

# 知识蒸馏驱动的卫星网络流量轻量化预测模型

何杏宇<sup>1,2</sup>,李念慈<sup>1</sup>,杨桂松<sup>1</sup>

<sup>1</sup>(上海理工大学 光电信息与计算机工程学院,上海 200093)

<sup>2</sup>(上海理工大学 出版学院,上海 200093)

E-mail: gsyang@aliyun.com

**摘要:**传统流量预测模型计算复杂度高,难以适应卫星网络的资源限制,同时对卫星流量的时空特性建模不足,无法有效表征其拓扑变化和流量分布,导致预测精度与计算效率难以兼顾,限制了其在卫星网络中的实际应用.为此,提出一种由知识蒸馏驱动的轻量化卫星网络流量预测模型,旨在捕捉卫星网络流量的时空特征并降低模型复杂度.首先,该模型通过构建结合图卷积网络(GCN)和Transformer的教师模型,以及融合GCN和门控循环单元(GRU)的学生模型,实现时空特征提取.其次,设计包含软损失、硬损失、特征损失和时序动态加权KL(Kullback-Leibler)散度损失的蒸馏机制,利用多层次特征提取与动态序列权重调整,引导学生模型高效学习教师模型中的时空依赖关系与流量动态.实验结果表明,蒸馏后的学生模型在MSE和MAE上分别降低26.1%和17.4%,接近教师模型性能,同时参数量、计算量和推理时间分别减少75.6%、78.6%和3.6倍,展示了其在资源受限环境下的高效性.

**关键词:**卫星网络;流量预测;时空特征;知识蒸馏;动态加权损失

中图分类号: TP393

文献标识码: A

文章编号: 1000-1220(2026)04-0974-08

## Lightweight Prediction Model of Satellite Network Traffic Driven by Knowledge Distillation

HE Xingyu<sup>1,2</sup>, LI Nianci<sup>1</sup>, YANG Guisong<sup>1</sup>

<sup>1</sup>(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

<sup>2</sup>(College of Publishing, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** Most traditional traffic prediction models are unsuitable for resource-constrained satellite networks, due to their high computational complexity. Furthermore, they also fail to capture the spatial-temporal features of satellite traffic, such as topological changes and traffic distribution, which hampers the balance between prediction accuracy and computational efficiency, limiting their practical applicability in satellite networks. This paper proposed a lightweight traffic prediction model for satellite networks driven by knowledge distillation to capture the spatial-temporal features of satellite network traffic and reduce the complexity. The model first constructed a teacher model that integrates GCN (Graph Convolutional Neural Network) and Transformer, and a student model combining GCN and GRU (Gated Recurrent Units), to extract spatial-temporal features of satellite network traffic. Further, this model designed a distillation mechanism that incorporates soft loss, hard loss, feature loss, and dynamic weighted KL (Kullback-Leibler) divergence loss. This mechanism employed multi-level feature extraction and dynamic sequence weight adjustment, guiding the student model to learn spatial-temporal dependencies and traffic dynamics from the teacher model. Experimental results showed that, the distilled student model reduced MSE by 26.1% and MAE by 17.4%, approaching the performance of the teacher model. Furthermore, the proposed model achieved a 75.6% reduction in parameters, a 78.6% reduction in computational complexity, and a 3.6-fold decrease in inference time, which demonstrates its efficiency in resource-constrained environments.

**Keywords:** satellite network; traffic prediction; spatial-temporal feature; knowledge distillation; dynamic weighted loss

## 0 引言

近年来,随着卫星通信技术的迅猛发展,卫星网络在全球通信、导航定位、环境监测以及军事国防等诸多领域的应用愈发广泛,已然成为现代信息基础设施的关键组成部分<sup>[1,2]</sup>.据相关统计数据显示,全球卫星通信市场规模预计在未来数年内将以每年超过10%的增长率持续扩张,越来越多的用户终端接入卫星网络,促使卫星网络流量呈现出爆炸式增长态势.

面对如此海量且复杂多变的卫星网络流量,如何实现精准、高效的流量预测,对于优化卫星网络资源配置、保障通信服务质量具有举足轻重的意义<sup>[3]</sup>.

精准的流量预测在卫星网络中能支持资源分配、路由优化和服务质量保障<sup>[4,5]</sup>.运营商可根据预测合理调配带宽,避免拥塞,提高效率;通过流量趋势预测优化传输路径,减少延迟和信号损耗,降低成本;同时确保关键业务在高流量时段稳定运行,满足用户需求,推动卫星通信在应急、军事和偏

收稿日期:2025-03-10 收稿修改日期:2025-04-11 基金项目:国家自然科学基金项目(61602305,61802257)资助;上海市自然科学基金项目(18ZR1426000,19ZR1477600)资助. 作者简介:何杏宇,女,1984年生,博士,副教授,CCF会员,研究方向为卫星网络、无人系统和群智计算;李念慈,女,1999年生,硕士研究生,研究方向为卫星网络流量预测;杨桂松(通信作者),男,1982年生,博士,副教授,CCF会员,研究方向为物联网、卫星网络、边缘计算与普适计算等.

远地区的应用,促进产业发展。

目前,流量预测的研究方法主要分为传统统计学模型和基于机器学习的方法。传统统计学模型,如自回归差分移动平均模型(Autoregressive Integrated Moving Average, ARIMA)<sup>[6]</sup>,能有效捕捉简单的趋势和周期性变化,但难以应对卫星网络流量的非线性特性。虽然机器学习方法如支持向量机(Support Vector Machine, SVM)<sup>[7]</sup>和随机森林(Random Forests, RF)拓展了传统模型的适用范围,能够更好地捕捉数据中的非线性特征,但依然面临卫星网络流量中的多维动态性、长时间依赖性 & 复杂的时空关联性挑战。为解决这些问题,近年来,深度学习技术逐渐成为流量预测领域的研究热点。模型如长短期记忆网络(Long Short-Term Memory, LSTM)<sup>[8]</sup>、门控循环单元(GRU)和 Transformer 等,凭借其强大的特征提取和时序建模能力,在一定程度上显著提升了卫星网络流量预测的精度。如 Cai 等<sup>[9]</sup>基于 LSTM 结合人口密度调整覆盖范围的流量趋势,提升了预测精度;Zhu 等<sup>[10]</sup>引入注意力机制增强 LSTM 模型,通过加强自相关性特征,加快了模型的收敛速度并验证了其有效性。Liu 等<sup>[11]</sup>结合注意力机制与 GRU 模型,挖掘流量数据序列的自相似性和长相关性,从而提升了预测能力。考虑卫星网络流量的自身特性, Li 等<sup>[12]</sup>提出基于迁移学习的 GRU 神经网络,有效表达卫星网络流量的非线性、自相似性及长相关性特征,同时降低了模型的复杂度。Yang 等<sup>[13]</sup>结合图卷积神经网络(GCN)与 GRU,并利用卫星网络的时空特征及节点关联性,实现了更高精度的预测。Cao 等<sup>[14]</sup>提出 TLS-Net 模型,结合时间卷积网络(Temporal Convolutional Network, TCN)和 LSTM,捕捉卫星网络流量数据的长短期特征,进一步提升了预测精度。此外,针对特定场景的优化策略也取得了一定成效,如 Zhao 等<sup>[15]</sup>通过区域划分与径向基函数训练降低了计算复杂度并提升了区域流量预测精度。上述研究在卫星网络流量预测中取得了一定进展,但仍面临以下局限性:

