



研究与开发

基于机器学习的SRv6动态性能优化方法研究

田东阳, 王硕凡, 祁璜, 王凯

(中国移动通信集团江苏有限公司, 江苏 南京 210029)

摘要: 随着互联网和通信技术的迅速发展, 段路由IPv6 (segment routing over IPv6, SRv6) 作为一项革新性网络技术, 简化了网络结构, 提高了网络拓扑的灵活性和可编程性, 可满足更多新业务的多种需求。而传统的静态策略无法实时响应庞大且复杂的网络状态变化, 同时网络流量难以实时监测和预测, 且网络优化通常需要在多个相互冲突的目标之间权衡。为此, 提出了一种基于时空混合神经网络与增强型强化学习相结合的SRv6动态优化框架, 利用长短期记忆 (long short-term memory, LSTM) 网络和图卷积网络 (graph convolutional network, GCN) 捕捉网络流量的时空特征, 进一步通过特征融合得到网络性能指标。再使用增强型Q学习算法与环境交互, 获得最优策略。实验结果表明, 所提算法较基本Q学习的算法速度有所提升, 且保持较高的带宽利用率, 为提升SRv6效率提供了技术参考。

关键词: 段路由IPv6; 强化学习; 动态优化; 网络流量

中图分类号: TP393.08; TN915.05

文献标志码: A

doi: 10.11959/j.issn.1000-0801.DXKX250630

SRv6 dynamic performance optimization method based on machine learning

Tian Dongyang, Wang Shuofan, Qi Huang, Wang Kai

China Mobile Group Jiangsu Co., Ltd., Nanjing 210029, China

Abstract: With the rapid development of the Internet and communication technology, segment routing over IPv6 (SRv6), as an innovative network technology, simplifies network architecture and enhances the flexibility and programmability of the network topology, thereby meeting the diverse demands of various new services. In contrast, traditional static policies are unable to respond in real-time to the vast and complex changes in the network states, while network traffic is difficult to monitor and predict in real-time, and network optimization often requires trade-offs among multiple conflicting objectives. To address these problems, an SRv6 dynamic optimization framework that combined spatio-temporal mixed neural networks with enhanced reinforcement learning was proposed. It utilized long short-term memory (LSTM) and graph convolutional network (GCN) to capture the spatio-temporal features of network traffic, subsequently merging these features to derive network performance metrics. The enhanced Q-learning algorithm was then employed to interact with the environment, obtaining the optimal strategy. Experimental results show that compared with the basic Q-learning, this approach shows improved learning speed while maintaining a high

bandwidth utilization rate, providing a technical reference for enhancing the efficiency of SRv6.

Key words: SRv6, reinforcement learning, dynamic optimization, network traffic

0 引言

云计算、边缘计算、工业互联网等各种新网络应用场景对网络高带宽、低时延和高可靠性的传输要求越来越高。段路由 IPv6 (segment routing over IPv6, SRv6) 作为一种新的 IP 承载协议逐渐得到青睐。SRv6 以全新角度应对各种场景需求, 使用 IPv6 扩展头定义网络层可编程参数, 可以大大地简化网络协议栈并易于运维管理。SRv6 最大的特点在于灵活可编程, 在 128 位段标识 (segment identifier, SID) 上承载扩展字段, 灵活定义业务路径, 按需调度流量和编排各种增值服务, 可以在云专网、软件定义广域网 (software-defined wide area network, SD-WAN) 等多个场景实现差异化服务。段路由体系结构为 SRv6 的可编程转发与路径编排提供了基础支撑, 可在不显著增加网络状态维护复杂度的前提下实现显式路径控制与业务链编排^[1]。SRv6 的另一个特点是能够保证数据流的不确定性, 具体体现为提供多个 “Live-Live” (多条存活状态路由), 可以在拥塞发生后依然实现吞吐量不变及尾时延减小的效果, 因此, 也成为远程控制、在线游戏、工业互联网等场景必不可少的核心支撑^[2]。AutoSRv6 等通过将系统建模与约束求解技术相结合作为高级别抽象, 在不考虑具体底层操作和语法规则的情况下自动生成 SRv6 配置, 成功解决了传统配置语言烦琐以及错误的问题, 促进了运营商大规模部署 SRv6^[3]。

尽管 SRv6 具有灵活可编程等优势, 并已在多种网络场景中得到验证, 但在卫星—地面融合及动态业务环境中仍面临新的挑战。一方面, 网络状态随业务波动和链路变化频繁变动, 而现有 SRv6 策略多依赖预定义规则或被动调整, 策略

更新通常滞后于网络变化, 易导致路径振荡和端到端时延升高^[4-5]。另一方面, 现有的监测机制虽可提供丰富的状态信息, 但更偏向于 “事后分析”, 难以及时指导策略优化^[6]。此外, 一些引入机器学习的方法通常聚焦于单一指标预测或局部优化, 尚未与 SRv6 策略形成协同设计^[7-11]。因此, 本文聚焦的问题是如何在 SRv6 框架下引入对未来网络状态的感知, 并结合自适应策略优化机制, 使系统能够在状态变化发生之前完成路径调整, 从而提升动态环境中的时延与稳定性。挑战的本质并非 SRv6 无法支持卫星网络, 而是在 SRv6 框架下如何结合未来状态预测与自适应策略优化, 以有效地应对动态链路变化。

1 本文贡献

在前述分析的基础上, 本文围绕 SRv6 在动态网络环境中的性能优化问题, 构建了一种面向卫星—地面融合场景的 SRv6 动态优化框架。该框架将网络状态感知、性能预测与策略优化进行协同设计, 使 SRv6 由以往的依赖静态配置和被动调整的模式, 逐步转变为具备自适应优化能力的策略体系。为此, 本文在状态建模方面引入时空联合预测思想, 利用长短期记忆 (long short-term memory, LSTM) 与图卷积网络 (graph convolutional network, GCN) 对链路带宽、时延和丢包率等关键性能指标进行联合建模与预测, 从而在复杂拓扑环境中获得更准确的未来网络状态信息; 在策略优化方面, 构建增强型 Q 学习机制, 通过经验回放与双 Q 学习来减少值函数过估计, 同时结合流量亲和与动作平滑策略, 降低路径振荡并实现多目标性能之间的权衡。此外, 本文基于轻量级网络仿真工具 Mininet 构建仿真平台, 将本文方法与基础强化学习 (reinforcement



learning, RL) 以及两类经典基线方法——传统静态路由、传统等价多路径负载均衡 (equal-cost multi-path, ECMP) 进行了对比试验。实验结果表明, 所提方法在端到端时延与收敛稳定性方面均具有明显改进, 为 SRv6 智能化优化提供了一种可行的思路。

2 相关工作

2.1 SRv6 性能优化研究现状

随着软件定义网络 (software-defined network, SDN) 与可编程转发技术的发展, 网络控制与路由决策的灵活性显著提升^[12]。当前非机器学习的 SRv6 性能优化研究主要聚焦于架构设计、算法改进和协议扩展 3 个方向, 传统方法普遍依赖预定义策略或者静态规则, 所以在面对实时变化的网络环境时会有点难以招架。例如, 在混合 IP/段路由 (segment routing, SR) 网络部署场景中, 分域式段路由渐进部署架构通过域划分优化过渡期性能, 但是其采取的静态路径策略会导致在面对突发流量时吞吐量急剧衰减^[13]。Wu 等^[6]提出的多域 SFC 动态编排算法虽然利用改进蚁群优化技术将端到端时延降低 22%, 但由于依赖预设拓扑信息, 在拓扑频繁变化的卫星网络中仍频繁出现路由失效问题。G-SRv6 策略压缩技术^[14]通过逻辑拆分段功能提升调度精度, 但压缩率仅 15%~20%, 突发流量下路径计算效率显著降低。

