

一种基于深度增强学习的智能路由技术

孙鹏浩, 兰巨龙, 申 涓, 胡宇翔

(解放军战略支援部队信息工程大学, 河南郑州 450002)

摘 要: 随着网络规模的不断增大以及网络复杂度的不断提高, 传统路由算法面对网络流量在时空分布上的剧烈波动难以兼顾计算复杂度和算法效率. 近年来, 随着软件定义网络和人工智能技术的兴起, 基于机器学习的自动路由策略生成逐渐受到关注. 本文提出一种基于深度增强学习的智能路由技术 SmartPath, 通过动态收集网络状态, 使用深度增强学习自动生成路由策略, 从而保证路由策略能够动态适应网络流量变化. 实验结果表明, 本文所提出的方案能够不依赖人工流量建模动态更新网络路由, 在测试环境下比当前最优方案减少至少 10% 的平均端到端传输时延.

关键词: 路由优化; 软件定义网络; 人工智能; 深度增强学习

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2020)11-2170-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.11.011

An Intelligent Routing Technology Based on Deep Reinforcement Learning

SUN Peng-hao, LAN Ju-long, SHEN Juan, HU Yu-xiang

(PLA Strategic Support Force Information Engineering University, Zhengzhou, Henan 450002, China)

Abstract: With the expansion of network scale and network complexity, traditional routing algorithms cannot ensure both the calculation complexity and performance under the large fluctuation of spatial-temporal distribution of network traffic. In recent years, with the development of Software-Defined Networking (SDN) and Artificial Intelligence (AI), AI-based methods of automatic routing strategies are gaining attention. In this paper, we propose an intelligent network routing technology called SmartPath based on Deep Reinforcement Learning (DRL). With dynamic collection of network status, we can use DRL to generate routing policies automatically, thus ensuring that the routing policy can dynamically adapt to the change of network traffic. Experiment result shows that the proposed scheme can adjust the routing strategy dynamically without human experience on traffic analysis and can reduce the average end-to-end transmission delay by at least 10% compared with the state-of-art schemes.

Key words: routing optimization; software-defined networking (SDN); artificial intelligence (AI); deep reinforcement learning (DRL)

1 引言

近年来, 信息通信网络经历了飞速的发展, 网络连接规模的增大和网络应用种类的增多造成了网络流量复杂度不断提高, 流量的时空分布也呈现出剧烈的波动性. 面对流量的时空分布波动性, 现有传输网络采用的路由算法灵活性不足, 无法动态适应流量特征, 易造成网络资源利用不充分和网络服务质量难以保障的困境. 为解决上述问题, 学术界开展了大量基于流量工程和路由优化的

研究^[1,2]. 然而, 传统路由算法需要事先分析网络流量特征, 依赖大量人工先验性经验开展路由计算, 难以动态适应流量时空分布的波动性, 因此难以实现路由策略的灵活性. 近年来, 随着人工智能技术的蓬勃发展, 基于机器学习的网络策略自动化生成逐渐展现出了潜力. 例如, 2003 年 David 等人就提出了网络“知识平面”的概念, 前瞻性地描绘了基于人工智能技术实现网络自动化管理的蓝图^[3]. 随后, Thomas 等于 2005 年提出了认知网络^[4], Derbel 等于 2009 年提出了网络自动化管理架构 ANE-

收稿日期: 2019-09-23; 修回日期: 2019-12-23; 责任编辑: 孙瑶

基金项目: 国家自然科学基金 (No. 61521003, No. 61702547, No. 61872382); 国家重点研发计划课题 (No. 2017YFB0803204); 广东省重点领域研发计划项目 (No. 2018B010113001)

MA^[5], Michele 等于 2016 年提出了智能网络管理架构 COBANETS^[6]. 上述架构更多地论述了网络架构于人工结合的概念性构想,并没有提出具体实现方案. 2017 年, Albert 等人提出了知识定义网络的概念^[7],详细论述了基于软件定义网络实现的知识定义网络架构,并展示了多个演示场景,引起了业界的广泛关注. 然而,目前机器学习技术在网络场景下的充分使用仍然面临许多困难, Agoulmine 等人列举的网络智能化管理面临的相关问题仍然没有得到较好的解决^[8].

软件定义网络技术 (Software-Defined Networking, SDN) 的发展和成熟为机器学习与网络管理的结合提供了新的契机. 在 SDN 中,网络的控制层和数据层分离,网络管理者可以在控制层集中收集底层网络的全局信息,部署全局性策略. SDN 的集中控制逻辑一方面在数据层提供了开放的数据层编程接口,为灵活部署新型网络功能、提供细粒度网络监测提供了可能^[9,10],另一方面也为网络管理者实时监测网络状态、获取网络全局视图提供了便利,有利于部署全局优化的网络策略. 然而,充分利用控制层提供的网络全局视图和实时的网络状态信息来动态制定路由策略等网络管理策略也对信息的规模化实时处理能力提出了较高的要求. 面对这个问题,传统的优化算法无法兼顾算法运行效率和算法效果,因此无法充分利用 SDN 的集中控制逻辑优势进行网络路由.