1) 传统深度学习模型通常计算复杂且需要大量数据,而卫星作为空间飞行器,其计算和存储资源非常有限,难以直接部署大规模、复杂的深度学习模型<sup>[16]</sup>,这使得流量预测任务的计算复杂性和资源消耗成为重大挑战;

2) 卫星网络的节点间距大、信号传播时延长且拓扑结构因卫星高速运动而持续变化,增加了流量时空相关性的复杂性<sup>[17]</sup>。此外,卫星网络流量受地球自转、公转以及人类活动周期等因素的影响,呈现明显的周期性规律<sup>[18]</sup>,这对预测精度提出了更高要求。

因此,在确保预测精度的前提下,如何在有限资源下设计轻量化并能够有效考虑卫星流量时空特性的预测模型,成为当前卫星网络领域亟待攻克的关键难题。

针对深度模型轻量化问题,研究者们提出了多种优化方法以提升计算效率和存储性能,广泛应用于资源受限环境中的多个领域。例如,剪枝、量化、迁移学习以及知识蒸馏(Knowledge Distillation, KD)等方法被用于简化模型结构、减少计算开销。Zhang 等人<sup>[19]</sup>通过结构化剪枝来消除不重要的冗余元素,利用强化学习自动执行剪枝过程,以寻找最佳压缩率。Thaker 等人<sup>[20]</sup>通过正则化损失关注通道级参数组来优化剪枝过程,从而修剪卷积层以降低计算成本。Li 等人<sup>[21]</sup>通过

剪枝和量化操作,降低了深度学习模型的复杂度和存储空间占用,并以最小熵评估进行改进模型,提高有效性。He 等人<sup>[22]</sup>设计了一种纯整数标量二次幂量化的模型压缩方法,通过分布损失正则化器减少了量化误差和训练干扰。Lin 等人<sup>[23]</sup>通过解耦时间序列数据中的周期性来简化预测任务,提出 SparseTSF 以跨周期稀疏预测技术为核心轻量化预测模型,降低计算复杂度。Tsanakas 等人<sup>[24]</sup>利用多种知识蒸馏方法对深度学习模型进行轻量化处理,成功应用于边缘设备,有效缩短推理时间。Zhang 等人<sup>[25]</sup>提出了基于知识蒸馏的时空多层感知器(Spatial-Temporal Multilayer Perceptron, ST-MLP)模型,通过将复杂模型的知识迁移至简化结构,实现了实时交通流量预测中的高效性和精准性。Chen 等人<sup>[26]</sup>进一步提出了基于连续知识蒸馏的模型,聚焦于天地一体化网络的流量预测,显著提高了模型精度与适应性。

然而,尽管这些技术在边缘计算、交通流量预测等领域取得了显著成效,但在卫星网络流量预测领域少有相关应用研究。另外,相比传统的剪枝或量化方法,知识蒸馏能够将复杂模型(教师模型)的知识迁移到简化模型(学生模型)中,保留对卫星网络流量的时空动态和周期性规律的有效感知。这种方法不仅在提升计算效率的同时,避免了过度简化带来的性能损失,还能针对卫星网络流量的特殊挑战,优化模型的预测精度。

综上所述,如何结合卫星网络流量的时空特性,设计知识蒸馏机制进一步优化轻量化模型,是解决卫星网络流量预测问题的关键所在。为此,本研究提出了一种知识蒸馏驱动的卫星网络流量轻量化预测模型,旨在深入挖掘卫星网络流量的时空相关性和周期性等内在特征,并利用知识蒸馏技术实现模型的轻量化与性能优化。具体而言,本研究的目标主要包括以下 3 个方面:首先,设计适用于卫星网络流量预测的教师模型与学生模型架构,充分挖掘流量数据中的时空规律,为知识蒸馏提供高质量的模型基础;其次,构建高效的蒸馏模块,优化蒸馏损失函数,以提高知识从教师模型向学生模型迁移的精度与效率;最后,基于实际卫星网络流量数据集对所提模型进行训练与验证,并通过实验对比分析模型的性能表现,进而为卫星网络流量预测任务提供可靠且高效的解决方案。

## 1 卫星流量特征分析与问题定义

### 1.1 卫星网络流量时空特征

卫星网络流量的时空相关性是其显著特征,深入分析这一特性对于精确构建流量预测模型至关重要。从时间维度来看,单颗卫星的流量表现出明显的周期性波动和趋势性变化。例如,在工作日的白天,随着用户活动的增加,流量逐步上升,而在夜间流量则相对平稳,处于较低水平。卫星网络流量的周期性变化可以分为小周期和大周期:小周期通常与卫星的轨道周期相关,通常为 90 ~ 120 分钟,流量随卫星覆盖不同地理区域而波动;大周期则反映了卫星轨道与地球自转的相互作用,流量在较长时间尺度上呈现重复模式。

由于流量的时序性波动和周期性,需要采用能够建模时序特征的模型,例如,循环神经网络(RNN)、长短时记忆网络(LSTM)或 Transformer 等。然而传统的时间序列预测方法可

能不足以捕捉流量变化的复杂模式.因此,如何精确预测流量的周期性变化,并在突发事件发生时快速响应,成为预测模型设计中的重点问题.

从空间维度来看,卫星网络流量的分布受到地理位置、用户分布等多种因素的影响,表现出复杂的空间相关性.不同地区的经济发展水平、人口密度及通信需求差异导致流量的空间分布不均.例如,城市地区由于用户密集和需求旺盛,流量较大,而偏远地区的流量则较少.此外,卫星节点之间的流量常常呈现协同变化,某一区域发生突发事件时,流量波动可能会通过星间链路传播到相邻节点,导致相邻节点流量的波动.

流量的空间分布不均和节点之间的协同波动要求模型能够处理复杂的空间依赖性.传统的流量预测模型往往忽视了卫星网络中的拓扑结构和节点间的依赖性,因此,设计预测模型时应采用能够处理卫星节点间动态连接关系和复杂网络拓扑的模型,如图神经网络(GNN)、图卷积网络(GCN)等方法,从而能够更准确地捕捉卫星流量的空间相关性,提高预测准确性.