现有技术对网络负载的动态预测能力提出了一定的要求, 但仍有一定的提升空间。例如, SRv6 性能监控框架段路由 IPv6 性能监控 (SRv6 performance monitoring, SRv6-PM) 可以通过流量着色实现丢包率测量, 但是其分析机制的侧重点在事后数据处理, 无法实现对策略的提前调整^[15]。如果在链路故障场景下, ECMP 需要一定时间完成重新收敛, 其间可能出现吞吐下降与时延抖动, 从而影响业务连续性。但是 Linux 内核中的 SRv6 实现在引入基于哈希的消息认证码

(hash-based message authentication code, HMAC) 安全机制后, 可能会带来额外的处理开销, 在突发流量下对吞吐能力产生不利影响。所以通过分析这些数据, 可以得出结论: 当前 SRv6 性能优化技术在实时响应能力、预测性机制以及动态适应性方面存在瓶颈, 亟须突破传统方法的局限性。综上所述, 现有 SRv6 优化技术更多依赖静态规则或预设拓扑, 当网络负载或拓扑发生突变时, 策略难以及时适配, 可能引发路径失效与性能波动。缺乏面向未来状态的预测能力是这类方法在高动态场景中表现受限的重要原因。

2.2 机器学习在网络性能优化中的应用

随着网络规模的扩大, 在当前网络中的业务需求逐渐增加以及新技术的产生, 传统优化网络性能方法已经不能再胜任。SRv6 具有段路由灵活可扩展的特点、强大的业务部署能力, 并结合新的技术可以提供新的面向用户或面向应用层的能力给 SDN 更高级的网络控制, 以带来更好的网络性能优化方法^[10]。与此同时, 这种灵活动态和强大功能特点给性能优化带来的挑战也随之增加, 随着机器学习的出现, 针对 SRv6 的性能优化也提供了新方向。其本质是利用机器学习强大的数据分析和模型辨识能力来在复杂多样的环境中做出智能化的网络性能优化决策。其中, SRv6 网络中的段路由问题优化是基于 SRv6 网络性能机器学习的重要环节。该问题需要考虑在复杂的网络环境中如何实现数据平面段列表构建、路径选择、服务链优化等。运用机器学习来解决这个复杂问题时需要首先学习网络拓扑结构和节点间语义, 这促使基于图神经网络的 SRv6 网络能够更好地利用语义信息进行高效路由, 从而研究者们提出了一系列网络算法方案来实现 SRv6 网络中的更高效合理的路由策略, 该策略学习整个 SRv6 网络中的各节点以及每个具体段的信息从而为 SRv6 段的优化决策提供依据^[9-10]。

在当前的机器学习情形中,网络可以从数量巨大且瞬息万变的运行数据中,并且在不需要人进行刻意干预的前提下,自主地识别和挖掘隐藏在背后的更深层次规律。也可以针对流量特征、链路状态,以及丢包率等诸多方面的实际情况探查,并预测未来趋势。在了解相关细节的基础上做出智能决策。具体来说,将机器学习用于改善SRv6的性能并加以优化已经深入人心,目前已经被成功地应用到多种重要的领域。例如,利用时间序列分析模型开展流量的趋势预测活动,LSTM可较精准地预测网络流量,为事先分配资源或者防止出现拥堵现象提供相应的依据^[13]。借助分类以及异常探测这类机器学习算法,如支持向量机(support vector machine, SVM)、随机森林、深度学习等模型来及时发现网络中存在的诸多问题,并对其他潜在发生的故障予以提前预警,为提高网络自身的安全性和可靠性作出相应的贡献^[16]。Q学习和双Q网络(deep Q-network, DQN)等算法在网络上也得到了广泛的使用,可以依照网络当前的状态和各项指标随时对不同的资源分配进行实时调整,从而最大化网络效率,并尽量将传输时延降到最低,使重要的业务得以畅通无阻^[17]。与此相似的,上述聚类算法能够更好地诠释流量模式以及用户的行为,而深度学习也可以为端到端地对网络性能进行建模并优化其表现作出巨大的贡献^[9]。各种不同的应用形式汇集在一起展现出了这样的一个进程,在机器学习所获得的能量之下,网络终于达成了一种新的形式——从被动变为主动、从静态变为动态、从单独的变化演进变成整体协作优化的智能转型,为应对SRv6新型网络架构在更复杂的环境中所面临的动态性能优化难题提供了良好的方法论根基以及技术上的强有力支撑。然而,这些基于机器学习的方法大多集中在某一性能指标(如时延预测或带宽估计),对多目标之间的权衡与策略层面的协同优化关注不足。尤其是在SRv6架构下,

如何将预测结果与路径策略动态联动,使系统在保持可编程的同时具备自适应能力,仍有待进一步研究。

2.3 现有性能优化方法的局限性

通过对现有工作的梳理可以看出,当前SRv6性能优化在高动态网络环境中仍存在一定的局限性。首先,现有策略大多依赖实时监测信息进行事后调整,缺乏对未来链路负载和拓扑变化的有效预测能力,当网络即将进入拥塞或发生切换时,难以及时做出策略响应。其次,在策略层面仍以静态配置或规则驱动为主,当链路状态频繁波动时,策略更新通常滞后于网络变化,容易导致路径振荡与端到端时延抖动。与此同时,许多研究更多关注单一性能指标的优化,而在时延、丢包率与负载均衡等多个目标之间缺乏统一建模和权衡机制,这在复杂业务场景中进一步限制了系统整体性能的提升。基于上述问题,本文提出了一种基于时空预测与强化学习协同设计的SRv6动态优化方法,使路由策略能够在状态变化发生之前完成调整,从而在保证可靠性的同时提升网络的时延性能与稳定性。

3 基于机器学习的动态性能优化方法

3.1 整体框架设计

本文所提的整体架构采用3层协同设计,由下至上依次为数据采集层、智能决策层与策略执行层。数据采集层作为数据预处理的平台,通过分布式探针实时捕获SRv6数据平面的关键性能指标,其中,包括吞吐量、时延、丢包率等网络状态数据。智能决策层作为系统核心,采用多模式融合的智能路由决策引擎。该层集成了时空混合神经网络和增强型Q学习代理。其中,时空混合神经网络扮演的角色为负责网络状态的预测分析,增强型Q学习代理承担路由决策的核心功能,而该层设计的自适应奖励函数充当决策质量的评价标准,三者协同工作共同实现SRv6网络的智能化路由优



化。策略执行层通过可编程控制接口将决策指令实时下发至SRv6转发节点,执行路径切换、负载均衡参数调优等操作。3层结构通过闭环反馈机制形成有机整体,其中,策略执行层产生的网络状态变化数据将重新注入数据采集层,驱动机器学习模型迭代更新,最终实现SRv6网络性能的自适应优化循环。整体框架如图1所示。

3.2 网络负载预测模型设计

网络流量具有明显的时间相关性和空间分布特性,为了全面捕捉SRv6网络中的负载变化模式,本文提出了一种时空特征融合的混合预测(spatio-temporal hybrid prediction network, ST-HybridNet)模型。ST-HybridNet采用多模块协同的设计思想,利用LSTM网络在时序建模方面的优势和GCN在空间关系学习方面的能力来完成设计。模型的输入包括历史窗口内的网络负载数据序列以及网络拓扑结构信息,输出为未来多个时刻的网络性能指标预测值。整个模型通过端到端的训练方式,实现时空特征的有效融合和多维度性能指标的联合预测。网络负载预测模型架构如图2所示,主要由时间特征提取模块、空间特征提取模块和多层融合预测模块组成。

时间特征提取模块采用LSTM网络,核心计算单元为LSTM单元,能够有效地捕获网络流量

的长短期时序依赖关系。该模块接收历史时间窗口内的网络负载数据序列 $\{X_{t-n}, X_{t-n+1}, \dots, X_{t-1}\}$,其中, t 为当前时间步, n 为历史时间窗口长度, X 为对应时间步的网络负载特征向量,模块最终输出对未来时刻负载的高级时序特征 H_t 。LSTM的核心优势在于解决了传统循环神经网络(recurrent neural network, RNN)中的梯度消失问题,使模型能够学习长期依赖关系,特别适合处理网络流量这类具有周期性和突发性特征的时序数据。

