

一种结合DRL与OTN-OSU的节点索引排名路由优化算法

王庆福¹,康丽峰^{2*},冯英伟²,王兴国¹

(1.辽宁行政学院,沈阳 110161;2.河北建筑工程学院,河北 张家口 075000)

摘要:针对传统启发式路由方案在网络性能上的局限,以及现有深度强化学习(DRL)方法在大规模软件定义卫星光网络(SDSO)中面临的可扩展性不足和拓扑敏感性高等挑战,提出了一种结合DRL与光传送网光业务单元(OTN-OSU)的节点索引排名路由优化(NIR2OD)算法。该算法通过定义节点索引排名(NIR)指标筛选驱动节点(DN),并将深度确定性策略梯度(DDPG)框架引入路由决策过程,利用智能体学习流量负载与网络性能的关联,动态调整链路权重以逼近最优解。实验结果表明,相较于Scaleddeep算法,NIR2OD算法在抑制链路拥塞方面表现优异,将端到端时延降低了约25.09%,同时吞吐量提升了22.54%。

关键词:软件定义卫星光网络;深度强化学习;节点索引排名;路由优化

中图分类号:TN256 **文献标志码:**A **文章编号:**1002-5561(2026)03-0113-06

DOI:10.13921/j.cnki.issn1002-5561.2026.03.019

Node index ranking-based routing optimization algorithm combining DRL and OTN-OSU

WANG Qingfu¹, KANG Lifeng^{2*}, FENG Yingwei², WANG Xingguo¹

(1. Liaoning Academy of Governance, Shenyang 110161, China; 2. Hebei University of Architecture, Zhangjiakou Hebei 075000, China)

Abstract: To address the limitations of traditional heuristic routing schemes in network performance, as well as the challenges of insufficient scalability and high topology sensitivity faced by existing deep reinforcement learning (DRL) methods in large-scale software-defined satellite optical networks (SDSON), this paper proposes a node index ranking-based routing optimization (NIR2OD) algorithm combining DRL with optical transport network-optical service unit (OTN-OSU). The algorithm screens destination nodes (DN) by defining a node index ranking (NIR) metric and introduces the deep deterministic policy gradient (DDPG) framework into the routing decision process. By leveraging agents to learn the correlation between traffic load and network performance, it dynamically adjusts link weights to approximate optimal solutions. Experimental results demonstrate that, compared to the Scaleddeep algorithm, NIR2OD algorithm exhibits superior performance in mitigating link congestion, reducing end-to-end latency by approximately 25.09% while increasing throughput by 22.54%.

Key words: software-defined satellite optical networks, deep reinforcement learning, node index ranking, routing optimization

0 引言

软件定义网络(SDN)通过解耦数据平面与控制平面,为应对传统网络架构在满足多样化、严格服务质

量(QoS)需求方面的僵化问题,提供了高度灵活的网络管理范式^[1-2]。这一架构优势与卫星光网络新兴场景天然契合,通过构建软件定义卫星光网络(SDSO),有望凭借敏捷的流量调度与动态资源管理实现定制化网络服务^[3-4]。然而,如何使SDN控制器依据具体应用需求做出精细化转发决策,并利用其集中式全局视图实现路由策略的自动计算与部署,依然是该领域面临的棘手问题^[5-6]。

