



研究与开发

# 基于长短期记忆深度Q网络的卫星网络多业务路由算法

温京龙<sup>1</sup>, 张怡<sup>1</sup>, 魏德宾<sup>1</sup>, 潘成胜<sup>2</sup>

(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:** 为了解决卫星网络中多业务服务质量保障与负载均衡的协同优化难题, 提出了一种融合长短期记忆网络和深度Q网络的智能多业务路由算法。该算法把长短期记忆深度Q网络作为智能体部署在当前卫星节点, 将当前节点与周围节点的链路时延、带宽、丢包率、流量、网络拓扑、业务类型作为网络状态输入智能体进行训练, 输出动作是下一跳节点, 使用带宽、时延、丢包率的加权和与最大链路带宽利用率的加权和作为奖励函数调整动作。智能体训练收敛后, 卫星进行多业务传输。通过仿真实验和性能评估, 研究结果表明所提出的算法在不同性能指标上均取得了显著成效, 并且在平衡网络负载方面表现优异。

**关键词:** 卫星网络; 深度强化学习; 负载均衡; 多业务路由

**中图分类号:** TP393; TN297

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.DXKX250576

## Multi-service routing algorithm for satellite networks based on deep Q-network with long short-term memory

Wen Jinglong<sup>1</sup>, Zhang Yi<sup>1</sup>, Wei Debin<sup>1</sup>, Pan Chengsheng<sup>2</sup>

1. College of Information Engineering, Dalian University, Dalian 116622, China

2. National Local Joint Engineering Research Center for Communication and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

**Abstract:** To address the challenge of collaborative optimization between multi-service quality of service (QoS) guarantee and load balancing in satellite networks, an intelligent multi-service routing algorithm integrating long short-term memory with deep Q-Network was proposed. The algorithm deployed LSTM-DQN (LDQN) as an agent at the current satellite node, where the network state input to the agent includes link latency, bandwidth, packet loss rate, traffic, network topology and service type between the current node and its neighboring nodes. The output action was the next-hop node selection for the current node. The reward function, designed to guide action adjustment, combined the weighted sum of bandwidth, latency, and packet loss rate with the weighted sum of maximum link bandwidth utilization. After the agent's training convergence, multi-service transmission was performed in the satellite network.

收稿日期: 2025-09-30; 修回日期: 2026-01-21

通信作者: 潘成胜, panacs@sina.com

基金项目: 国家自然科学基金资助项目 (No.61931004)

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

Simulation experiments and performance evaluations demonstrate that the proposed algorithm achieves significant improvements across various performance metrics while exhibiting outstanding load balancing capabilities.

**Key words:** satellite network, deep reinforcement learning, load balancing, multi-service routing

## 0 引言

在数字化浪潮的推动下，地面通信基础设施不断完善，用户对通信服务的期望值不断提升，不仅要求实现毫秒级的超低时延传输和 Gbit/s 级的超高带宽接入，更期望获得稳定可靠、安全可控的端到端服务质量保障。卫星网络凭借其独特的空间优势，能够有效突破地理屏障的限制，实现全球范围内的无缝覆盖，而且不受自然灾害和突发事件的影响，能够确保通信服务的持续性和稳定性，并逐步从传统的补充性角色向基础性通信手段转变。此外，随着新一代互联网应用的快速发展，网络业务类型日趋多元化，从传统的语音、数据业务扩展到实时视频传输、物联网通信、云计算服务等新兴应用场景，不同的业务类型具有不同的 QoS (quality of service, 服务质量) 需求，这就要求卫星网络路由必须突破传统单一业务模式的限制，向支持多业务差异化服务质量保障的方向发展。

在保障多业务 QoS 需求方面，文献[1]在软件定义卫星网络下提出了基于拓扑感知的多业务路由 (topology awareness based on demand routing, TAODR) 算法，该算法结合了软件定义网络 (software defined network, SDN) 架构和链路属性量化方法，通过构建拓扑感知模型和多业务策略模型、考虑动态链路属性，实现了对动态拓扑及多元业务的自适应路由。文献[2]提出了基于业务调度的软件定义卫星网络混合路由算法，根据业务的不同 QoS 需求，提出了轨道内优先转发的低复杂度路由算法和基于链路状态的加权路由算法。该算法降低了卫星网络的时延和丢包率，且提高了吞吐量，保障了不同业务的 QoS 需求，但

是该算法对星上处理能力要求较高。文献[3]提出了一种区间 Type-2 模糊集路由策略，对软件定义卫星网络中的多服务质量路由问题进行建模，然后进行模糊决策，满足多业务需求。文献[4]提出了一种 QoS 路由优化方案，该算法通过建立多维度业务分类模型，将不同特征的数据流按照其服务质量需求进行精细化区分。然而，在保障业务 QoS 需求的同时，该方案导致网络流量在特定链路上过度集中，其他可用链路却处于低利用率状态，这种资源分配的不均衡不仅降低了网络的整体吞吐性能，还造成了带宽资源的浪费。文献[5]提出了一种多目标遗传的卫星网络多业务 QoS 路由算法，针对不同业务类型，设计了不同的 QoS 目标函数，并通过加权系数进行综合优化，但该算法的计算复杂度和资源消耗较大。文献[6]提出一种基于优先级的 SDRA (software defined routing algorithm, 软件定义路由算法)，针对不同轨道卫星节点，提出了高轨道节点路由策略和低轨道节点路由策略，但是该算法仅针对单一业务场景进行优化，未能充分考虑不同类型业务对传输特性的差异化需求。

针对巨型星座多业务路由问题，文献[7]提出分层聚类多准则路由，通过飞行方向-纬度联合聚类降低拓扑管理复杂度，并通过定义属性一致性准则实现无权重分配的多业务 QoS 保障。文献[8]提出将 SDN 集中控制与多维链路评估模型结合，通过业务感知的权重矩阵和瓶颈节点抑制机制，在保障多业务 QoS 需求的同时，实现低复杂度、高稳定性的星间路由。文献[9]结合 SDN、AI 技术和模糊逻辑等优势，提出了基于模糊卷积神经网络的多任务路由方法，具有较好的流量控制性能和寻路灵活性。