空间特征提取模块通过GCN实现,将SRv6网络拓扑表示为图 $G=(V, E)$,其中,节点 V 代表网络设备,边 E 代表链路连接。GCN能够有效地学习节点间的空间相关性,捕捉流量在网络中的传播特性和拓扑依赖关系,最终输出网络的空间结构特征 Z 。在SRv6网络中,本文对标准GCN进行了以下改进。首先,通过引入边权重机制,根据链路带宽、时延等属性为边分配不同权重,使模型能够区分不同质量的网络连接;其次,采用注意力机制动态地调整邻居节点的贡献权重,可以得到对当前节点流量预测更重要的邻居信息;最后,设计多尺度图卷积层,分别捕获1跳、2跳等不同距离邻居的影响,全面反映网络中的空间依赖关系。

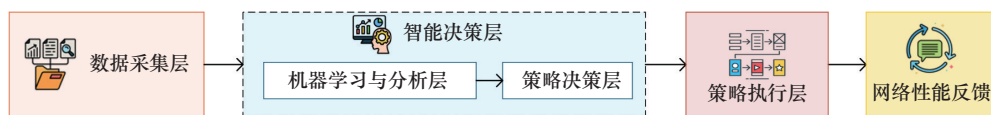


图1 整体框架

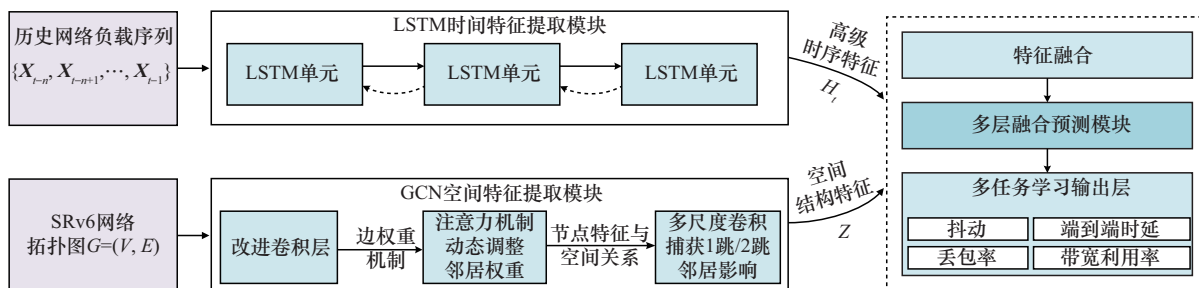


图2 网络负载预测模型架构

为全面评估网络性能状态, 本模型同时预测多个关键性能指标, 包括链路带宽利用率、端到端时延、丢包率和抖动。多维度预测采用多任务学习框架, 通过共享底层特征表示同时输出多个目标值, 既保证了预测效率, 又增强了模型的泛化能力。

综上所述, ST-HybridNet 通过联合建模网络流量的时间演化规律和拓扑空间依赖关系, 在统一框架下同时输出带宽利用率、时延、丢包率和抖动等多维性能指标的未来值。与仅依赖当前监测数据或单一时间序列模型的方法相比, 该模型能够提前感知网络即将进入的高负载或异常状态, 为后续的策略优化提供前瞻性的状态输入, 从根本上缓解了现有 SRv6 性能优化中“缺乏预测能力、决策滞后”的问题。

3.3 SRv6 动态策略调整机制

仅依靠性能监测和状态预测仍不足以保证整体网络性能, 因为不同链路之间存在复杂耦合关系, 局部最优的路径选择通常会带来全局拥塞或时延反弹。为此, 本文在 SRv6 框架下设计了动态策略调整机制, 将路由选择问题建模为一个序列决策过程, 由 RL 智能体在长期收益意义下自动学习最优策略。具体而言, 系统首先利用前文的 ST-HybridNet 获得带宽利用率、时延、丢包率等关键指标的预测结果, 并将其与当前拓扑状态共同构成环境状态; 智能体在此基础上从地面、卫星及混合路径集合中选择具体转发策略, 网络环境根据实际转发效果返回综合考虑时延、丢包和负载均衡的奖励信号。通过持续交互与迭代更新, 策略由人工预设规则转变为数据驱动的自适应决策。

本文的动态策略调整机制开始于利用 LSTM 来提取数据的时序特征, 如式 (1) 所示。

$$\mathbf{F}_{\text{temporal}} = \text{LSTM}(\mathbf{X}_{t-n:t-1}) \quad (1)$$

其中, $\mathbf{X}_{t-n:t-1}$ 表示过去 n 个时间步的网络状态数

据, $\mathbf{F}_{\text{temporal}}$ 表示提取到的时序特征向量。与此同时, GCN 提取空间特征, 如式 (2) 所示。

$$\mathbf{F}_{\text{spatial}} = \text{GCN}(\mathbf{X}_{\text{spatial}}, \mathbf{A}) \quad (2)$$

其中, $\mathbf{A}$ 表示网络拓扑的邻接矩阵, $\mathbf{X}_{\text{spatial}}$ 表示当前各节点的状态特征, $\mathbf{F}_{\text{spatial}}$ 表示提取到的空间特征向量。在时间、空间特征提取完成后, 会在融合层将提取到的时间和空间特征进行拼接和联合建模, 如式 (3) 所示。

$$\mathbf{F}_{\text{fused}} = \text{ReLU}(\mathbf{W}_f[\mathbf{F}_{\text{temporal}}; \mathbf{F}_{\text{spatial}}] + \mathbf{b}_f) \quad (3)$$

其中, ReLU 为修正线性激活函数, $\mathbf{W}_f$ 为融合权重矩阵, $\mathbf{b}_f$ 为偏置向量。在该过程结束后会输出对带宽利用率 (式 (4))、时延 (式 (5))、丢包率 (式 (6))、抖动 (式 (7)) 4 项关键指标的预测结果。

$$y_{\text{bandwidth}} = \sigma(\mathbf{W}_b \mathbf{F}_{\text{fused}} + \mathbf{b}_b) \quad (4)$$

$$y_{\text{latency}} = \mathbf{W}_l \mathbf{F}_{\text{fused}} + \mathbf{b}_l \quad (5)$$

$$y_{\text{loss}} = \sigma(\mathbf{W}_{\text{loss}} \mathbf{F}_{\text{fused}} + \mathbf{b}_{\text{loss}}) \quad (6)$$

$$y_{\text{jitter}} = \mathbf{W}_j \mathbf{F}_{\text{fused}} + \mathbf{b}_j \quad (7)$$

其中, $y_{\text{bandwidth}}$ 、 y_{latency} 、 y_{loss} 、 y_{jitter} 分别为模型输出的链路带宽利用率预测值、端到端时延预测值、链路丢包率预测值及时延抖动预测值, $\mathbf{W}_b$ 、 $\mathbf{W}_l$ 、 $\mathbf{W}_{\text{loss}}$ 、 $\mathbf{W}_j$ 分别为带宽利用率、时延、丢包率、抖动 4 个预测分支对应的全连接层权重矩阵, $\mathbf{b}_b$ 、 $\mathbf{b}_l$ 、 $\mathbf{b}_{\text{loss}}$ 、 $\mathbf{b}_j$ 分别为 4 个预测分支对应的全连接层偏置向量。式 (4) 与式 (6) 中 σ 为 sigmoid 函数, 作用是输出 0~1 范围。以上便是在生成关键性能指标的状态向量后的特征提取阶段。这种设计能够获得低维、去噪、全面的网络信息, 还能根据历史数据预测未来的网络状态, 在问题发生之前进行决策。

在特征提取阶段结束之后, 动态策略调整机制的第 2 阶段可以总结为智能路由决策与路径选择阶段, 在该阶段, 根据上个阶段提取的状态特征和输出的预测结果, 通过改进的 ε -贪



婪策略在地面/卫星/混合路径之间动态切换，实时适应网络环境的变化。这便是增强型Q学习算法中的分支——双Q学习算法，如式（8）所示。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, \arg\max_a Q(s_{t+1}, a)) - Q(s_t, a_t)] \quad (8)$$

