

面向多模态网络业务切片的虚拟网络功能资源容量智能预测方法

兰巨龙, 朱棣, 李丹

(信息工程大学信息技术研究所, 河南 郑州 450001)

摘 要: 为解决多模态网络场景中网络切片上的 VNF 部署方案调整不及时造成的切片性能下降问题, 提出了一种基于时空特征提取的 VNF 资源容量预测方法, 旨在通过预测为即将到来的切片需求进行 VNF 的预部署。所提方法首先对用于预测的数据流时间序列进行加权处理, 然后把处理后的时间序列及其依赖的空间拓扑信息输入网络模型中, 通过图卷积网络重组时间序列的空间分布特征, 再由门控循环单元感知输入数据的时序依赖关系, 最后基于数据流序列与 VNF 实例数量的映射关系, 由前馈神经网络最终输出 VNF 资源需求预测视图。实验结果表明, 所提方法比对照方法中预测精度最高的方法提高了 6.54% 的需求预测精度。

关键词: 多模态网络; 虚拟网络功能; 资源容量; 时空特征提取

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2022098

Intelligent prediction method of virtual network function resource capacity for polymorphic network service slicing

LAN Julong, ZHU Di, LI Dan

Institute of Information Technology, Information Engineering University, Zhengzhou 450001, China

Abstract: To solve the problem of network slicing performance degradation caused by untimely adjustment of VNF deployment scheme on network slicing in polymorphic network scenario, a VNF resource capacity prediction method based on spatial-temporal feature extraction was proposed, which aimed to pre-deploy VNF by forecasting for upcoming slicing needs. Firstly, the data flow time series used for prediction were weighted, and then the processed time series and its dependent spatial topology information were inputted into the network model. Then, the spatial distribution features of the time series were reorganized through the graph convolution network, and then the timing dependence of the input data was perceived by the gated recurrent unit. Finally, based on the mapping relationship between the data flow sequence and the number of VNF instances, the feed forward neural network outputted the VNF resource demand prediction view. The experimental results show that the proposed method improves the demand prediction accuracy by 6.54% over the comparison method with the highest prediction accuracy.

Keywords: polymorphic network, virtual network function, resource capacity, spatial-temporal feature extraction

0 引言

随着云计算、虚拟化等技术的迅速发展, 网络结构灵活性及资源能力都得到了极大提升, 网络从传统的刚性架构向多模态融合体制持续演进^[1]。多

模态网络 (PINet, polymorphic network)^[2]预期形成多模态的网元形态、多模态的管理控制、多模态的寻址路由和多模态的传输协议等, 以支持网络的个性化、多元化、智慧化、安全化发展需求。现阶段 PINet 借助网络功能虚拟化 (NFV, network function

收稿日期: 2022-01-04; 修回日期: 2022-04-02

通信作者: 朱棣, mailbox_judy@126.com

基金项目: 国家重点研发计划基金资助项目 (No.2020YFB1804803); 国家自然科学基金资助项目 (No.62002382)

Foundation Items: The National Key Research and Development Program of China (No.2020YFB1804803), The National Natural Science Foundation of China (No.62002382)

virtualization)和人工智能等技术的蓬勃发展和广泛应用,使网络管理控制模式尽可能灵活适配多样化的用网场景。基于 NFV 网络架构,一组有序虚拟网络功能(VNF, virtual network function)及其物理资源组成的端到端逻辑网络被称为网络切片(NS, network slicing)^[3],该技术进一步灵活化了网络服务的管理。VNF 作为网络切片的核心要件,其性能发挥直接影响着网络服务水平的保证与提升。然而,如何根据业务用户实时的网络资源需求动态调整 VNF 占用的物理资源规模,弹性伸缩 VNF 的处理能力,实现网络资源按需匹配业务需求的 VNF 容量调整,对于网络业务的精细化响应是一个巨大的挑战。

VNF 容量调整^[4]有水平放缩和垂直放缩 2 种机制。VNF 资源需求变化时,水平放缩指分配给每一个 VNF 实例的资源容量保持不变,通过 VNF 实例数量的变化调节整体资源分配情况,提高 VNF 的处理能力。垂直放缩指保持运行中的 VNF 实例的数量不变,通过调整分配给每一个 VNF 实例的资源容量来调节整体资源分配情况,提高 VNF 的处理能力。然而,现有的虚拟化资源存储技术均是状态敏感的,不支持运行中网元的资源容量调整。因此 VNF 容量的水平放缩成为目前的研究重点。

根据业务数据流是否可预测,主动的 VNF 容量调整又可划分为数据流可预测的离线容量调整和数据流不可预测的在线容量调整。通过准确的数据流预测,在当前时间为未来的业务需求进行 VNF 实例的预部署,可以实现对 VNF 容量调整无感的连续性网络业务供给。数据流可预测的 VNF 资源容量水平调整问题可以划分为 2 个阶段:1) VNF 资源需求视图生成阶段,根据流经切片上 VNF 节点的数据流预测将来一段时间的 VNF 实例数量需求;2) VNF 需求视图映射阶段,进行 VNF 实例化,使上一阶段生成的资源需求视图与物理网络视图相匹配。

本文研究 VNF 资源容量水平调整的第一阶段,并将该问题建模为时间序列预测问题,旨在通过预测即将到来的数据流,重新配置 VNF 实例的数量以高效利用有限的网络资源。本文提出了一种基于图卷积网络(GCN, graph convolutional network)和门控循环单元(GRU, gated recurrent unit)的 VNF 容量需求预测(VNFPre, prediction for VNF capacity demand)方法,把节点的历史数据流负载序列用于

未来节点 VNF 资源容量的预测。该方法具有 2 个特征提取模块和一个特征重组模块,首先通过 GCN 模块在节点分类任务上进行预训练,提取待测节点周围网络的拓扑特征,之后输入 GRU 模块提取承载 VNF 实例的节点上历史资源负载的时序特征,接着基于数据流负载与 VNF 资源容量的映射关系,利用前馈神经网络(FNN, feedforward neural network)模块输出对该节点未来的 VNF 资源容量的预测值,以在充分获取节点数据流序列的时空特征后,更加精准地预先调整 VNF 实例的数量,进而不需要任何人工干预就能实现资源灵活调整。该方法旨在实现主动高效的 VNF 水平放缩,达到网络与多元业务的高度适配。与其他方法相比,本文的主要贡献如下。

1) 本文提出了一种数据驱动的 VNF 容量需求预测方案,结合 GCN 与 GRU 充分提取数据流的时空特征,提升了数据流预测精度。

2) 本文将数据流预测结果送入 FNN 进行进一步计算,并充分利用了数据流与 VNF 资源容量的映射关系,获得了准确度很高的 VNF 资源容量需求预测视图。本文所提方法与文献[5]提出的 VNF 容量预测方法相比,容量需求预测精度提升了 6.54%。

3) 本文将数据流序列进行加权处理后输入预测模型,有效避免了网络中的突发流对数据流预测结果的影响,比不处理输入数据的方案降低了 3.54%的预测误差。

1 相关工作

在实际的网络服务提供过程中,对于服务中断的容忍程度并不高,服务中断会造成用户体验的大幅下降以及某些连续性要求高的服务无法上线。目前,VNF 的实例化时间最短为“分钟级”,采取被动的反应式策略进行 VNF 容量放缩操作最少也需要数十秒量级的完成时间,而过长的功能部署调整时间会损害网络业务的服务质量(QoS, quality of service)水平。为了避免服务中断的情况,可以将 VNF 的资源容量调整建模为时间序列的预测问题,从而进行主动的、预测性的 VNF 水平放缩。基于流量预测的 VNF 容量调整方法,通过预测输入 VNF 集合的数据流波动趋势,预判 VNF 集合中 VNF 实例数量的变化,提前部署或移除相应的 VNF 实例,实现资源容量按需分配。

