



研究与开发

基于图卷积神经网络的分解可解释性的 蜂窝网络流量预测模型

张子天, 温之馨, 诸葛斌, 吕智豪, 董黎刚, 蒋献
(浙江工商大学, 浙江 杭州 310018)

摘要: 随着智能互联应用在城市场景中的普及, 城市网络流量的激增带来了新挑战。基站蜂窝网络中的流量预测是资源分配与调度等关键应用的核心, 准确预测蜂窝流量对于高效分配网络资源尤为重要。然而在蜂窝网络流量预测研究中, 复杂的城市蜂窝流量往往有着深层次的时间和空间特征需要挖掘, 为了解决这个问题, 提出一个基于图卷积神经网络的分解可解释性时空图卷积神经网络 (DISTGCN)。该神经网络利用蜂窝网络流量分解以及时空相关性, 提高了流量预测的准确性, 同时分解后的流量特征增强了预测结果的可解释性。在真实的经典数据集上的实验结果表明, DISTGCN的预测性能优于传统深度学习预测模型和图神经网络模型。

关键词: 蜂窝流量预测; 图神经网络; 流量分解; 可解释性; 深度学习

中图分类号: TP301

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025120

A decomposable and interpretable cellular network traffic prediction model based on graph convolutional neural network

ZHANG Zitian, WEN Zhixin, ZHUGE Bin, LYU Zhihao, DONG Ligang, JIANG Xian
Zhejiang Gongshang University, Hangzhou 310018, China

Abstract: With the rapid proliferation of intelligent connected applications in urban environments, the dramatic increase in urban network traffic presents significant challenges. Accurate traffic prediction in cellular base station networks is critical for optimizing resource allocation and scheduling in various applications. However, the complexity of urban cellular traffic necessitates the extraction of deep temporal and spatial features. To address this issue, a novel deep learning framework based on graph convolutional neural networks (GCN) was proposed, named decomposable

收稿日期: 2024-12-13; 修回日期: 2025-01-06

通信作者: 张子天, zitian.zhang@mail.zjgsu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62301488, No.W2421086); 浙江工商大学省属高校基本科研业务费项目 (No.QRK23010)

Foundation Items: The National Natural Science Foundation of China (No.62301488, No.W2421086), Zhejiang Gongshang University Basic Research Operation Cost Project for Provincial Universities (No.QRK23010)



and interpretable spatiotemporal graph convolutional neural network (DISTGCN). The decomposability of cellular network traffic and its spatiotemporal correlations were leveraged by DISTGCN to enhance prediction accuracy. Furthermore, the decomposed traffic features were shown to provide improved interpretability for the prediction results. Extensive experiments conducted on real-world benchmark datasets demonstrate that DISTGCN significantly outperforms conventional deep learning models and state-of-the-art GNN-based models in terms of prediction performance.

Key words: cellular traffic prediction, graph neural network, traffic decomposable, interpretable, deep learning

0 引言

随着近些年来4G网络向5G网络的应用普及,全球5G移动用户数量激增。根据思科年度互联网报告(2018—2023)白皮书^[1],全球移动设备数量将从2018年的88亿增长到2023年的131亿。现在已远远超越了这个数量,这种惊人的增长趋势将在城市万物互联和AI的发展中继续下去,城市蜂窝塔高效的资源分配需求变得尤为重要。精确预测蜂窝塔流量是实现网络智能化和高效化管理的关键手段,它直接关系到网络资源的优化利用和用户体验的提升。蜂窝网络的核心技术涉及用户与基站之间的链路调度、基站之间的频谱分配、移动网络中的传输方式等多个方面。由于城市中的蜂窝网络流量变化迅速,且容易在短时间内出现网络拥堵,需要进行快速的短期资源分配来应对这些挑战。通过对城市蜂窝流量的精确预测,能够提前识别并解决这些问题,从而更高效地进行网络资源的分配和调度。此外,准确的蜂窝流量预测在网络规划、资源调度以及故障诊断等领域也具有重要的应用价值。