其中，状态 s_t 表示当前网络性能指标，动作 a_t 表示选择路径（地面/卫星/混合），奖励 r_{t+1} 为路径性能的好坏评分， $Q(s_t, a_t)$ 为当前状态 s_t 下执行动作 a_t 的价值， α 用来控制学习速度， γ 为折扣因子表示未来奖励的重要性， $\arg\max_a Q(s_{t+1}, a)$ 为下一状态的最佳动作。在高动态的卫星—地面融合网络中，若直接采用传统Q学习，样本序列高度相关和目标值过估计容易带来训练不稳定、收敛缓慢的问题。针对上述问题，本文在标准算法框架的基础上进行了3方面改进。首先，利用经验回放缓冲区打破样本之间的时间相关性，通过随

机小批量采样提升数据利用率和训练稳定性；其次，引入双Q学习将动作选择与价值评估分离，有效地抑制值函数的系统性过估计，使学习到的策略更可靠；最后，通过设计动作平滑与流量亲和机制约束路径切换频率，在保证自适应性的同时避免频繁振荡，从而提升网络运行的连续性和用户体验。在传统Q学习的基础上增加了经验回放用以存储过往经验，如，当前状态、执行的动作、获得的奖励、执行动作后到达的新状态等信息，用来提高数据的利用率、提升训练稳定性。本阶段使用的动作平滑机制减少了智能体输出的剧烈波动，具体表现为通过检查最近3个时间步的动作历史，在检测到近期有路径切换时以70%的概率保持当前路径不变。以此来达到让网络策略更稳定、收敛更快的目的。流量亲和性机制的引入，如式（9）所示。

$$\text{Affinity}(\text{flow_id}) = \begin{cases} \text{historical_path, Available}(\text{historical_path, load}) = \text{True}; \\ \text{min_cost_path, 其他} \end{cases} \quad (9)$$

其中，如果之前的路径仍可用且性能良好，会复用历史路径，如果历史路径不可用或性能下降，会计算新路径。基于流量亲和性的路径选择方法可以保证在同一数据流性能稳定的前提下，实现网络负载的均衡与优化。具体地，首先查询历史记录，如果某条路径还可用，优先复用，减少频繁切换，否则计算 k 条候选路径。

在选定路由决策后，会进入该机制的第3个阶段，可以总结为策略执行与网络配置阶段。在该阶段中，将抽象的动作映射为具体的SRv6段列表配置。系统也会利用实现的 K 最短路径算法^[18] 计算多条候选路径，该算法可在给定的SRv6网络拓扑中，求解出源节点到目的节点之间多条代价最优的无环转发路径，为路由决策提供充足的可选路径集。后续根据流量和负载感知实时选择当前性能最优的路径，从而避免拥塞、

降低时延、提升整体网络效率。确定了最优路径之后，更新相应的SID列表，进而实现实际的路径切换和流量重定向。为了确保网络服务的连续性，本机制对整个策略下发过程提出了毫秒级的要求。

第4阶段性能评估与奖励反馈阶段的重点在奖励函数设计。为了解决网络拥塞、流量突增等问题，并在提升系统稳定性的同时保障业务服务质量（quality of service, QoS），本文在奖励函数中引入多目标优化设计，以同时兼顾高带宽利用率、低时延和低丢包率，从而在全局层面提升策略的稳定性。具体地，建立多维度QoS指标加权奖励函数，如式（10）所示。

$$R(s_t, s_t^{-1}, a_t) =$$

$$\begin{aligned}
& w_b \cdot R_{\text{bandwidth}}(u_t) + w_l \cdot R_{\text{latency}}(l_t) + \\
& w_p \cdot R_{\text{packetloss}}(p_t) + \\
& w_s \cdot R_{\text{stability}}(s_t, s_t^{-1}) + \\
& w_c \cdot R_{\text{consistency}}(a_t, a_t^{-1})
\end{aligned} \quad (10)$$

其中, s_t^{-1} 表示上一时间步的网络状态, 用于刻画相邻时间步之间网络状态的变化情况; a_t 表示当前时间步选择的路由动作, a_t^{-1} 表示上一时间步选择的路由动作, 用于判断相邻时间步路由动作是否保持一致。 u_t 、 l_t 、 p_t 分别表示当前时间步的带宽利用率、归一化时延和归一化丢包率。 w_b 、 w_l 、 w_p 、 w_s 、 w_c 为各奖励项对应的权重系数, 分别用于调节带宽利用率、时延、丢包率、状态稳定性和动作一致性在综合奖励中的影响程度, 并满足权重系数之和为 1, 即 $w_b + w_l + w_p + w_s + w_c = 1$ 。 $R_{\text{stability}}$ 表示状态稳定性奖励。 $R_{\text{consistency}}$ 表示动作一致性奖励。设计带宽最优区间奖励以保证带宽利用率, 如式 (11) 所示。

$$R_{\text{bandwidth}}(u) = \begin{cases} 20, & 0.6 \leq u \leq 0.8; \\ 10, & 0.5 \leq u < 0.6; \\ 5, & 0.8 < u \leq 0.9; \\ -15, & \text{其他} \end{cases} \quad (11)$$

其中, u 表示带宽值。为了避免提高通信质量而过度使用带宽的问题, 设计时延奖励最小化全局平均时延, 如式 (12) 所示。

$$R_{\text{latency}}(l) = \begin{cases} 15, & l < 0.3; \\ 10, & 0.3 \leq l < 0.5; \\ 0, & 0.5 \leq l < 0.75; \\ -15, & l \geq 0.75 \end{cases} \quad (12)$$

其中, l 是归一化的时延值。丢包率的设计奖励将目标设置为小于 1%, 如式 (13) 所示。

$$R_{\text{packetloss}}(p) = \begin{cases} 25, & p < 0.2; \\ 10, & 0.2 \leq p < 0.4; \\ 0, & 0.4 \leq p < 1.0; \\ -25, & p \geq 1.0 \end{cases} \quad (13)$$

其中, p 是归一化的丢包率, 并加入稳定性奖励与动作一致性奖励。针对环境与状态转移, 利用

网络仿真工具 Mininet 生成近似真实的 QoS 指标, 环境接收并应用路由动作、获取新的网络状态、计算链路并更新历史记录保证状态平滑变化。

为了使整个学习过程形成闭环反馈机制, 最后一个阶段总结为持续学习与策略优化阶段。在该阶段中, 通过经验回放机制将每次交互的状态、动作、奖励、下一状态等信息存储到缓冲区中, 系统定期进行小批量采样和 Q 值更新。与此同时, 系统采用双 Q 学习算法减少价值函数的过估计偏差, 通过与网络环境的持续交互和策略迭代, 逐步优化路由决策质量。因此, 在 5 个阶段的共同作用下, SRv6 动态策略调整机制最终完成了“感知-决策-执行-反馈”的闭环, 也实现了 SRv6 网络性能的自适应优化。

综上, SRv6 动态策略调整机制通过“预测-决策-执行-反馈”的闭环设计, 将时空特征预测模型与增强型 Q 学习有机结合。前者提供面向未来的多维网络状态感知, 后者在此基础上利用经验回放、双 Q 学习、动作平滑和多维奖励函数实现对时延、丢包率和带宽利用率等指标的协同优化。与此同时, K 最短路径算法与流量亲和性机制共同完成具体 SID 列表的生成与更新, 在保证单流路径稳定性的前提下实现全局负载均衡。相较于依赖静态规则或单一优化目标的传统方法, 该机制能够在复杂动态场景中生成更稳定、具有自适应能力的 SRv6 路由策略。

3.4 算法复杂度分析

由于本文在 SRv6 策略优化框架中引入了时空预测模型和增强型 Q 学习机制, 有必要对整体算法的复杂度进行分析。从时间复杂度来看, ST-HybridNet 主要由 LSTM 与 GCN 构成。在每个预测周期内, LSTM 负责对链路历史状态序列进行建模, 计算复杂度与序列长度和隐藏层维度近似呈线性关系; GCN 在网络拓扑图上进行特征传播和聚合, 计算量与链路数和节点度数近似成正比。综合来看, 单次预测的时间开销随网络节点



