] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [14] 温京龙, 张怡, 魏德宾, 等. 基于长短期记忆深度 Q 网络的卫星网络多业务路由算法[J]. 电信科学, 2026, 42(5): 48-59.
- [15] WEN J L, ZHANG Y, WEI D B, et al. Multi-service routing algorithm for satellite networks based on deep Q-network with long short-term memory[J]. Telecommunications Science, 2026, 42(5): 48-59.
- [16] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016: 2094-2100.
- [17] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]//Proceedings of the International Conference on Machine Learning. New York:



- PMLR, 2016: 1995-2003.
- [17] XU G, ZHAO Y, RAN Y, et al. Spatial location-aided fully-distributed dynamic routing for large-scale LEO satellite networks[J]. IEEE Communications Letters, 2022, 26(12): 3034-3038.
- [18] LI S, WU Q, WANG R, et al. Efficient multipath differential routing and traffic scheduling in ultra-dense LEO satellite networks: a DRL with Stackelberg game approach[J]. IEEE Transactions on Mobile Computing, 2025, 24(11): 12424-12440.
- [19] LIAO H, ZHANG X, ZHOU J, et al. Real-time routing design for LEO satellite networks: an enhanced multi-agent DRL approach[C]//Proceedings of 2024 IEEE/CIC International Conference on Communications in China Workshops. Piscataway: IEEE Press, 2024: 547-552.

[作者简介]



王欣睿，出生于2001年，男，硕士，大连大学，主要研究方向为队列管理、网络路

由等。



魏德宾，出生于1978年，男，博士，大连大学，副教授，主要研究方向为空间信息网络传输技术、流量工程和网络优化。

杨晨，出生于2001年，男，大连大学，硕士研究生，主要研究方向为网络路由、网络流量预测等。；



潘成胜，出生于1962年，男，博士，南京邮电大学，教授，主要研究方向为一体化智能网络流量理论与关键技术。