为提升SDN的智能决策水平,研究者们广泛探索了基于机器学习的方法^[7-9]。其中,深度强化学习(DRL)方法融合了深度神经网络的表征能力与强化学

收稿日期:2025-10-28。

作者简介:王庆福(1979—),男,硕士,副教授,主要从事计算机应用技术、网络通信、物联网和大数据等方面研究工作。主持和参与省级课题多项,发表学术论文多篇。

***通信作者:**康丽峰(1981—),女,硕士,副研究员,主要研究方向为人工智能、数据分析与处理、信息系统设计与开发等。



王庆福,康丽峰,冯英伟,等:一种结合DRL与OTN-OSU的节点索引排名路由优化算法

习的序贯决策优势,在路由优化任务中展现出显著潜力^[10-11]。DRL方法将全局网络状态数据直接输入智能体以生成路由策略,但面临网络规模扩大时数据量激增所导致的训练不稳定与收敛困难。Song Yu等人^[12]设计了多维度奖励函数以引导深度确定性策略梯度(DDPG)算法进行路由决策优化,在一定程度上提升了QoS保障能力^[13-14]。然而,现有基于DRL的流量调度研究多局限于小规模或仿真环境,在大型网络中难以同时兼顾降低传输损耗与避免网络拥塞^[15-17],于是,Liu Qiang等人^[18]提出局部 h 指数中心性(LH-index)方法,通过融合节点自身及其直接邻居的 h 指数信息,更为精确地识别和排序网络关键节点,有助于缓解该问题。但是,此类现有驱动节点(DN)选择策略在面临链路严重重叠时,仍可能导致网络性能下降,无法有效减少路由重叠与抑制链路拥塞。

针对上述挑战,本文提出一种结合DRL与光传送网光业务单元(OTN-OSU)的节点索引排名路由优化(NIR2OD)算法(下文简称NIR2OD算法),以减少路由重叠,抑制链路拥塞。

1 算法原理

NIR2OD算法的核心动机在于突破现有DN策略在处理链路重叠场景时的局限性,通过重新定义节点重要性的评估与索引排名机制,提升大规模网络环境下智能路由优化的稳定性与实效性。

1.1 网络与业务模型

本文研究的卫星光网络场景涵盖多个业务传输任务的并行调度。其承载网络为由星间弹性光管道构建的卫星光网络 $G=(V, E)$,其中 V 和 E 分别代表卫星传输节点集合以及星间链路集合。

光层建模与资源抽象:假设每个波长信道的传输容量为 c_{ch} Gbit/s。所有卫星算力节点均作为业务的入口节点,具备数据交换与转发能力。光业务单元(OSU)的容量设定为 c_{OSU} Mbit/s,作为最小业务映射粒度单位;卫星传输节点在数据转发过程中的最小交换单元为光通道数据单元(ODU k), $k \in \{0, 1, 2, 3, 4\}$,其对应速率分别约为1.25、2.5、10、40、104 Gbit/s。

业务请求定义:设第 i 种业务请求为 $r_i=(c_i, s_i, b_i, d_i)$,其中 c_i 表示业务源节点, s_i 表示目的节点, b_i 表示第 i 种业务的传输

带宽要求, d_i 表示第 i 种业务的端到端传输时延要求。

在此模型中,卫星光网络需为每条业务动态分配传输节点并规划端到端路径。传输数据所需的OSU数量计算公式为

$$n_{OSU}^i = b_i / c_{OSU} \quad (1)$$

1.2 NIR2OD算法核心机理

基于SDSON架构的NIR2OD算法原理框图如图1所示。该架构依托知识定义网络(KDN)范式构建,由知识平面、控制平面和数据平面3个功能平面协同运作^[19]。其中,控制平面与数据平面保持传统SDN架构的功能不变;仅在知识平面引入DRL算法进行改进,设计并训练一个具备策略学习能力的智能体。该智能体以多维网络状态为输入,通过与环境交互不断优化其策略函数的参数,在复杂、时变的业务负载条件下自主生成高效且鲁棒性强的路由决策。

为应对大规模SDSON中DRL路由方法所面临的可扩展性受限与拓扑敏感性高等挑战,本文提出的NIR2OD算法采用“分而治之”策略。该策略的核心思想是将复杂的全网路由优化问题解耦为2个紧密耦合但计算可管理的子任务:

1)通过基于节点索引排名(NIR)的关键驱动节点(DN)筛选算法(下文简称筛选算法),从全网中识别出对路由性能具有主导作用的DN;

