

引用本文: 袁帅, 张慧, 蔡安亮, 等. 基于自注意力深度强化学习的特定流路由选择算法[J]. 光通信技术, 2024, 48(3): 7-12.

基于自注意力深度强化学习的特定流路由选择算法

袁帅¹, 张慧², 蔡安亮¹, 沈建华^{1*}

(1. 南京邮电大学 通信与信息工程学院, 南京 210003; 2. 深圳赛柏特通信技术有限公司, 广东 深圳 518000)

摘要: 为了有效降低传统流量工程机制中重路由对网络带来的负面影响, 基于软件定义网络的全局网络视角和管理能力, 提出一种基于自注意力深度强化学习的特定流路由选择算法, 以重新路由少量流量达到接近最优的性能。通过多尺度融合注意力机制的神经网络模型来提取流量的特征, 并采用集中式训练-分布式执行架构, 根据观测网络状态做出实时决策。理论研究和实验结果表明, 与传统深度强化学习算法与启发式算法相比, 所提算法在平均负载和端到端延迟性能方面均有显著改进。

关键词: 软件定义网络; 多智能体深度强化学习; 流量工程; 负载均衡

中图分类号: TN91 **文献标志码:** A **文章编号:** 1002-5561(2024)03-0007-06

DOI: 10.13921/j.cnki.issn1002-5561.2024.03.002

开放科学(资源服务)标识码(OSID): 

Specific flow routing selection algorithm based on Self-Attention deep reinforcement learning

YUAN Shuai¹, ZHANG Hui², CAI Anliang¹, SHEN Jianhua^{1*}

(1. School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. CypressTel Shenzhen Communication Technology Company, Shenzhen Guangdong 518000, China)

Abstract: To effectively mitigate the negative impact of rerouting on the network in traditional traffic engineering mechanisms, this paper proposes a specific flow routing selection algorithm based on Self-Attention deep reinforcement learning, leveraging the global network perspective and management capabilities of software-defined networking, to reroute a small amount of traffic and achieve near-optimal performance. A neural network model with multi-scale fusion attention mechanism is used to extract features of traffic, and a centralized training distributed execution architecture is adopted to make real-time decisions based on the observed network state. The theoretical research and experimental results show that compared with traditional deep reinforcement learning algorithms and heuristic algorithms, the proposed algorithm has significant improvements in average load and end-to-end delay performance.

Key words: software-defined networking, multi-agent deep reinforcement learning, traffic engineering, load balancing

0 引言

流量工程(TE)是软件定义网络(SDN)中的关键技

收稿日期: 2024-01-19。

基金项目: 国家自然科学基金青年基金项目(62301284)资助; 南京邮电大学企业委托研发重点课题(KH0020322072)资助。

作者简介: 袁帅(1999—), 男, 硕士研究生, 现就读于南京邮电大学电子信息专业, 主要研究方向是软件定义网络中的负载均衡和路由优化, 参与了南京邮电大学横向科研课题“新一代智能 SD-WAN 体系架构及其关键技术”研发。

***通信作者:** 沈建华(1972—), 男, 工学博士, 教授, 硕士生导师, 主要研究方向为高速率大容量光纤通信系统、智能光网络等。



术, 其通过管理和调整骨干网络中的路由配置来分配和调控网络中的流量, 对于优化网络性能和提高资源使用效率至关重要。传统的 TE 方法需要定期重新路由网络中的流, 这些流被定义为数据传输的路径, 旨在平衡链路上的负载并减少网络拥堵的可能性^[1-6]。流路由问题的定义通常基于特定的性能指标, 如延迟、带宽利用率等, 并致力于通过优化目标函数来寻找最佳的路由方案。在给定流量矩阵的情境下, 目标常常是重新路由网络中所有的流, 以最小化网络中的最大链路利用率。尽管这种方式能有效提升网络性能, 但它往往忽视了重路由可能给网络带来的负面效应, 特

袁帅,张慧,蔡安亮,等:基于自注意力深度强化学习的特定流路由选择算法

别是当重路由导致流量过大时,可能会引发网络不稳定或服务中断等问题;同时,无限制的重路由流量会对 SDN 控制器造成很大的负担。相关文献研究表明,在重新路由多个传输控制协议(TCP)流之后,吞吐量可能会下降一半^[7]。传统的等价多路径(ECMP)算法使用哈希函数分配网络流,但无法感知网络拥塞,因此缺乏实时性。当网络中存在大流时,ECMP 可能会将多条大流分配在同一路径上,导致网络拥塞,从而使得网络吞吐量大幅下降。