数与链路数线性增长, 适合在当前 SRv6 控制器或边缘控制节点上以周期性方式运行。强化学习部分采用基于经验回放的双 Q 学习, 每一步交互主要包括状态转换、从候选路径集合中选择动作以及更新对应的 Q 值, 时间复杂度与候选路径数量线性相关。需要强调的是, 策略训练过程可以在离线或准实时环境中持续进行, 并不直接位于数据平面转发路径上, 因此对在线转发时延的影响较小。

从空间复杂度来看, ST-HybridNet 的参数规模主要由 LSTM 与 GCN 的层数和隐藏单元数决定, 在给定网络规模下为常数级; Q 学习部分需要维护“状态-动作”价值表或其近似表示, 并存储一定规模的经验回放缓冲区, 其空间开销随状态维度和候选路径数量线性增长。综合考虑, 本文所提方法的存储需求随网络规模和候选路径规模线性扩展, 能够在典型的边缘服务器或控制器硬件条件下满足部署需求。总体而言, 引入 LSTM 与增强型 Q 学习虽在训练阶段带来一定的额外开销, 但在线决策阶段每次路径选择仅涉及一次预测结果读取和一次策略查表/更新, 其时间与空间复杂度均处于可接受范围内, 在实际 SRv6 网络中的工程部署具备可行性。

4 实验设计与结果分析

4.1 实验环境与配置

本实验设计采用仿真环境进行 SRv6 网络动态性能优化的验证。在软件环境方面, 采用 Ubuntu 20.04 LTS 操作系统, 使用 Python 3.8.10 版本作为主要开发语言, 而机器学习框架采用 TensorFlow 2.6.0 和 PyTorch 1.9.0, 网络仿真则通过 Mininet 2.3.0 平台实现, 最后 SRv6 功能的模拟实现依赖于 Linux 内核原生支持 (kernel 5.4.0 版本)。

在创建模拟卫星-地面混合架构的 SRv6 网络中, 本文构建了包含 10 个交换节点的拓扑, 其中,

地面骨干节点为 $s_1 \sim s_5$, 卫星节点为 $s_6 \sim s_{10}$, 并在边缘接入侧挂接 8 个主机用于业务产生与测量。链路带宽配置为 100 Mbit/s $\sim$ 1 Gbit/s。地面骨干链路时延为 5 $\sim$ 12 ms、卫星链路时延为 120 $\sim$ 135 ms、天地互联链路时延为 145 $\sim$ 155 ms, 用以体现天地融合网络的高时延异构特性。该实验中网络拓扑设计为了使 K 最短路径算法的路径选择能有充足的决策空间, 同时也要确保路由优化算法能够在多样化的网络环境中进行有效的性能验证, 因而设计能够支持多条候选路径的并存。最终, 实验成功实现了 K 最短路径算法与 SRv6 路由的结合, 通过流量亲和性机制减少了路径振荡现象。本文实验规模主要用于验证方法的有效性, 而非极限性能评估。

为保证过程可复现, 本文将路由控制过程离散为固定时间步。每个时间步间隔设为 $\Delta t = 0.1$ s, 控制器采集链路带宽利用率、端到端时延与丢包率等状态特征构成状态 s_t 。智能体在 K 最短路径候选集中选择动作 a_t , 并将所选路径以 SRv6 段列表形式下发执行。为抑制频繁切换导致的路径振荡, 设置最小动作保持窗口为 5 个时间步, 并在奖励中引入切换抑制项, 从而提升策略稳定性。

4.2 模型性能指标与评估方法

本文所提的 SRv6 动态路由优化模型从学习曲线、QoS 性能指标以及多目标优化能力 3 个维度对实验仿真结果进行分析。

(1) 学习曲线

奖励函数是强化学习策略优化的核心信号, 其变化趋势直接反映了智能体对动态网络环境的适应能力。强化学习训练过程如图 3 所示, 随着训练轮次的增加, 平均奖励曲线随训练轮次增加呈现出稳定上升的态势, 并在约第 40 个训练轮次后达到并稳定在高位平台, 这表明智能体逐步学习到稳定且高效的路由策略。与基础强化学习方法相比, 本文所提方法在奖励曲线上表现出更平

滑的上升过程和更快的收敛速度,这主要得益于LSTM-GCN混合模型提供的前瞻性网络状态预测,减少了策略探索的随机性;同时,双Q学习与经验回放机制缓解了值函数过估计问题,奖励函数中引入的动作平滑惩罚有效地抑制了路径频繁切换导致的策略振荡,从而实现了学习效率与稳定性的显著提升。与此同时,带宽利用率曲线逐步上升并稳定在理想区间内,其中,带宽利用率的最优运行区间设为 $0.65\sim 0.75$,区间外将施加更严格惩罚,以避免链路过载或资源闲置,体现了本文方法在资源利用方面的优越性。时延曲线从初始的约 110 ms 持续下降,显著低于设定的目标值,且在整个训练过程中保持稳定下降趋势。丢包率曲线同样呈现稳步下降特征。这些指标变化趋势表明,本文方法通过协同优化机制,能够在训练过程中同步提升多个关键性能指标。

(2) QoS 性能表现

网络性能时序变化如图4所示,展示了网络动态环境中的实时性能,可以看出,累计奖励曲线整体保持在较高正值区间,仅在少数时间步出现短暂波动,表明智能体在长期运行中能够维持稳定且高效的决策质量,证明了智能体在长期动态环境中,策略具有持续有效性和鲁棒性,能够不断适应变化并积累正向收益。在网络性能指标方面,如图4所示,随着训练进行,端到端时延逐步降低并趋于稳定,最终维持在较低且平稳的水平,满足实时业务的传输要求。该结果表明,所提方法能够在动态网络环境中有效地抑制端到端时延的增长,为实时业务提供稳定保障。在收敛阶段,丢包率保持较低水平,波动范围小,虽有小幅波动,但能快速恢复至低水平,从未发生持续性的高丢包率灾难。这证明了所提方法具备强大的抗干扰和自恢复能力,表明模型具备良好