## 1.2 问题定义

本研究聚焦于卫星网络流量预测任务,旨在构建由知识蒸馏驱动的轻量化预测模型,精准预测卫星网络流量变化趋势,为资源优化配置、网络高效运行提供有力支撑.在卫星网络流量预测模型的设计中,关键是准确捕捉流量的时序性和空间特征.时间上,流量的周期性波动和突发事件要求模型处理长短期时序依赖,准确预测突发波动;空间上,流量的不均匀分布和卫星节点间的协同波动要求模型捕捉节点间的动态连接和复杂拓扑,从而精准反映空间相关性.具体而言,设卫星网络在一定时间段内的历史流量数据为  $X = \{x_1, x_2, \dots, x_T\}$ ,其中  $x_t$  表示在时刻  $t$  的流量观测值, $T$  为历史数据长度.预测任务目标是利用这些历史数据,通过构建的模型预测未来  $n$  个时刻的流量值,即得到预测序列  $\hat{Y} = \{\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+n}\}$ ,使预测值尽可能接近真实未来流量值  $Y = \{y_{T+1}, y_{T+2}, \dots, y_{T+n}\}$ .

然而,卫星网络通常面临计算能力和带宽等资源的限制,因此模型需要进行轻量化处理.轻量化不仅有助于减小计算负担和资源消耗,还能确保模型在低延迟和低带宽环境中高效运行,从而提升实时性并适应卫星网络的动态变化.因此,在卫星网络流量预测任务中,如何在保证预测准确性的基础上,设计一个能够高效运行、低资源消耗的轻量化模型,是当前面临的核心问题.

为此,模型需在时空特征的复杂性与计算资源的限制之间取得平衡,以实现高效、实时的流量预测.这要求设计一种能够提取卫星流量时空特征的教师和学生模型的架构,充分利用知识蒸馏方法,构建适应卫星网络时空特性的蒸馏机制,实现跨层次特征提取与高效知识传递,提升学生模型的学习能力,得到兼顾预测精度与计算效率的轻量化模型.在减少模型复杂度和计算负担的同时,尽可能保持其预测精度和泛化能力,以确保模型能够在低带宽、低算力的卫星环境中高效运行,实现精准的流量预测.

## 2 轻量化预测模型框架

本节设计了一种知识蒸馏驱动的轻量化卫星网络流量预

测模型,结合卫星网络流量的时空特性,构建了教师模型和学生模型,通过知识蒸馏机制实现知识迁移,并设计了综合性蒸馏损失函数以提升学生模型的预测性能.在满足卫星设备资源限制的同时,确保模型的预测精度.

### 2.1 教师模型

在卫星网络中,节点流量具有显著的时空相关性和动态性,空间特征反映了不同卫星节点间的流量关联,而时间特征揭示了流量的周期性和时序依赖.为了充分挖掘这些特性,教师模型设计融合了GCN和Transformer,如图1(a)所示,旨在充分挖掘卫星网络流量数据的时空特征,为学生模型提供精准且丰富的知识引导.其中,GCN用于提取卫星网络流量数

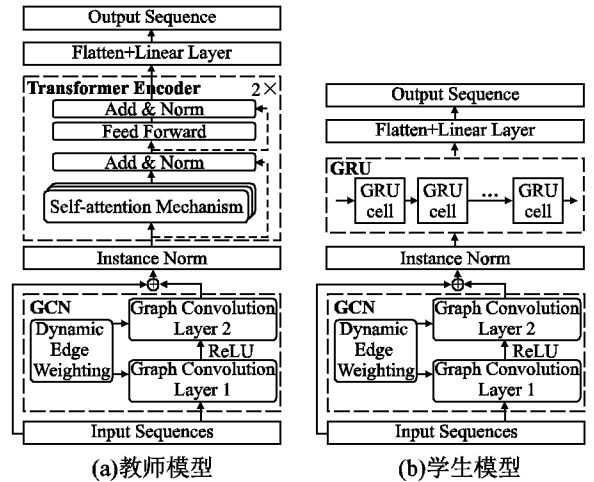


图1 教师模型和学生模型框架图

Fig. 1 Teacher model and student model framework

据的空间特征,通过对网络拓扑图的卷积操作,聚合节点间信息,捕捉流量在不同区域的空间相关性;Transformer则专注于挖掘流量数据的时间动态特性,利用其多头注意力机制,对序列流量数据中的复杂依赖关系进行建模,学习不同时段流量变化模式.教师模型凭借强大的学习与表征能力,学习到流量数据中的隐含知识,为后续知识迁移奠定基础.

#### 2.1.1 空间特征提取模块

卫星网络的节点通过星间链路和星地链路将节点连接起来,形成一个高度动态的通信网络.在卫星网络中,节点的流量不仅与自身负载相关,还受邻近节点流量和网络拓扑变化的影响.由于卫星网络的拓扑结构和节点流量会随时间变化,捕捉这些动态性和空间依赖性对流量预测至关重要.GCN作为一种专门处理图结构数据的深度学习模型,能够有效建模卫星网络中的空间关系及其时变特性.GCN通过构建基于网络拓扑变化的邻接矩阵,表示节点间的连接关系,并通过卷积操作聚合邻居节点的信息,从而提取卫星网络的空间特征.通过多层卷积操作,GCN逐层聚合来自多个跳数邻居的信息,精确建模流量的空间分布及其依赖关系,灵活适应卫星网络的动态拓扑变化,为流量预测提供强有力的空间特征表示.

卫星网络的拓扑结构可以用图  $G = (V, E)$  表示,其中  $V$  是卫星节点集合, $E$  是节点间的链路集合.卫星节点的流量特征矩阵记为  $X \in \mathbb{R}^{N \times F}$ ,其中  $N$  是节点数, $F$  是流量特征维度.在卫星网络的实际场景中,邻接矩阵  $A(t)$  的边权重并非固定不变,而是动态变化的.这种动态性主要由相邻卫星之间的距

离决定,随着卫星在轨道上的不断移动,星间链路的通信质量和权重会随时间发生变化.为了更加准确地反映卫星网络的时空特性,邻接矩阵的边权重可以根据相邻卫星之间的距离  $d_{ij}(t)$  定义为:

$$a_{ij}(t) = \exp\left(-\frac{d_{ij}(t)}{\delta}\right) \quad (1)$$

其中  $a_{ij}(t)$  为边权重,值越大表示节点间的联系越紧密,  $d_{ij}(t)$  是时刻  $t$  卫星  $i$  和  $j$  之间的欧几里得距离;  $\delta$  是平滑参数,用于控制权重对距离的敏感程度.

这种设计使得邻接矩阵动态调整权重,能够在每个时间步动态捕捉卫星间的物理位置变化对流量相关性的影响.在此基础上,基于邻接矩阵的图卷积操作表达如下:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (2)$$

其中,  $\tilde{A} = I + A$  为添加了自连接的邻接矩阵,  $I$  是单位矩阵,确保每个节点都能包含自身信息;  $\tilde{D}$  是  $\tilde{A}$  的度矩阵,其对角元素  $\tilde{d}_i$  等于  $\tilde{A}$  第  $i$  行元素之和;  $H^{(l)}$  为第  $l$  层的节点特征矩阵;  $\sigma$  为 ReLU 激活函数,用于引入非线性特性;  $W^{(l)}$  是第  $l$  层的可学习权重矩阵.