对于网络中的时间序列预测问题，机器学习模型已经成为经典统计模型的主要竞争方法。文献[6]设计了一种基于深度学习的虚拟机工作负载预测方法，但该方法不能解决跨虚拟机场景的预测问题。文献[7]基于业务请求动态变化的 5G 网络切片应用场景，设计了一种基于长短期记忆（LSTM, long short-term memory）网络的流量感知模型，该模型考虑到了动态时变的业务请求可能造成的虚拟机缓冲区队列积压，但仅将计算资源作为虚拟化的载体，因此预测准确度并不理想。文献[8]设计了一种数据中心场景下基于深度学习的数据流预测方法，然而该方法难以应用于其他用网场景。文献[9]设计了一种基于循环神经网络（RNN, recurrent neural network）流量预测方法估计 5G 核心网中上行链路的未来流量，然而该方法并非同类最优，精度有限。文献[10]针对云数据中心中的多级虚拟网络嵌入，首先使用历史监督数据以及二进制监督分类器确定实时传入虚拟网络的可接受性，然后设计了一个基于深度强化学习（DRL, deep reinforcement learning）的预测模型，使用虚拟机类型的信息提高云资源利用率，最后通过径向基回归模型用于预测衍生特征。但该算法复杂度较高，场景限制大。

对于 VNF 容量调整问题，机器学习模型同样被研究者广泛应用。绝大多数的研究者都采用 RNN 中的 LSTM 和 GRU 来解决 VNF 容量调整本质涉及的时间序列预测问题。文献[11]将 VNF 容量调整问题建模为通过准确的 CPU 资源预测实现的低成本 VNF 实例配置，提出了一种基于 LSTM 的深度学习资源预测模型，并通过一种分布式稳定匹配的策略生成有效的 VNF 实例扩展决策，降低 VNF 的部署成本。文献[12]分别设计了基于 LSTM 和 GRU 的数据驱动方法预测虚拟机中的流量，并将预测结果作为一个二进制优化模型的输入，该模型的输出为未来所需的 VNF 实例的预测数量。该方法虽然提升了流量预测精度，但不能解决多虚拟机场景的预测问题。文献[13]提出了一种基于 GRU 的流量预测模型，根据预测结果提前缩放 VNF 实例，并将 VNF 的缩放结果输入名为 A3C 的 DRL 算法来训练代理，然后获得部署新实例的最佳策略。文献[14]提出了一种具有主动流量预测的自适应 NFV 资源分配方法，设计了基于 LSTM 的用户需求预测模型。

LSTM 和 GRU 作为 RNN 的变体，能够有效解

决长期时间序列的预测问题，但其仅对单一节点上随时间变化的情况进行考虑，并未将网络中其他节点对待测节点的影响纳入考虑范围。此外，还有一些 VNF 容量调整的研究考量了网络中节点之间的相互影响。文献[15]提出了一种基于 GCN 的算法预测每个 VNF 的资源使用情况，该算法基于拓扑信息，使用图上相邻的 VNF 实例来预测目标 VNF 实例在未来的资源需求。但由于 VNF 内部结构一般不公开，且同一 VNF 的不同实例可能处理不同的数据流负载，即使处于同一拓扑中也可能具有完全不同的资源需求，因此无法保证其预测准确性。文献[16]提出了一种图神经网络（GNN, graph neural network）增强的异步 DRL 的系统架构，通过拓扑感知预测动态 NFV 环境中的 VNF 资源需求。该方案使用 GNN 对服务功能链（SFC, service function chain）中节点之间的拓扑依赖关系进行建模，并使用生成的编码网络作为异步 DRL 代理的环境，帮助形成 VNF 资源的预测性扩展策略。但该方案主要针对高度聚合的 POP 级别的流量和 VNF 需求预测，在超大吞吐量的流量场景下，预测精度和时间效率将会大打折扣。

在一些最新的 VNF 容量调整研究中，主要采用 LSTM 和卷积神经网络（CNN, convolutional neural network）实现预测。文献[17]提出了一种基于上下文和方便嵌入依赖目标的 LSTM 模型以预测 VNF 资源需求，并通过 SFC 中邻居 VNF 的已用资源数据来提高预测准确性。由于该模型利用了其他 VNF 的信息，因此不需要为每个 VNF 单独进行预测，通用性更好。然而，该模型仅考虑同一个 VNF 仅用于一条 SFC 的情况。文献[5]讨论了基于深度神经网络（DNN, deep neural network）和基于 LSTM 的 VNF 资源需求预测机制，还探讨了 LSTM 变体（如双向 LSTM）和结合拓扑信息和时间序列的 CNN-LSTM 机制在 VNF 需求预测方面的性能，是目前最新颖、考虑最全面的关于 VNF 需求预测的研究。

VNF 需求预测问题本质上是特定拓扑结构约束下的时间序列预测问题，输入信息的空间特征和时间特征能够被准确提取是影响预测结果的核心要素。综上所述，上述基于机器学习的 VNF 需求预测研究工作存在以下几点缺陷：1) 大部分研究工作仅考虑提取输入信息的时间特征，忽略了其空间特征的提取；2) 部分研究工作具有明确的场景限

制,通用性有待考察;3) 在同时考虑输入信息的时空特征的研究工作中,虽然能够采用 LSTM、GRU 等先进的 RNN 结构有效提取时间特征,但往往采用 CNN 提取空间特征,不能准确反映非欧氏结构拓扑的特点。

在一些解决其他网络问题的研究工作中,已经有一些工作对提取网络拓扑的空间特征做出了有益尝试。文献[18]设计了一种专用于网络路由场景的图感知卷积结构,该结构由一个图形内核和一个 CNN 结构组成,从网络图中提取拓扑信息并根据提取的拓扑处理输入数据,最后对处理后的数据应用卷积产生输出,能够有效改善 CNN 的空间特征提取效率。文献[19]提出了一种无线信道图卷积网络,用来解决大规模的无线网络资源管理问题,并通过仿真证明该方法对无线网络的功率控制和波束成形问题的资源优化效果高于没有领域知识的经典优化算法。

2 系统架构与问题建模

2.1 网络切片部署架构

网络切片是一组 VNF 集合及其映射到物理网络中的物理资源构建的端到端、个性化定制和相互隔离的逻辑网络。网络切片技术通过合作建网的模式,为 5G 时代移动运营商与行业客户的双赢提供了无障碍支撑,突破了 4G 移动通信网适配多样化移动用户业务的瓶颈,成为移动运营商迈向万物互联的使能平台的关键^[20]。

多模态网络中的网络切片主要用于用户业务需求与网络服务能力之间的按需匹配,实现网络运行、功能编排等的自适应承载^[2]。如图 1 所示,网络切片的基本组件包括切片选择器、切片编排管理器以及 VNF 集合与其对应的物理资源 3 个部分。其中,切片选择器根据不同的业务和通信场景选择不同的切片;切片编排管理器收到被选择的切片需求后负责切片资源的分配调度、性能监测和生命周期的管理,完成 VNF 的实例化和切片的创建;VNF 集合与其对应的物理资源负责提供不同的网络能力,来满足各种业务的需求。网络切片需求方请求的业务数据流通过切片选择器进入适合的切片,切片编排管理器根据业务需求制定路由策略,引导数据流通过切片上的 VNF 集合进行流水线式的数据处理,完成处理后的数据流将直接接入目的节点。