尽管在混合整数规划和启发式算法驱动的卫星网络多业务路由研究中已取得进展,但仍然面临着一些待解决的挑战与局限性。例如,在考虑网络多业务路由时,上述方法通常是基于历史数据和模型进行优化,对于快速变化的网络状态可能反应不及时。另外,启发式算法往往侧重于局部最优解,可能在全局范围内无法找到最优解。而对于大规模网络,混合整数规划和启发式算法还可能受到计算资源和时间的限制,难以在合理的时间内找到最优解。这使得在实际应用中,算法的效率和可扩展性成为制约因素。为了提高网络性能和适应多样化应用需求,大量研究通过结合监督学习、强化学习等人工智能方法,致力于打破传统路由算法的限制,提供更灵活、智能的SDN路由方案。

基于深度强化学习的路由算法,文献[10]提出一种基于深度强化学习的智能可靠路由方案,在软件定义卫星网络架构下学习自动选择路由路径的策略,旨在提高传输过程的效率和可靠性。文献[11]提出一种融合图神经网络的卫星网络动态路由算法,不仅降低传输时延和丢包率,还提升了卫星网络吞吐量。为了解决网络负载不均衡的问题,文献[12]提出了一种基于深度强化学习的低轨(low earth orbit, LEO)卫星负载均衡路由算法,降低了路径的平均队列利用率,有效地平衡了网络负载。文献[13]提出了一种基于软件定义网络架构和人工智能算法的低轨卫星网络路由算法,该算法通过机器学习算法优化链路权重,实现网络流量的智能调度,从而降低链路的最大负载,提升网络性能。文献[14]提出了一种基于深度强化学习的状态感知路由算法,该算法中每个节点考虑两跳邻居范围内的链路状态,输出最优下一跳节点。文献[15]提出了一种用于大规模卫星网络的分布式路由算法,该算法能最大限度地缩短平均传送时间。对于超大型卫星星座来说,实现全局路由并保证卫星间传输性能一直

是一个巨大的挑战。针对这一问题,文献[16]提出了一种基于深度强化学习的鲁棒路由策略。该策略提升了信息传输的有效性,对干扰的鲁棒性更强。此外,文献[17]基于Q-learning强化学习模型规划业务流路径,提出了Q-routing路由算法,而翟凡妮<sup>[18]</sup>提出了一种基于深度Q网络(deep Q-network, DQN)的多业务流QoS路由算法(multi-service QoS routing method based on deep Q-network, MDQN),上述方法分别从多路径和多业务角度提升了网络性能。

综上所述,目前国内外针对业务服务质量保障和基于深度强化学习的卫星路由算法已取得显著进展,但在设计思路仍存在明显局限性,其优化目标往往局限于单一业务场景,难以有效满足多业务并行的复杂网络环境。并且部分采用遗传算法计算路由,虽然在一定程度上提升了路由选择的灵活性,但往往伴随着较高的计算复杂度和星上处理开销。同时现有算法普遍缺乏对网络整体拥塞状况的动态感知能力,加剧了网络负载的不均衡性。因此,亟须设计一种具备多目标优化能力的路由算法,能够在保障差异化服务质量需求的同时,实现网络负载均衡,从而提升卫星网络的整体性能和服务质量。

为了有效解决卫星网络中多业务服务质量保障与负载均衡的协同优化难题,本文提出了一种基于长短时记忆深度Q网络的智能多业务路由算法(a routing algorithm based on long short-term memory deep Q-network, LDQN-RA)。该算法综合考虑链路时延、带宽和丢包率建立多业务QoS路由优化模型。智能体在路由节点收集周围卫星节点的链路时延和负载状态信息,在周围节点中根据奖励函数选择一个节点作为下一跳路由节点。智能体训练收敛后,选择最优路径进行分布式转发。

## 1 问题建模

将卫星运动周期进行切片,在每个切片中采

用虚拟拓扑方法对一个时间切片内网络模型的路由进行分析研究。假设将卫星运行周期  $T$  进行等间隔划分, 将其分割为  $n$  个连续的时间片段, 记为  $(t_0=0, t_1], (t_1, t_2], (t_2, t_3], \dots, (t_{n-1}, t_n=T]$ 。每个时间片段内的网络拓扑结构被视为静态不变, 而拓扑的动态变化, 如链路切换和节点连接关系的改变仅发生在各个时间片段的边界点  $t_0, t_1, \dots, t_n$  处。基于这种假设, 卫星网络的拓扑特征可以用图  $G(V, E)$  来表示, 其中, 节点集  $V = \{1, 2, \dots, N\}$  表示网络中共有  $N$  颗卫星, 边集  $E$  则描述了节点间的连接关系, 用  $e_{i,j} \in E$  表示节点  $i$  到  $j$  的链路, 其中,  $i, j \in V$ 。

根据业务 QoS 的需求将业务类型分为以下 3 个类别: 时延敏感型业务, 需要网络保持较低的时延, 端到端时延阈值  $D_{\max}$ ; 带宽敏感型业务, 需要网络保持较大的带宽, 最小带宽需求  $B_{\min}$ ; 丢包率敏感型业务, 需要网络保持较低的丢包率, 最大丢包率  $\text{Loss}_{\max}$ 。

设  $p(s, d)$  表示从源节点  $s$  到目的节点  $d$  的路径, 路径时延表示如下:

$$D_{p_{sd}}(t) = \sum_{e_{i,j} \in p(s,d)} d_{ij}(t) \quad (1)$$

其中,  $d_{ij}(t)$  为链路  $e_{i,j}$  的时延。源节点  $s$  到目的节点  $d$  路径剩余带宽表示如下:

$$B_{p_{sd}}(t) = \min_{e_{i,j} \in p(s,d)} b_{ij}(t) \quad (2)$$

其中,  $b_{ij}(t)$  为链路  $e_{i,j}$  的剩余带宽。源节点  $s$  到目的节点  $d$  路径丢包率表示如下:

$$L_{p_{sd}}(t) = 1 - \prod_{e_{i,j} \in p(s,d)} (1 - \text{Loss}_{ij}(t)) \quad (3)$$