通过迭代式的聚合与更新操作,GCN 能够逐步融合每个节点与其邻居节点的信息,进而提升节点特征的表达力,从而提取卫星网络流量中的空间特征  $X_{\text{space}}$ . 提取出的空间特征可与原特征进行融合,形成时间特征提取的输入  $X' = \text{context}(X, X_{\text{space}})$ ,为后续流量预测提供强有力的支持.

### 2.1.2 时间特征提取模块

Transformer 模块专注于学习卫星网络流量的时间特征,捕捉复杂的时间序列动态变化.针对卫星网络中实时通信与数据回传等多样化业务的需求,Transformer 利用其多头注意力机制并行建模不同时段流量数据间的依赖关系.在处理长时间序列流量数据时,它为每个时间步分配多个注意力头,分别关注不同子序列的短期波动和长期周期性变化,深度挖掘流量的时间规律,为精准预测提供支撑.

在教师模型中,Transformer 模块采用编码器结构,由多头注意力机制和前馈神经网络构成.多头注意力机制计算过程如下:首先,将原始序列特征与空间特征的融合结果作为输入序列  $X' = \{x'_1, x'_2, \dots, x'_T\}$ ,通过线性变换得到查询矩阵  $Q$ 、键矩阵  $K$ 、值矩阵  $V$ :

$$Q = X' W^Q, K = X' W^K, V = X' W^V \quad (3)$$

其中,  $W^Q$ 、 $W^K$ 、 $W^V$  为可学习的权重矩阵.

然后,计算注意力得分:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

其中,  $d_k$  为键向量的维度,用于缩放注意力得分,避免因维度过高导致梯度消失或梯度爆炸问题.

最后,通过多头注意力并行计算并拼接结果:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(h_1, h_2, \dots, h_h) W^O \quad (5)$$

其中,  $h$  为头的数量,  $h_i$  为第  $i$  个头的注意力计算结果,  $W^O$  为拼接后的线性变换权重矩阵.

当输入特征经过多头注意力计算后,首先会通过残差连接(Add)与层归一化(Norm)进行初步整合,得到中间表示  $y'$ ,这一步骤能够保留原始输入信息并稳定特征分布.随后,中间特征被送入前馈神经网络(FFN),通常由两个全连接层

和一个激活函数(ReLU)组成,用于进一步对特征进行变换:

$$\text{FFN}(y') = W_2 \cdot \text{ReLU}(W_1 \cdot y' + b_1) + b_2 \quad (6)$$

其中,  $W_1$  和  $W_2$  分别是两个全连接层的权重矩阵,用于线性变换特征维度,而  $b_1$  和  $b_2$  是相应的偏置向量,用于调整神经元激活阈值和特征分布.最后,FFN 的输出再次与中间特征  $y'$  进行残差连接,并通过层归一化处理,形成模块的最终输出  $\hat{y}$ .

通过多层的编码器堆叠,Transformer 能够逐步提升对流量序列特征的提取能力,有效捕捉流量在时间维度上的复杂变化规律,为卫星网络流量预测提供精准的时间序列特征信息.将 GCN 与 Transformer 相结合,教师模型能够充分发挥两者优势,对卫星网络流量的时空特征进行全面、深入的学习,为后续知识蒸馏提供高质量的知识源.

## 2.2 学生模型

学生模型的设计目标是在资源受限的卫星环境下构建一个高效、轻量化的架构,能够充分利用教师模型的知识,同时精准预测卫星网络流量.为此,学生模型采用 GCN 和 GRU 相结合的结构,如图 1(b) 所示.其中,GCN 用于提取网络的空间特征,GRU 负责捕捉时间序列的动态特性,实现了时空特征的高效融合.

### 2.2.1 空间特征提取模块

学生模型复用教师模型的空间特征提取部分,主要由于该模块已充分捕捉了卫星网络流量的空间相关性.具体而言,教师模型中的 GCN 通过卷积操作聚合邻居节点信息,提取节点之间的依赖关系,从而反映区域间流量的协同变化特征.此外,GCN 在优化过程中结合卫星网络的特点,引入了带权邻接矩阵和多层堆叠设计,以增强模型的表现力和适应性.

复用设计的主要优势在于,学生模型能够继承教师模型在空间特征学习中的有效性,避免了重新设计和训练高复杂度的特征提取模块.通过复用优化后的 GCN 模块,学生模型能够在较短时间内实现高效的时空特征提取,显著降低计算资源消耗.该设计还减少了模型参数量,减轻了计算负担,并通过一致的特征表示提高了知识蒸馏效果,避免了信息丢失,从而确保了流量预测的准确性和稳定性.

### 2.2.2 时间特征提取模块

卫星网络流量具有显著的时间序列特性,随着时间的推移,流量数据呈现出复杂的动态变化,受到多种因素影响,如地球自转、区域内的人类活动规律、气象条件变化等.学生模型采用 GRU 作为时间特征提取模块,凭借其高效的门控机制,在保留时间序列信息的同时降低了计算复杂度. GRU 包含更新门和重置门两个关键机制,其计算过程包括:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x'_t] + b_z) \quad (7)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x'_t] + b_r) \quad (8)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x'_t] + b_h) \quad (9)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (10)$$

其中,  $x'_t$  为原始序列特征与空间特征的融合结果;  $z_t$  和  $r_t$  分别为更新门和重置门;  $W_z$ 、 $W_r$ 、 $W_h$  为可学习权重矩阵;  $b_z$ 、 $b_r$ 、 $b_h$  为偏置项;  $\odot$  表示逐元素乘法.

在处理时间序列流量数据时,选择 GRU 而非 LSTM 或普通循环神经网络(Recurrent Neural Network, RNN),主要考虑卫星终端的资源限制和流量预测精度需求. RNN 因梯度消失问题难以学习长期依赖,导致预测精度较低; LSTM 虽通过

门控机制缓解梯度问题,但结构复杂、参数量大,计算成本高,不适合资源受限的卫星终端.相比之下,GRU结构更为精简,仅包含更新门与重置门.更新门 $z_t$ 用于控制前一时刻隐藏状态信息 $h_{t-1}$ 保留到当前时刻的程度,重置门 $r_t$ 则决定了如何将新输入信息 $x_t$ 与前一时刻隐藏状态 $h_{t-1}$ 相结合.在卫星实时流量监测场景下,GRU能够利用其门控机制高效筛选信息,保留关键流量变化趋势,遗忘冗余噪声信息.面对突发通信高峰的流量突变时,重置门 $r_t$ 会增大新输入流量信息的权重,使模型快速响应突变;更新门 $z_t$ 则适当保留历史信息,以参考正常流量的基准水平,预测峰值后流量回落的趋势,以较小计算代价实现高精度流量预测,有效适配卫星资源受限环境,在保证一定预测精度的同时,降低模型复杂度,提升整体运行效率.