解决这类问题的有效方法之一是采用 ECMP 来转发大部分的流量,并利用 SDN 控制器有针对性地重新规划一小部分特定流量的路由,以平衡网络链路的利用率。这些特定的流量通常是对网络性能产生重要影响的流量^[8-9]。ECMP 通过平均分配等成本路径上的流量来降低拥塞概率,而重新特定流路由则旨在通过减少网络干扰来实现更高的性能提升。然而,ECMP 无法感知网络拥塞状态,容易发生哈希冲突,从而导致网络拥塞和性能下降。为此,本文提出一种基于自注意力深度强化学习的特定流路由选择算法(下文简称 SFRS 算法),并通过应用线性规划(LP)方法^[10]来优化流量分配,从而有效地平衡网络的链路利用率,实现最优的系统性能。

1 SFRS 算法

1.1 系统建模

智能体将状态 $s_t=M_t$ 作为输入, M_t 是时间步 t 的流量矩阵,矩阵的横纵维度分别代表源节点与目的节点编号,具体元素代表每个流的流量需求信息。

对于每个状态 s_t ,算法将选择 K 个特定的流。在具

有 N 个节点的网络中总共有 $N(N-1)$ 种不同源节点与目的节点的流,需要大小为 $C_{N(N-1)}^K$ 的动作空间。将动作空间定义为 $\{0,1,\dots,N(N-1)\}$,并允许智能体在每个时间步 t 中采样 K 个不同的动作 $(a_t^1, a_t^2, \dots, a_t^K)$ 。

奖励值用于评估神经网络提供的动作质量,它能够反映出智能体在当前网络环境下所做出的动作的优劣。本文算法以最大链路利用率作为综合评价指标,在执行动作流量选取之后,利用 LP 重路由方式来重新路由流量并获得相应奖励^[10]。奖励值计算如式(1)所示。

$$r = \frac{1}{U} \quad (1)$$

其中, r 表示奖励值, U 表示最大链路利用率。在计算出奖励值后,该值会被反馈给智能体,以便智能体根据这个奖励值来调整多路径路由中的链路权重和流量分配。随着模型训练过程的进行,奖励值会随着训练步骤的增加而逐步累积。通过观察累积奖励值的增长趋势和总奖励值,可以评估训练模型的收敛效率。