其中,  $\text{Loss}_{ij}(t)$  为链路  $e_{i,j}$  的丢包率。

本文提出了一种多业务 QoS 路由优化模型, 综合考虑了时延、带宽和丢包率 3 个服务质量指标。为了进一步平衡网络负载, 模型特别引入了最大链路带宽利用率 RU。将最大链路带宽利用率与多业务 QoS 保障机制有机融合, 构建了一个多维度的优化目标函数, 其数学模型为:

$$\begin{aligned} & \max \omega_1 \times (\alpha_1 \times \text{Delay}_{sd} + \alpha_2 \times \text{Bw}_{sd} + \alpha_3 \times \text{Loss}_{sd}) + \\ & \omega_2 \times (1 - \text{RU}) \\ & \text{s.t. } \text{Delay}_{sd} \leq D_{\max}, \\ & \text{Bw}_{sd} \geq B_{\min}, \\ & \text{Loss}_{sd} \leq \text{Loss}_{\max}, \\ & \text{RU} = \max_{e_{(i,j)} \in E} \left\{ \frac{\text{bw}_{ij}^{\text{used}}}{\text{bw}_{ij}} \right\} < 0.9, \quad e_{i,j} \in p(s, d) \end{aligned} \quad (4)$$

其中,  $\alpha_i (i=1, 2, 3)$  表示路径业务类别的优先系数, 且  $\alpha_1 + \alpha_2 + \alpha_3 = 1$ , 用于量化不同业务类型对各类服务质量指标的差异化需求。这些优先系数反映了特定业务对时延、带宽、丢包率 QoS 指标的敏感程度和重要程度, 其取值因业务类型而异。 $\text{Delay}_{sd}$  和  $\text{Loss}_{sd}$  分别为按成本型变量归一化后的路径时延和路径丢包率,  $\text{Bw}_{sd}$  为按效益型变量归一化后的路径剩余带宽。 $\omega_1, \omega_2$  为动态权重, 可根据网络拥塞程度调整。

## 2 基于 LDQN 的多业务路由算法

LDQN 在 DQN 的基础上进行了改进, 通过整合循环神经网络 (recurrent neural network, RNN) 的隐藏层机制, 实现了对历史状态信息序列的自动存储和处理。DQN 单帧输入丢失历史信息, 可能会导致策略震荡或收敛困难。在处理当前状态时, RNN 可以自动调用之前存储的历史状态信息, 避免了传统方法中需要重复处理历史数据的计算开销, 从而大幅减少了训练时间。同时, 通过捕捉状态之间的时序依赖关系, RNN 增强了算法对状态转移模式的学习能力, 提高了策略优化的效果。这种架构设计不仅降低了计算复杂度, 还通过更好地利用历史信息来提升决策质量。

长短期记忆 (long short-term memory, LSTM) 网络是一种特殊设计的循环神经网络, 可以用于解决传统 RNN 在处理长序列数据时的长期依赖问题。LSTM 记忆单元 cell 的内部结构如图 1 所示。

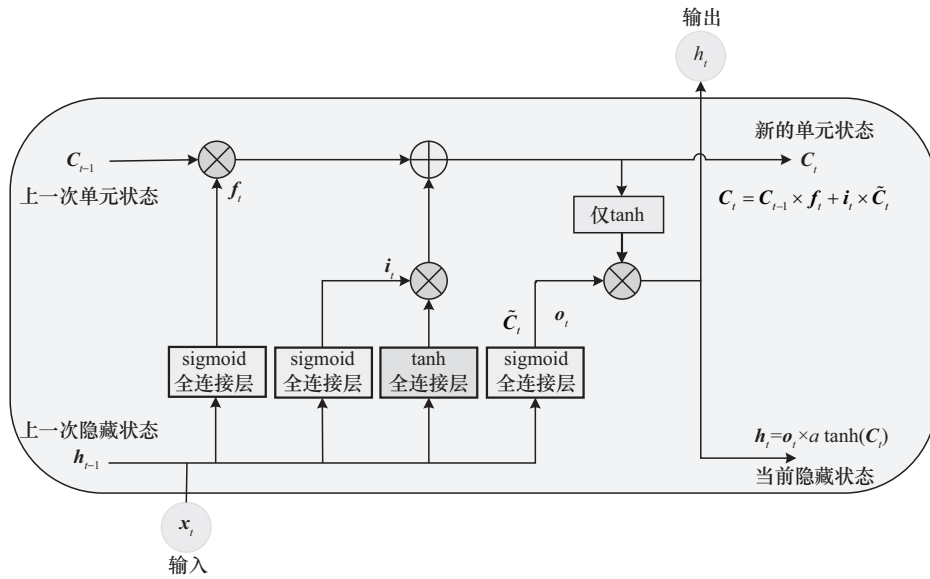


图1 LSTM记忆单元cell的内部结构

对于每个时间步  $t$ , LSTM 执行以下操作: 首先计算遗忘门  $f_t$ 、输入门  $i_t$  和输出门  $o_t$ 。

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (6)$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (7)$$

其中,  $\sigma$  是 sigmoid 函数,  $W$ 、 $b$  分别表示全连接层的权重和偏置,  $h_{t-1}$  表示上一次的隐藏状态,  $x_t$  表示当前时间步的输入。

其次, 计算候选记忆单元值  $\tilde{C}_t$ 。

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \quad (8)$$

然后结合遗忘门和输入门的输出, 更新记忆单元  $C_t$ 。

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (9)$$

其中,  $C_{t-1}$  是上一次单元状态。最后计算当前时间步的隐藏状态:

$$h_t = o_t \times \tanh(C_t) \quad (10)$$

本文的 LSTM 网络采用 2 层堆叠结构, 双层 LSTM 网络如图 2 所示。底层 LSTM1 先提取基础的序列特征, 为上层提供预处理后的中间信息。上层 LSTM2 同时结合原始输入和底层特征, 能更全面地捕捉序列的长短期依赖关系, 提升对复

杂序列模式的识别精度。

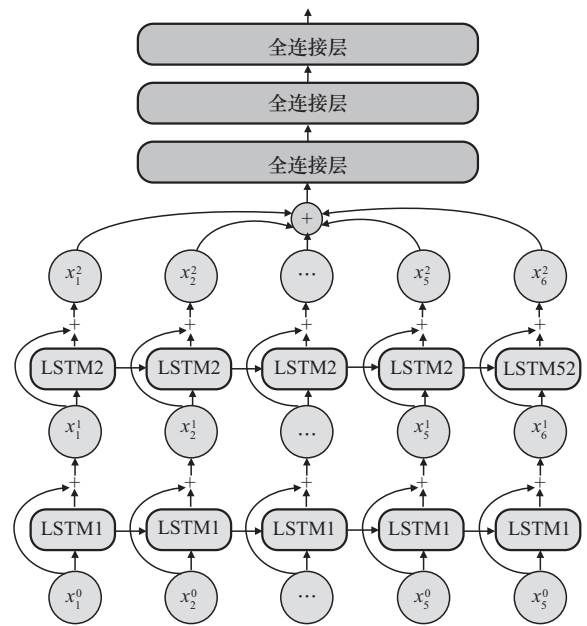


图2 双层 LSTM 网络

本文将 LSTM 作为核心组件整合到 DQN 的 Q 值网络和目标网络结构中。LSTM 网络通过其独特的门控机制, 能够有效捕捉和处理具有时序依赖性的状态数据, 实现对历史状态信息的长期记忆和选择性遗忘。这一特性使算法在卫星网络场景中表现出显著优势: 当某节点状态短暂不可见