然而,由于多种因素的影响,准确预测蜂窝通信量是一项具有挑战性的任务。首先,蜂窝通信具有高度的动态性和显著的时间依赖性。研究表明,不同类型的蜂窝流量随时间的推移会发生显著变化。同时,尽管高峰时段的蜂窝通信具有高度的相似性,但其间隔较长,这表明捕捉长期时间依赖性至关重要。在时间维度上,某一时间段的流量与前一时间段的流量关

联被称为局部时间依赖性,而与更长时间跨度假流量的关系则定义为全局时间相关性。此外,蜂窝流量受跨域信息的分布、社会活动水平的影响,蜂窝流量在空间上呈现出均衡性和复杂的相关性。

近年来,为解决蜂窝流量预测中的挑战,研究者进行了大量探索与实践。统计方法^[2-3]以及传统的机器学习技术^[4-5],首次被应用于该领域,取得了显著的进展。然而,这些方法在早期的预测中有明显缺陷,这些方法在捕捉蜂窝流量中隐藏的复杂关系时表现效率低下。在人工智能的推动下,各种深度学习技术开始应用于各种蜂窝预测当中。像循环神经网络(RNN)^[6]、长短期记忆(LSTM)神经网络^[7-9]这类深度学习网络已被证明在蜂窝流量的预测问题上有着很好的结果。对于预测城市蜂窝流量而言,仅在时间维度捕捉依赖性进行预测忽略了蜂窝流量拥有的空间特征。为了解决捕捉空间特征的问题,近年来,图神经网络(GNN)^[10]作为一种新兴的深度学习方法被应用于空间信息的提取,它能够处理不规则的图结构数据,进而揭示蜂窝流量中的复杂依赖关系,从而提高预测精度。

虽然这类方法提升了预测性能,但仍然有其局限性。加入空间信息之后,模型变得更为复杂,需要学习的特征数量增多,学习时间变长,在模型训练过程中对硬件的需求更为严格。另一个缺陷是目前深度学习模型的可解释性较差。端到端的架构通常使基于深度学习模型表示为一个黑盒,这意味着很难判断预测依赖于什么,以及在给定任务中哪些特征或表示发挥更重要的作

用^[33]。不了解模型的决策规则和预测结果时，往往只能通过预测设定的评价指标去推测，可靠性大大降低，这样就无法为发现预测时出现偏见和错误决策的问题提供帮助。

本文提出了一种 DISTGCN，用于解决上述在蜂窝流量预测中出现的问题。因为发现了蜂窝流量可分解的特点，所以可以将真实的蜂窝流量分解为基波流量和矩阵系数的线性组合。蜂窝流量有着明显的周期性和季节性，通过随机梯度下降去捕获每个基站流量的基波流量，再去对原流量进行拟合得到矩阵系数。此时的基波系数和流量有着强相关性。将矩阵系数作为数据集放入本文的预测模型中，得到的输出结果与开始的基波流量拟合就可以得到预测流量。本文的主要贡献如下。

(1) 对蜂窝网络流量时空分布特性进行分析，发现流量业务是可以分解的，可以将流量表示为基波流量和矩阵系数的线性组合。在这种情况下，基波系数与流量高度相关，并且系数会随流量的变化而变化，通过实验也验证了随着基波数量的增加，流量的拟合精度越来越高，在到达一定数量时趋于稳定。

(2) 提出了一种新的 DISTGCN 模型，该预测模型利用图卷积神经网络 (GCN) 对空间信息进行捕捉将空间特征加入时空图层中，并结合时间域预测模块，可以对未来一段时间内多个蜂窝塔的流量进行预测。

(3) 预测模型有很好的可解释性，提出一个流量分解模块将流量分解得到基波流量和系数矩阵，可以清晰地分析流量变化的趋势，和邻近蜂窝之间流量的相似性。最后通过对基波流量和系数矩阵的拟合得到预测结果，此预测结果有着良好的可解释性。