针对上述问题,本文提出了一种基于深度强化学习的智能路由技术 SmartPath. SmartPath 基于 SDN 网络架构结合深度强化学习技术,通过控制器动态收集网络状态信息,在控制器上运行智能路由应用生成动态路由策略跟踪网络流量分布,从而达到动态智能路由效果. 本文主要贡献如下:

(1) 提出了基于机器学习的网络路由策略自动生成的具体实现方案,并且在真实网络设备上验证了所提架构的有效性.

(2) 提出了基于门控循环单元 (Gate Recurrent Unit, GRU) 的深度强化学习路由策略生成算法,实现了高效精确的模型训练过程.

(3) 提出了区分流量类型的网络路由策略智能机制,进一步提升了智能路由的服务质量.

2 相关研究

目前,智能路由技术在快速、规模化的数据处理上展现出了优势,并且借助 SDN 网络架构的集中控制逻辑优势得到了业界广泛的关注. 其中,基于机器学习的智能路由研究主要分为监督学习和无监督学习两类. 基于监督学习的智能路由以神经网络技术分类为主,例如 Barabas 等人提出了基于 QoS 效果评估的智能路

由管理策略^[11],初步描述了基于 SDN 架构的智能路由定制方案,然而该研究在实用性方面尚有欠缺;文献 [12] 采用深度神经网络对节点流量特征进行归类分析,然而应用场景受限. 文献 [13] 和文献 [14] 分别使用了深度生成网络和他神经网络实现了网络流量特征的建模分析,但方案为静态网络流量特征分析,难以应对流量变化. 总体来说,由于监督学习需要大量训练数据集进行训练,而现有网络环境中难以获得大量正确标记的数据集,因此基于监督学习的智能路由生成机制研究进展缓慢. 另一方面,基于增强学习的智能路由生成机制克服了监督学习这一缺点. 文献 [15] 基于强化学习提出了一种安全路由机制,初步实现了强化学习在路由生成的安全性方面的应用. 文献 [16] 提出了基于强化学习的物联网路由算法,解决了物联网中分布式节点的路由选择问题. 文献 [17] 提出了基于深度强化学习的数据中心相关流调度技术,提升了数据中心的数据流处理效率. 文献 [18, 19] 将增强学习与路由生成结合,旨在提高网络服务质量 (Quality of Service, QoS); 文献 [20, 21] 提出了基于 Q-learning 的视频传输路由算法用以提高视频服务质量,然而上述方案难以解决网络环境中高维的输入输出空间问题,因此在实际应用中难以取得较好效果. DeepConf^[22] 提出了基于深度增强学习的数据中心流量管理架构,并且通过通过拓扑优化验证了所提架构的有效性. 文献 [23] 提出了基于深度增强学习 (Deep Reinforcement Learning, DRL) 的智能网络路由架构,然而该方案只是对架构进行了论证,没有提出实用化算法. DRL-TE^[24] 提出了基于 DRL 的流量工程方案,相比于传统流量工程算法取得了优势,然而该方案直接控制每个源-目的通信数据流实现流量调度,算法适用规模小.

3 SmartPath 架构

SmartPath 基于 SDN 架构实现,其主要模块包括数据层、控制层及运行于控制器之上的智能路由生成应用,如图 1 所示. 其中,数据层由软件定义转发设备组成,包括 OpenFlow 交换机等. 数据层的软件定义能力能够方便控制器实时检测网络状态、灵活收集所需信息,并且根据需求灵活改变转发规则. 控制层由控制器组成,控制器收集并整理所需网络状态信息,将全局信息视图传输到智能路由生成应用,并且根据智能路由生成应用下发的路由更新策略动态更新数据层的路由规则. 智能路由生成应用基于深度增强学习算法实现,主要负责生成动态路由规则以适应网络中不断变化的流量分布特征,其具体实现详见第 4 节.

3.1 SmartPath 工作流程

SmartPath 运行过程包括信息采集、策略生成及策

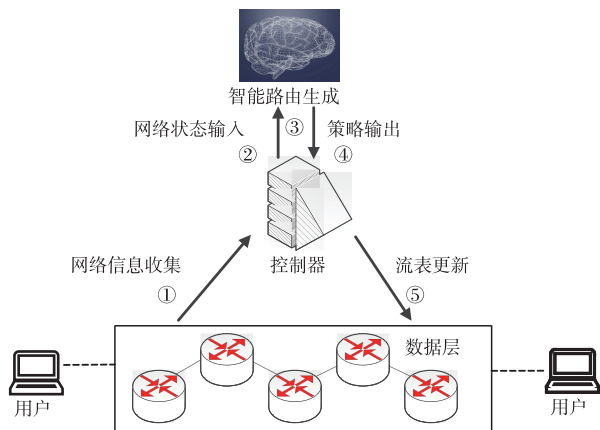


图1 SmartPath基本架构

略部署三个环节。三个逻辑环节相互配合形成逻辑闭环,不断监测网络状态、更新路由策略从而完成智能路由过程。

信息采集环节完成网络信息由底层向上传输的过程,始于数据层交换设备,终于智能路由生成应用,如图1中过程①②所示。在信息采集环节,控制器通过南向接口收集各种网络状态信息,包括各个数据流的流量统计、交换机资源利用率等。控制器通过南向接口收集到网络状态信息后,进行汇总和格式处理,提取智能路由生成应用所需状态信息(详见4.2节),由北向接口上传至智能路由生成应用,从而完成信息采集环节。