时, LSTM可基于历史数据推测其当前负载。在训练过程中, LSTM网络通过迭代更新不断优化参数, 逐步学习到状态-动作值函数的准确估计, 从而在保持较高学习效率的同时, 显著提升了算法对序列决策问题的处理能力。LDQN-RA结构如图3所示。

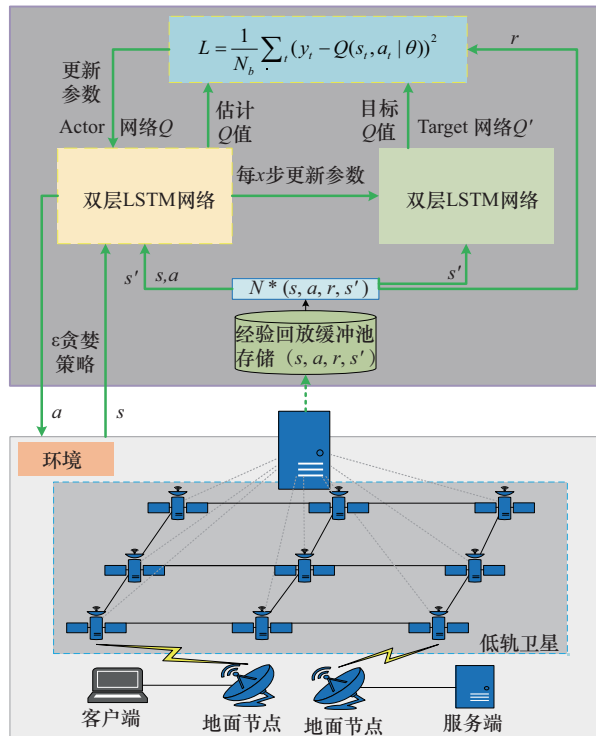


图3 LDQN-RA结构

状态空间  $S$ : 网络状态空间包括通信网络的各种性能指标和流量要求。  $s_t \in S$  是在时间片  $[t - \Delta t, t)$  采样得到的网络状态, 包含链路带宽  $B_t$ 、链路时延  $D_t$ 、链路丢包率  $L_t$ 、流量需求  $F_t$ 、网络拓扑  $V_t$ 、业务类型  $P_t$ 。

$$s_t = (D_t, B_t, L_t, F_t, V_t, P_t) \quad (11)$$

其中,  $B_t = (\tilde{B}_t(e_{i,j}))_{N \times N}$ 、 $D_t = (\tilde{D}_t(e_{i,j}))_{N \times N}$ 、 $L_t = (\tilde{L}_t(e_{i,j}))_{N \times N}$ ,  $\forall e_{i,j} \in E$  分别表示在时间片  $[t - \Delta t, t)$  内的网络节点间归一化的链路带宽、链路时延和链路丢包率。  $F_t = (\tilde{F}_t(s w_{src}, s w_{des}))_{N \times N}$ ,  $\forall s w_{src}, s w_{des} \in V, s w_{src} \neq s w_{des}$  是控制器在时间片  $[t -$

$\Delta t, t)$  内收集到的流量需求信息。其中,  $\tilde{F}_t(s w_{src}, s w_{des})$  表示源-目的节点对  $(s w_{src}, s w_{des})$  之间, 归一化处理后的流量传输需求。

动作空间  $A$ : LDQN 智能体的目标是将多维信息路由特征映射到业务 QoS 路由, 即 LDQN 根据网络环境所采取的动作是对节点间不同业务的最优下一跳  $a_t$ 。动作空间的大小等于邻接节点的数目。

奖励函数: LDQN 奖励  $r_t$  来评价路由策略  $a_t$  的有效性, 并将其反馈给智能体以进一步改进策略。不同的动作通常意味着不同的奖励。

$$r_t = \omega_1 \times (\alpha_1 \times \text{Delay}_{sd} + \alpha_2 \times \text{Bw}_{sd} + \alpha_3 \times \text{Loss}_{sd}) + \omega_2 \times (1 - \text{RU}) \quad (12)$$

本文使用 LDQN 来训练智能体并做出最优的路由决策。当某个卫星接收到卫星服务请求时, 当前卫星 (智能体) 的状态 (输入) 由周围卫星的状态信息确定, 动作 (输出) 是选择其中一个卫星作为下一跳路由。此时, 在 DQN 中,  $Q(s, a)$  通过估计  $Q$  值网络, 根据当前卫星状态  $s$  和当前路由决策  $a$  进行计算。计算出  $Q(s, a)$  后, 智能体采取使该值最大化的动作, 并选择具有最高  $Q$  值的卫星作为服务请求的下一跳。然后, 所选的卫星更新其状态信息, 并将服务请求发送到路由中的下一跳, 直到到达目的地。

在训练开始时, 首先对经验回放池、 $Q$  值网络和目标网络进行参数初始化。训练以迭代方式进行, 每次迭代随机选取一组源节点和目标节点作为起点和终点。其中, 源节点作为智能体的初始状态启动路由决策过程。在决策机制设计上, 算法采用动态贪心策略来平衡探索与利用: 在训练初期倾向于探索环境, 随着训练的推进逐步转向选择  $Q$  值最大的最优决策。每次决策执行后, 智能体会基于预设的奖励函数计算即时奖励, 并将  $(s_t, a_t, s_{t+1}, r)$  存储在经验回放池中以备。  $Q$  值网络通过梯度下降法实时更新权重参数, 而目



标网络则采用周期性更新策略，每隔  $d$  个时间步从估计网络同步参数，这种设计有效提高了训练的稳定性。当智能体到达目标节点时，单次路由过程结束，智能体会计算该路径的累积奖励值。为了充分探索网络环境，训练过程会重复进行多次，每次随机生成新的源-目标节点对，通过大量迭代使智能体逐步学习到最优路由策略。