(4) 在真实的数据集上进行预测，并和多个深度学习模型进行比较后，验证了此模型有着良好的预测性能，并显著提升了预测精度。与现有

的图神经模型训练时间进行对比，此模型在训练时间上消耗的时间也更短。

1 研究现状

流量预测在网络研究中被视为一个重要的问题，本节主要对近些年来在蜂窝流量预测方面的研究进行介绍。这里将详细介绍统计学模型、机器学习模型和深度学习预测模型，包括近些年来一些传统时间序列模型在蜂窝流量预测的应用和最新的图神经网络在蜂窝网络流量预测方面的研究进展。

在统计学模型中，ARIMA 模型^[11]凭借其广泛的适用性和成熟的理论体系，成为蜂窝流量预测中的经典工具。该模型通过结合自回归 (AR)、差分 (I)、移动平均 (MA) 3 个部分，捕捉时间序列数据中的模式，从而帮助推断出数据的变化趋势。ARIMA 模型具备明确的数学理论优势，预测误差的分布相对可控。然而，这种统计学模型对数据的要求较为严格，需要满足诸如正态分布、线性关系等假设条件。因此，在复杂的实际应用场景中，模型的预测性能可能会有限制。

机器学习在很大程度上解决了统计模型的局限性。如今，数据流通常包含多个特征，机器学习模型可以自动从流量中提取有价值的特征，因而在处理高维数据时表现优于传统的统计模型。此外，机器学习对冗余特征具有较强的鲁棒性，能够有效降低噪声特征的影响。然而，浅层机器学习模型由于其结构简单，能够捕获的业务流量特征较为有限，并且对数据不平衡和噪声较为敏感，容易受到异常数据的干扰^[12]。此外，浅层机器学习模型有时依赖先验知识进行特征提取，增加了实际应用中的复杂性。例如，支持向量机 (SVM) 在应用中对核函数的选择至关重要，不同核函数的选择直接影响模型的预测性能，因此近些年来一些研究提出了自动化核函数优化技术



去提升预测性能^[13]。

深度学习在过去数十年的一系列时间序列预测问题上取得了巨大成功,随着对蜂窝流量特征的深入了解,许多先进的神经网络被用来预测蜂窝流量^[14]。在处理时间序列数据问题上,提出了基于RNN的方法来捕获时间依赖性进行预测,RNN中引入了隐藏状态来记忆历史信息。然而,由于会产生反向传播过程中梯度爆炸和梯度消失的问题,这类深度学习模型在处理长期的依赖关系性能方面效率低下^[15]。因此后续提出了LSTM^[16]这种RNN变体,LSTM在单个细胞中引入了3个门的机制,用来更好地控制存储信息,以维护更长的历史信息。门控单元循环网络(GRU)^[17]也是RNN的一种变体,它可以根据需要添加隐藏层,以适应不同的任务需求,也可作为模块加入其他的神经网络中。这些RNN变体适用于具有输入时间序列的时间预测问题,但是不适合加入空间信息的时空预测问题。ConvLSTM^[18]是LSTM的重要扩展,它结合了卷积神经网络(CNN)和LSTM的深度学习模型,ConvLSTM已被用于时空预测问题,它以一系列二维矩阵作为输入,可以处理时空特征的数据,捕获数据的空间相关性。注意力机制^[19]和Transformer网络^[20]也被证明在蜂窝流量预测问题中是有效的。

图神经网络作为一种新兴的深度学习方法,可以呈现更加全面的空间信息。城市空间信息多为非欧几里得的不规则图结构,相比较于网格规则结构,非欧几里得图结构可以展现更多的空间信息。图神经网络(GNN)可以聚合连接节点和边特征信息,通过消息传递去接收邻居节点的消息。文献[21]将蜂窝通信量分解为塔内通信量和塔间通信量,以表示蜂窝塔之间的空间依赖性。由于他们共同考虑了蜂窝塔之间的时间和空间依赖性,故提出的GNN模型实现了城市区域范围内的细粒度流量预测。文献[22]利用基于基站之间地

理距离构建的静态预定义图来实现图卷积网络。文献[23]首先提出了时空图卷积网络(STGCN)来解决交通流域的时间序列问题。他们在图上表达问题,并使用完整的卷积结构构建模型,这使得训练速度更快,参数更少。文献[24]提出了一种类似于变压器结构的基于注意力的空间临时图神经网络(ASTGNN)。文献[25]提出了一种用于蜂窝交通预测的多视图时空图网络(MVSTGN),该网络使用精细的全局注意力机制提取潜在的全局相关性,然后使用密集连接的CNN捕获隐藏在全局编码中的局部时空依赖关系。