策略生成环节运行于智能路由生成应用内,图1中以③表示。由信息采集环节得到的状态信息作为深度增强学习中神经网络的输入,由神经网络计算得到输出结果,神经网络输出经过格式处理后作为深度增强学习输出策略(详见4.2节),从而完成策略生成环节。

策略部署环节始于智能路由生成应用的输出策略,终于数据层交换设备流表更新,图1中以过程④⑤表示。在策略部署环节,控制器通过北向接口获得智能路由生成应用输出的路由策略,基于特定路径算法将其转化为数据流在网络中的路径信息并将路由信息转化为流表项,通过南向接口完成交换设备的流表更新,从而改变网络中数据流的路径。

3.2 SmartPath 网络交互的实现

首先,信息采集环节中,控制器通过南向接口收集网络信息并通过北向接口将网络状态传递到智能路由生成应用中。在SDN网络中,目前存在多种网络状态监测方法,通常这些方法可以归类为被动型^[25]、主动型^[26]及带内网络检测技术^[26]等。在SmartPath中,综合考虑试验场景及实验需求,本文采用主动式信息收集方法收集网络流量信息,其过程如图2所示。控制器从数据层收集链路流量信息后,将流量信息分类、归一化处理并且按照特定链路顺序排列以对应智能路由生成

策略应用中神经网络输入神经元,最终通过北向接口向上传递。其中,SmartPath中流量信息可以根据数据流类型进行区分,从而同时产生多种流的网络流量视图。

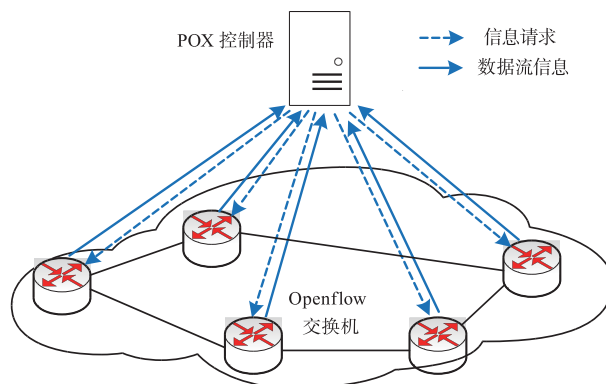


图2 SmartPath主动信息采集示意

智能路由生成应用输出的策略为数据层链路对应权重信息,该信息通过北向接口传递到控制器后,由控制器负责生成具体路由策略。其中,对应于信息采集过程中的两种流量视图,智能路由生成应用的输出策略也包含两组链路权重信息,分别适用于大象流及老鼠流。SmartPath中,控制器通过Floyd-Warshall算法同时计算网络中所有端到端的加权最短路径,其中所使用的链路权重即智能路由生成应用的输出策略。通过加权最短路径计算,控制器能够根据智能路由策略动态调整大象流及老鼠流的路由,进而实现网络的智能路由。

4 基于DRL的智能算法

SmartPath的智能路由生成应用中使用DRL进行网络流量视图信息分析及策略产生。其中,针对输出的链路权重为连续值的需求,SmartPath基于Twin Delayed Deep Deterministic Policy Gradient (TD3)^[27]算法得到动态化链路权重更新值。

4.1 TD3 算法模型

TD3算法属于深度增强学习模型。典型的增强学习模型由智能体(agent)与环境(environment)组成,智能体感知环境状态(state)后,做出决策(action)输出到环境中,继而从环境变化中得到此次决策的奖励值(reward)。其学习过程以马尔科夫过程表示,基本包括五个要素: S, A, R, P, γ 。其中, S 表示由状态 s 组成的状态空间, A 表示由动作 a 组成的动作空间, R 表示由奖励值 r 组成的奖励值空间, P 表示转移概率函数($p(s_{t+1}, r_t | s_t, a_t) \in P$), $\gamma \in [0, 1]$ 表示折扣因子。智能体根据观测到的状态 s 通过特定策略选择动作 a ,该策略在TD3中以 $a = \mu(s)$ 表示^[27]。同时,为保证在学习过程中智能体能够探索不同策略进行对比,最终的输出动作需要增加

随机性,即 $a = \mu(s) + N$,其中 N 为噪声.