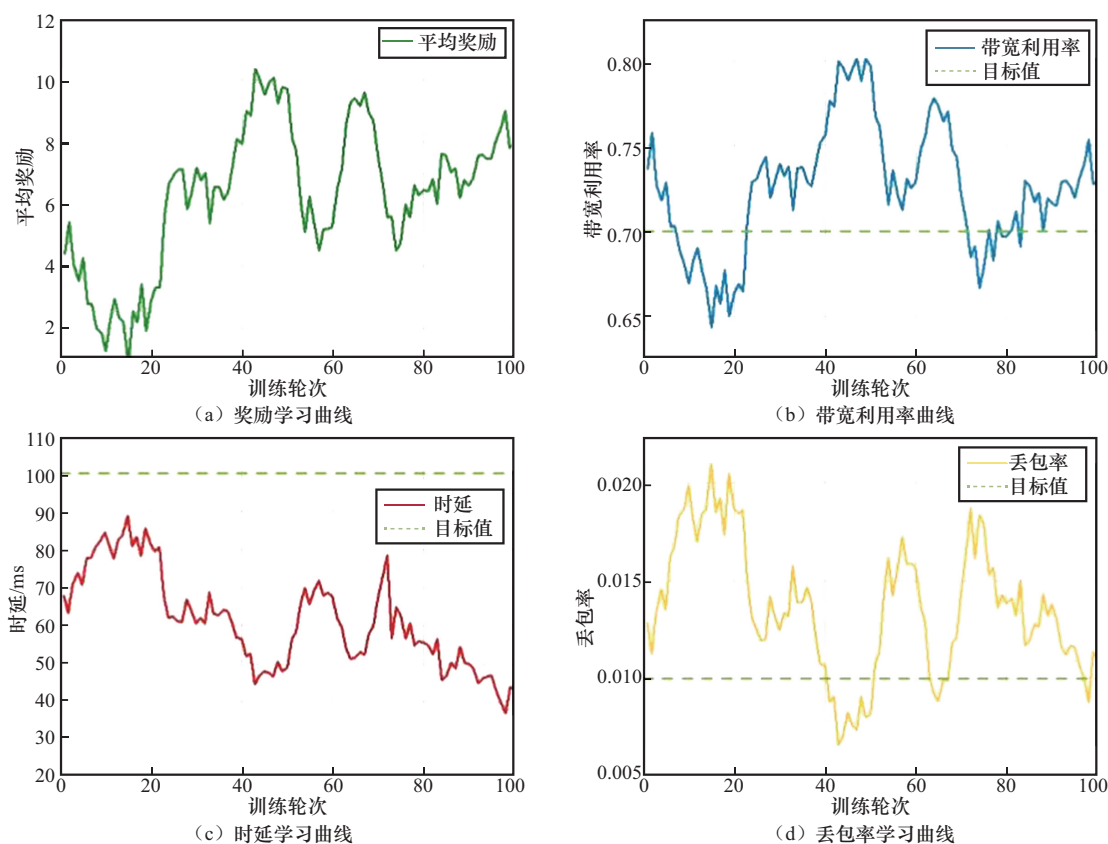


图3 强化学习训练过程



的拥塞控制与路径可靠性保障能力。带宽利用率曲线在理想区间内平稳波动,既避免了链路过载,也减少了资源闲置,实现了高效的负载均衡。

(3) 多目标优化能力

实验结果表明,本文所提算法成功地解决了网络优化中的多目标权衡问题,如图4所示,系统在保持高带宽利用率的同时实现了低时延和低丢包率,打破了传统方法中高利用率导致高时延的困境。这种优化效果源于深度强化学习算法对环境的深度感知和自适应决策能力。

4.3 实验结果对比与分析

4.3.1 算法性能对比分析

为验证所提的SRv6动态优化方法的有效性,本文将所提方法与基础RL算法(DQN)以及两类经典基线算法(传统静态路由、ECMP)进行对比。实验在相同网络拓扑与业务负载条件下开展,并采用ST-HybridNet架构,同时结合 K 最短

路径候选集与流量亲和性约束,使不同方法在统一环境中接受一致的评估。算法性能对比结果如图5所示,可以看出4种算法在平均奖励、收敛速度、稳定性和整体性能方面的对比。

需要说明的是,静态路由与ECMP属于固定策略,不存在训练迭代过程,因此不具备“收敛轮次”的可比含义。为实现统一比较,本文将“奖励/得分”定义为统一的QoS评价函数:在测试阶段对每个时间步的带宽利用率、时延、丢包率及稳定性进行综合打分并求平均,因此静态路由与ECMP在图5中表现为固定基线,而图5的收敛轮次仅对学习型方法给出(静态路由与ECMP记为N/A)。如图5所示,本文方法整体平均奖励高于DQN算法,并显著优于静态路由与ECMP算法,说明其能够更有效地学习网络状态并实现多指标权衡优化。图5给出了学习效率对比:在相同收敛判据下(移动平均进入稳定区