2)构建基于深度确定性策略梯度(DDPG)的智能路由决策算法(下文简称智能路由决策算法),专注于对所选DN实施智能路由决策,从而显著降低策略空间维度并提升泛化能力。

算法的基本流程如下:

a)SDSON控制器采集各DN的网络状态信息,包

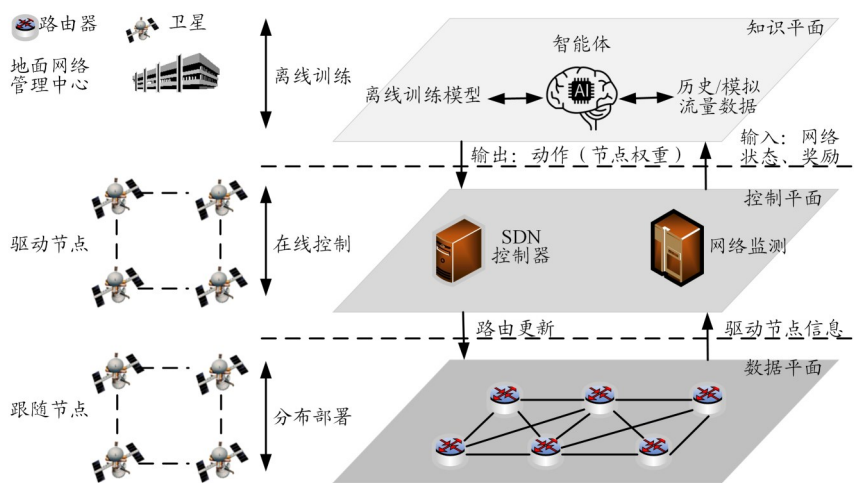


图1 基于SDSON的NIR2OD算法原理框图

括状态 S 与即时奖励 R ;

b) DN 基于当前状态,通过本地部署的 DRL 智能体计算动作 A (路由控制指令);

c) 利用动作 A 生成端到端的光路路由路径;

d) 更新后的路由配置由 SDSON 控制器统一下发至对应 DN,完成闭环调控。

1.2.1 基于 NIR 的 DN 选择

在复杂网络理论中,网络的动态行为往往由少数关键节点主导。因此,NIR2OD 算法将全网节点划分为2类:DN与从属节点(FN)。通过将路由决策聚焦于高影响力的DN,可显著压缩DRL智能体的状态空间维度,从而有效提升算法在大规模SDSON中的可扩展性与拓扑鲁棒性。

为量化节点的局部影响力,对于网络中任意节点 v ,其NIR值被定义为其一阶邻居集合中度数(连接数)严格大于 v 自身度数的邻居节点数量。基于此定义,NIR值为零的节点即为局部区域内的“局部最大度节点”,因为其所有邻居的度数均不高于自身。NIR2OD算法最终选取所有NIR值为零的节点构成DN集合。

该选择策略具有明确的网络科学依据。在现实网络(尤其是无标度网络)中,节点度数通常服从重尾分布,高度数节点倾向于相互连接(存在“富人俱乐部”现象)。因此,局部最大度数节点往往位于网络的核心位置,对信息流具有强引导作用。此外,所选DN之间的链路重叠较少,有助于均衡负载、缓解拥塞。

针对无标度网络,DN数量 D 可近似表示^[18]为

$$D \approx e^{\frac{1}{2}(1-\frac{1}{\lambda-1})} |k| \quad (2)$$

式中: λ 表示网络度指数, k 表示网络中节点的平均度。

基于上述理论,本文设计了筛选算法(见算法1)。其中, v_i 表示卫星节点集合中的单个节点, $NIR(v_i)$ 表示卫星节点 v_i 的节点指数排名, $N(v_i)$ 表示卫星节点 v_i 的邻居节点数量, V 、 E 分别表示网络 G 中的卫星节点集合、边集合。