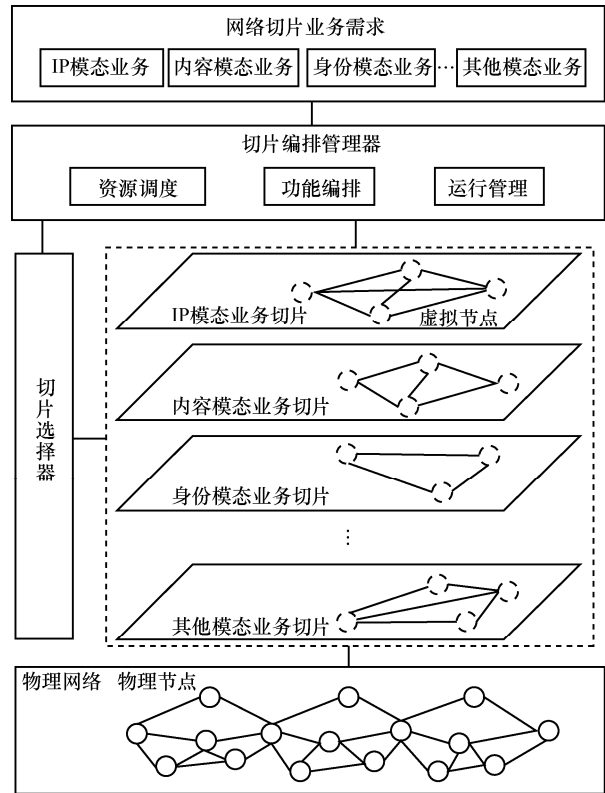


图 1 多模态业务切片架构

2.2 VNF 资源容量预测问题建模

多模态网络旨在通过网络功能要素的全维度开放和可定义,解决业务承载的发展瓶颈和网络资源的无谓浪费。目前,多模态网络利用网络切片技术,结合现有的组网情况整合各类业务网络,但仍难以在各类服务的各阶段为用户提供无差别的良好体验,需要提前规划和调整 VNF 的位置分布、资源容量等要素,才能真正具备全业务承载能力。如图 2 所示,在只考虑单一模态的单个网络切片的场景下,VNF 资源容量预测问题的研究内容是根据不同时刻流经切片上各虚拟节点的数据流历史信息,预测未来某一时刻或某段时间内相应虚拟节点上的 VNF 实例数量需求,生成 VNF 资源容量需求视图。其中,切片上单一虚拟节点在某时刻的数据流信息为该虚拟节点映射在物理网络中的物理节点上流经的数据流信息。一个虚拟节点可能映射一个或多个物理节点,若有多个物理节点映射单一虚拟节点,则这些物理节点的数据流吞吐量之和为对应虚拟节点的数据流信息。

VNF 资源容量预测的目标是为下一阶段 VNF 实例的预部署提供依据,提前调整切片虚拟节点上的 VNF 实例数量,在保证服务供给的 QoS 水平的

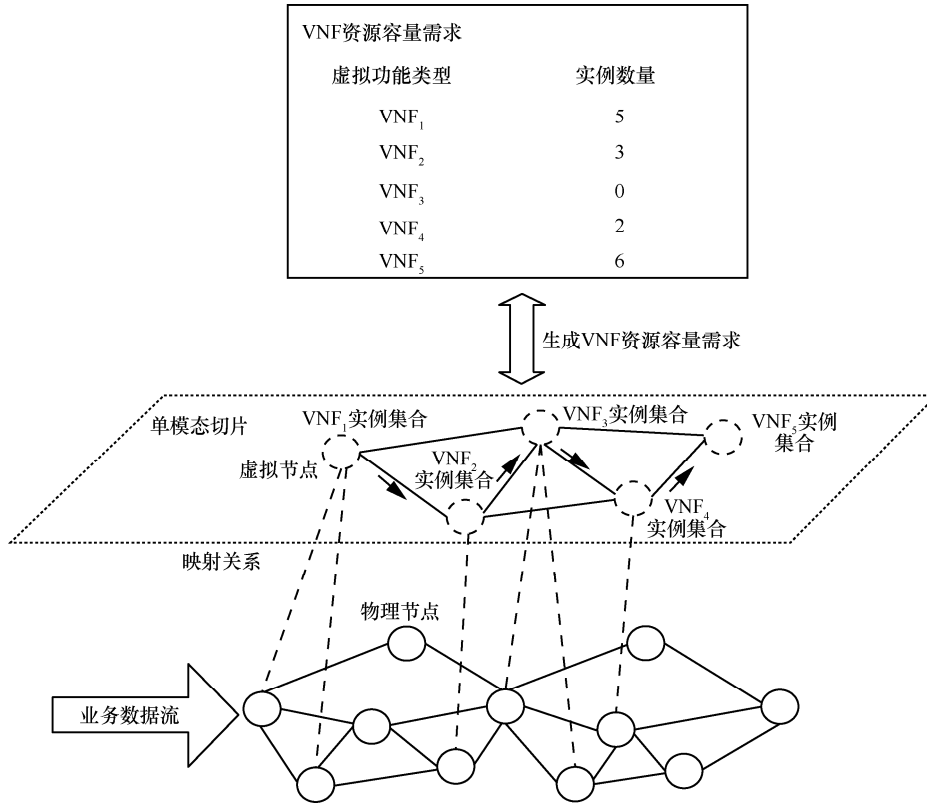


图 2 单模态单一切片场景下 VNF 资源容量预测模型

同时提高物理资源利用率。

本文将切片中虚拟网络的节点数量记为 n ，节点上的 VNF 实例数量记为该节点的特征。其中，时刻 i 单个节点的 VNF 实例数量记为 X_i^n 。则时刻 i 整个网络所有节点的特征可以用 n 维向量 $\mathbf{X}_i = (X_i^1, X_i^2, \dots, X_i^n)$ 表示， $\mathbf{X}_i \in \mathbf{R}^n$ 。网络中每个节点上 VNF 实例数量历史序列的长度记为 l ， l 表示单一节点输入模型的特征数量，则整个网络的节点特征由矩阵 $\mathbf{X}^{n \times l}$ 表示。本文将 $f(\cdot)$ 记为待学习的预测模型， T 记为需要预测的时间序列的长度，当 $T=1$ 时，仅预测该节点下一时刻的 VNF 实例数量。则 VNF 资源容量预测问题可表示为

$$[\mathbf{X}_{t-l}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t] \xrightarrow{f(\cdot)} [\mathbf{X}_{t+1}] \quad (1)$$

$$[\mathbf{X}_{t-l}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t] \xrightarrow{f(\cdot)} [\mathbf{X}_{t+1}, \dots, \mathbf{X}_{t+T}] \quad (2)$$

由以上分析可得，VNF 资源容量预测的步骤为：1) 将流量数据输入预测模型；2) 得到预测模型输出的流量预测结果；3) 将流量预测结果映射为 VNF 资源容量需求，即节点的 VNF 实例数量需求。

3 方案设计

由第 2 节的分析可得，成功的 VNF 资源容量预测的关键有两点：1) 获得准确的数据流预测结果；2) 选择适合的方法将数据流预测结果映射为 VNF 资源容量需求。

如图 3 所示，本文提出一种多模态网络场景下面向网络切片的 VNF 资源容量智能预测方法 VNFPre：1) 提出 GCN 联合 GRU 的数据流预测模型，利用 GCN 模型作为空间分布特征提取器，充分获取嵌入在数据流中的拓扑依赖性，并将特征重组后的数据继续输入 GRU 模型中，根据历史流量数据的时序关联特征预测未来时间的数据流；2) 提出一种基于 FNN 的数据流吞吐量与 VNF 实例数量的映射方法实现 VNF 资源容量预测。

网络中的数据流预测看似为一个单纯的时间序列分析问题，但网络中数据流的传输行为建立在特定的网络拓扑之上，因此相邻节点的数据流吞吐量、传输方向等具有一定的相关性。本文方法的研究目标是将单一切片上的 l 个历史数据流序列及其依赖的网络拓扑数据输入 VNFPre 中，在获得涵盖

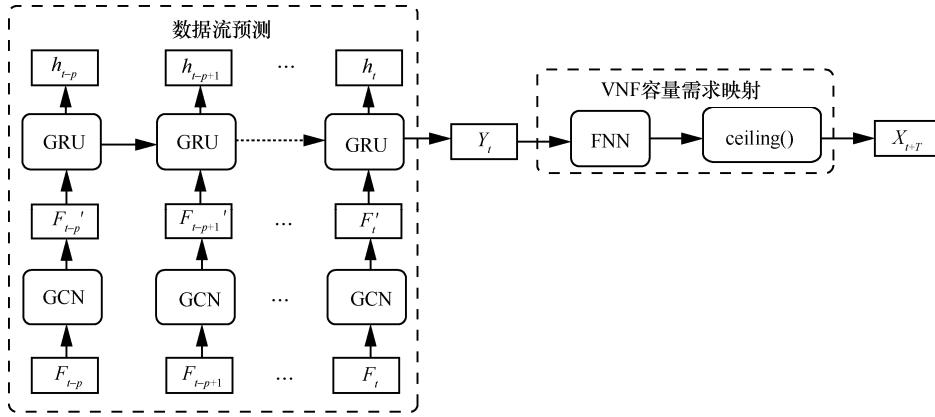


图 3 VNFPre 模型架构

时空特征 $\{h_{t-l}, \dots, h_{t-1}, h_t\}$ 的 l 个隐藏状态 (h) 的基础上进一步获得未来 $t+T$ 时刻的 VNF 资源容量需求视图。