### 2.3 知识蒸馏模块构建

知识蒸馏的核心思想是将教师模型中学到的知识迁移到轻量化的学生模型,减少模型复杂度和计算资源消耗,同时保证预测性能.在卫星网络流量预测中,教师模型通过复杂结构有效捕捉时空特征,而学生模型则简化以适应资源有限的卫星环境.由于卫星流量的动态性,教师模型捕捉的时空依赖关系具有时变性,学生模型需在有限资源下有效学习这些特征,确保高效预测.

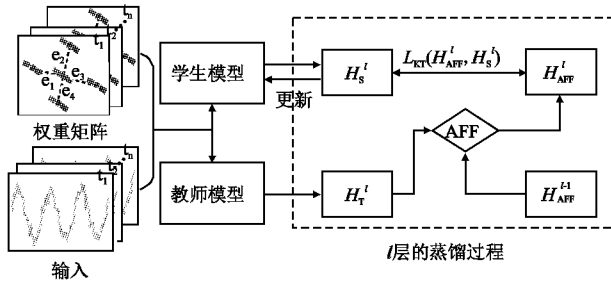


图2 基于中间层特征对齐的递归蒸馏方法

Fig.2 Recursive distillation method based on intermediate layer feature alignment

为避免学生模型在简化过程中丢失关键特征,本研究提出基于中间层特征对齐的递归蒸馏方法,如图2所示,详细说明了教师模型与学生模型在不同层级的特征对齐机制及知识传递方式,确保学生模型能够准确学习教师模型中的时空依赖关系.首先,输入数据包括流量时序特征和卫星网络拓扑信息,分别被教师模型和学生模型处理以生成特征表示.教师模型通过其强大的表征能力生成高维特征 $H_T^l$ ,而学生模型则以轻量化设计生成低维特征 $H_S^l$ ,后者需要通过特征对齐过程来优化.特征对齐由注意力特征融合模块(Attention Feature Fusion, AFF)完成. AFF模块接收教师模型当前层 $l$ 的特征 $H_T^l$ 和上一层对齐后的特征 $H_{AFF}^{l-1}$ ,通过 $L1$ 范数加权求和的方式实现自适应融合,生成对齐后的特征表示 $H_{AFF}^l$ ,用于下一层更新,其公式为:

$$AFF(H_T^l, H_{AFF}^{l-1}) = \frac{\|H_T^l\|}{\|concat(H_T^l, H_{AFF}^{l-1})\|_1} \cdot H_T^l + \frac{\|H_{AFF}^{l-1}\|}{\|concat(H_T^l, H_{AFF}^{l-1})\|_1} \cdot H_{AFF}^{l-1} \quad (11)$$

其中,  $concat(H_T^l, H_{AFF}^{l-1})$  表示特征的拼接操作;  $\|\cdot\|_1$  表示

$L1$  范数,用于计算加权比例.

这种融合机制不仅能够提取每一层的关键特征,还能有效抑制浅层特征中的噪声,从而为学生模型提供更加准确和高效的指导.这种递归对齐机制从深层到浅层逐步传递特征,实现多层级时空特征的优化提取.

蒸馏过程中,特征蒸馏损失 $L_{KT}$ 使用均方误差衡量学生特征 $H_S^l$ 与对齐特征 $H_{AFF}^l$ 的差异,公式如下:

$$L_{KT} = \sum_{l=L}^1 MSE(H_S^l, H_{AFF}^l) \quad (12)$$

其中, $H_S^l$ 是学生模型在第 $l$ 层的输出特征, $H_{AFF}^l$ 是第 $l$ 层经过AFF生成的对齐特征, $L$ 是模型中间层数.

### 2.4 蒸馏损失函数设计

针对卫星网络流量预测任务的特性,构建了多成分融合的蒸馏损失函数,以全面衡量教师模型与学生模型之间的差异,并引导学生模型高效学习教师模型所蕴含的知识.

#### 2.4.1 软损失

软损失的设计核心是通过对比教师模型输出的软标签与学生模型预测概率分布,迁移教师模型的知识.软标签由Softmax函数生成,蕴含丰富的类别间概率信息,与真实标签相比,能够提供更全面的类别关联信息.通过学习这些软标签,学生模型可以弥补参数量不足的问题,提升其性能.软损失采用带有温度参数 $\tau$ 的KL散度(Kullback-Leibler Divergence)进行衡量,公式如下:

$$L_{soft} = \tau^2 \cdot KL(P_T(y|x) \| P_S(y|x)) \quad (13)$$

其中, $P_T(y|x)$ 和 $P_S(y|x)$ 分别为教师模型与学生模型的预测概率分布; $KL$ 表示KL散度,用于量化两个概率分布之间的差异; $\tau$ 是温度参数,用于平滑软标签的分布.

在蒸馏的初期,较大的温度参数 $\tau$ 会使软标签分布更加平滑,让学生模型关注类别间的相对关系,从而高效地提取整体趋势知识;随着训练的推进,逐渐降低 $\tau$ 的值,使学生模型更加专注于精确的类别预测,如区分高峰与低谷流量的类别,最终贴近教师模型的表现.

#### 2.4.2 硬损失

硬损失旨在通过对比学生模型预测值与真实流量值的差异,提升学生模型对真实流量的拟合能力.采用均方误差(Mean Square Error, MSE)作为硬损失的衡量指标,公式为:

$$L_{hard} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

其中, $y_i$ 为真实流量值; $\hat{y}_i$ 为学生模型的预测值; $n$ 为样本数量.

这一损失项确保学生模型在实际流量数值预测上不偏离真实情况,对卫星网络资源分配提供可靠依据.