算法1 筛选算法

输入:卫星光网络拓扑图 $G=(V,E)$

输出:卫星光网络中的DN列表和FN列表

- 1) 将卫星节点按度数升序排列;
- 2) 使网络中的卫星节点达到最小度(min-degree);
- 3) for each nodes v in the vertices V
- 4) if degree == min-degree then
- 5) Append the node v to the FN

- 6) for each node are the neighbours of v do
- 7) if the neighbour nodes' degree == min-degree then
- 8) Append the neighbour node to the FN
- 9) end if
- 10) end for
- 11) end if
- 12) end for
- 13) 将FN中的卫星节点从 V 中移除并更新 G ;
- 14) 计算卫星节点索引排名 $NIR(v_i)$;
- 15) 如果 $NIR(v_i)=0$,则将卫星节点 v 添加到DN列表中 v_i ;
- 16) 将剩余的卫星节点添加到FN列表

1.2.2 基于 DRL 的智能路由决策

在DN集合确定后,NIR2OD算法利用DRL框架执行动态路由优化。该过程被建模为1个马尔可夫决策过程(MDP),旨在通过智能体与网络环境的持续交互,学习最优的链路权重调整策略。本文采用DDPG算法作为DRL框架,具体的DDPG智能体训练流程包括经验回放、网络参数更新等细节,如算法2所示。

算法2 智能路由决策算法

输入:网络状态 s_t 、奖励 r_t

输出:动作 a_t

1) 初始化 Actor 网络、Critic 网络的 online 策略网络参数;

Actor网络基于行为策略参数 β ,在 t 时刻为状态 s_t 选择动作 a_t ,并作用于环境,环境得到 t 时刻的奖励 r_t ,并生成 $t+1$ 时刻的状态 s_{t+1} ;

2) 将状态、动作、奖励集(s_t, a_t, r_t, s_{t+1})作为经验回放缓存数据集;

3) 随机选取 $N^*(s_t, a_t, r_t, s_{t+1})$ 数据集作为 mini-batch 数据,用于训练 Actor 网络、Critic 网络的 online 策略网络;

4) 计算 Critic 网络的 online 策略网络参数梯度;

5) 使用类似监督学习的方法将Q网络的损失定义为均方误差(MSE);

6) 使用优化器更新 Critic 网络的 online 策略网络参数 θ^Q ;

7) 为 Actor 网络的 online 策略网络参数 θ^A 计算梯度;

8) 使用优化器更新 Actor 网络的 online 策略网络参数 θ^A ;

9)更新 Actor 网络、Critic 网络的 target 策略网络参数 θ' 、 θ'' ;

10)end for step

11)end fore pisode

1.2.3 基于 OSU 的带宽动态调整机制

为了支撑 NIR2OD 算法的高效执行,本文利用 OSU 技术的支路端口(TPN)承载机制,实现了轻量级的无损带宽调整。与传统 OTN 需逐跳配置中间节点不同,OSU 允许目的端自动感知源端有效载荷块数量的变化并同步更新,无需人工干预传输路径上的中间节点。基于此特性,本文设计了 DN 与跟随节点的协同调整功能:NIR2OD 算法的 DRL 智能体仅需向控制平面中的 DN 下发带宽调整指令,随后由 DN 利用 OSU 带内开销与跟随节点自动交互调整信息。这种机制将复杂的端到端配置简化为单点指令下发,显著降低了运维复杂度与信令开销。