在每个交互环节,智能体需要评价当前策略质量,因此引入了价值函数来表示策略质量.目前增强学习中使用最多的价值函数是 Q 值.在状态 s 下,智能体选择动作 a 的策略价值表示为:

$$Q(s_t, a_t) = E\left[\sum_{k=0}^{\infty} \gamma^k R(s_{t+k}, a_{t+k})\right] \quad (1)$$

式(1)也可以用 Q 值迭代的形式表示,即

$$Q(s_t, a_t) = E[R_t + \gamma Q(s_{t+1}, a_{t+1})] \quad (2)$$

离散环境中, $Q(s, a)$ 可以通过查表法记录,而连续环境中,DRL采用神经网络来表示 Q 值.此时, $Q(s, a)$ 可以进一步表示为 $Q(s, a|\theta)$,其中 θ 为神经网络参数.在 DRL 中,策略的更新即为神经网络参数的更新,此时,神经网络采用反向传播的方式将误差信息更新到神经元连接权重上.在 TD3 中,误差函数定义为:

$$L(\theta) = E[(Q(s_t, a_t|\theta) - y_t)^2] \quad (3)$$

其中 y_t 定义如下:

$$y_t = R(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{t+1})|\theta) \quad (4)$$

为减少策略估值偏差问题,TD3 引入了两个 Q 值网络 $Q(s, a|\theta_1)$ 和 $Q(s, a|\theta_2)$,并且对应地设计了 $y_{t,1}$ 与 $y_{t,2}$,分别由 $Q(s, a|\theta_2)$ 和 $Q(s, a|\theta_1)$ 进行交叉计算.TD3 中神经网络的策略更新方法可进一步定义为:

$$\begin{aligned} \nabla_{\theta^{\mu}} J &\approx E[\nabla_{\theta} Q(s, \mu(s|\theta^{\mu})|\theta^Q)] \\ &= E[\nabla_a Q(s, a|\theta^Q) \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu})] \end{aligned} \quad (5)$$

其中, θ^{μ} 为动作网络的神经网络参数,用以生成输出策略 $a = \mu(s|\theta^{\mu})$; θ^Q 为评价网络的神经网络参数,用以评估当前策略价值.整体算法流程如算法 1 所示.

算法 1 路由策略智能学习算法

输入:网络状态信息、路由性能反馈

输出:路由调整策略

- 1 初始化神经网络参数 θ^Q, θ^{μ} 和 θ^{σ} ;
- 2 对应动作网络、评价网络初始化相应目标网络;
- 3 初始化缓存空间 B ;
- 4 for episode = 1 to M ;
- 5 初始化随机过程 N 用于策略探索;
- 6 for $t = 1$; STEP_NUM
- 7 收集网络状态信息 s_t ;
- 8 动作网络产生 $a_t = \mu(s_t|\theta^{\mu}) + N_t$;
- 9 网络执行 a_t 并收集 r_t 和 s_{t+1} ;
- 10 在缓存 B 中保存 (s_t, a_t, r_t, s_{t+1}) ;
- 11 在缓存 B 中获取训练数据;
- 12 $y_t = r_t + \gamma \min_{i=1,2} Q_{\theta_i}(s_{t+1}, a_t)$;
- 13 以 $L = \frac{1}{N} \arg \min_{\theta} \sum (y_t - Q_{\theta}(s_t, a_t))^2$ 更新 θ^Q ;
- 14 以 $\nabla_{\theta^{\mu}} J \approx \frac{1}{N} \sum [\nabla_a Q(SS_t, a_t|\theta^Q) \nabla_{\theta^{\mu}} \mu(SS_t|\theta^{\mu})]$ 更新 θ^{μ} ;
- 15 更新目标网络:
- 16 $\theta^Q \leftarrow \tau \theta^Q + (1 - \tau) \theta^Q$;

- 17 $\theta^{\sigma} \leftarrow \tau \theta^{\sigma} + (1 - \tau) \theta^{\sigma}$;
- 18 end for
- 19 end for

4.2 神经网络接口定义

4.1 节论述了 SmartPath 所采用的 DRL 算法框架,本节重点论述 DRL 算法框架中神经网络的实现方案. SmartPath 中以循环神经网络(Recurrent Neural Network, RNN)作为 DRL 网络的神经网络.目前,基于 RNN 的神经网络已经形成多种成功案例,其中包括长短期记忆网络(Long Short-Term Memory, LSTM)、门控循环单元(Gated Recurrent Unit, GRU)等.典型 RNN 的基本结构如图 3 所示,其中网络主要分为三部分:输入层、输出层和隐藏层. SmartPath 采用 GRU 作为 RNN 神经网络的具体实现方案,其中,GRU 的输出层连接到两层前馈神经网络(Feedforward Neural Network, FNN),经过再次计算后得到最终输出结果. SmartPath 算法实现逻辑如图 4 所示.

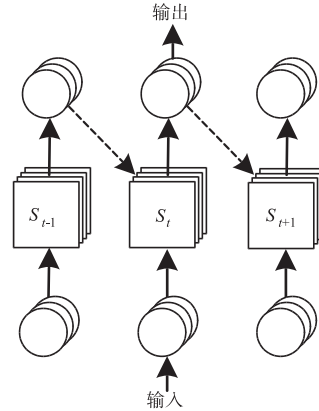


图3 RNN基本结构示意图

其中,对应于 DRL 的各个接口数据的具体定义如下.