#### 2.4.3 特征损失

为了进一步强化学生模型对流量特征的学习,除了关注模型输出层的蒸馏损失外,还考虑引入中间层的特征蒸馏损失.如3.2节中所阐述的,特征损失通过比较教师模型与学生模型中间层的特征差异,促使学生模型在特征表征上逐步向教师模型对齐,确保学生模型继承教师模型的全局特征信息.特征损失的计算基于欧氏距离,公式如下:

$$L_{feature} = L_{KT} + \frac{1}{m} \sum_{i=1}^m \|H_{AFF}^l[i] - H_S^l[i]\|_2^2 \quad (15)$$

其中, $H_{AFF}^l = AFF(H_T^l, H_{AFF}^{l-1})$ ,  $H_{AFF}^{l-1}$ 为教师模型在第 $l-1$ 层的

对齐特征,  $H_T^l$  和  $H_S^l$  分别为教师模型与学生模型的第  $l$  层的特征表示;  $m$  为特征元素个数。

#### 2.4.4 时序动态加权 KL 损失

在卫星网络流量预测模型中,流量分布随时间变化显著,特别是在高峰期与低谷期,流量的波动模式存在较大差异。因此,学生模型在蒸馏过程中需要精准对齐教师模型的输出分布,以提升预测精度。利用双向 KL 损失进一步从信息论角度强化学生模型与教师模型输出的相似性,通过同时衡量两者的正向和反向相对熵,避免单向学习带来的偏差。单向 KL 损失具有非对称性,可能导致学生模型过度关注教师模型的高置信度预测区域,而忽略低置信度区域的信息。相较而言,双向 KL 损失同时计算,确保信息在教师与学生之间双向流动,使得学生模型不仅能学习教师模型的高置信度预测,还能适当关注低置信度区域,提高预测的稳定性和泛化能力。为此,本文采用双向 KL (Kullback-Leibler) 散度作为损失函数,以更全面地衡量教师模型  $P_T$  与学生模型  $P_S$  之间的分布差异,并引入时序动态权重  $w_t$ , 以适应不同时间段的流量变化特征,其数学表达式如下:

$$L_{KL} = \sum_{t=1} w(t) \cdot (D_{KL}(P_T(t) \| P_S(t)) + D_{KL}(P_S(t) \| P_T(t))) \quad (16)$$

其中,  $D_{KL}(P_S(t) \| P_T(t)) = P_S(t) \log(P_S(t)/(P_T(t) + \epsilon))$ ,  $\epsilon$  是一个小常数,用于避免分母为零的情况;  $w(t)$  是时间步  $t$  的时序动态权重,用于调节不同时间段的 KL 损失贡献,通过计算流量的波动幅度来确定。

时序权重  $w_t$  的设计基于流量的时段特性:在高峰时段,流量变化剧烈,类别间关联复杂。此时应增大权重,强化学生模型对教师模型分布的学习能力。在低谷时段,流量较为平稳,类别间关联较弱。此时减小权重,避免对教师分布的过度拟合,具体的时序动态权重可以设计为:

$$w(t) = 1 + \alpha \cdot \frac{|F_t - \bar{F}|}{\max(F) - \min(F)} \quad (17)$$

其中,  $F_t$  表示时刻  $t$  的流量,  $\bar{F}$  表示历史平均流量,  $\max(F)$  和  $\min(F)$  分别表示历史最大和最小流量,  $\alpha$  是一个超参数,用于控制权重的波动幅度。该权重设计能够自适应地根据流量变化情况进行调整,在高峰时段强化蒸馏学习,在低谷时段减少蒸馏影响,从而提高预测的准确性和泛化能力。

最终,总蒸馏损失函数为:

$$L_{\text{distill}} = \lambda_1 L_{\text{soft}} + \lambda_2 L_{\text{hard}} + \lambda_3 L_{\text{feature}} + \lambda_4 L_{\text{KL}} \quad (18)$$

其中,  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  是权重参数,用于平衡各部分损失的贡献。通过上述损失函数的优化,学生模型能够充分吸收教师模型的知识,在卫星资源受限的环境下,实现对流量动态变化的高效预测。

### 3 实验

为了验证提出的知识蒸馏驱动的轻量化卫星网络流量预测模型的有效性,本文进行了一系列实验。实验从数据集与预处理、模型训练与超参数设置,以及实验结果对比与分析三方面展开,评估模型的预测性能、计算效率及蒸馏损失设计的效果。

#### 3.1 实验数据集与预处理

本研究使用了真实的卫星网络流量数据集 WetLinks<sup>[27]</sup>,

该数据集包含了 Starlink 的测量数据。数据采集历时 6 个月,从两个位于欧洲的优越位置同步进行。数据集由约 140,000 个测量值组成,涵盖了上传吞吐量、下载吞吐量等关键网络参数。为了确保数据的代表性和可靠性,所有测量值经过严格的筛选与预处理。

卫星网络流量数据呈长尾分布,导致不同时间段的流量波动较大。为平稳波动并符合模型分布假设,数据进行了变换,避免高波动数据对模型的负面影响。同时,为减少特征量级差异对模型训练的影响,所有数据进行了最大最小值归一化,确保各特征在相同量级下处理,提升了训练的稳定性 and 收敛速度。数据按时间顺序划分为训练集(70%)、验证集(15%)和测试集(15%),并按小时粒度生成包含 720 个时间步长的流量序列数据,确保实验结果的稳定性与可复现性。

#### 3.2 模型训练与超参数设置

在实验中,本文采用基于 Transformer 和 GCN 的教师模型和经过知识蒸馏的轻量化学生模型进行对比评估。在模型训练过程中,采用 Adam 优化器进行参数优化,初始学习率设为 0.001,并通过余弦退火策略动态调整学习率以加速收敛。训练采用的批量大小为 32,总训练轮次为 200,同时应用早停策略以避免过拟合。本文比较了教师模型和学生模型的训练损失和性能表现,如图 3 所示。教师模型在训练初期收敛速度较快,并在一定训练轮次后趋于稳定,而学生模型由于模型结构轻量化,训练初期的损失略高于教师模型,但在融合了知识蒸馏的动态损失设计后,逐步逼近教师模型的损失水平。此外,学生模型的收敛过程更加平滑,表明知识蒸馏对模型优化过程起到了显著的指导作用。

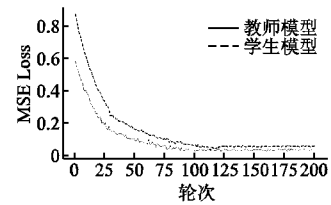


图 3 对比教师模型和学生模型的训练损失  
Fig. 3 Comparison of training loss between teacher model and student model

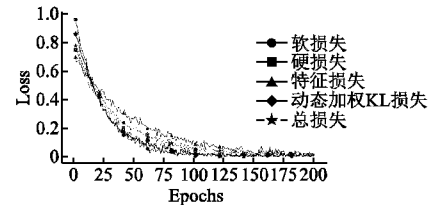


图 4 不同损失项在训练中的变化趋势  
Fig. 4 Variation trend of different loss terms during training

在损失函数方面,采用软损失、硬损失、特征损失以及时序动态加权 KL 损失的组合。本文进行实验对比了不同损失项在训练过程中的变化趋势,进一步验证了损失设计的合理性,如图 4 所示。其中,软损失和硬损失在初期均呈现较高值,随着训练的进行逐步下降,表明学生模型逐渐学习到教师模型的类别分布特征和真实标签信息。特征损失下降速度较慢,整体变化相对平稳,最终达到稳定状态,这表明学生模型在中

间层特征学习上的优化过程较为稳定. 而动态加权 KL 损失的权重根据时段的调整有效适应了卫星网络高峰与低谷流量的变化特点,使模型对复杂的时空动态规律表现出更好的适应性. 综合损失则通过平衡多种损失项,使学生模型在整体上表现出稳定且快速的收敛特性.