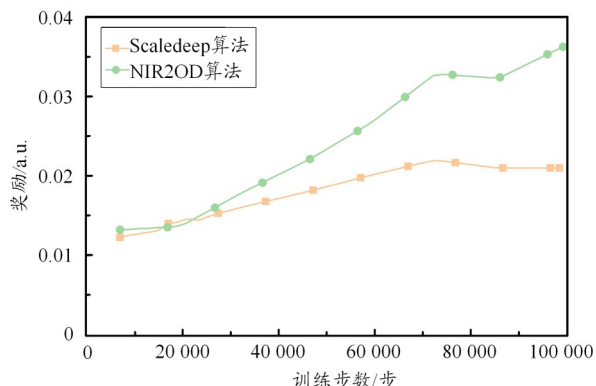
2 实验设计与性能分析

本文在 Ubuntu 18.04 TLS 系统下,利用 OMNeT++ 6.1.0 和 STK 13.0 对包含 12 个节点的类 ABILENE 及 40 个节点的类 GEANT 2 种卫星光网络拓扑(星间链路带宽均为 20 Gbit/s,拓扑以邻接矩阵输入)进行仿真实验,并从训练奖励、丢包率、时延和吞吐量 4 个方面对比了 NIR2OD 算法与基准 Scaleddeep 算法^[20-23]的性能。

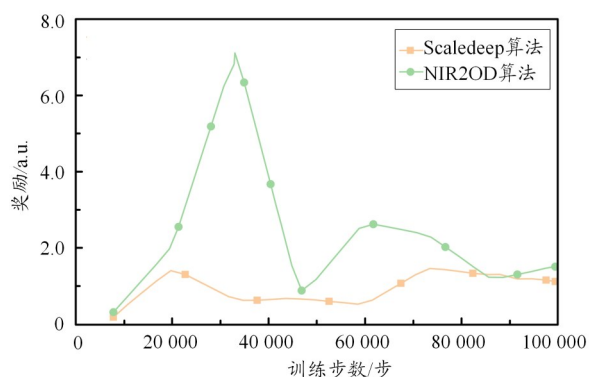
2.1 训练奖励对比

在 2 种网络拓扑结构下,2 种算法的训练奖励随训练周期的变化趋势如图 2 所示。可以看出,相较于 ScaleDeep 算法,NIR2OD 算法在 2 类拓扑中均获得了更高的累积训练奖励,表明其智能体能够更有效地学习到针对核心 DN 的权重调度策略,进而提升整体路由效率。

此外,网络拓扑结构的复杂性对 2 种算法的收敛过程均产生显著影响,但影响程度存在差异:在类 ABILENE 网络拓扑中,由于任意两节点间仅存在 2~3 条不相交路径,DN 之间路径耦合效应较强,NIR2OD 算法因权重微调机制仍会引发轻微策略振荡,收敛速度约为 80 000 步;而 ScaleDeep 算法受限于离散式权重更新与 DN 链式聚集特性,振荡幅度更大、收敛更缓慢,且最终奖励值明显低于 NIR2OD 算法。在类 GEANT 网络拓扑中,高连接冗余度提供了更多可选路径,有效抑制了奖励信号波动,具体表现如下:NIR2OD 算法凭借连续动作空间与流量隔离机制,收敛速度显著提升(约 60 000 步)且奖励曲线平稳;ScaleDeep 算法虽因



(a) 类 ABILENE 网络



(b) 类 GEANT 网络

图 2 2 种算法训练过程中的累积奖励

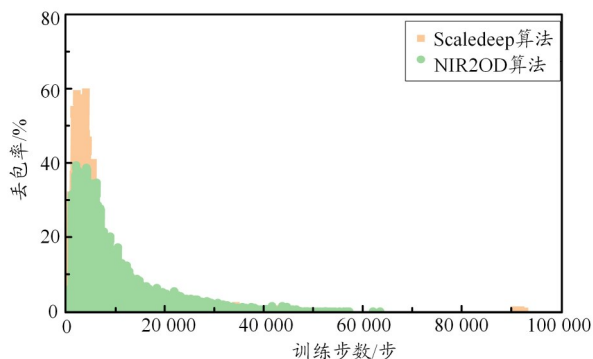
路径冗余而加速收敛,但其奖励值波动范围仍较大,稳定后的累积奖励仍低于 NIR2OD 算法。