状态:状态信息对应于神经网络输入层数据,本文中状态矩阵 $s^{d \times n}$ 表示,其中, s 表示网络链路上的流量值, d 表示数据流分类数量, n 表示网络的链路数量. SmartPath 中支持将流量分类,例如数据中心中流量可分为大象流和老鼠流,则此时 d 取值为 2.

动作:SmartPath 采用神经网络的输出层数据直接作为输出动作.其中,输出层神经元数量为 $d \times n$,因此输出动作以 $a^{d \times n}$ 表示. SmartPath 以每个神经元的输出值 a 作为链路权重,并且以此链路权重为依据计算网络中所有数据流的加权最短路径.由于大象流及老鼠流使用不同的链路权重,因此其计算所得路径不同.

奖励:对于 DRL 训练过程中的奖励值(reward), SmartPath 根据需要考虑的网络性能指标计算奖励值.通常,网络性能指标考虑吞吐率(throughput)、时延(de-

lay) 和丢包率 (loss) 等参数, 因此奖励值设为 $r = f(\text{throughput}, \text{delay}, \text{loss})$. 在不同的应用需求下, SmartPath 可以根据需求灵活调整奖励值从而达到预期效果, 例如, 可根据式(6)计算奖励值:

$$r = w_1 \times \text{throughput} + w_2 \times \text{delay} + w_3 \times \text{loss} \quad (6)$$

其中, w 为权重系数.

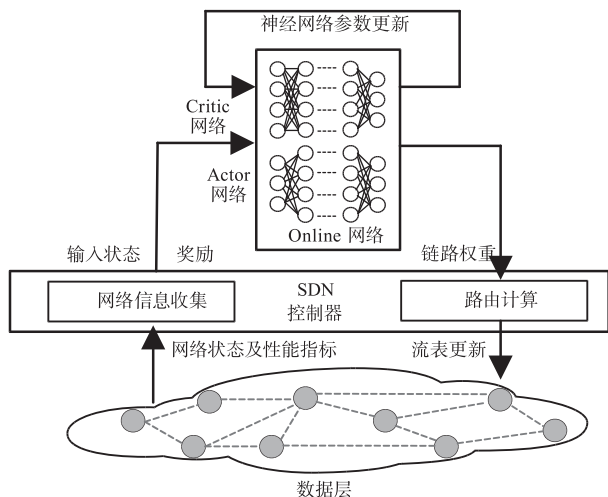


图4 SmartPath算法实现逻辑

5 实验及分析

本节主要介绍 SmartPath 的实验环境并且重点分析 SmartPath 的性能指标.

5.1 实验环境

本文通过实体网络环境搭建测试并验证 SmartPath 的性能. SmartPath 验证环境的网络拓扑使用了 OS3E^[28] 的基本拓扑结构, 如图 5 所示. 其中, 为保证实验内容与实验环境的硬件设备相适应, 实验方案调整了链路的带宽、设备等参数(所有链路的带宽为 10Mbps, 时延等参数设置默认相同). 实验的网络拓扑共有 38 个路由节点, 其中每个路由节点连接一个虚拟主机用以产生通信流量及接受信息. 实验拓扑中数据层的路由节点为 OpenFlow 交换机, 采用 PICA8 交换机实现; 用以产生端到端通信的虚拟主机通过运行于服务器之上的虚拟机实现. 控制层采用 POX 控制器实现, 智能路由生成应用以 Python 2.7 编写, 运行于基于 Tensorflow 的 Keras 平台. 控制器及智能路由生成应用均运行于 DELL 塔式工作站, 其硬件配置采用 Intel Xeon E5-2600 v4 处理器及 128G DDR4 内存.

端到端通信流量按照 Lakhina 等人^[29] 所统计的网络流量特征生成. Lakhina 等人指出, 网络中的源-目的通信流量中, 周期性流量占据了主要成分, 以数据集 Sprint-1^[30] 中的流量为例, 其源-目的通信流量中超过 90% 流量为周期性流量. 同时, 不同网络中不同流的数量和带宽呈

现出明显的分化特征, 例如, 数据中心网络中可将流量分为大象流(elephant flow)和老鼠流(mouse flow), 其中大象流的数量较少, 却占据了大部分带宽(80%以上). 因此, 本文的流量按照两个指标产生: 流量随机性和大小流比例. 其中, 流量随机性为随机性流量(以 RF 表示)和周期性流量(PF)的比值, 以 RP 表示:

$$RP = \frac{RF}{PF} \quad (7)$$

而大小流的比例按照“二八”原则设置, 即网络中总流数量 20% 的大流占据整个网络 80% 的带宽.