### 3.3 实验结果对比与分析

#### 3.3.1 知识蒸馏对学生模型性能的提升

为评估知识蒸馏在提升学生模型预测性能中的作用,实验分别对教师模型、学生模型以及未采用知识蒸馏的学生模型进行了预测性能评估,指标包括均方误差(MSE)、平均绝对误差(Mean Absolute Error, MAE)和决定系数(R-Square, R2). 如图5所示,学生模型在知识蒸馏后性能显著提升,相较于未蒸馏的学生模型,蒸馏后的学生模型 MSE 降低了 26.1%,MAE 降低了 17.4%,且 R2 值更接近于教师模型. 这表明通过知识蒸馏,学生模型能够有效学习教师模型对时序模式和空间依赖关系的建模能力.

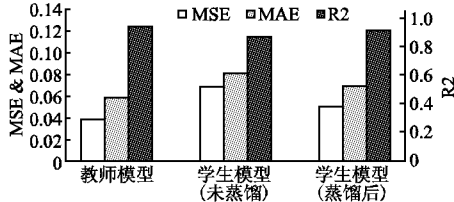


图5 对比知识蒸馏前后模型的性能  
Fig.5 Compare the performance of models before and after knowledge distillation

为验证轻量化学生模型在计算效率上的优势,实验对模型的参数量、浮点运算次数(FLOPs)以及推理时间进行了统计和对比. 结果如表1所示,轻量化学生模型的参数量和FLOPs分别减少了75.6%和78.6%,推理时间加速了约3.6倍,显著优于教师模型. 这一结果表明,轻量化学生模型能够在性能接近教师模型的前提下,大幅降低计算复杂度,适合部署在计算资源有限的卫星网络边缘设备上.

表1 教师模型与学生模型的效率评估

Table 1 Efficiency evaluation of teacher versus student models

模型	参数量(M)	浮点运算次数(G)	推理时间(ms/batch)
教师模型(GCN-Transformer)	12.7	38.4	15.6
学生模型(GCN-GRU-无蒸馏)	3.1	8.2	4.3
学生模型(GCN-GRU-蒸馏后)	3.1	8.2	4.3

尽管学生模型简化了网络结构,但通过知识蒸馏的软损失、硬损失、特征损失以及时序动态加权 KL 损失的设计,使其能够学习到教师模型的深层特征表达和时空动态规律. 这种平衡了预测性能与计算效率的设计,为卫星网络流量预测的实际应用提供了一个切实可行的解决方案,尤其是在对实时性和资源受限有高要求的场景中.

#### 3.3.2 不同预测模型的精度和计算效率对比

为验证所提模型的有效性,本文将蒸馏后的学生模型与传统时间序列预测模型(如 LSTM、GRU、TCN、STGCN、iTransformer)以及最新的轻量化预测模型(如 SparseTSF)进行了对比分析,结果如图6所示. 从预测精度来看,蒸馏后的学生模型在 MAE 和 MSE 指标方面与除 iTransformer 之外的

深度学习模型相当. iTransformer 通过自注意力机制捕捉长期依赖,在部分场景下具备更高精度,但其计算复杂度和参数量

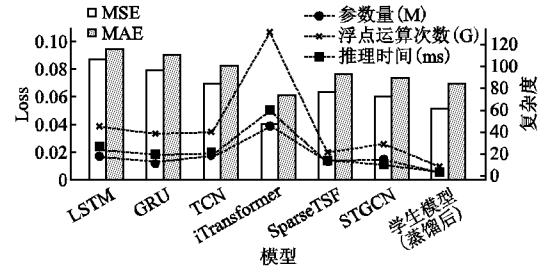


图6 不同模型的性能评估

Fig.6 Performance evaluation of different models

较大,推理时间较长,不适用于计算资源受限的环境. 相比之下,蒸馏后的学生模型在计算复杂度大幅降低的同时,保持较高预测精度. 此外,与最新的轻量化预测模型相比,该模型在卫星流量预测任务中表现出更高的预测精度,这主要源于模型在设计过程中充分考虑卫星流量的时空特性,并结合卫星网络的动态拓扑变化信息,从而提高对卫星流量时空模式的学习能力和泛化性能.

#### 3.3.3 不同蒸馏损失函数设计的影响分析

为了进一步探讨不同损失设计对模型性能的影响,本文进行了对比实验,结果如表2所示. 单独采用软损失时,模型性能较好,但对复杂卫星流量特征的捕捉能力较弱. 增加特征损失后,学生模型在中间层特征学习方面得到了强化,性能有

表2 不同蒸馏损失函数的影响

Table 2 Effects of different distillation loss functions

模型	软损失	特征损失	时序动态加权 KL 损失	MSE	MAE	R2
仅软损失	✓	×	×	0.0264	0.0408	0.873
软损失 + 特征损失	✓	✓	×	0.0235	0.0376	0.895
完整蒸馏设计	✓	✓	✓	0.0212	0.0354	0.912

了明显提升. 特别是引入动态加权 KL 损失后,学生模型能够自适应调整对不同时段流量分布的学习权重,进一步优化了预测效果. 这表明损失函数中的各项设计具有互补作用,能够更好地提高学生模型的整体性能.

#### 3.3.4 时序动态加权 KL 损失效果分析

为了验证时序动态加权 KL 损失对模型的影响,本文分别比较了高峰时段和低谷时段的预测性能.

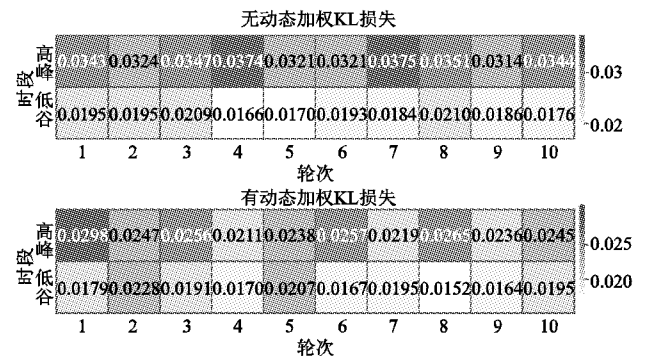


图7 动态加权 KL 损失对模型的影响

Fig.7 Effects of dynamically weighted KL loss on the model

如图7所示,无动态加权 KL 损失的数据显示,高峰时段

的 MSE 较高,表明模型在复杂流量波动下的适应能力有限;低谷时段 MSE 较低,但依然存在一定误差。有动态加权 KL 损失的数据显示,高峰时段 MSE 减少了 22.6%,这表明动态加权机制能够更好地捕捉高峰期复杂的流量变化,提升了模型对流量突发波动的适应能力。低谷时段 MSE 也有所降低,表明动态加权 KL 损失并未引入额外的误差,而是进一步提升了整体预测性能的平衡性。通过自适应调整损失权重,学生模型能够根据不同流量时段的特点进行有效学习,使得模型在复杂时空场景下表现出更好的稳定性和准确性。