在路由算法的决策执行阶段，首先利用训练完成的决策模型，通过输入源节点、目的节点以及当前网络拓扑结构等信息来初始化路由路径。决策过程采用迭代式搜索策略：在每个决策点，智能体基于网络环境的感知和分析，运用学习到的决策策略选择最优的下一跳节点。选定下一跳节点后，智能体会将其动态添加到当前路由路径中，同时将下一跳节点作为当前节点。随后，算法会进行目标检测，判断当前节点是否为目的节点。若到达目的节点，则终止决策过程并输出最终确定的路由路径；若未到达目的节点，则继续执行下一轮决策迭代，直至找到满足条件的完整路径。

在 LDQN 算法中， $Q$  值网络通过多层非线性变换，将高维连续状态映射为对应动作的价值估计。 $Q$  值网络使用均方误差进行更新，该误差是通过 TD 偏差计算得出的。TD 偏差是使用目标  $Q$  网络和在线  $Q$  网络计算得出的。

TD 偏差的计算式如下：

$$L = \frac{1}{N_b} \sum_t (y_t - Q(s_t, a_t | \theta))^2 \quad (13)$$

其中， $N_b$  是随机采样的容量， $\theta$  是  $Q$  值网络的参数， $y_t$  表示目标  $Q$  值，它代表在目标网络中计算得到的在下一个状态  $s_{t+1}$  中通过计算动作  $a_{t+1}$  获得的  $Q$  值，而在下一个状态  $s_{t+1}$  中产生最大目标  $Q$  值的动作将作为下一个动作  $a_{t+1}$ 。  $y_t$  的计算式如下：

$$y_t = r_t + \gamma \max_{a_{t+1}} \hat{Q}(s_{t+1}, a_{t+1} | \hat{\theta}) \quad (14)$$

其中， $r_t$  表示智能体在状态  $s_t$  下采取动作  $a_t$  所获得的奖励值。目标网络的参数为  $\hat{\theta}$ ，而  $\gamma \in [0, 1]$  表示折扣因子。

因此， $Q$  值网络参数的更新方式如下：

$$\theta_{t+1}^Q = \theta^Q + \alpha [r_t + \gamma \max_{a_{t+1}} \hat{Q}(s_{t+1}, a_{t+1} | \hat{\theta}) - Q(s_t, a_t | \theta)] \nabla Q(s_t, a_t | \theta) \quad (15)$$

本文所提算法的伪代码如下：

LDQN-RA 算法伪代码

初始化经验回放池  $D$

初始化带有初始权重  $\theta$  的  $Q$  值网络

初始化带有初始权重  $\hat{\theta}$  的  $\hat{Q}$  值网络

基于拓扑发现模块的网络视图  $G=(V, E)$

执行  $M$  个回合循环，对每个回合

初始化环境，统计得到初始状态  $s_t$

循环  $T$  个时间步，对每个时间步

使用  $\varepsilon$  贪心策略选择动作  $a_t$ ，根据  $a_t$  计算路由策略

网络运行一段时间  $\Delta t$

根据  $a_t$  收集新的网络信息并计算  $r_t$  和  $s_{t+1}$

将样本  $(s_t, a_t, r_t, s_{t+1})$  存入经验回放池  $D$

更新策略如下：

从  $D$  中随机采样一个小批量  $N_b$  个样本  $(s_t, a_t, r_t, s_{t+1})$

计算实际  $Q$  值

$$y_j = \begin{cases} r_j, & \text{回合在 } t \text{ 时不 } j+1 \text{ 终止} \\ y_t = r_t + \gamma \max_{a_{t+1}} \hat{Q}(s_{t+1}, a_{t+1} | \hat{\theta}), & \text{其他} \end{cases}$$

对损失函数  $L = \frac{1}{N_b} \sum_t (y_t -$

$Q(s_t, a_t | \hat{\theta}))^2$  关于参数做随机梯度下降

$$\theta_{t+1}^Q = \theta^Q + \alpha [r_t + \gamma \max_{a_{t+1}} \hat{Q}$$

$(s_{t+1}, a_{t+1} | \hat{\theta}) - Q(s_t, a_t | \theta)] \nabla Q(s_t, a_t | \theta)$

每  $C$  步重置  $\hat{Q} = Q$

### 3 仿真分析

#### 3.1 仿真参数设置

在实验中使用 Mininet 模拟卫星网络拓扑，具体表现为放置交换机和控制器模拟低轨卫星和高轨卫星，连接交换机和控制器模拟卫星间链路。Ryu 作为 SDN 控制器负责路径规划和流表安装，具体为使用 Ryu 编写应用程序获取 Mininet 中的链路信息计算路径，以及向交换机下发流表项来指导数据包的转发。在 Ubuntu20.0.4 中采用 Python 的 PyTorch 框架，并使用 Python3.8 编程语言进行模拟实验和性能分析。Mininet 与 Python 的结合主要通过其提供的 Python API 来实现。具体来说，本文定义了一个继承自 mininet.topo.Topo 的拓扑类，在其 build() 方法中添加主机、交换机以及带宽、时延等参数化链路，随后在脚本中以 Mininet 的方式创建仿真网络，并通过相应方法完成网络的启动、参数设置和运行控制。

仿真采用铱星星座，由平均分布在 6 条轨道上的 66 个卫星节点组成，控制器部署在 3 个地球静止轨道 (geostationary earth orbit, GEO) 卫星节点上。仿真时间 90 min，每个时隙 2 min。卫星轨道参数见表 1。

表 1 卫星轨道参数

参数名称	值
LEO 卫星数目	66
LEO 轨道高度/km	780
LEO 轨道倾角/(°)	86.4
GEO 卫星数目	3
GEO 轨道高度/km	35 786
GEO 轨道倾角/(°)	0
LEO 卫星之间链路带宽/(Mbit·s <sup>-1</sup> )	20
LEO 与 GEO 卫星之间链路带宽/(Mbit·s <sup>-1</sup> )	20

同时，本文设计了一个简化的多业务流量模型来模拟真实网络环境中的业务分布特征。在构建的铱星系统拓扑结构中，随机选取源节点和目的节点对进行数据传输实验。每个数据流的数据包生成服从泊松分布<sup>[19]</sup>，发送速率均值设定为 200~2 000 kbit/s，以模拟不同强度的业务流量。为了反映实际网络中多样化的业务需求，每个数据流被分配特定的业务类型标签，这些标签对应着不同的服务质量需求。业务类型的比例配置如下：时延敏感型业务占 50%，带宽敏感型业务占 30%，丢包率敏感型业务占 20%。各类业务的权重系数是通过判断矩阵法计算得出的，3 种业务的相对权重情况见表 2<sup>[20]</sup>。