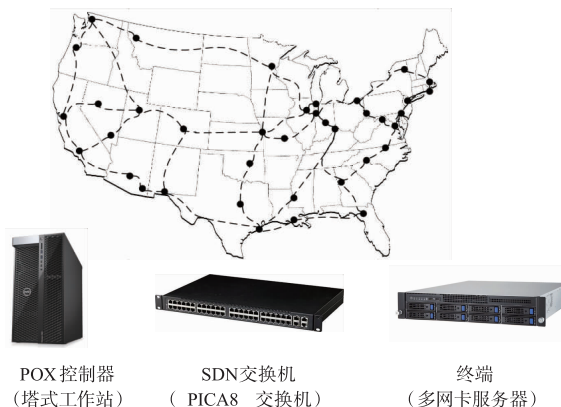


图5 实验网络拓朴示意

5.2 实验结果分析

本文从传输时延、丢包率等指标方面将 SmartPath 与典型路由策略及当前最优的智能路由策略进行对比. 对比方案有: 传统 OSPF 协议中使用的最短路径优先策略(Shortest Path First, SPF)、等价多路径(Equal-Cost Multipath Routing, ECMP)和当前基于机器学习进行流量工程的最优方案 DRL-TE^[18].

图 6 展示了不同方案的时延与丢包率综合性能的对比. 其中, 网络中总流量设置为 100Mbps, 其中 $RP = 0.1$, 流量均匀分为两组, 即输出动作 $a^{d \times n}$ 中 $d = 2$, 算法的训练奖励值(式(6))设置为相同权重的吞吐率与丢包率归一化平均值. 由图 6 可以看出, SmartPath 的流平均端到端传输时延为 7.6ms, 明显低于 SPF 的 11.3ms、ECMP 的 10.6ms 和 DRL-TE 的 9.7ms. 同时, 由于交换机端口缓存有限, 流量拥塞会引起交换机端口缓冲队列溢出而造成丢包. 由图 6 可以看出, SmartPath、DRL-TE 和 ECMP 丢包率明显低于 SPF, 说明其网络拥塞更少. 而相比于丢包率, ECMP 和 DRL-TE 的平均流端到端时延比 SPF 更大主要因为 DRL-TE 采用了多条备用路径的方式进行了流量工程, 多条路径的数据流端到端时延受限于最慢路径; 并且由于实验中通信流数量过多, DRL-TE 方案需要为每条流分配动作, 其神经网络输出动作空间过大, 容易引起维度爆炸(Curse of Dimensionality)问题, 因此算法难以收敛. 图 7 展示了

DRL-TE 和 SmartPath 在本实验中的训练过程. 由图 7 可以看出,在整个训练过程(共 150 个训练周期,每个周期包含约 1000 步交互过程)中,DRL-TE 没有取得明显进步,而 SmartPath 具有明显的训练提升. 因此,算法收敛性不足也在很大程度上造成了 DRL-TE 在实际实验中性能的不足.

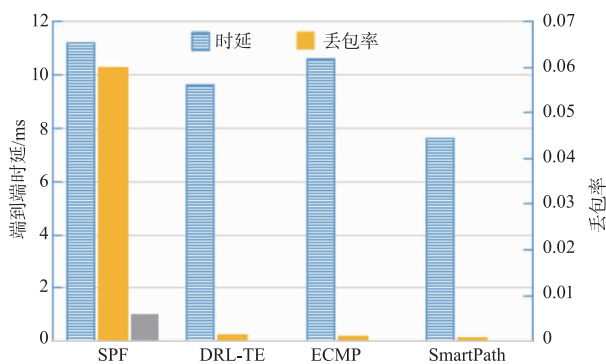


图6 不同方案时延与丢包率对比

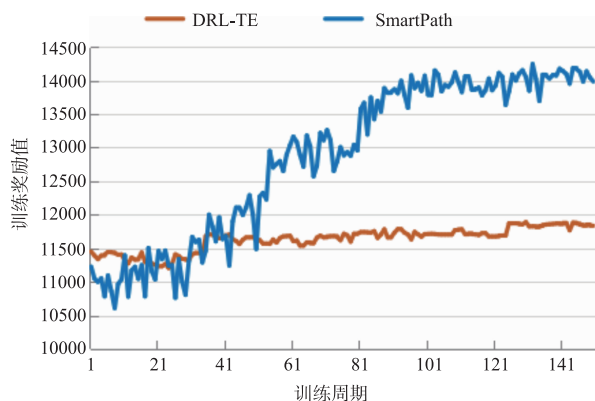


图7 训练过程示意

同时,本文测试了在固定流量总量前提下,不同流量随机性(RP)下不同方案的性能对比,实验结果如图8所示. 由图8可以看出,对于 SPF 和 ECMP,由于其路由策略固定,因此在总流量固定的情况下,流量成分对其路由效果的影响不明显.

而对于 DRL-TE 和 SmartPath,由于其算法效果依赖于神经网络对流量时空分布关联性的分析,因此随着流量中随机成分的增多(RP 值增大),DRL-TE 和 SmartPath 的神经网络提取流量特征的难度增大,所实现的数据流端到端时延也随之增大. 同时,由于 SmartPath 的算法收敛性更好,其端到端时延明显低于 DRL-TE,在相同 RP 值下至少减少 10% 端到端时延.