2.2 丢包率对比

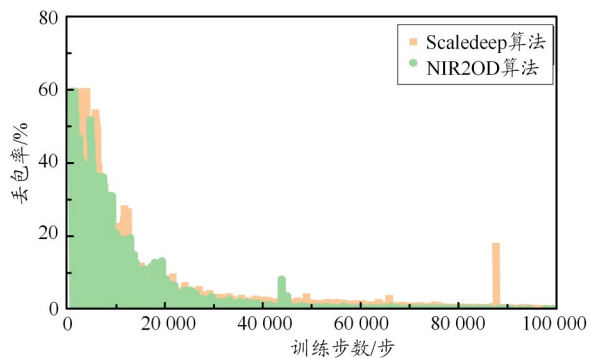
2 种算法在训练过程中的丢包率变化趋势如图 3 所示。可以看出,在类 ABILENE 网络拓扑中,NIR2OD 算法经约 60 000 步训练后,丢包率由 40% 降至 1%;在类 GEANT 网络拓扑中,经约 80 000 步训练,丢包率由 60% 降至 2%。相比之下,ScaleDeep 算法在 2 种网络中丢包率波动均幅度较大,收敛性较差。主要原因是 2 种算法在 DN 选择机制上的本质不同:NIR2OD 算法在筛选 DN 时,确保所选节点之间不存在直接邻居关系,从而实现有效的流量隔离;而 ScaleDeep 算法基于度中心性选取的节点倾向于在高连接区域形成链式聚集,导致多条高负载业务流被迫共享同一链路,显著增加了资源冲突概率。

2.3 端到端时延对比

2 种算法在训练过程中的端到端时延对比变化趋势如图 4 所示。可以看出,相较于 ScaleDeep 算法,NIR2OD 算法的平均端到端时延降低了约 25.09%。ScaleDeep 算法的时延表现相对稳定,但在训练完成后