3.1 模型输入设计

现有的大部分数据流预测方案直接提取当前时刻之前一段时间内的流量数据作为模型输入，但由于实网中的数据流往往由绝大部分的常规流和小部分时延敏感的突发流^[21]组成，单纯地利用一段时间内的流量序列作为输入容易受到突发流的影响，不能充分反映流量的实际变化规律。并且预测模型在提取数据流特征时仅考虑了其在时间上的顺序特征，忽略了不同数据流的不同承载拓扑赋予其在空间中的分布特征。

本文在利用时序回归的数据流预测方法解决 VNF 资源容量预测问题时，为提高数据流预测准确度，使用数据流依赖的网络拓扑数据和加权历史数据流共同作为 VNFPre 模型的输入。

1) 网络拓扑数据

本文将承载数据流的物理网络的拓扑结构记为 $G=(P,E)$ ，其中 $P=\{p_i | i=1, \dots, n\}$ 为网络节点的集合， n 为节点数； $E=\{e_{ij} | i, j=1, \dots, n\}$ 为网络中节点 p_i 和 p_j 之间链路的集合。VNFPre 模型将上述网络 G 的拓扑结构信息用 $n \times n$ 的无向图邻接矩阵 A 表示，该矩阵定义为

$$A \in R^{n \times n}, A_{ij} = \begin{cases} a_{ij}=1, & e_{ij} \in E \\ 0, & \text{其他} \end{cases} \quad (3)$$

如果节点 p_i 和 p_j 之间存在一条链路 e_{ij} ，则 $a_{ij}=1$ （无向无权图），否则为 0。

2) 加权历史数据流

本文将当前时刻记为 t 时刻， t 时刻经过网络中单一节点的数据流记为 L_t ，输入 VNFPre 的历史序列的长度记为 l ，采用 $[t-l, t]$ 时段和 $[t-2l-1, t-l-1]$ 时段内的加权历史数据流作为 VNFPre 模型的输入数据。则 VNFPre 的数据流输入序列 $[F_{t-l}, F_{t-l+1}, \dots, F_t]$ 可表示为 $a_1[L_{t-l}, L_{t-l+1}, \dots, L_t] + a_2[L_{t-2l-1}, L_{t-2l}, \dots, L_{t-l-1}]$ ，其中 a_1 和 a_2 为非负权重值，且 $a_1 + a_2 = 1$ 。

基于上述定义，可以将式(1)和式(2)中的 VNF 需求预测模型重新定义为基于网络拓扑 G 和网络节点的特征矩阵 X 学习映射函数 f 的过程。从当前时刻起，到未来 $t+T$ 时刻的 VNF 实例数量计算式为

$$(G; [X_{t-l}, \dots, X_{t-1}, X_t]) \xrightarrow{f(\cdot)} [X_{t+1}] \quad (4)$$

$$(G; [X_{t-l}, \dots, X_{t-1}, X_t]) \xrightarrow{f(\cdot)} [X_{t+1}, \dots, X_{t+T}] \quad (5)$$

3.2 模型输出设计

VNFPre 的 VNF 资源需求预测分为数据流预测阶段和数据流预测结果映射为 VNF 资源容量需求阶段。经过数据流预测阶段，模型获得涵盖时空特征 $\{h_{t-l}, \dots, h_{t-1}, h_t\}$ 的 l 个隐藏状态 (h)。本文设计利用 FNN 将隐藏状态降维，使用基于归一化指数函数 softmax 的回归模型进行归一化分类，输出流量预测序列 $[Y_{t+1}, \dots, Y_{t+T-1}, Y_{t+T}]$ 。本文将数据集的单个节点数据流负载的最大上升值记为 u Gbit/s，假设在每个 VNF 实例集合的部署中需要处理的最大数据流负载相当于 u Gbit/s，并且每个 VNF 实例可以处理最大 1 Gbit/s 的数据流而不会降低 QoS，则有

$$\begin{aligned} & [X_{t+1}, \dots, X_{t+T-1}, X_{t+T}] = \\ & \left[\text{ceiling} \left(\frac{Y_{t+1}}{u} \right), \dots, \text{ceiling} \left(\frac{Y_{t+T-1}}{u} \right), \text{ceiling} \left(\frac{Y_{t+T}}{u} \right) \right] \end{aligned} \quad (6)$$

其中, ceiling 表示向上取整。

整个 VNF 资源容量预测方法的训练过程如算法 1 所示。

算法 1 VNFPre 方法训练过程

- 1) 初始化网络拓扑 G 和特征矩阵 X
- 2) 初始化历史数据流权重值 a_1, a_2
- 3) 初始化历史数据流序列长度 l
- 4) $[F_{t-l}, F_{t-l+1}, \dots, F_t] \leftarrow a_1 [L_{t-l}, L_{t-l+1}, \dots, L_t] + a_2 [L_{t-2l-1}, L_{t-2l}, \dots, L_{t-l-1}]$
- 5) while $a_1 + a_2 = 1$ do
- 6) for $a_1 = 0.5$ 到 1 :
- 7) $([F_{t-l}, F_{t-l+1}, \dots, F_t]; G) \xrightarrow{\text{GCN}}$
 $[F'_{t-l}, F'_{t-l+1}, \dots, F'_t]$
- 8) $([F'_{t-l}, F'_{t-l+1}, \dots, F'_t]; h_{t-l-1}) \xrightarrow{\text{GRU}}$ h_t
- 9) $h_t \xrightarrow{\text{FNN}} h'_t \xrightarrow{\text{ceiling()}}$ X_{t+T}
- 10) 计算最优拟合 X_{t+T} 值对应的 a_1 值
- 11) end for
- 12) end while

3.3 实现细节

1) GCN 空间特征提取

GCN 把图结构融入权重进行训练, 能够表现非欧几里得结构, 擅长节点分类、图分类或链路预测任务。本文研究的 VNF 资源容量预测问题的空间特征提取就是一种节点分类任务。

VNFPre 结构使用 GCN 层对 GRU 的输入数据进行特征提取, 以支持序列预测。首先将承载数据流负载的网络拓扑图建模为一个无向图, 使用频谱图卷积捕捉拓扑的空间特征。文献[22]的实验研究

发现, 2~3 层的 GCN 就能取得良好的特征提取效果。如图 4 所示, VNFPre 构造了一个两层的 GCN, 第一层激活函数采用修正线性单元 (ReLU, rectified linear unit), 第二层激活函数采用 softmax。

$$Z = f(F, A) = \text{softmax} \left(\hat{A} \text{ReLU} \left(\hat{A} X W^{(0)} \right) W^{(1)} \right) \quad (7)$$

其中, F 是特征矩阵; A 是邻接矩阵, $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ 是预处理步骤; $W^{(0)} \in \mathbf{R}^{K \times H}$ 是从输入层到隐藏层的权重矩阵, K 是时间跨度, H 是隐藏单元数; $W^{(1)} \in \mathbf{R}^{H \times T}$ 是从隐藏层到输出层的权重矩阵; $f(F, A) \in \mathbf{R}^{n \times T}$ 是预测长度为 T 的输出。最后, VNFPre 针对所有带标签的节点计算交叉熵损失函数

$$\text{Loss}(\hat{Y}_y, Y_y) = - \sum_{l \in y_t} \sum_{f=1}^F Y_{yf} \ln \hat{Y}_{yf} \quad (8)$$

该损失函数对应节点分类的模型, 其中, Y_y 为标签矩阵, \hat{Y}_y 为模型的输出值。至此, VNFPre 通过 GCN 模型确定了待测节点周围节点之间的拓扑。

2) GRU 时间特征提取

GRU 模型是 VNFPre 实现序列预测的关键部分。GRU 是一种特殊的 RNN 结构, 只有重置门和更新门 2 个门, 比一般 RNN 和 LSTM 结构更简单, 具有更少的参数和更短的训练时间。该结构继承了一般 RNN 隐藏层将输入信息回传自身以保留序列历史信息的技巧, 并且通过门控机制克服了一般 RNN 可能导致的梯度消失或爆炸的问题, 能够学习长序列的数据依存关系。