针对不同网络流量负载强度,本文进行了相关实验对比不同方案的性能,如图9所示. 由图9可以看出,随着网络流量负载的增加,SPF 的平均传输时延上升最明显,因为 SPF 为单一路径传输,负载增加更容易引起链路拥塞. DRL-TE、ECMP、SmartPath 三种方案随着网

络流量负载的提高端到端传输时延缓慢上升,其中 SmartPath 时延最低.

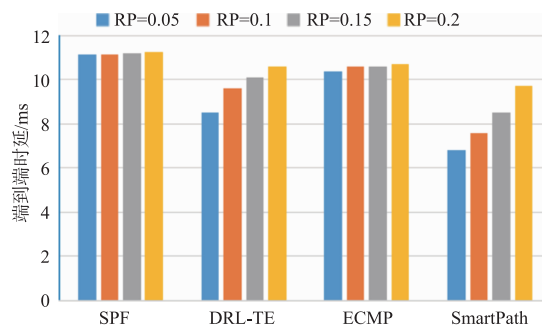


图8 不同流量特征下性能对比

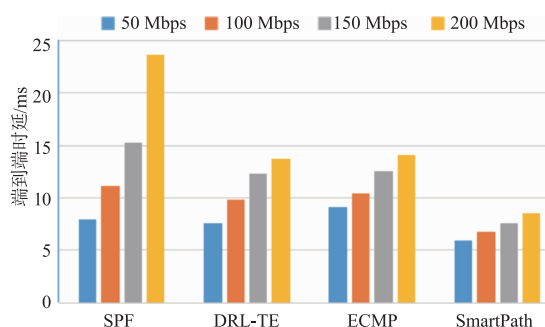


图9 不同负载强度下性能对比

6 总结

本文提出了一种基于深度增强学习的智能路由技术 SmartPath,通过动态实时收集网络中的流量状态,使用目前业界最新的 TD3 深度增强学习算法自动生成路由策略,从而保证路由策略能够动态适应网络流量变化. 为验证 SmartPath 在真实网络中的有效性,本文基于 PICA8 交换机搭建交换网络、基于 PC 机实现 POX 控制器,并且采用服务器模拟用户端产生网络流量. 实验结果表明,SmartPath 能够不依赖人工流量分析动态更新网络路由,在测试环境下比当前最优方案 DRL-TE 节省至少 10% 的平均端到端传输时延,证实了人工智能与网络控制相结合的技术发展潜力,为后续相关研究提供了参考方向.

参考文献

- [1] Egilmez H E, Dane S T, Bagci K T, et al. OpenQoS: An OpenFlow controller design for multimedia delivery with end-to-end quality of service over software-defined networks [A]. Signal & Information Processing Association Summit and Conference [C]. USA: IEEE, 2012. 1-8.
- [2] Liu J, Shroff N B, Xia C H, et al. Joint congestion control and routing optimization: An efficient second-order distributed approach [J]. IEEE/ACM Transactions on Networking,