综上所述,由于目前的研究要么是简单地利用局部蜂窝流量的时间相关性,要么是混合模型不能充分地利用时空信息,如何准确有效地建模时空相关性仍然是一个挑战。所以,本文提出了一个DISTGCN的深度学习框架,利用流量分解特点捕捉更多的时空特征,以高精度和低复杂度预测蜂窝网络。

2 分析

2.1 数据集

为了评估本文所提模型,在数据集上,本文选择了最广泛使用的意大利米兰^[26]真实的蜂窝流量数据集进行实验。该数据集共收集了米兰市区62天的互联网活动的详细数据,它将整个区域划分成(100×100)的网格,每个网格大小约为0.05 km²。数据集中每10 min采集每个区域网格中不同属性,包括地区ID、短信、呼叫和互联网使用数据等,为时间和空间提供了足够的分析样本,本文对其中的蜂窝流量数据进行预测。

在现实的蜂窝流量的交互过程中,往往伴随着大量的网络流量变化。不同的蜂窝信号塔之间有着自己的覆盖范围,随着城市人群一天中不同地点的动态变化,不同地区在不同时间段出现流量的差异。不同时间段米兰蜂窝信息空间分布如图1所示,对同一天所有地区中午12时和晚

上9时的网络流量活跃程度做了热力图对比,中午,可知城市的蜂窝流量分布较为集中,此时城市中心网络流量密集。然而在晚上9时,城市网络流量的集中程度不会那么高,在城市的各个居住区域会有流量的高点,但流量分布较为分散。这样的流动性往往伴随着流量周期性的特点,基于这样的特点,对流量进行分解,探索更多数据特征抓取的可能性。

2.2 问题定义

本文的蜂窝流量预测问题是一个时空预测问

题。整个地区的蜂窝通信量可以表示为数据点 $X_s = \{X_{s,t} | t=0, 1, 2, \dots, T\} \in R^{T \times U \times V}$ 的时间序列,其中 T 是时间间隔的总数, $X_{s,t}$ 是 t 时刻的蜂窝流量矩阵,矩阵用如下计算式描述。

$$X_{s,t} = \begin{bmatrix} x'_{s(1,1)} & x'_{s(1,2)} & \dots & x'_{s(1,V)} \\ x'_{s(2,1)} & x'_{s(2,2)} & \dots & x'_{s(2,V)} \\ \dots & \dots & \dots & \dots \\ x'_{s(U,1)} & x'_{s(U,2)} & \dots & x'_{s(U,V)} \end{bmatrix} \quad (1)$$

其中, $s_{(u,v)}$ 表示坐标, $x'_{s(u,v)}$ 表示该地点时间间隔 t 内的蜂窝流量在本文中,这里要解决的问题是在给定的历史步长下,基于过去 α 个时间间隔的观测值,去预测未来 $x'_{s(u,v)}^{t+1}$ 的蜂窝数据流量。预测问题可表示为:

$$x'_{s(u,v)}^{t+1} = f(X_S^{t-\alpha+1}, X_S^{t-\alpha+2}, \dots, X_S^t) \quad (2)$$

2.3 网络流量分解

对蜂窝流量的分解是本文的一个重要工作,为了验证时间序列的蜂窝流量可分解性,利用分解模块对原始数据处理,并对分解的基波向量数量和拟合效果进行研究以验证猜想。首先,通过滑动窗口对原始数据进行分段切割。随后,定义一个与窗口长度相匹配的数据结构,并结合同一位置在不同时段的观测值,采用随机梯度下降方法进行迭代优化,最终得到目标基波流量和对应的系数矩阵。将蜂窝流量分解为多个基波向量和系数矩阵的线性组合,模型的时间序列流量 y 可通过基向量 $\{b_1, b_2, \dots, b_j\}$ 和相应的系数矩阵 $\{w_1, w_2, \dots, w_k\}$ 的加权和:

$$y = \sum_{j=1}^k w_j \cdot b_j \quad (3)$$

为了得到好的拟合效果,定义了损失函数为均方误差,最终的优化问题可表示为:

$$\arg \min_{b_j, w_j} \sum_{i=1}^N \left\| y_i - \sum_{j=1}^k w_j \cdot b_j \right\|^2 \quad (4)$$