如图 5 所示, 在 VNFPre 模型中, GRU 时间特征的计算过程为

$$z_t = \sigma \left(W^{(z)} f(A, F_t) + U^{(z)} h_{t-1} \right) \quad (9)$$

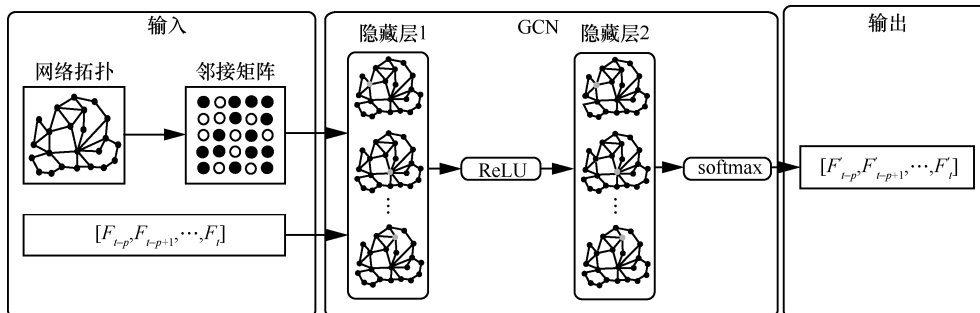


图 4 GCN 提取序列空间特征的实现细节

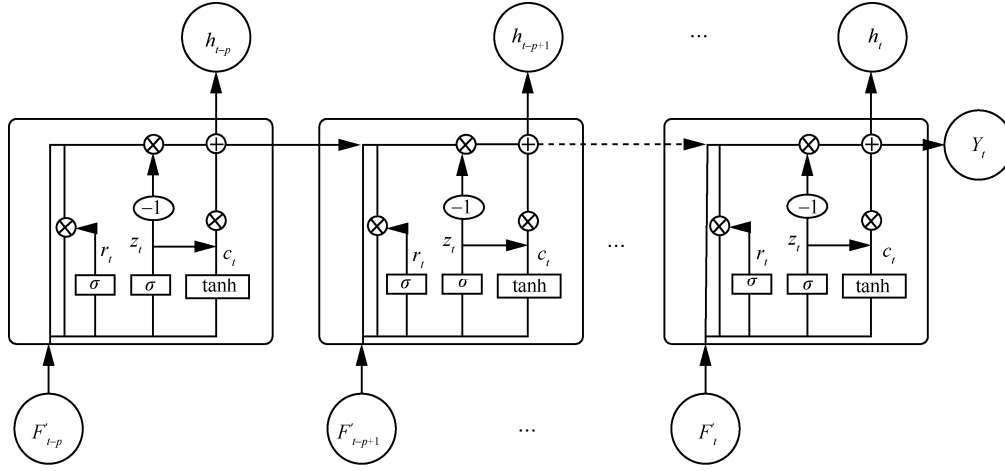


图 5 GRU 提取序列时间特征的实现细节

其中, z_t 是更新门, 定义保存到当前时间步长的历史信息量, 即决定有多少已知信息需要被传递到未来; $f(\mathbf{A}, \mathbf{F}_t)$ 是图卷积过程, 为第 t 个时间步的输入向量, 它与权重矩阵 $\mathbf{W}^{(z)}$ 相乘进行非线性处理; h_{t-1} 保存的是前一个时间步 $t-1$ 的信息, 它也需要与权重矩阵 $\mathbf{U}^{(z)}$ 相乘进行非线性处理。

$$r_t = \sigma(\mathbf{W}^{(r)} f(\mathbf{A}, \mathbf{F}_t) + \mathbf{U}^{(r)} h_{t-1}) \quad (10)$$

其中, r_t 是重置门, 决定新的输入与历史信息的结合情况, 即有多少历史信息需要被遗忘。如果某时刻的数据对应的门控值为 0, 该数据将被完全遗忘; 如果某时刻的数据对应的门控值为 1, 该数据将被完全代入下一时刻的计算。

$$c_t = \tanh(\mathbf{W}^{(c)} f(\mathbf{A}, \mathbf{F}_t) + (r_t \odot h_{t-1}) \mathbf{U}^{(c)}) \quad (11)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot c_t \quad (12)$$

其中, c_t 是当前时刻的记忆内容; h_t 是当前时间步长的最终记忆, 即当前时刻的输出内容; $\mathbf{W}^{(c)}$ 和 $\mathbf{U}^{(c)}$ 是用来进行非线性处理的权重矩阵; \odot 是矩阵乘积。

GRU 模型利用前一时刻的隐藏状态和当前时刻的流量数据作为输入来确定当前时刻的节点特征。通过门控机制, 在捕获当前节点特征时, 也保留过去的节点特征的变化趋势。因此, VNFPre 通过 GRU 模型有效地提取了待测数据的时间特征。

VNFPre 模型测数据流预测部分的损失函数定义为

$$\text{Loss} = \|\mathbf{V}_t - \hat{\mathbf{V}}_t\| + \lambda L_{\text{reg}} \quad (13)$$

其中, \mathbf{V}_t 和 $\hat{\mathbf{V}}_t$ 分别是时刻 t 的实际数据流序列和预测数据流序列, L_{reg} 是归一化项, λ 是一个超参数。

4 性能分析

4.1 实验方案

为了测试 VNFPre 模型的性能, 本文使用 Tensorflow 1.4、Python 3.6 实施了该预测模型。整个测试过程运行在一个配备 Intel Core i7-9700CPU、32 GB DDR4 内存和一个 GTX 1080Ti 显卡的电脑上。

1) 数据集策略

本文中用来评估模型的数据集为来自 TOTEM 项目^[23]的真实数据, 该项目提供了 XML 格式的 GÉANT 网络拓扑和其域内数据流矩阵。鉴于该数据集的每个节点仅包含数据流负载, 可视为单变量数据集, 且现实中部署的 NFV 网络数据流数据难以取得, 因此本文用该网络数据流数据集作为 VNFPre 模型和对比模型的输入数据。

首先, 本文将数据集划分为训练和测试数据集, 将 80% 的数据用作训练集, 其余 20% 用作测试集。下一步, 进行数据平稳化处理并转换为监督学习数据集, 以训练集为例来取样本。本文定义一个长度固定为 $S+1$ 的滑动窗口, 预测 $t+1$ 时刻的 VNF 实例数量。长度 $S+1$ 为一个训练样本的大小, 包含长度为 S 的历史数据输入和长度为 1 的 $t+1$ 时刻的输出。最后把这个窗口每次移动一个时间单位, 得到所有的训练样本, 再用同样的方法得到测试样本。

2) 实验设置

经过实验, VNFPre 模型的数据流预测部分网络优化器定为 ADAM。训练超参数选取如下: 将一

次训练所选取的样本数设置为 32, 学习率设置为 0.001, 训练回合数设置为 2 000, 神经网络隐藏单元数设置为 115, FNN 设置为 2 层结构。

4.2 对比方法

本文将 VNFPre 的性能与如下方法进行对比。

1) VNFPre-input。VNFPre-input 方法与本文提出的 VNFPre 方法结构相同, 但输入 VNFPre-input 的数据流序列仅采用当前时刻之前一段时间的数据流, 即 $[t-l, t]$ 时段的数据流作为流量预测模型的输入。

2) CNN-LSTM^[5]。CNN-LSTM 是同时使用时空特征进行 VNF 需求预测的最新方法。该解决方法每次只输入一个节点和其周边信息 (为网络中的每一个节点构建图), 通过 CNN 层提取特征, 经过全连接层变成向量, 结合其他因素, 然后输入 LSTM 网络中, 并对输入数据进行时空特征提取以支持序列预测。

3) LSTM^[7]。LSTM 方法使用 LSTM 模型来解决 VNF 资源容量需求预测问题。具体来说, LSTM 模型基于数据流负载来估计 VNF 的需求, 以实现 NFV 资源的动态管理。但是, 使用输入数据流作为估算 VNF 需求的唯一因素并不总是可靠的, 这可能是因为 NFV 环境中的其他因素 (例如网络拓扑结构) 正在影响 VNF 资源需求, 所以应该进行更深入的分析以实现更好的预测效果。

4.3 实验结果

1) 评价指标^[24]