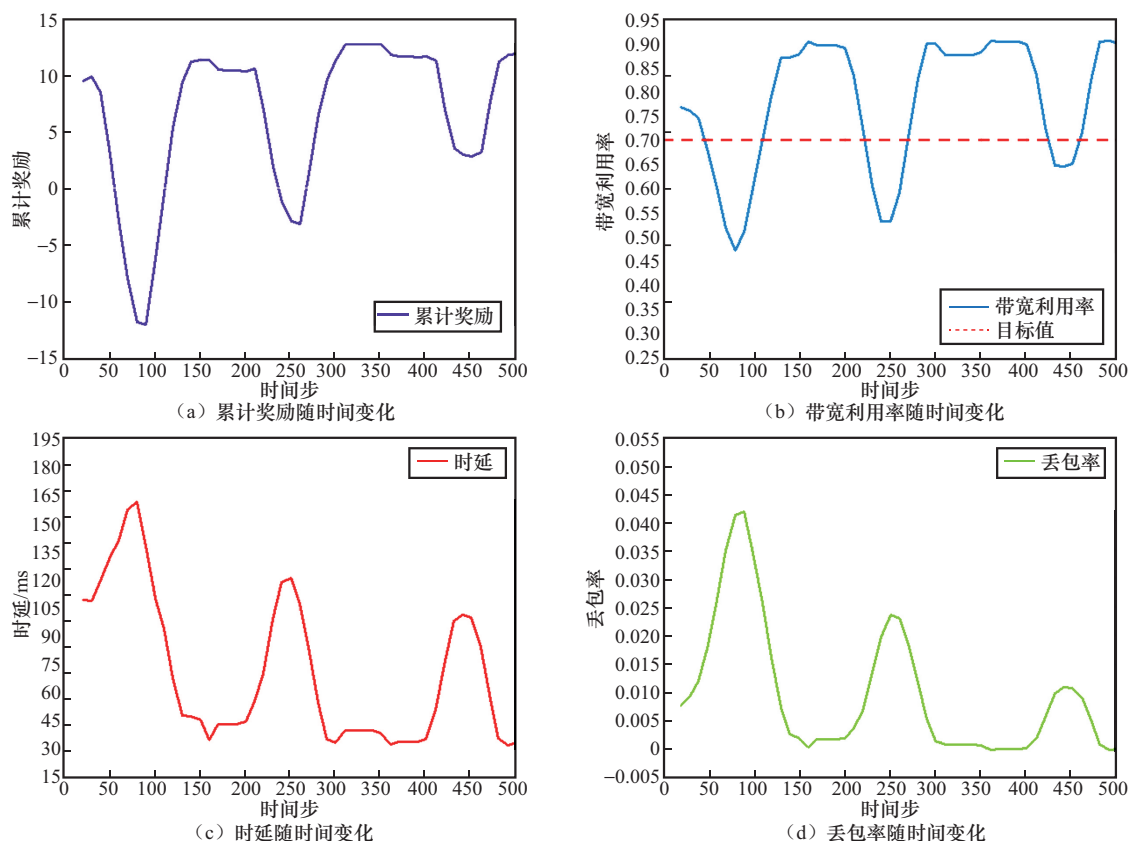


图4 网络性能时序变化

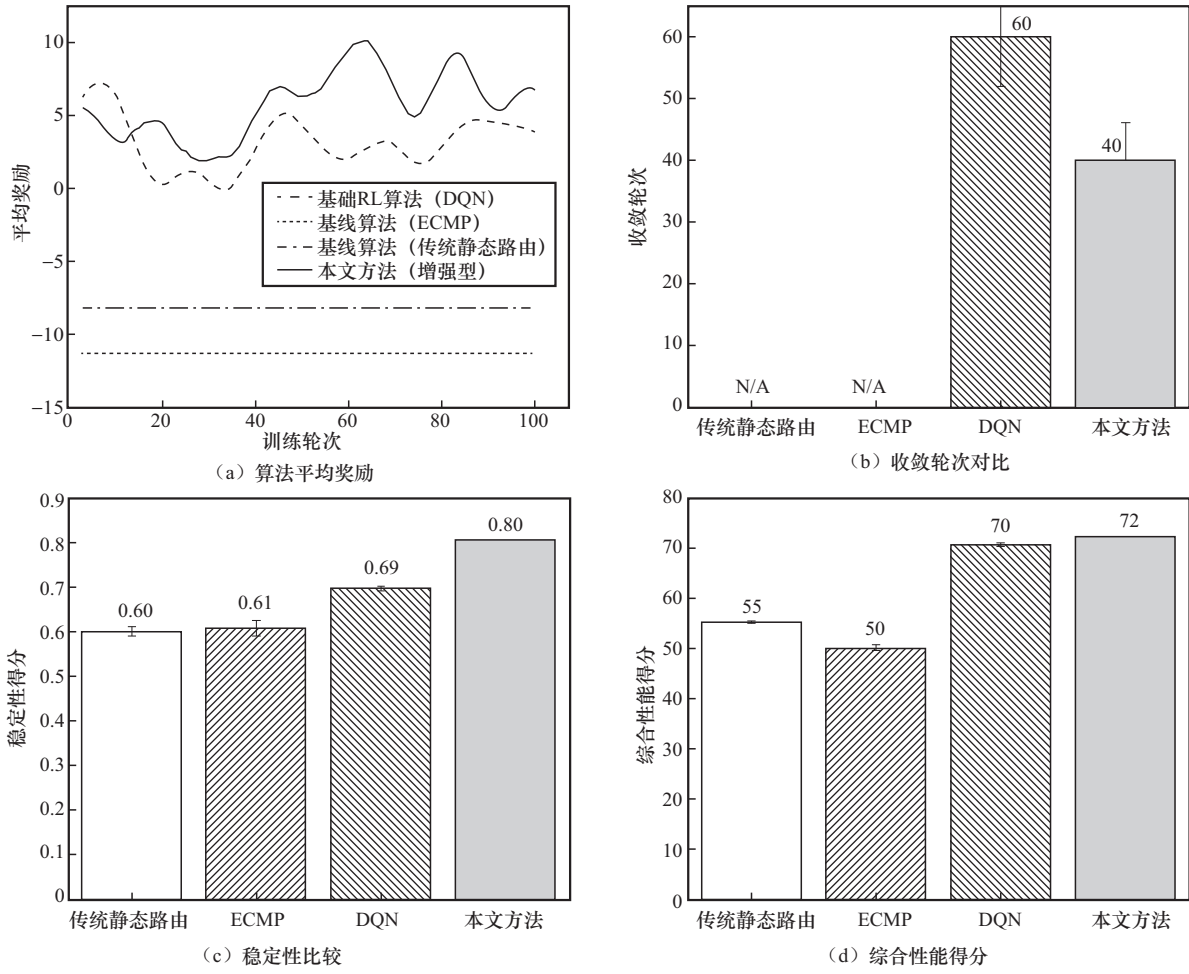


图5 算法性能对比结果

间并保持一段连续轮次不再显著上升), DQN的收敛轮次约为60, 而本文方法约为40, 减少了20轮, 表明本文方法能更快地获得稳定策略。

因为系统稳定性是评价网络优化算法的重要评价指标, 所以本实验引入了稳定性评分机制。设评估窗口内动作(路径)序列为 $a_1, a_2, \dots, a_T$, 其中, a_t 表示第 t 个时间步所选择的动作或路径, T 表示评估窗口内的时间步总数。路径切换总次数定义为

$$N_{sw} = \sum_{t=2}^T \mathbb{I}(a_t \neq a_{t-1}) \quad (14)$$

其中, $\mathbb{I}()$ 为指示函数, 当 $a_t \neq a_{t-1}$ 时取值为1, 表示相邻两个时间步之间发生了一次路径切换, 否则取值为0。 N_{sw} 表示评估窗口内的路径切换总次

数。进一步定义稳定性得分为

$$S = 1 - \frac{N_{sw}}{T-1} \quad (15)$$

其中, $S \in [0, 1]$, $\frac{N_{sw}}{T-1}$ 表示路径切换比例。由于 T 个时间步最多发生 $T-1$ 次相邻路径切换, 因此, S 为路径保持稳定的程度, S 越大表示路径切换次数越少, 策略越稳定。图5显示本文方法稳定性得分为0.80, 高于DQN算法的0.69, 明显优于静态路由与ECMP算法(0.60/0.61), 说明动作平滑与流量亲和性约束能有效降低不必要的切换与路径振荡。

进一步地, 本文给出综合性能得分(0~100分)用于统一评价多指标: 对带宽利用率采用正



向归一化、对时延与丢包率采用反向归一化，并与稳定性得分加权求和得到。

$$\text{Score} = 100 * (w_U \hat{U} + w_D \hat{D} + w_P \hat{P} + w_S S) \quad (16)$$

其中， $w_U + w_D + w_P + w_S = 1$ ，权重与奖励函数中的QoS权重保持一致。图5中本文方法综合得分为72，高于DQN的70，且显著优于综合得分55的静态路由与综合得分50的ECMP，表明本文方法在多指标权衡意义下取得了最佳的综合表现。

4.3.2 不同方法关键指标随时间变化对比

为进一步验证本文方法在动态业务环境中的实际优化效果，本文对不同方法在同一动态流量场景中的关键指标随时间变化进行了对比，不同方法关键指标随时间变化对比如图6所示。图6给出了带宽利用率、丢包率、奖励值与网络时延4个指标在时间维度上的变化曲线，能够直观地反映各方法对突发流量与链路状态扰动的自适应

能力。为保证统计可靠性，所有方法在相同拓扑与相同动态流量序列下重复运行3次，图中曲线为3次结果的均值。

如图6所示，本文方法在带宽利用率方面整体优于DQN与传统静态路由：其曲线在大多数时间步保持较高水平，同时波动幅度相对更小，说明本文方法能够更有效地实现链路负载分担与资源利用提升。在丢包率方面，本文方法整体维持在更低水平，并能在负载变化较大的阶段保持更好的鲁棒性，表明其通过合理的路径选择与负载分配机制有效地缓解了拥塞引起的丢包。在网络时延方面，传统静态路由在部分区间呈现较高的平台式时延，而本文方法能够在多数时间步维持更低的时延水平，说明其具备更强的拥塞规避能力与路径优化能力。与此同时，奖励值曲线亦反映出本文方法在多指标联合权衡目标下取得了

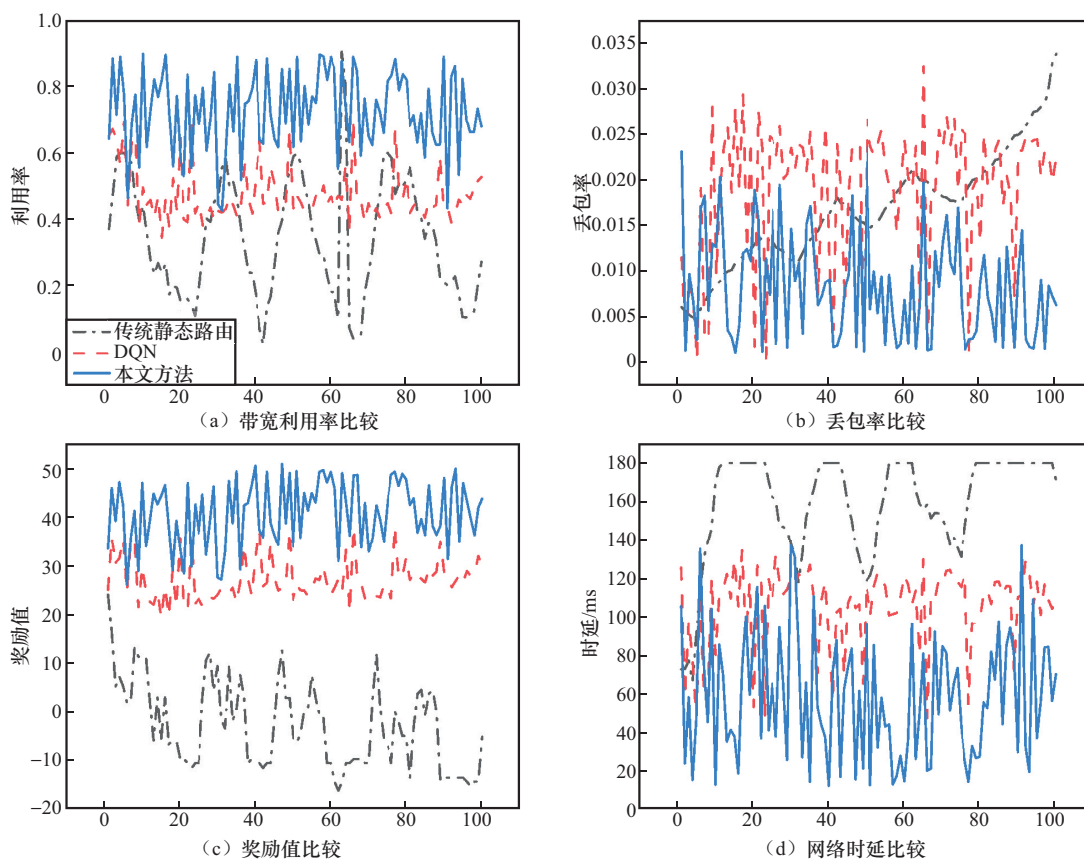


图6 不同方法关键指标随时间变化对比

更高的整体收益。

需要指出的是,由于实验场景引入了动态流量突发与链路状态扰动,各指标曲线呈现一定的非平稳波动属于正常现象。本文更关注在同一扰动条件下各方法的相对表现:相比传统静态路由与 DQN,本文方法在多数时间步同时实现更高带宽利用率、更低丢包率与更低时延,并体现出更好的稳定性与自适应能力。