- 2015,24(3):1-17.
- [3] Clark D D, Partridge C, Ramming J C, et al. A knowledge plane for the internet[A]. Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications [C]. USA: ACM, 2003. 68-73.
- [4] Thomas R W, Dasilva L A, Mackenzie A B. Cognitive networks [A]. First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks [C]. Baltimore, MD, USA: IEEE, 2005. 11-50.
- [5] Derbel H, Agoulmine N, Salaün M. ANEMA: Autonomic network management architecture to support self-configuration and self-optimization in IP networks [J]. Computer Networks, 2009, 53(3):418-430.
- [6] Zorzi M, Zanella A, Testolin A, et al. COBANETS: A new paradigm for cognitive communications systems [A]. International Conference on Computing [C]. USA: IEEE, 2016. DOI:10.1109/ICCNC.2016.7440625.
- [7] Mestres A, Rodrigueznatal A, Carner J, et al. Knowledge-defined networking [J]. ACM Sigcomm Computer Communication Review, 2016, 47(3):2-10.
- [8] Agoulmine N, Balasubramaniam S, Botvitch D, et al. Challenges for autonomic network management [A]. The 1st IEEE International Workshop on Modelling Autonomic Communications Environments (MACE) [C]. USA: IEEE, 2006. 87-92.
- [9] Nick McKeown, Tom Anderson, Hari Balakrishnan, et al. OpenFlow: Enabling innovation in campus networks [J]. ACM SIGCOMM Computer Communication Review, 2008, 38(2):69-74.
- [10] Sivaraman A, Kim C, Krishnamoorthy R, et al. DC. p4: Programming the forwarding plane of a data-center switch [A]. ACM SIGCOMM [C]. USA: ACM, 2016. 1-8.
- [11] Barabas M, Boanea G, Andrei Bogdan R, Dobrota V. Congestion control based on distributed statistical QoS-aware routing management [J]. Przegląd Elektrotechniczny, 2013, 89(2b):251-256.
- [12] Huang W, Song G, Hong H, Xie K. Deep architecture for traffic flow prediction: Deep belief networks with multi-task learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5):2191-2201.
- [13] Justin A Boyan, Michael L Littman. Packet routing in dynamically changing networks: A reinforcement learning approach [A]. Proceedings of the 6th International Conference on Neural Information Processing Systems [C]. USA: ACM, 1994. 671-678.
- [14] Xiao S, He D, Gong Z. Deep-q: Traffic-driven qos inference using deep generative network [A]. Proceedings of the 2018 Workshop on Network Meets AI & ML [C]. USA: ACM, 2018. 67-73.
- [15] Rusek K, Suárez-Varela J, Mestres A, et al. Unveiling the potential of Graph Neural Networks for network modeling and optimization in SDN [A]. Proceedings of the 2019 ACM Symposium on SDN Research [C]. USA: ACM, 2019. 140-151.
- [16] 章洋, 范植华, 何晓新, 等. 移动自组网络中多径路由的匿名安全 [J]. 电子学报, 2005, 33(11):2022-2030. ZHANG Yang, FAN Zhi-Hua, HE Xiao-Xin, et al. Anonymous secure multipath routing in mobile ad-hoc networks [J]. Acta Electronica Sinica, 2005, 33(11):2022-2030. (in Chinese)
- [17] 张德干, 葛辉, 刘晓欢, 等. 一种基于 Q-Learning 策略的自适应移动物联网路由新算法 [J]. 电子学报, 2018, 46(10):23-30. ZHANG De-gan, GE Hui, LIU Xiao-huan, et al. A kind of new routing algorithm with adaptivity for mobile IOT based on Q-learning [J]. Acta Electronica Sinica, 2018, 46(10):2325-2332. (in Chinese)
- [18] 马腾, 胡宇翔, 张校辉. 基于深度增强学习的数据中心网络 coflow 调度机制 [J]. 电子学报, 2018, 46(7):84-91. MA Teng, HU Yu-xiang, ZHANG Xiao-hui. Deep reinforcement learning based Coflow scheduling in data center networks [J]. Acta Electronica Sinica, 2018, 46(7):1617-1624. (in Chinese)
- [19] Shih-Chun Lin, Ian F Akyildiz, Pu Wang, Min Luo. QoS-aware adaptive routing in multi-layer hierarchical software defined networks: A reinforcement learning approach [A]. In 2016 IEEE International Conference on Services Computing (SCC) [C]. USA: IEEE, 2016. 25-33.
- [20] Jiang J, Hu L, Hao P, et al. Q-FDBA: Improving QoE fairness for video streaming [J]. Multimedia Tools & Applications, 2017, (2):1-20.
- [21] Giorgio Stampa, Marta Arias, David Sánchez-Charles, et al. A deep-reinforcement learning approach for software-defined networking routing optimization [J]. arXiv Preprint, 2017, arXiv:1709.07080.
- [22] Haipeng Y, Tianle M, Xiaobin X, et al. NetworkAI: An intelligent network architecture for self-learning control strategies in software defined networks [J]. IEEE Internet of Things Journal, 2018, 5(6):4319-4327.
- [23] Salman S, Streiffer C, Chen H, et al. DeepConf: Automating data center network topologies management with machine learning [A]. Proceedings of the 2018 Workshop on Network Meets AI & ML [C]. USA: ACM, 2018. 8-14.
- [24] Zhiyuan Xu, et al. Experience-driven networking: A deep reinforcement learning based approach [A]. In IEEE INFOCOM2018 [C]. USA: IEEE, 2018. 1871-1879.
- [25] Yu C, Lumezanu C, Zhang Y, Singh V, Jiang G, Madhyas-

- tha H V. FlowSense: Monitoring network utilization with zero measurement cost [A]. International Conference on Passive and Active Network Measurement [C]. Berlin: Springer, 2013. 31 – 41.
- [26] Adrichem van N L M, Doerr C, Kuipers F A. OpenNet-Mon: Network monitoring in OpenFlow software-defined networks [A]. Network Operations and Management Symposium (NOMS) [C]. USA: IEEE, 2014. 1 – 8.
- [27] Kim C, Sivaraman A, Katta N, et al. In-band network telemetry via programmable dataplanes [A]. In the Demo Session at SIGCOMM [C]. USA: INT, 2015. 1 – 15.
- [28] Fujimoto S, van Hoof H, Meger D. Addressing function approximation error in actor-critic methods [J]. arXiv Preprint, 2018, arXiv: 1802. 09477.
- [29] OS3E [OL]. [https://www. internet2. org/news/detail/4865/](https://www.internet2.org/news/detail/4865/). 2019.
- [30] Lakhina A, Papagiannaki K, Crovella M, et al. Structural analysis of network traffic flows [A]. Joint International Conference on Measurement and Modeling of Computer Systems [C]. USA: ACM, 2004. 61 – 72.

作者简介



孙鹏浩 男, 1992 年生于山东即墨. 现为国家数字交换系统工程技术研究中心硕士研究生. 主要研究方向为可编程网络数据平面.
E-mail: sphshine@126. com



兰巨龙 男, 1962 年生于河北张北. 现为国家数字交换系统工程技术研究中心教授、博士生导师. 主要研究方向为新一代信息网络关键理论与技术.
E-mail: ndsclj@163. com