强化学习的目标是找到所有可用动作的概率分布 $\pi(a_t|s_t)$ 并输出最优策略。由 θ 参数化的随机策略 $\pi(a_k|s_t)$ 可表示为

$$\pi(a_k|s_t;\theta) \approx \prod_{i=1}^K \pi(a_t^i|s_t;\theta) \quad (2)$$

其中, a_k 表示 K 个采样动作的组合, $a_k = (a_t^1, a_t^2, \dots, a_t^K)$ 。

1.2 基于多智能体的强化学习算法

SFRS 算法采用集中式训练与分布式执行(CTDE)框架^[11],旨在优化训练与执行的速度与效率,其结构如图 1 所示。每个智能体的价值网络负责拟合全局状态

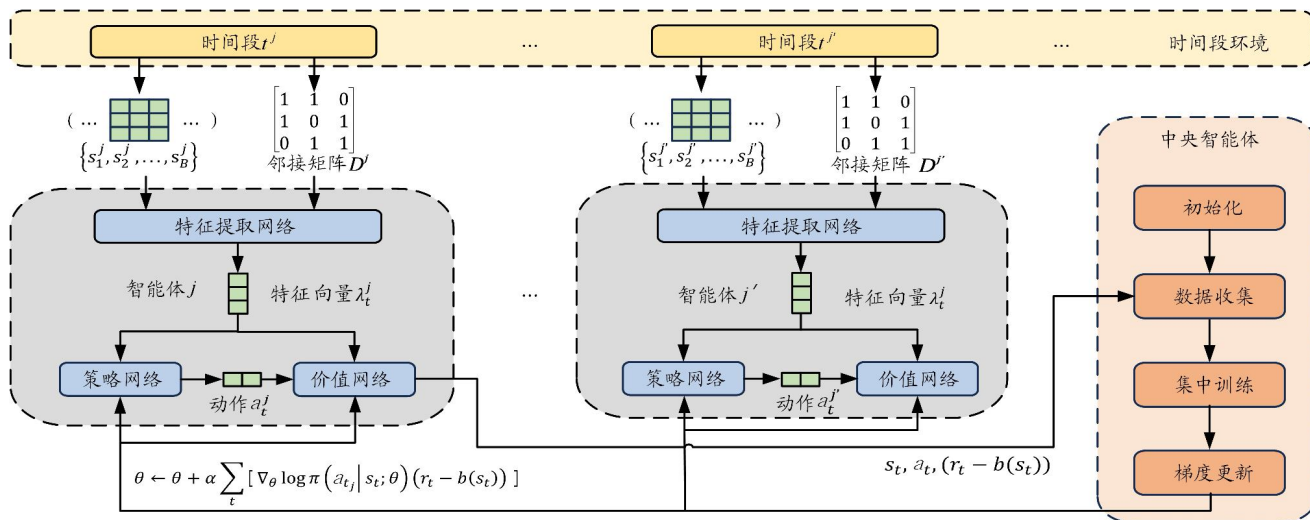


图 1 SFRS 算法中的 CTDE 结构

价值函数以评估环境状态,而策略网络则根据智能体的局部观测来输出动作,并作出决策。具体流程如下:首先,以图1左侧灰色模块智能体 j 为例,智能体 j 将一组流量矩阵 $\{s_1^j, s_2^j, \dots, s_B^j\}$ (绿色部分为流量矩阵)与邻接矩阵 D^j 输入到特征提取网络;然后,将提取出的特征向量为 λ^j 输入策略网络以获得动作 a_t^j 和奖励 r_t ,并将这些经验存储在缓冲区中。邻接矩阵 D^j 的横纵维度分别代表源节点与目的节点编号,其中节点相连的对应矩阵元素1,不相连的为0。其余智能体的工作流程均与智能体 j 相同,在训练期间,智能体通过与环境的持续交互来获取经验。

在一个回合(episode)结束后,中央智能体使用这些收集到的经验进行集中式训练。训练目标是在不同流量矩阵上优化网络性能,即要使预期奖励 $E[r_t]$ 最大化。为此,采用了包含基线 $b(s_t)$ 的蒙特卡洛策略梯度(REINFORCE)算法^[12],并通过梯度上升方法来提升 $E[r_t]$ 。策略参数 θ 根据式(3)进行更新。

$$\theta \leftarrow \theta + \alpha \sum_{i=1}^T \{ [\nabla_{\theta} \log \pi(a_k | s_i; \theta) (r_i - b(s_i))] \} \quad (3)$$

其中, $\nabla_{\theta} \log \pi(a_k | s_i; \theta)$ 表示策略的对数似然函数对参数 θ 的梯度, α 表示学习率,控制参数更新的大小, T 表示最终时间步。为了保证智能体在训练过程中能够充分探索行动空间,以便找到有效的策略,策略 π 的熵被加入到式(3)中。该熵项有助于避免过早地收敛到次优的确定性策略,从而提升探索效果^[13]。因此,式(3)可以进一步地修改为以下形式:

$$\theta \leftarrow \theta + \alpha \sum_{i=1}^T \{ \nabla_{\theta} \log \pi(a_k | s_i; \theta) (r_i - b(s_i)) + \beta \nabla_{\theta} H[\pi(\cdot | s_i; \theta)] \} \quad (4)$$

其中, H 代表策略的熵(行动概率的分布),超参数 β 用于调整熵正则化项的作用强度。

1.3 基于多尺度融合注意力机制的特征提取网络

卷积神经网络(CNN)凭借其出色的特征提取能力,能够提供一种既可扩展又富有表现力的方式,有效地将这些多样化的流量矩阵整合到选择策略中。现有基于强化学习的关键流重路由(CFR-RL)算法^[10]巧妙地运用了CNN作为策略网络和价值网络的特征提取器,这不仅简化了训练过程,还极大地便利了实现过程。然而,CNN提取的特征类型相对单一,这在一定程度上限制了其在多节点、大规模网络复杂场景中的应用。特别是在进行流量选择时,其效果并不理想。