表 2 3 种业务的相对权重情况

业务类型	相对权重
时延敏感型业务	(0.56, 0.14, 0.3)
带宽敏感型业务	(0.2, 0.51, 0.29)
丢包率敏感型业务	(0.38, 0.15, 0.47)

本文中的 LDQN 算法配置了一个包含输入层、输出层和 2 个隐藏层的循环神经网络。使用 Adam 优化器和 ReLU 激活函数，模型参数设置见表 3。

表 3 模型参数设置

参数名称	值
算法训练步数	1 200
经验回放缓冲池大小	1 000
训练批量大小 $N$	128
折扣因子 $\gamma$	0.99
更新率 $\tau$	0.001
奖励值权重因子 $\alpha_i (i=1, 2, 3)$	0~1

#### 3.2 仿真结果

LDQN 算法使用生成的数据流对网络进行训练，并在收敛后生成了一个路由决策模型。奖励值训练收敛图如图 4 所示，表示奖励随着训练数据流的变化趋势，训练步数为 400 时奖励收敛，证实所提算法具有收敛性。

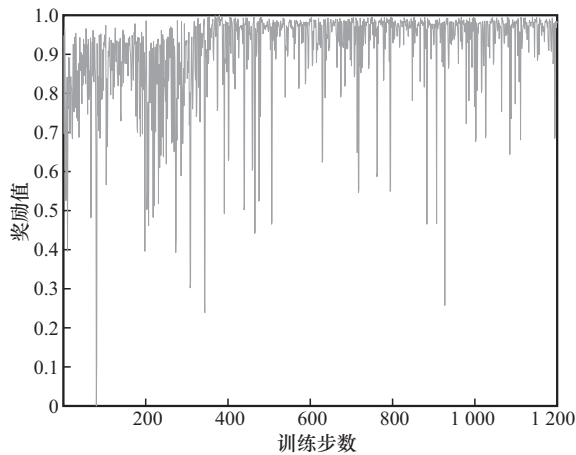


图4 奖励值训练收敛图

为了验证本文所提算法LDQN-RA的有效性,将LDQN-RA与TAODR<sup>[1]</sup>、Q-routing<sup>[17]</sup>、MDQN<sup>[18]</sup>和DQN-RA进行比较,分析端到端平均时延、丢包率、流量分布指数3个性能指标,取3次实验的平均值作为仿真结果进行分析。其中,TAODR结合了SDN架构和链路属性量化方法,通过构建拓扑感知模型和多业务策略模型、考虑动态链路属性保障多业务需求。Q-routing中使用Q-learning强化学习模型规划业务流路径。MDQN提出了一种基于DQN的多业务流QoS路由算法。DQN-RA是基于本文算法的基线算法,其核心是将原网络模型替换为DQN结构。

不同流量传输速率的平均端到端时延如图5所示,比较了5种路由算法在不同流量传输速率下的平均端到端时延,本文所提路由算法在不同流量传输速率下的平均端到端时延均为最低,其次是TAODR、DQN-RA、MDQN和Q-routing算法,平均端到端时延最高的是Q-routing。TAODR算法考虑到每个节点处理数据包的能力,减少节点的排队及丢包现象。因此,具有比DQN-RA和MDQN更低的时延。LDQN-RA一直保持最低时延的原因是采用LSTM作为神经网络,可利用对历史信息的存储能力,通过训练得到流量规律,且LSTM神经网络具有较优的收敛速度。

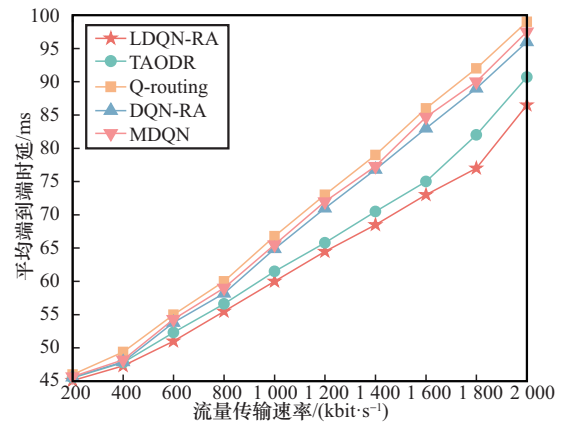


图5 不同流量传输速率的平均端到端时延

不同流量传输速率的网络丢包率如图6所示,比较了5种路由算法在不同流量传输速率下的网络丢包率。LDQN-RA具有最低的丢包率,因为LSTM神经网络具有存储历史信息 and 及时获取网络状态变化的优势,并且算法中考虑了链路负载情况,避免选择负载较重的链路传输业务,极大减少了链路拥塞导致的丢包。TAODR因为考虑了链路的稳定性和网络中流量分布,所以具有较低的丢包率。DQN-RA比Q-routing的丢包率低是因为考虑了链路利用率,充分利用网络中的链路可以避免链路负载过重导致的丢包。

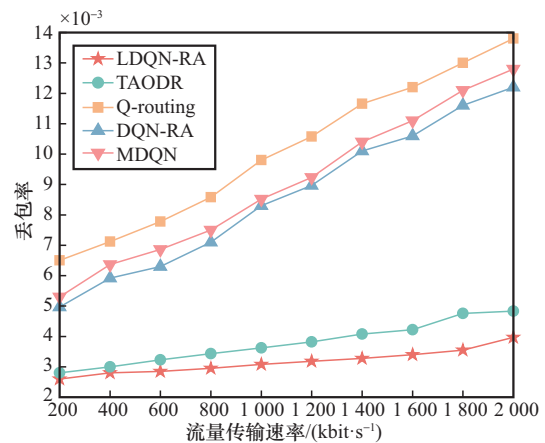


图6 不同流量传输速率的网络丢包率

不同流量传输速率的流量分布指数如图7所示,比较了4种路由算法在不同流量传输速率下的流量分布指数。流量分布指数的值越大,网络负载

越均衡。4种路由算法的流量分布指数都是随着流量变大先逐渐变小，然后变大。这是因为随着流量变大，网络中会出现某些链路负载急剧增加，而其他可用链路却处于空闲的情况。但是这4种路由算法会根据节点或链路的负载去调整网络不均衡，因此，流量更大时，流量分布指数会逐渐变大。由于TAODR考虑每个节点负载情况，相比于Q-learning和MDQN能够利用更多的节点进行多任务寻路，因此流量分布指数高于Q-learning和MDQN。由于DQN-RA考虑了链路利用率，可以更好实现网络负载均衡，因此流量分布指数比TAODR高。LDQN-RA的流量分布指数高的原因是不仅考虑了链路负载，而且使用的神经网络更容易提取流量特征，可以提前避免选择负载重的链路。