#### 4 结束语

本研究聚焦于卫星网络流量预测这一领域,提出了一种知识蒸馏驱动的轻量化预测模型。该模型构建了融合 GCN 与 Transformer 的教师模型,以深入挖掘流量的时空依赖性;同时设计了结合 GCN 与 GRU 的学生模型,通过知识蒸馏将教师模型的知识迁移至学生模型,在显著降低模型复杂度和计算资源消耗的同时,保持了接近教师模型的预测性能。通过中间层特征对齐的递归蒸馏方式,引入时序动态加权 KL 损失来动态调整学习权重,从而有效提升对复杂流量特征的适应性。实验结果验证了知识蒸馏方法在提高预测精度的同时,提供了适用于资源受限卫星网络设备的高效解决方案。未来研究将优化模型结构与训练过程,结合多种模型压缩方式,探索深度模型压缩以降低推理延迟,推动流量预测在资源受限环境中的应用。

#### References:

- [1] Zhu X M, Jiang C X. Integrated satellite-terrestrial networks toward 6G: architectures, applications, and challenges [J]. *IEEE Internet of Things Journal*, 2021, 9(1): 437-461.
- [2] Xia Z G, Ren L W, Zhu W W, et al. Integrated emergency communication system based on low earth orbit satellite and beidou satellite technology [C]//International Conference on Integrated Intelligence and Communication Systems (ICIICS), 2023: 1-7.
- [3] Guo B Q, Li H Y, Zhang Z, et al. Online network slicing for real time applications in large-scale satellite networks [C]//International Conference on Communications (ICC), 2023: 5564-5569.
- [4] Xiao A L, Chen Z M, Wu S, et al. Collaborative long-short term bandwidth allocation for satellite-terrestrial networks [J]. *IEEE Communications Letters*, 2022, 26(5): 1121-1125.
- [5] Kawamoto Y, Takahashi M, Verma S, et al. Traffic prediction-based dynamic resource control strategy in HAPS-mounted MEC-assisted satellite communication systems [J]. *IEEE Internet of Things Journal*, 2023, 11(8): 13824-13836.
- [6] Wang H, Zhang B B. Research on ARIMA model for short-term traffic flow prediction based on time series [C]//8th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2023: 92-95.
- [7] Urkan O D, Cetin M. Short-term traffic prediction based-on support vector regression [C]//International Conference on Data Analytics for Business and Industry (ICDABI), 2022: 228-233.
- [8] Chu Q L, Li G Z, Zhou R J, et al. Traffic flow prediction model based on LSTM with finnish dataset [C]//6th International Conference on Intelligent Computing and Signal Processing (ICSP), 2021: 389-392.
- [9] Cai J Y, Song S T, Zhang H P, et al. Satellite network traffic prediction based on LSTM and GAN [C]//IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), 2023: 175-178.
- [10] Zhu F Y, Liu L X, Lin T. An LSTM-based traffic prediction algorithm with attention mechanism for satellite network [C]//Proceedings of the 3rd International Conference on Artificial Intelligence and Pattern Recognition, 2020: 205-209.
- [11] Liu Z G, Li W J, Feng J X, et al. Research on satellite network traffic prediction based on improved GRU neural network [J]. *Sensors*, 2022, 22(22): 8678.
- [12] Li N, Hu L, Deng Z L, et al. Research on GRU neural network satellite traffic prediction based on transfer learning [J]. *Wireless Personal Communications*, 2021, 118(1): 815-827.
- [13] Yang L, Gu X X, Shi H F. A novel satellite network traffic prediction method based on GCN-GRU [C]//International Conference on Wireless Communications and Signal Processing (WCSP), 2020: 718-723.
- [14] Cao M X, Liu J, Zhi J M, et al. TLS-net: a hybrid time series prediction model combining TCN and LSTM for ship-satellite network traffic [C]//7th International Conference on Transportation Information and Safety (ICTIS), 2023: 1168-1173.
- [15] Zhao Y, Wang N, Chen Q, et al. Satellite coverage traffic volume prediction using a new surrogate model [J]. *Acta Astronautica*, 2022, 193: 357-369, doi: 10.1016/j.actaastro.2022.01.026.
- [16] Jin C, He X, Ding X J. Traffic analysis of LEO satellite internet of things [C]//15th International Wireless Communications & Mobile Computing Conference (IWCMC), 2019: 67-71.
- [17] Na Z Y, Gao Z H, Guo Q. Performance analysis of self-similar traffic in LEO satellite network [C]//International Conference on Machine Learning and Cybernetics, 2007: 2649-2652.
- [18] Judice A, Venusamy K, Livin J. Multilayer LEO satellite constellation coverage analysis and its current research directions [C]//IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2022: 1-5.
- [19] Zhang X, Sun J K, Wang J X, et al. PAOLtransformer: pruning-adaptive optimal lightweight transformer model for aero-engine remaining useful life prediction [J]. *Reliability Engineering & System Safety*, 2023, 240: 109605, doi: 10.1016/j.res.2023.109605.
- [20] Thaker P, Mohan B R. Enhancing deep compression of CNNs: a novel regularization loss and the impact of distance metrics [J]. *IEEE Access*, 2024, 12: 172537-172547, doi: 10.1109/ACCESS.2024.3498901.
- [21] Li H, Zhang J, Li Z, et al. Improvement of min-entropy evaluation based on pruning and quantized deep neural network [J]. *IEEE Transactions on Information Forensics and Security*, 2023, 18: 1410-1420.
- [22] He F, Ding K, Yan D, et al. A novel quantization and model compression approach for hardware accelerators in edge computing [J]. *Computers, Materials & Continua*, 2024, 80(2): 3021-3045.
- [23] Lin S S, Lin W W, Wu W T, et al. SparseTSF: modeling long-term time series forecasting with 1k parameters [C]//Proceedings of the 41st International Conference on Machine Learning, 2024: 30211-30226.
- [24] Tsanakas S, Hameed A, Violos J, et al. A light-weight edge-enabled knowledge distillation technique for next location prediction of multitude transportation means [J]. *Future Generation Computer Systems*, 2024, 154: 45-58, doi: 10.1016/j.future.2023.12.025.
- [25] Zhang J F, Xie C, Cai H M, et al. Knowledge distillation-based spatio-temporal MLP model for real-time traffic flow prediction [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(11): 18122-18135.
- [26] Chen X, Huang Q L, Yang L. Spatial-temporal traffic prediction of space-integrated-ground networks based on continuous knowledge distillation [J]. *IEEE Sensors Journal*, 2025, 25(3): 5547-5563.
- [27] Laniewski D, Lanfer E, Meijerink B, et al. Wetlinks: a large-scale longitudinal starlink dataset with contiguous weather data [C]//8th Network Traffic Measurement and Analysis Conference (TMA), 2024: 1-9.