自注意力机制^[14]可以对输入序列中的任意位置进行建模,能捕捉到输入序列中不同位置之间的依赖关系。在模型训练中,不同的流量特征对最终的流量选择的影响程度并不相同。因此,本文通过策略-价值网络结构结合自注意力机制来找到每个特征的相对重要性,实现全局和局部信息的交互,以更好地捕捉流量特征。首先,将自注意力编码器和神经网络融合,然后设计一种融合多尺度的自注意力流量编码器,以期实现流量矩阵的多尺度特征融合^[15],从而在不同场景下更好地实现动作选择和价值评估。本文提出的策略-价值网络结构如图2所示。将融合多尺度的自注意力流量编码模块直接应用在控制器收集的原始流量矩阵上,其中的流量矩阵和邻接矩阵均为三节点网

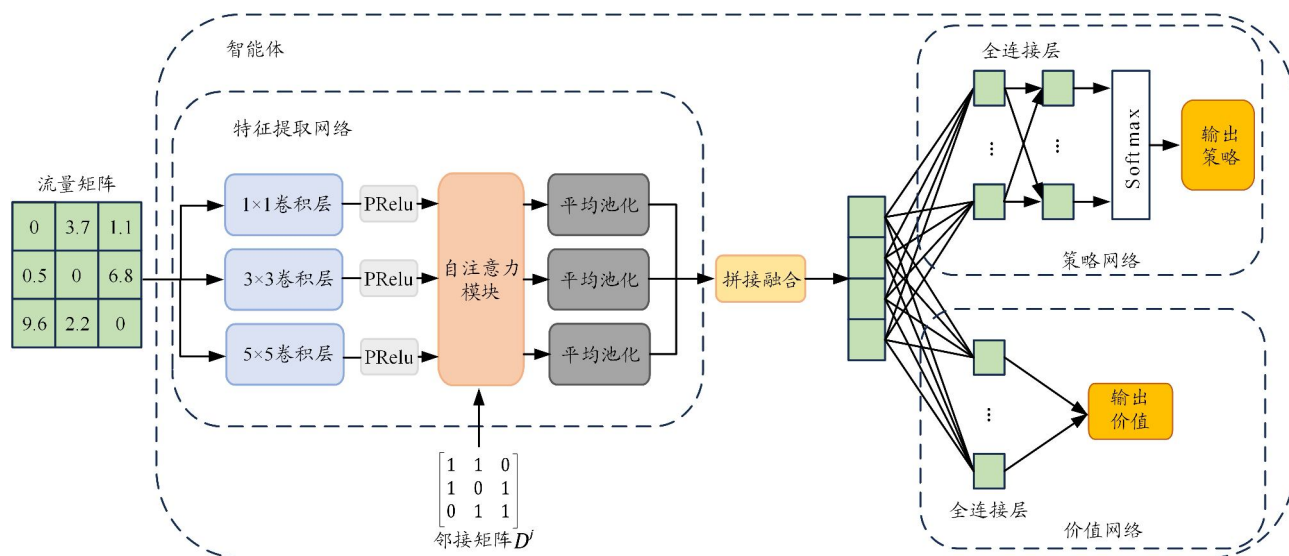


图2 策略-价值网络结构

袁帅,张慧,蔡安亮,等:基于自注意力深度强化学习的特定流路由选择算法

络矩阵。首先,在特征提取网络中,由不同尺寸卷积核的卷积层捕捉不同尺度的特征。小尺寸卷积核的卷积层可以捕捉相邻节点的流量细节特征,而较大卷积核的卷积层可以提取更广泛的流量上下文信息。这样的设计使得模型不仅能够深入理解网络流量数据的局部细节,还能够全面把握其全局特征和动态变化。然后,卷积层通过 PReLU 激活函数,其学习参数 γ 设置为 0.25;在经过激活函数之后,不同尺度的特征信息分别通过带掩码的自注意力模块,获取各个尺度下的全局上下文信息。自注意力模块中的掩码为网络节点的邻接矩阵 D^j 。最后,通过平均池化拼接融合特征,进入相应的策略和价值网络中进行特征解码。在策略网络中,先通过 2 层 128 个神经元的全连接层,再通过 1 个 Softmax 层以输出所有可用动作的概率。在价值网络中,先通过 1 个 128 个神经元的全连接层,再通过 1 个神经元个数为 1 的全连接层以输出此轮评估的价值。