本节以可复现的经典基线(静态路由、ECMP)和学习型基线(RL/DQN)为对比对象,用于第 1 阶段验证本文方法在动态场景中的性能与稳定性优势。现有部分动态流量工程机制(如,基于阈值触发的自适应重路由、开放式最短路径优先流量工程扩展(open shortest path first traffic engineering extensions, OSPF-TE)^[19]等)亦可在一定程度上缓解拥塞,但此类方法通常依赖周期性状态统计与规则触发,且在多指标联合权衡与参数配置一致性方面存在较大差异。为保证对比的可复现性与评估口径的一致性,本文暂未在本轮实验中展开一一对应的量化对比。后续工作将补充与典型动态流量工程机制(包括 OSPF-TE 与 ECMP 扩展策略等)的系统性对比实验,以更全面地验证本文方法的性能优势与适用边界。

5 结束语

在面向动态网络环境的 SRv6 性能优化问题上,传统方法普遍依赖静态策略或规则触发机制,难以及时反映网络状态的突发变化,并且在时延、丢包率与负载均衡等多目标之间缺乏统一权衡。针对上述问题,本文提出了一种面向卫星-地面融合场景的 SRv6 动态优化框架,将时空预测模型与增强型 Q 学习相结合,实现了“提前感知-自适应决策-闭环优化”的一体化设计。实验结果表明,该方法在端到端时延、丢包率稳定性及收敛速度等方面均优于传统静态与负载均衡策略,验证了预测引导与强化学习协同设计在

SRv6 场景中的有效性。

在算法可部署性方面,本文对核心模型的时间与空间复杂度进行了分析。结果表明,预测与策略训练阶段主要运行在控制平面,可采用离线或准实时方式持续优化;而在线决策仅须基于预测结果和策略表进行快速选择,整体复杂度随网络规模线性增长,具备良好的扩展性与工程可行性。

当然,当前研究仍存在一定局限。首先,实验主要基于仿真平台进行验证,未来需要在真实网络或大规模测试床环境中进一步评估鲁棒性与稳定性。其次,不同动态路由算法(如 OSPF-TE、ECMP 扩展等)尚未在统一平台下开展系统性对比,后续将补充更多交叉评估与应用场景实验。此外,对于网络失效、异常流量与极端负载等复杂场景,算法容错与策略安全性仍须进一步加强。

总体而言,本文工作从理论设计与实验验证两个层面证明了机器学习驱动的 SRv6 动态优化具有现实可行性与应用潜力,为卫星-地面融合网络中 SRv6 的智能化演进提供了一种可行思路,也为后续研究奠定了基础。

参考文献:

- [1] Filsfils C, Dukes D, Previdi S, et al. Segment routing architecture[J]. IEEE Communications Magazine, 2015, 53(7): 38-46.
- [2] Polverini M, Cianfrani A, Caiazzini T, et al. SRv6 meets DetNet: a new behavior for low latency and high reliability[J]. IEEE Journal on Selected Areas in Communications, 2025, 43(2): 448-458.
- [3] Liang B C, Li F L, Zheng N G, et al. AutoSRv6: configuration synthesis for segment routing over IPv6[J]. IEEE Journal on Selected Areas in Communications, 2025, 43(2): 473-483.
- [4] 赵肖河,宋丽华,王晨.基于 SRv6 的卫星网络业务承载策略[J]. 无线电通信技术, 2025, 51(2): 302-310.
Zhao X H, Song L H, Wang C. Traffic carrying strategy for satellite networks based on SRv6[J]. Radio Communications Technology, 2025, 51(2): 302-310.
- [5] Cianfrani A, Listanti M, Polverini M. Incremental deployment of segment routing into an ISP network: a traffic engineering



- perspective[J]. IEEE/ACM Transactions on Networking, 2017, 25(5): 3146-3160.
- [6] Wu Y T, Zhou J H. Dynamic service function chaining orchestration in a multi-domain: a heuristic approach based on SRv6[J]. Sensors, 2021, 21(19): 6563.
- [7] Hao P, Yang F. SRv6 strategy based on network state and behavior prediction in 5G network[C]//Proceedings of the 2023 9th International Conference on Computer and Communications (ICCC). Piscataway: IEEE Press, 2023: 1044-1048.
- [8] Kreutz D, Ramos F M V, Verissimo P E, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2015, 103(1): 14-76.
- [9] Rusek K, Suárez-Varela J, Almasan P, et al. RouteNet: leveraging graph neural networks for network modeling and optimization in SDN[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(10): 2260-2270.
- [10] Chen B, Zhu D, Wang Y W, et al. An approach to combine the power of deep reinforcement learning with a graph neural network for routing optimization[J]. Electronics, 2022, 11(3): 368.
- [11] Gunavathie M A, Umamaheswari S. Traffic-aware optimal routing in software defined networks by predicting traffic using neural network[J]. Expert Systems with Applications, 2024, 239: 122415.
- [12] Yalda K, Hamad D J, Tapus N. Network traffic prediction performance Using LSTM[J]. Romanian Journal of Information Science and Technology, 2024, 27(3/4): 336-347.
- [13] 王语乔. 基于机器学习的城域网流量预测与业务重构技术研究[D]. 北京: 北京邮电大学, 2019.
- Wang Y Q. Traffic prediction and service reconstruction in metropolitan area networks based on machine learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [14] 程伟强, 李振斌, 徐洪磊, 等. G-SRv6: 面向IPv6网络的增强型分段路由技术[J]. 电信科学, 2020, 36(10): 12-20.
- Cheng W Q, Li Z B, Xu H L, et al. G-SRv6: enhanced segment routing technology for IPv6 networks[J]. Telecommunications Science, 2020, 36(10): 12-20.
- [15] Loreti P, Mayer A, Lungaroni P, et al. SRv6-PM: a cloud-native architecture for performance monitoring of SRv6 networks[J]. IEEE Transactions on Network and Service Management, 2021, 18(1): 611-626.
- [16] Bakhare S, Mohod D S W. Evaluating the performance and challenges of machine learning models in network anomaly detection[J]. International Journal of Scientific Research in Science, Engineering and Technology, 2024, 11(3): 42-52.
- [17] Shalini K S. Decentralized machine learning for dynamic resource optimization in wireless networks using reinforcement learning[J]. Journal of Electrical Systems, 2024, 20(5s): 2025-2033.
- [18] Eppstein D. Finding the k shortest paths[J]. SIAM Journal on Computing, 1998, 28(2): 652-673.
- [19] Katz D, Kompella K, Yeung D. Traffic engineering extensions to OSPF[S]. RFC 3630, 2003.

[作者简介]



田东阳 (1985-), 男, 中国移动通信集团江苏有限公司网络管理中心工程师, 主要研究方向为算力网络、确定性网络。



王硕凡 (1998-), 男, 中国移动通信集团江苏有限公司网络管理中心工程师, 主要研究方向为人工智能、算力网络。



祁璜 (1982-), 男, 中国移动通信集团江苏有限公司网络管理中心工程师, 主要研究方向为算力网络、数据通信。



王凯 (1984-), 男, 中国移动通信集团江苏有限公司网络管理中心工程师, 主要研究方向为算力网络、互联网技术。