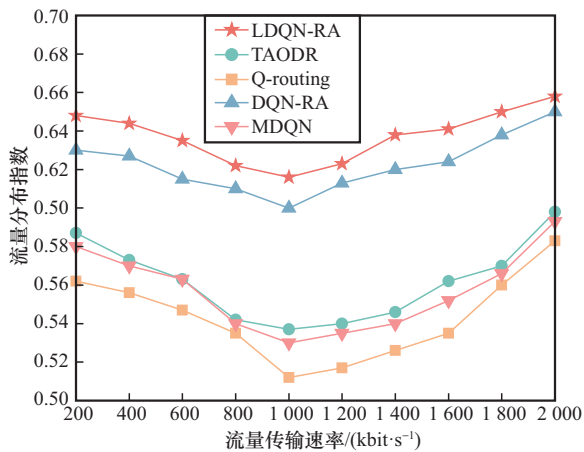


图7 不同流量传输速率的流量分布指数

面对多样化业务类型的QoS需求时，采用基于业务特征的按需路由策略能够显著提升资源利用效率。前面的仿真实验仅对比了不同路由算法的性能指标，而LDQN-RA在不同业务场景下的QoS保障能力存在显著差异，这种差异正是源于针对不同业务类型得到多样的路由策略。因此，本节进一步评价LDQN-RA各业务类型的平均端到端时延、丢包率及带宽满意度。

图8比较了LDQN-RA中不同业务的平均端到端时延。由于时延敏感型业务对时延的要求较

高，因此平均端到端时延相比其他类型业务更低。丢包率敏感型业务次之。带宽敏感型业务对时延要求低，选择多条链路来满足带宽，因此具有更高的时延。

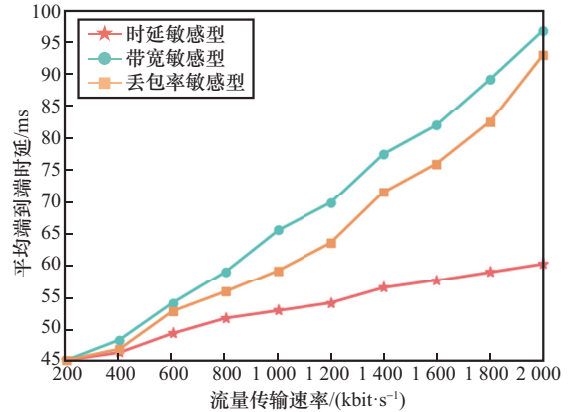


图8 不同业务类型的平均端到端时延

不同业务类型的丢包率如图9所示，比较了LDQN-RA中不同业务类型的丢包率。在流量较小的时候，所有业务类型的丢包率差别比较小。随着网络流量逐渐增加，由于丢包率敏感型业务在传输过程中追求较低的丢包率，因此具有较优的丢包性能。带宽敏感型业务由于需要选择带宽容量大的链路，可以避免选择负载较重的链路，因此丢包性能优于时延敏感型业务。时延敏感型业务总是优先选择时延低的链路，容易造成链路拥塞导致丢包，因此丢包性能最差。

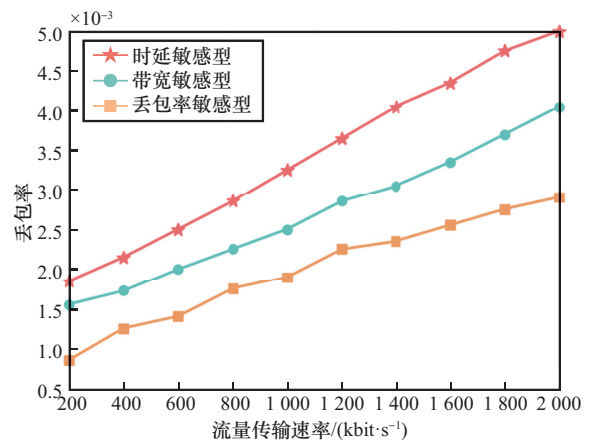


图9 不同业务类型的丢包率



带宽满意度 (bandwidth satisfaction, BWS) 作为衡量网络资源利用效率的重要指标, 其计算方法是网络中所有链路剩余带宽与链路容量比值进行连乘运算, 具体计算式见式 (16)。此性能指标能够有效反映网络路径的带宽利用潜力, 当 BWS 值较高时, 表明网络中存在充足的可用带宽资源, 能够满足更多业务的传输需求; 反之, 较低的 BWS 值则预示着网络带宽资源趋于紧张。

$$BWS = \prod_{e(i,j) \in E} \frac{b_{ij}}{bw_{ij}} \quad (16)$$

其中,  $b_{ij}$  为实际带宽,  $bw_{ij}$  为理论带宽。

不同业务类型的带宽满意度如图 10 所示。其中带宽敏感型业务的带宽满意度比其他 2 种业务类型高, 因为带宽敏感型业务在传输过程中需要选择剩余带宽大的链路。丢包率敏感型业务优先选择丢包率小的链路, 大部分是链路处于空闲状态, 因此带宽满意度在流量变大时高于时延敏感型业务。

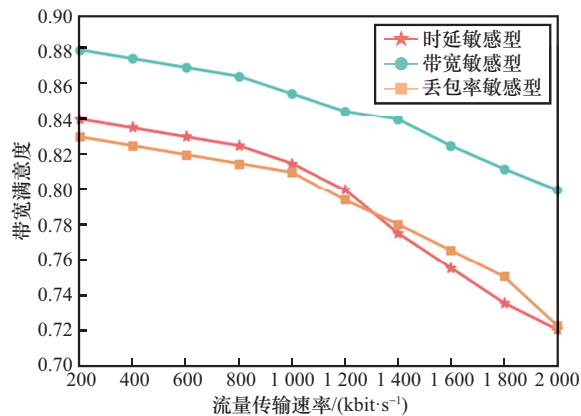


图 10 不同业务类型的带宽满意度

综上, LDQN-RA 算法通过深度强化学习, 能够动态感知网络拓扑变化, 实时监测链路通断状态和节点拥塞程度, 从而做出最优的路由决策。并且该算法能够为不同业务类型提供差异化的服务质量保障, 在时延、带宽、丢包率等关键