2 实验与结果分析

2.1 实验设置

本文使用评估网络流量指标常用的 2 个现实世界的网络拓扑,包括 Abilene 网络和 Ebone 网络^[16],以评估算法的性能。要获取 Abilene 网络的流量矩阵和网络拓扑信息(包括链路连接性、权重和容量等),可以参阅文献[17]。由于 Abilene 流量矩阵每 5 min 测量 1 次,因此每天总共有 288 个流量矩阵。为了评估本文模型的性能,本文选择第一周(从 2004 年 3 月 1 日开始)总共 2016 个流量矩阵作为训练集,第二周(从 2004 年 3 月 8 日开始)的流量矩阵作为测试集。对于 Ebone 网络而言,相关文献中只提供了链路成本,但未给出明确的链路容量。因此,本文将链路容量推断为链路成本的倒数,这是基于 Cisco 路由器中的默认链路成本设置^[8-9,18]。由于 Ebone 网络无法直接获得流量矩阵,本文使用流量矩阵生成工具^[19]生成 100 个均匀流量矩阵,并使用其中的 70%作为训练集,30%作为测试集。

本文算法的仿真实验中,神经网络的学习率 η 最初配置为 0.000 1,并以 0.96 为基数,每 500 次迭代衰减。使用 RMSprop 作为优化器,经过反复实验,最终确定训练 epoch 为 15 000。根据文献 [20] 设置参数 $K=10\%N(N-1)$ 。算法验证的实验配置参数如下:操作系统为 Ubuntu20.04;CPU 为 i9-11900H,GPU 为 RTX3080,深度学习框架采用 TensorFlow2.13.1,最终的数据处理

采用 Python3.8。

为了评估算法的性能,本文选择了负载均衡性能比 P_U 和端到端延迟性能比 P_Ω 作为评价指标。定义如下:

$$P_U = \frac{U^{\text{optimal}}}{U} \quad (5)$$

$$P_\Omega = \frac{\Omega^{\text{optimal}}}{\Omega} \quad (6)$$

其中, U^{optimal} 表示通过所有流的最佳显式路由^[10]实现的最大链路利用率, Ω^{optimal} 表示通过对所有流进行最佳显式路由实现的最小端到端延迟。 P_Ω 越高,算法实现端到端延迟的性能越好。 $P_U=1$ 表示算法达到了最优的性能, $P_U=1$ 表示算法实现了与最优路由一样好的负载均衡。定义式(6)中的端到端延迟 Ω ^[6]为

$$\Omega = \sum_{(i,j) \in E} \left(\frac{l_{i,j}}{c_{i,j} - l_{i,j}} \right) \quad (7)$$

其中, E 表示网络中有向边的集合, $c_{i,j}$ 表示链路 (i,j) 的容量, $l_{i,j}$ 表示链路 (i,j) 的负载。

2.2 基准模型设置

本文使用 3 个基准算法作为对比,分别是基于深度强化学习算法和 2 种启发式算法:

1)CFR-RL 算法:以优化网络性能为目标,使用单一的卷积层构建深度强化学习网络进行关键流选取,并使用 LP 方法重路由。

2)TOP-K 算法:从给定的流量矩阵中挑选出需求量最高的前 K 个流量。该算法的核心假设在于,较大的网络流量会对系统性能产生显著影响。

3)Top-K Critical 算法:该算法与 Top-K 算法类似,但它是从网络中最拥堵的链路选出需求量最大的前 K 个流量。这种算法的基本假设在于,最拥堵链路上的流量会对网络性能产生显著影响。

2.3 实验结果

本文从 Abilene 网络测试集中,按照时间顺序选择了 600 个流量矩阵作为研究数据,得到不同算法的仿真结果如图 3 和图 4 所示。

从图 3 可以看出,本文提出的 SFRS 算法的平均负载均衡性能相较于 CFR-RL 算法提升了约 1.5%,相较于 Top-K Critical 算法提升了约 2%,相较于 Top-K 算法提升了约 10%,因此所提算法在 Abilene 网络中的性能最佳。在多数流量矩阵的场景下,算法的负载均衡性能比明显优于 CFR-RL 算法,这表明本文设计的基于多尺度融合注意力的编码器能够提取到更多的特征,使得智能体在训练中获取到了更多的信息。但是,