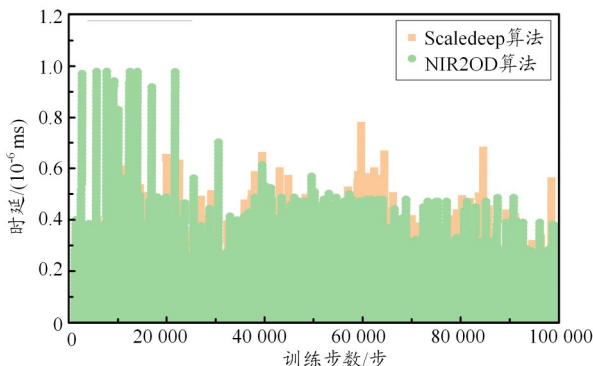


(a) 类 ABILENE 网络

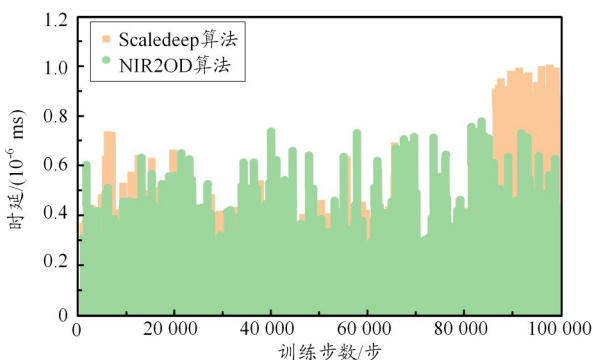


(b) 类 GEANT 网络

图3 2种算法的训练过程中的业务损失



(a) 类 ABILENE 网络



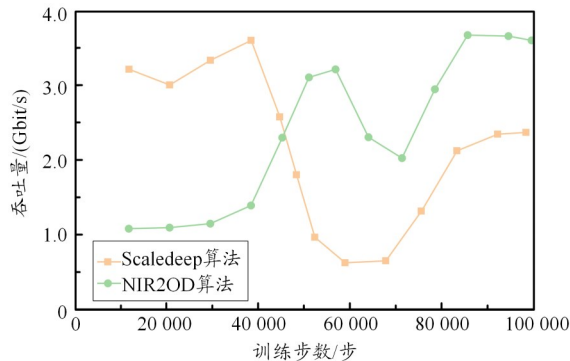
(b) 类 GEANT 网络

图4 2种算法的训练过程中的业务时延

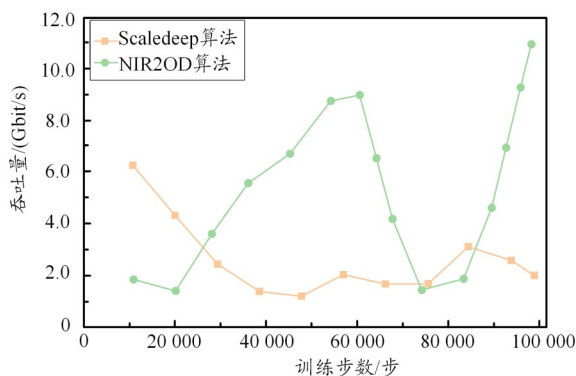
未呈现显著下降趋势。造成该差异的主要原因在于2种算法在链路权重调整策略上的根本区别:NIR2OD算法通过 Actor对DN的链路权重进行连续微调,调整幅度虽小但方向精确,有效引导时延单调递减并避免策略振荡;而ScaleDeep算法采用离散式权重更新机制,调整粒度较粗,难以实现精细优化。此外,NIR2OD算法的DN覆盖了更广泛的端到端路径集合,表明其权重调整能够直接影响更多源-目的流量对,从而提升全局优化效率。

2.4 吞吐量对比

2种算法训练过程中的业务吞吐量与训练步数的关系如图5所示。根据图形计算可得,在2种网络拓扑下,NIR2OD算法的平均吞吐量较ScaleDeep算法均提升了22.54%。该性能增益主要源于NIR2OD算法在DN选择与策略探索机制上的协同优势:所选DN承载了网络中较高比例的流量,Actor通过对这些关键节点链路权重的优化(如增大带宽分配权重),能够直接缓解全局瓶颈链路的拥塞;此外,NIR2OD算法引入的随机噪声在训练初期有效激励智能体探索高吞吐量的配置空间,增强策略多样性,而在训练后期则逐渐



(a) 类 ABILENE 网络



(b) 类 GEANT 网络

图5 2种算法的训练过程中的业务吞吐量与训练集关系图

收敛并稳定于最优解邻域,保障了学习过程的稳定性与收敛质量。

3 结束语

本文针对SDSON在大规模、高动态环境下路由优化面临的可扩展性差、收敛困难及链路拥塞严重等核心挑战,提出了一种NIR2OD算法。通过理论建模与仿真实验,得出以下结论:相较于Scaleddeep算法,NIR2OD算法通过DN流量隔离机制,将平均端到端时延降低了25.09%,吞吐量提升了约22.54%,且收敛速度更快,有效解决了大规模网络路由优化中的拥塞与收敛难题。

参考文献:

- [1] Ruffo D S V G, Lent D M B, Komarchesqui M, et al. Anomaly and intrusion detection using deep learning for software-defined networks: a survey[J]. Expert Systems with Applications, 2024, 256: 124982.
- [2] Maulana S, Anjani S A, Sanjaya Y P A, et al. Software-defined network: revolutionizing network management and optimization[J]. Journal of Computer Science and Technology Application, 2024, 1(2): 164–171.
- [3] 周伟, 罗青松, 谢晓兰, 等. 基于OISL-OSU的通算一体卫星路由优化算法[J]. 光通信技术, 2024, 48(5): 51–59.
- [4] 艾琪洋, 何荣希, 林润峰. 卫星光网络中基于DQN的匹配度感知波长路由算法[J]. 光通信技术, 2024, 48(3): 30–37.
- [5] 赵永利, 张会彬, 郑昆朋, 等. 大规模空间光传送网架构与关键技术(特邀)[J]. 光学学报, 2025, 45(13): 171–186.
- [6] Isyaku B, Bakar K A. Managing smart technologies with software-defined networks for routing and security challenges: a survey[J]. Computer Systems Science and Engineering, 2023, 47(2): 1839–1879.
- [7] Salau A O, Beyene M M. Software defined networking based network traffic classification using machine learning techniques[J]. Scientific Reports, 2024, 14(1): 20060.
- [8] Kumar C L, Betam S, Pustokhin D, et al. Metaparameter optimized hybrid deep learning model for next generation cybersecurity in software defined networking environment[J]. Scientific Reports, 2025, 15(1): 14166.
- [9] Serag R H, Abdalzaher M S, Elsayed H A E A, et al. Software defined network traffic classification for QoS optimization using machine learning[J]. Journal of Network and Systems Management, 2025, 33(2): 1–29.
- [10] Pei Xinglong, Sun Penghao, Hu Yuxiang, et al. Enabling efficient routing for traffic engineering in SDN with deep reinforcement learning[J]. Computer Networks, 2024, 241: 110220.
- [11] Sanchez L P A, Shen Yao, Guo Minyi. DQS: a QoS-driven routing optimization approach in SDN using deep reinforcement learning[J]. Journal of Parallel and Distributed Computing, 2024, 188: 104851.
- [12] Song Yu, Qian Xusheng, Zhang Nan, et al. QoS routing optimization based on deep reinforcement learning in SDN[J]. Computers, Materials & Continua, 2024, 79(2): 3007–3021.
- [13] Yu Changhe, Lan Julong, Guo Zehua, et al. DROM: optimizing the routing in software-defined networks with deep reinforcement learning[J]. IEEE Access, 2018, 6: 64533–64539.
- [14] Sumiea E H, Abdulkadir S J, Alhussian H S, et al. Deep deterministic policy gradient algorithm: a systematic review[J]. Heliyon, 2024, 10(9): e29658.
- [15] Sun Penghao, Guo Zehua, Li Junfei, et al. Enabling scalable routing in software-defined networks with deep reinforcement learning on critical nodes[J]. IEEE/ACM Transactions on Networking, 2021, 30(2): 629–640.
- [16] He Qiang, Wang Yu, Wang Xingwei, et al. Routing optimization with deep reinforcement learning in knowledge defined networking[J]. IEEE Transactions on Mobile Computing, 2023, 23(2): 1444–1455.
- [17] Kim G, Kim Y, Lim H. Deep reinforcement learning-based routing on software-defined networks[J]. IEEE Access, 2022, 10: 18121–18133.
- [18] Liu Qiang, Zhu, YuXiao, Jia Yan, et al. Leveraging local h-index to identify and rank influential spreaders in networks[J]. Physica A: Statistical Mechanics and its Applications, 2018, 512: 379–391.
- [19] Mestres A, Rodriguez-Natal A, Carner J, et al. Knowledge-defined networking[J]. ACM SIGCOMM Computer Communication Review, 2017, 47(3): 2–10.
- [20] Venkataramani S, Ranjan A, Banerjee S, et al. Scaleddeep: a scalable compute architecture for learning and evaluating deep networks[C]//ACM. Proceedings of the 44th Annual International Symposium on Computer Architecture. New York: ACM, 2017: 13–26.
- [21] He Xinlei, Xu Guowen, Han Xingshuo, et al. Artificial intelligence security and privacy: a survey[J]. Science China Information Sciences, 2025, 68(8): 1–90.
- [22] Zhang Feng, Zhang Chenyang, Guan Jiawei, et al. Breaking the edge: enabling efficient neural network inference on integrated edge devices[J]. IEEE Transactions on Cloud Computing, 2025, 13(2): 694–710.
- [23] Qararyah F, Maleki M A, Trancoso P. MCExplorer: exploring the design space of multiple compute-engine deep learning accelerators[J]. ACM Transactions on Architecture and Code Optimization, 2025, 22(4): 1–27.