QoS 指标上均优于其他路由算法, 有效提升了用户体验和网络资源利用率。

## 4 结束语

本文提出的融合 LSTM 与 DQN 的智能多业务路由算法 LDQN-RA, 通过动态感知卫星链路状态, 如时延、带宽、丢包率和业务类型差异化权重, 实现了低时延、低丢包率与高负载均衡的协同优化。实验验证了其相较于传统算法在 QoS 保障上的显著优势, 尤其在多业务差异化服务场景中表现突出。下一步将基于图神经网络进行拓扑特征提取, 同时探索联邦学习框架下的分布式训练策略, 以应对大规模星座的实时性需求。

## 参考文献:

- [1] 卢雪玉, 魏雯婷, 伏丽莹, 等. 基于拓扑感知的软件定义卫星网络多业务路由[J]. 西安电子科技大学学报(自然科学版), 2024, 51(6): 10-24.  
Lu X Y, Wei W T, Fu L Y, et al. Topology awareness based multi-service routing for the software-defined satellite network[J]. Journal of Xidian University (Natural Science), 2024, 51(6): 10-24.
- [2] 王朝炜, 杜嘉楠, 王程, 等. 软件定义卫星网络中基于业务调度的混合路由算法[J]. 电子学报, 2024, 52(5): 1506-1515.  
Wang C W, Du J N, Wang C, et al. A hybrid routing based on traffic scheduling in double-layer software defined satellite networks[J]. Acta Electronica Sinica, 2024, 52(5): 1506-1515.
- [3] Wu Y, Hu G Y, Jin F L, et al. Multi-objective optimisation in multi-QoS routing strategy for software-defined satellite network[J]. Sensors, 2021, 21(19): 6356-6372.
- [4] Jiang W J, Zong P. QoS routing algorithm based on traffic classification in LEO satellite networks[C]//Proceedings of the 2011 Eighth International Conference on Wireless and Optical Communications Networks. Piscataway: IEEE Press, 2011: 1-5.
- [5] 张慧卿, 彭宏利. 基于 Pareto 多目标遗传的 LEO 卫星网络多业务 QoS 路由算法[J]. 电子设计工程, 2017, 25(20): 116-120.  
Zhang H Q, Peng H L. LEO network multi-service QoS routing using Pareto multi-object genetic algorithm[J]. Electronic Design Engineering, 2017, 25(20): 116-120.

- [6] Xie T J. SDN: software-defined space networking: architecture and routing algorithm[J]. Mobile Networks and Applications, 2019, 24(5): 1542-1554.
- [7] Jiao J, Yang P L, Du Z, et al. Clustered multi-criteria routing algorithm for mega low earth orbit satellite constellations[J]. IEEE Transactions on Vehicular Technology, 2024, 73(9): 13790-13803.
- [8] 王奎宇, 宋晓勤, 缪娟娟, 等. 基于SDN的高性能QoS保障低轨道卫星间路由算法[J]. 计算机工程, 2022, 48(5): 185-190, 199.  
Wang K Y, Song X Q, Miao J J, et al. SDN-based high-performance and QoS guaranteed inter-satellite routing algorithm for low-earth orbit satellites[J]. Computer Engineering, 2022, 48(5): 185-190, 199.
- [9] Wang F, Jiang D, Wang Z, et al. Fuzzy-CNN based multi-task routing for integrated satellite-terrestrial networks[J]. IEEE Transactions on Vehicular Technology, 2021, 71(2): 1913-1926.
- [10] Wei W T, Fu L Y, Gu H X, et al. Iris: toward intelligent reliable routing for software-defined satellite networks[J]. IEEE Transactions on Communications, 2025, 73(1): 454-468.
- [11] 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.  
Wang H, Ran Y Y, Zhao L, et al. Dynamic routing algorithm for LEO satellite networks based on deep graph reinforcement learning[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35(4): 596-605.
- [12] 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): 2801-2821.
- [13] 李新桐, 张亚生. 一种适用于低轨卫星的SDN网络人工智能路由方法[J]. 电子测量技术, 2020, 43(22): 109-114.  
Li X T, Zhang Y S. Artificial intelligence routing method for SDN network suitable for LEO satellites[J]. Electronic Measurement Technology, 2020, 43(22): 109-114.
- [14] Wang C, Wang H, Wang W. A two-hops state-aware routing strategy based on deep reinforcement learning for LEO satellite networks[J]. Electronics, 2019, 8(9): 920-936.
- [15] Xu G L, Zhao Y Y, Ran Y 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.
- [16] Chu K, Cheng S, Zhu L. A robust routing strategy based on deep reinforcement learning for mega satellite constellations[J]. Electronics Letters, 2023, 59(11): 12820-12823.
- [17] Shi Y, Yuan Z, Zhu X, et al. An adaptive routing algorithm for inter-satellite networks based on the combination of multipath transmission and Q-Learning[J]. Processes, 2023, 11(1): 167-187.
- [18] 翟凡妮. SDN中基于强化学习的多业务智能QoS路由方法[J]. 智能物联技术, 2025, 57(1): 44-47.  
Zhai F N. Multi service intelligent QoS routing method based on reinforcement learning in SDN[J]. Technology of IoT&AI, 2025, 57(1): 44-47.
- [19] Wang M, Di X Q, Jiang Y M, et al. End-to-end stochastic QoS performance under multi-layered satellite network[C]//Proceedings of International Conference on Space Information Networks. Singapore: Springer, 2017: 182-201.
- [20] Wei D B, Cao Y, Teng Q X, et al. Software defined network-based multi service routing algorithm for spatial information network[J]. Journal of Physics: Conference Series, 2023, 2558(1): 012002.

## [作者简介]



温京龙 (2000-), 男, 大连大学硕士生, 主要研究方向为卫星网络路由、边缘计算和深度强化学习。



张怡 (2000-), 女, 大连大学硕士生, 主要研究方向为卫星网络路由、深度强化学习。



魏德宾 (1978-), 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。



潘成胜 (1962-), 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。