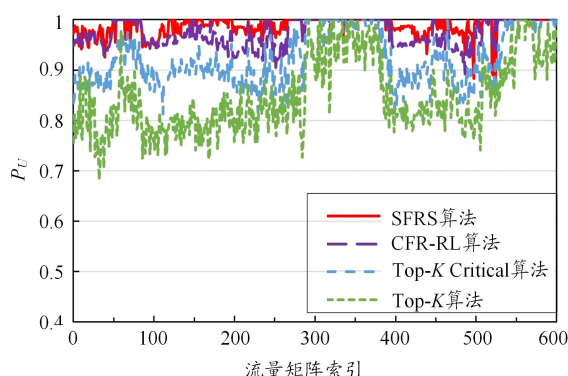


图3 Abilene网络中,不同算法负载性能比较

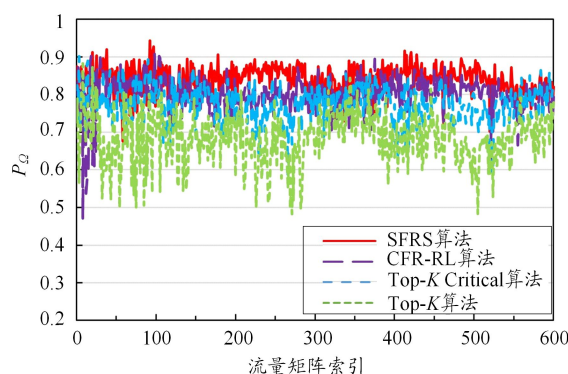


图4 Abilene网络中,不同算法端到端延迟性能比较

在个别情况下,SFRS算法仍然有陡降的情况发生,这说明该算法在应对突发情况方面仍有待加强,具备进一步优化的潜力。

从图4可以看出,CFR-RL算法在某些情况下会出现陡降的情况,导致延迟性能大幅下降。SFRS算法的端到端延迟性能相较于CFR-RL算法提升了约2.8%,相较于Top-K Critical算法提升了约3.7%,相较于Top-K算法提升了约15.6%,明显优于其它算法。

Ebone网络中,不同算法的仿真结果如图5和图6所示。由图5可知,本文提出的SFRS算法的负载性能在多数场景下优于其余3种算法,而启发式算法在流

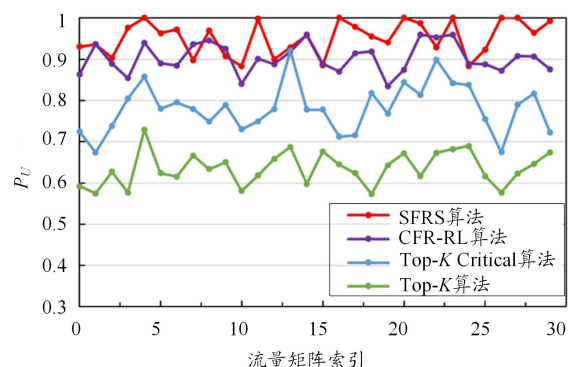


图5 Ebone网络中,不同算法负载性能比较

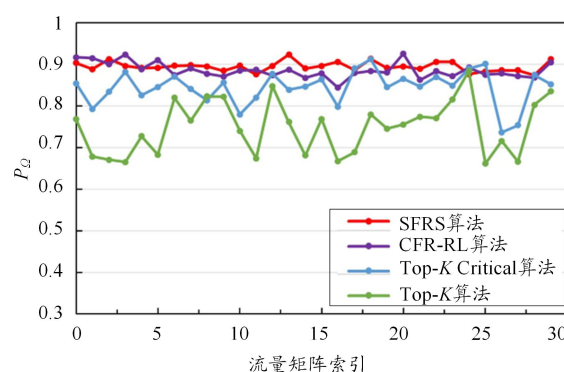


图6 Ebone网络中,不同算法端到端延迟性能比较

量较为均匀的场景中会体现出明显的劣势。这是因为流量差异不大,不存在明显的“大象”流,此时路径、传输延迟等其它因素对于流量选择的影响更大。

由图6可知,SFRS算法的端到端延迟性能总体上略有提升,但差距不大。总体而言,SFRS算法能通过为每个给定的流量矩阵智能地选择少量流量以实现接近最优的负载均衡性能。

3 结束语

本文针对SDN中的重路由问题,提出了一种基于自注意力深度强化学习的特定流路由选择算法。算法通过选择少量的流量重路由以达到接近最优的性能,能够有效解决重路由流量过大所引发的网络不稳定或服务中断等问题。仿真结果表明:融合多尺度的自注意力流量编码能够有效捕捉不同网络流量数据之间的依赖关系,与其它算法相比,该算法在平均负载和端到端延迟方面的性能最优。在未来的工作中,本文作者将尝试降低算法模型的复杂度,并综合考虑SDN网络中的链路带宽、抖动等指标,进一步提升算法性能。