均方根误差为

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (14)$$

其中, m 表示取样数量, y_i 表示真实值, \hat{y}_i 表示预测值。RMSE 反映模型预测值与真实值之间的误差程度, 取值范围为 $[0, \infty)$ 。RMSE 的值越小, 预测值越接近真实值, 误差越小。

准确度 R^2 为

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (\bar{y} - y_i)^2} \quad (15)$$

R^2 反映模型预测的准确程度, 取值范围为 $(0, 1]$ 。如果 $R^2 \rightarrow 0$, 说明该模型仅在盲猜预测结果; 如果 $R^2 = 1$, 说明该模型预测值和真实值一致。该指标的值越接近 1 则模型预测效果越好。

2) 性能分析

本节分析了几种方法的可用性, 从预测精度、预测误差和数据流预测结果的转换能力这 3 个方面对所提方法与对比方法进行分析和比较。

首先, 本文以模型评价指标为依据, 对 3.1 节输入数据流的权重值 a_1 和 a_2 的选择进行了实验分析。本文提出的数据流预测方法为实时预测, 因此 $[t-l, t]$ 时段的数据流序列应起到主要作用。进而, a_1 的取值范围应该为 $[0.5, 1]$ 。为节省训练时间, 本文分别对 a_1 的取值进行大步长和小步长结合的方式进行选择。如图 6 和图 7 所示, 在 a_1 较大步长取值时, VNFPre 的预测准确度和均方根误差在 $a_1 = 0.9$ 时达到最优。接着在 0.9 附近进行小步长测试, 最终确定当 $a_1 = 0.92$ 、 $a_2 = 0.08$ 时, VNFPre 的预测精度达到最优。

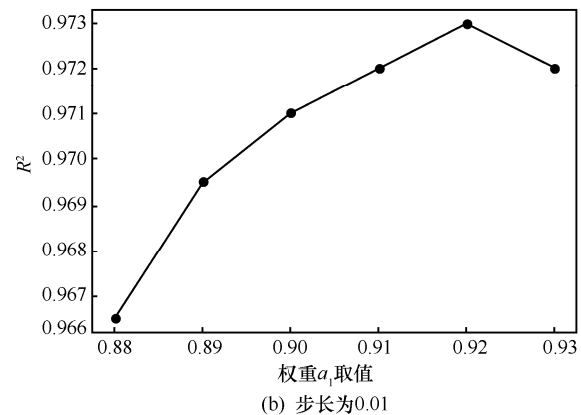
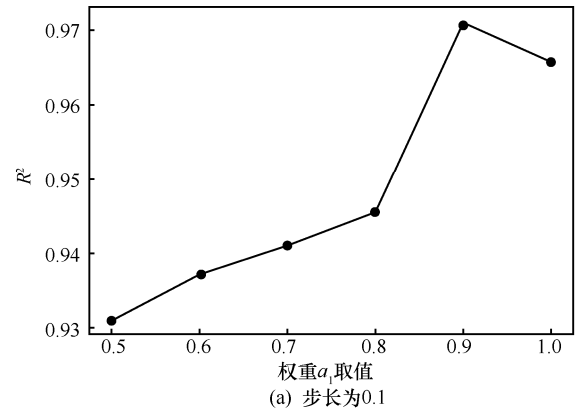


图6 不同取值步长下 a_1 对预测精度的影响

在输入数据流序列的权值选定且数据集中节点数据流最大负载为 6 Gbit/s 的情况下, 本文将 VNFPre 对数据流及 VNF 实例数量的预测结果和 VNFPre-input、CNN-LSTM 和 LSTM 模型的预测结果进行对比, 并从预测精度、预测误差和数据流预测结果的转换能力这 3 个角度分析了 VNFPre 模型性能。

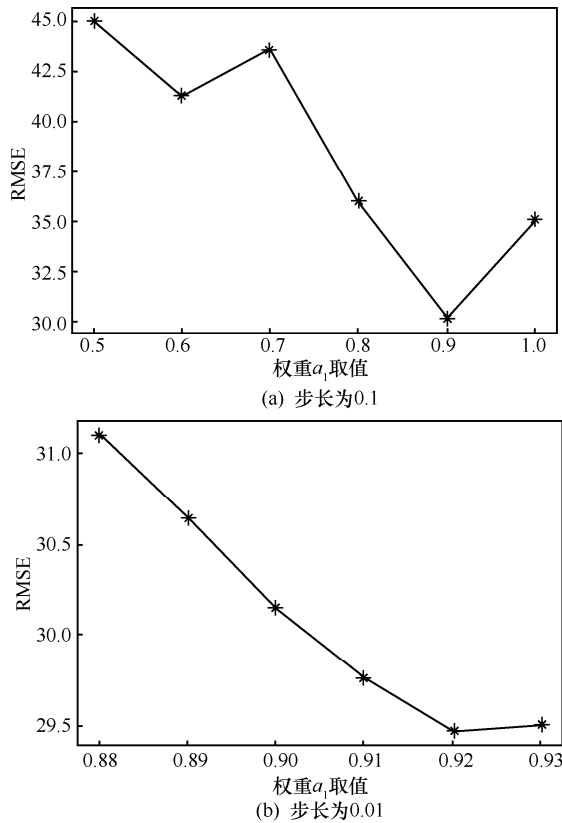


图 7 不同取值步长下 α_1 对预测误差的影响

各个方法预测准确度和误差如表 1 所示。其中，VNFPre 方法的预测准确度比 LSTM 方法提升了 18.04%，比 VNFPre-input 方法和 CNN-LSTM 方法分别提升了 8.41%和 6.54%。该结果证明了本文所提 VNFPre 方法在 VNF 资源容量需求预测问题上的有效性。

表 1 各方法预测准确度和误差值

方法	R^2	RMSE
VNFPre	0.973 0	29.142 0
VNFPre-input	0.897 5	30.212 5
CNN-LSTM	0.913 3	30.137 6
LSTM	0.824 3	32.082 1

本文提出的 VNFPre 方法和对比方法的模型预测误差对比如图 8 所示。由于 VNFPre 方法的输入和主体结构能够更加合理地预测数据流的变化情况，因而其预测误差比 VNFPre-input、CNN-LSTM 和 LSTM 方法分别低了 3.54%、3.30%和 9.16%，为下一阶段的 VNF 部署提供了更准确的 VNF 资源容量需求视图。

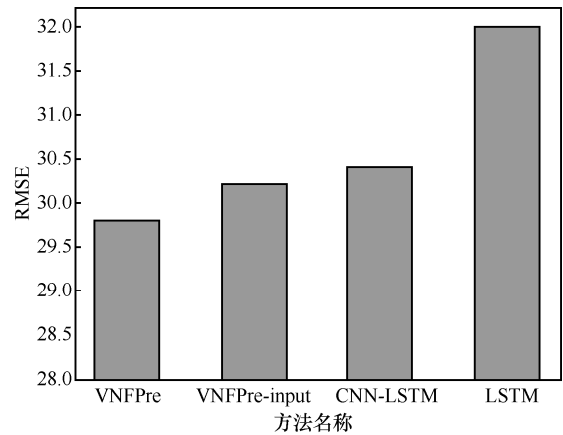


图 8 模型预测误差对比

图 9~图 12 为 VNFPre 和其他方法对数据集中 2005 年 1 月 16 日 24 小时的数据流预测结果。从图 9~图 12 中可以看出，与 VNFPre - input、CNN-LSTM 和 LSTM 这 3 种方法相比，VNFPre 方法在相同的数据流波动场景下与真实数据流波动趋势具有更好的拟合效果，其性能接近于理想状态。这是由于 VNFPre 中 GCN 和加权数据流输入序列共同作用的结果。其中，LSTM 方法的拟合效果相对其他方法较弱，并且在数据流波峰和波谷相距较近的位置难以拟合。其主要原因是 LSTM 无法根据节点间数据流之间潜在的相关性辅助预判数据流可能出现的突然变化。另外 3 种同时考虑了数据流时空相关性的 VNFPre、VNFPre-input 和 CNN-LSTM 方法对数据流波动的拟合情况整体好于 LSTM 方法。这是因为经过大量训练后，时空预测方案中的空间特征提取层能够迅速对网络的拓扑关系和数据流分布进行计算，从而得到更加准确的预测结果。其中，CNN-LSTM 方法将非欧几里得结构的数据流数据栅格化，变成网格结构的数据，而实际网络中的数据流数据不是处在规则的格子空间上，而是产生了高度不规则的偏移量。因此这种数据处理方式可能导致数据的空间稀疏性加剧，进而由于 CNN 权重共享的特性导致所有权重为零的极端情况^[25]。所以该模型可能不能准确地获取待测数据的空间特征，也就不能产生更准确的预测结果。