参考文献:

- [1] WANG Y, WANG Z. Explicit routing algorithms for internet traffic engineering [C]//IEEE. Proceedings of International Conference on Computer Communications and Networks'99. Boston: IEEE, 1999: 582-588.
- [2] OSBORNE E D, SIMHA A. Traffic engineering with MPLS [M]. Indianapolis: Cisco Press, 2002.
- [3] FORTZ B, THORUP M. Optimizing OSPF/IS-IS weights in a changing world [J]. IEEE Journal on Selected Areas in Communications, 2002, 20(4): 756-767.
- [4] HOLMBERG K, YUAN D. Optimization of internet protocol network design and routing [J]. Networks: An International Journal, 2004, 43(1): 39-53.
- [5] CHU J, LEA C T. Optimal link weights for ip-based networks support-

袁帅,张慧,蔡安亮,等:基于自注意力深度强化学习的特定流路由选择算法

- ing hose-model vpns[J]. IEEE/ACM Transactions on Networking, 2009, 17(3): 778–788.
- [6] ZHANG J, XI K, ZHANG L, et al. Optimizing network performance using weighted multipath routing[C]//IEEE. International Conference on Computer Communications and Networks'12. Munich: IEEE, 2012: 1–7.
- [7] CARPA R, ASSUNCAO M D, GLUCK O, et al. Evaluating the impact of SDN-induced frequent route changes on TCP flows [C]//IEEE. International Conference on Network and Service Management (CNSM). Tokyo: IEEE, 2017: 1–9.
- [8] ZHANG J, XI K, LUO M, et al. Dynamic hybrid routing: achieve load balancing for changing traffic demands[C]//IEEE. International Symposium on Quality of Service'14. Hong Kong: IEEE, 2014: 105–110.
- [9] ZHANG J, XI K, LUO M, et al. Load balancing for multiple traffic matrices using SDN hybrid routing [C]//IEEE. International Conference on High-Performance Switching and Routing'14. Vancouver: IEEE, 2014: 44–49.
- [10] ZHANG J, YE M, and GUO Z, et al. CFR-RL: Traffic engineering with reinforcement learning in SDN[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(10): 2249–2259.
- [11] 李斌. 基于多智能体强化学习的多无人机边缘计算任务卸载[J]. 无线电工程, 2023, 53(12): 2731–2740.
- [12] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning [J]. Machine Learning, 1992, 8(3): 229–256.
- [13] MNIH V, BADIA A P, MIRZA M et al. Asynchronous methods for deep reinforcement learning [C]//ICML. Proceedings of Machine Learning Research (PMLR). New York: ICML, 2016: 1928–1937.
- [14] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C]//ICML. International Conference on Machine Learning(PMLR). Los Angeles: ICML, 2019: 7354–7363.
- [15] 郭丹青,符颖,朱烨,等. 自注意力多尺度特征融合的遥感图像语义分割算法[J]. 计算机辅助设计与图形学学报, 2023, 35(8): 1259–1268.
- [16] SPRING N, MAHAJAN R, WETHERALL D. Measuring isp topologies with rocketfuel [J]. ACM SIGCOMM Computer Communication Review, 2002, 32(4):133–145.
- [17] ZHANG Y. ZHANG Y's abilene TM[EB/OL]. (2021–08–02)[2024–02–20]. <http://www.cs.utexas.edu/~yzhang/research/AbileneTM/>.
- [18] KODIALAM M, LAKSHMAN T, ORLIN J B, et al. Oblivious routing of highly variable traffic in service overlays and ip backbones[J]. IEEE/ACM Transactions on Networking, 2008, 17(2): 459–472.
- [19] VICTOR H. TMgen: traffic matrix generation tool [EB/OL]. (2020–08–10)[2024–02–20]. <https://tmgen.readthedocs.io/en/latest/>.
- [20] YE M, ZHANG J, GUO Z, et al. Date: disturbance-aware traffic engineering with reinforcement learning in software-defined networks[C]//IEEE. Proceedings of 29th International Symposium on Quality of Service (IWQOS). Tokyo: IEEE, 2021: 1–10.