接下来，本文将数据流吞吐量的预测结果映射为节点上未来 VNF 实例数量的变化情况。如图 13~图 16 所示，在 24 小时的 VNF 实例数量预测中，VNFPre 方法仅有两处预测结果与真实的实例数量有所偏差；VNFPre-input 方法有 11 处偏差，且在序列起始和结尾处偏差较大；CNN-LSTM 方法有 12 处偏

差；LSTM 方法有 16 处偏差。从图 13~图 16 可以看出，本文所提的 VNFPre 方法在 VNF 实例数量预测上具有明显优势，并且 VNFPre-input 方法在 VNF 实例数量预测上与 CNN-LSTM 和 LSTM 方法相比也具有一定的优势。这是因为本文提出的这 2 种方法更好地结合了数据流波动趋势与 VNF 实例数量变化趋势之间的映射关系，更加高效地实现了数据流预测与 VNF 资源容量需求预测的转化。

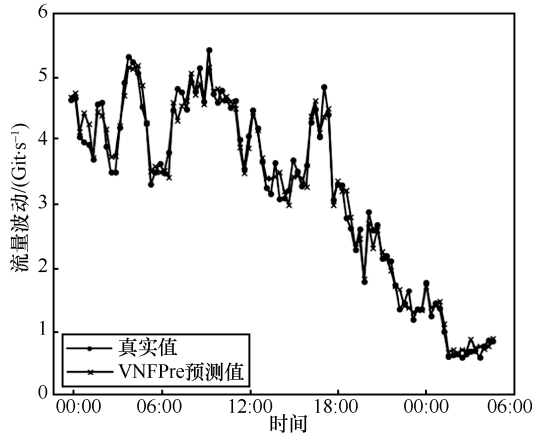


图 9 VNFPre 的数据流预测结果

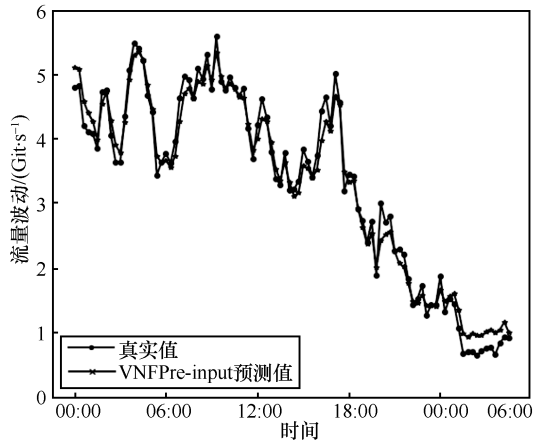


图 10 VNFPre-input 的数据流预测结果

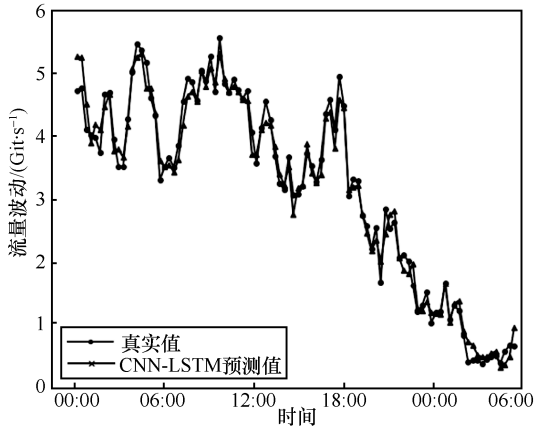


图 11 CNN-LSTM 的数据流预测结果

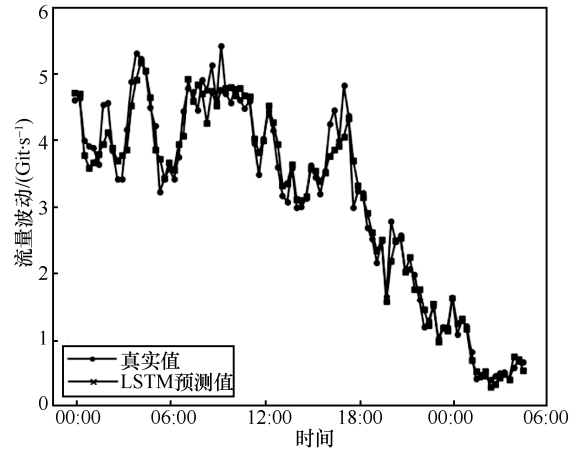


图 12 LSTM 的数据流预测结果

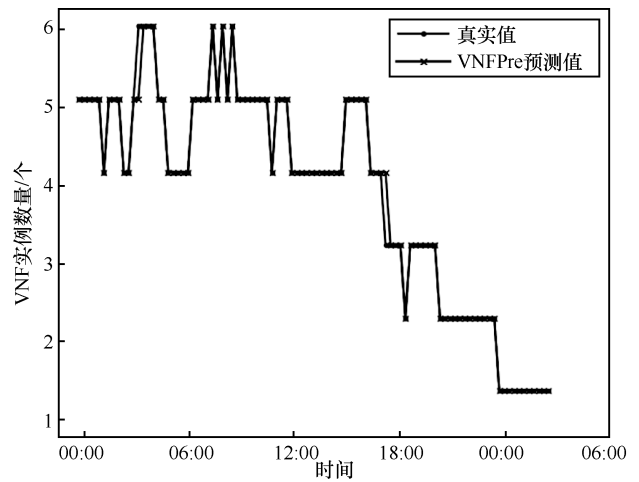


图 13 VNFPre 预测的 VNF 实例数量

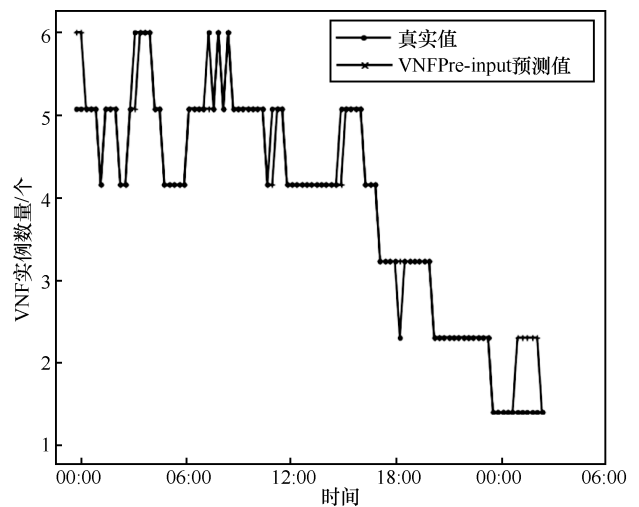


图 14 VNFPre-input 预测的 VNF 实例数量

5 结束语

多模态网络借助人工智能技术的蓬勃发展，为切片上的虚拟节点赋予计算、存储和传输能力，已经能够在自主感知数据吞吐和自主预测节点 VNF

的资源需求预测的基础上,通过业务需求数据的自适应流动,实现网络资源利用和用户体验的双重提升。然而,当前主流的机器学习解决方案存在模型参数与场景绑定、时间序列的空间特征提取不充分的问题,仍然在使用 CNN 提取非欧氏空间中的时间序列的地理分布特性。基于此,本文借助 GCN 和 GRU 提出了用于多种模态网络场景中 VNF 资源容量需求预测方法 VNFPre。实验结果说明,所提方法的预测精度比现有性能最优的 VNF 需求预测方法提高了 6.54%,对解决 VNF 资源容量需求预测问题有一定的实用价值。但所提方法采用两类成熟的神经网络进行组合,模型复杂度仍有较大的优化空间。未来可以考虑借鉴其他领域中进行时间序列预测的机器学习模型,将 2 个时空预测层进行融合与参数共享,降低模型复杂度。

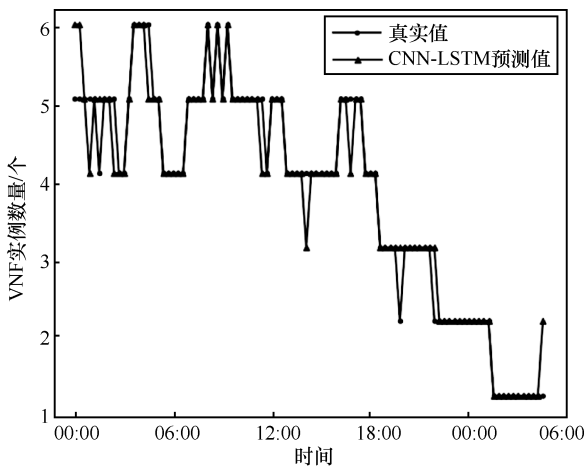


图 15 CNN-LSTM 预测的 VNF 实例数量

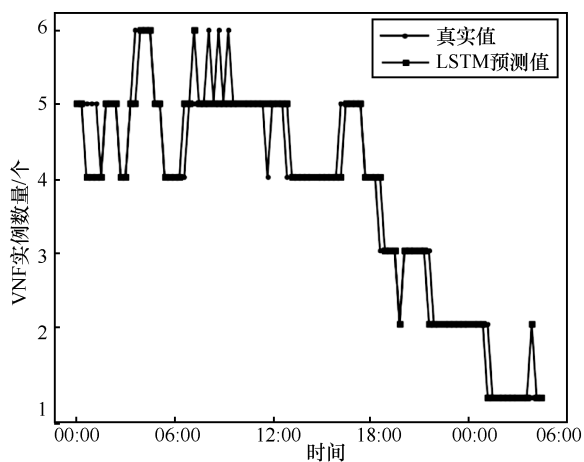


图 16 LSTM 预测的 VNF 实例数量

参考文献:

[1] 兰巨龙,程东年,胡宇翔.可重构信息通信基础网络体系研究[J].

通信学报, 2014, 35(1): 128-139.
 LAN J L, CHENG D N, HU Y X. Research on reconfigurable information communication basal network architecture[J]. Journal on Communications, 2014, 35(1): 128-139.
 [2] 胡宇翔,伊鹏,孙鹏浩,等.全维可定义的多模态智慧网络体系研究[J].通信学报,2019,40(8):1-12.
 HU Y X, YI P, SUN P H, et al. Research on the full-dimensional defined polymorphic smart network[J]. Journal on Communications, 2019, 40(8): 1-12.
 [3] NGMN Alliance. Description of network slicing concept[EB]. (2017-09-11) [2022-01-04].
 [4] AFOLABI I, BAGAA M, TALEB T, et al. End-to-end network slicing enabled through network function virtualization[C]//Proceedings of 2017 IEEE Conference on Standards for Communications and Networking. Piscataway: IEEE Press, 2017: 30-35.
 [5] ZAMAN Z, RAHMAN S, NAZMIN M. Novel approaches for VNF requirement prediction using DNN and LSTM[C]//2019 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.
 [6] QIU F, ZHANG B, GUO J. A deep learning approach for VM workload prediction in the cloud[C]//Proceedings of 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Piscataway: IEEE Press, 2016: 319-324.
 [7] 唐伦,周钰,杨友超,等.5G网络切片场景中基于预测的虚拟网络功能动态部署算法[J].电子与信息学报,2019,41(9):2071-2078.
 TANG L, ZHOU Y, YANG Y C, et al. Virtual network function dynamic deployment algorithm based on prediction for 5G network slicing[J]. Journal of Electronics & Information Technology, 2019, 41(9): 2071-2078.
 [8] TANG H, ZHOU D, CHEN D. Dynamic network function instance scaling based on traffic forecasting and VNF placement in operator data centers[J]. IEEE Transactions on Parallel and Distributed Systems, 2019, 30(3): 530-543.
 [9] ALAWE I, HADJADJ-AOUL Y, KSENTINI A, et al. Smart scaling of the 5G core network: an RNN-based approach[C]//Proceedings of 2018 IEEE Global Communications Conference. Piscataway: IEEE Press, 2018: 1-6.
 [10] THAKKAR H K, DEHURY C K, SAHOO P K. MUVINE: multi-stage virtual network embedding in cloud data centers using reinforcement learning-based predictions[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(6): 1058-1074.
 [11] PATEL Y S, VERMA D, MISRA R. Deep learning based resource allocation for auto-scaling VNFs[C]//Proceedings of 2019 IEEE International Conference on Advanced Networks and Telecommunications Systems. Piscataway: IEEE Press, 2019: 1-6.
 [12] SCALINGI A, ESPOSITO F, MUHAMMAD W, et al. Scalable provisioning of virtual network functions via supervised learning[C]//Proceedings of 2019 IEEE Conference on Network Softwarization (NetSoft). Piscataway: IEEE Press, 2019: 423-431.
 [13] TONG R M, XU S Y, HU B, et al. VNF dynamic scaling and deploy-

- ment algorithm based on traffic prediction[C]//Proceedings of 2020 International Wireless Communications and Mobile Computing (IWCMC). Piscataway: IEEE Press, 2020: 789-794.
- [14] TAO J, LU Z, CHEN Y, et al. Adaptive VNF scaling approach with proactive traffic prediction in NFV-enabled clouds[C]//ACM Turing Award Celebration Conference. New York: ACM Press, 2021: 166-172.
- [15] MIJUMBI R, HASIJA S, DAVY S, et al. Topology-aware prediction of virtual network function resource requirements[J]. IEEE Transactions on Network and Service Management, 2017, 14(1): 106-120.
- [16] JALODIA N, HENNA S, DAVY A. Deep reinforcement learning for topology-aware VNF resource prediction in NFV environments[C]//Proceedings of 2019 IEEE Conference on Network Function Virtualization and Software Defined Networks. Piscataway: IEEE Press, 2019: 1-5.
- [17] KIM H G, JEONG S Y, LEE D Y, et al. A deep learning approach to VNF resource prediction using correlation between VNFs[C]//Proceedings of 2019 IEEE Conference on Network Software Virtualization (NetSoft). Piscataway: IEEE Press, 2019: 444-449.
- [18] ZHUANG Z R, WANG J Y, QI Q, et al. Toward greater intelligence in route planning: a graph-aware deep learning approach[J]. IEEE Systems Journal, 2020, 14(2): 1658-1669.
- [19] SHEN Y F, SHI Y M, ZHANG J, et al. Graph neural networks for scalable radio resource management: architecture design and theoretical analysis[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(1): 101-115.
- [20] FOUKAS X, PATOUNAS G, ELMOKASHFI A, et al. Network slicing in 5G: survey and challenges[J]. IEEE Communications Magazine, 2017, 55(5): 94-100.
- [21] THOMAS R W, FRIEND D H, DASILVA L A, et al. Cognitive networks: adaptation and learning to achieve end-to-end performance objectives[J]. IEEE Communications Magazine, 2006, 44(12): 51-57.
- [22] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv Preprint, arXiv: 1609.02907, 2016.
- [23] UHLIG S, QUOITIN B, LEPROPRE J, et al. Providing public intradomain traffic matrices to the research community[J]. ACM SIG COMM Computer Communication Review, 2006, 36(1): 83-86.
- [24] 武静雯, 江凌云, 刘祥军. 基于特征选择的 VNF 资源需求预测方法[J]. 计算机应用研究, 2021, 38(10): 3131-3136, 3142.
- WU J W, JIANG L Y, LIU X J. VNF resource demand forecast method based on feature selection[J]. Application Research of Computers, 2021, 38(10): 3131-3136, 3142.
- [25] WANG B, LUO X Y, ZHANG F B, et al. Graph-based deep modeling and real time forecasting of sparse spatio-temporal data[J]. arXiv Preprint, arXiv: 1804.00684, 2018.

[作者简介]



兰巨龙（1962- ），男，河北张家口人，博士，信息工程大学教授、博士生导师，主要研究方向为未来信息通信网络关键理论与技术。



朱隼（1992- ），女，甘肃兰州人，信息工程大学硕士生，主要研究方向为新型网络体系结构、网络智慧化等。



李丹（1989- ），男，辽宁沈阳人，博士，信息工程大学副研究员，主要研究方向为新型网络体系结构、网络智慧化等。