具体的算法迭代过程将在下一章节的分解模块中详细介绍。首先对得到的基波流量进行特征

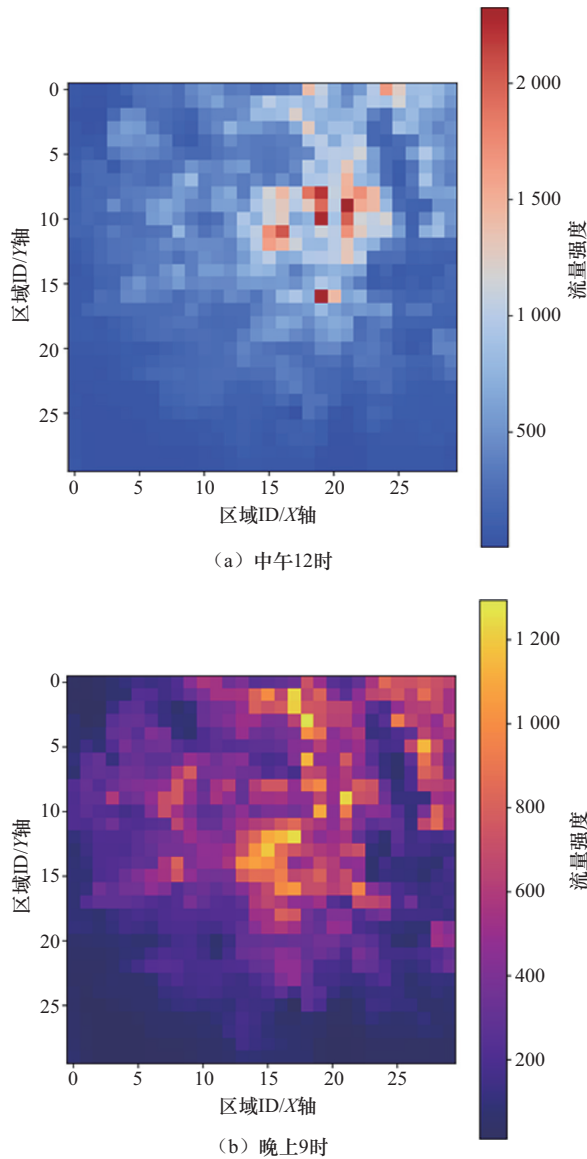


图1 不同时间段米兰蜂窝信息空间分布



学习，并对系数矩阵作为处理之后的输入进行预测，最后进行拟合，得到预测的蜂窝流量。再将划分的时间窗口长度定义为超参数 T 和生成的基波流量数量定义为超参数 N ，以适应不同的数据集设置，并便于根据实验需求灵活调整。这里介绍分解实验中所选用的参数设置。为了匹配后续实验的数据集，这里的基波流量长度 T 设置为 144，等同于该数据集中一天的流量时间点数。生成的基波流量和系数矩阵的数量将根据下列观测实验具体分析。

观测 1 基波流量的数量对最终的拟合精度有显著影响。为确保拟合精度在可接受范围内，尝试了调整基波流量数量 N ，并使用均方误差（MSE）作为评价指标，衡量不同基波流量数量下拟合的流量序列和真实流量序列之间的误差。不同数量基波流量拟合的精确度如图 2 所示，实验证明，随着基波流量的数量增加，拟合的精度会越来越好。在本次的数据集中，基波数量达到 8 的时候拟合精度趋于平稳，已经可以获得很好的拟合效果，在后续的实验，将基波流量数量 N 设置为 8。对基波流量损失函数使用的是窗口实际值和预测值之间的误差。

观测 2 本研究想要找到分解的单一系数矩阵和完整历史蜂窝流量的联系。基波流量对原始流量

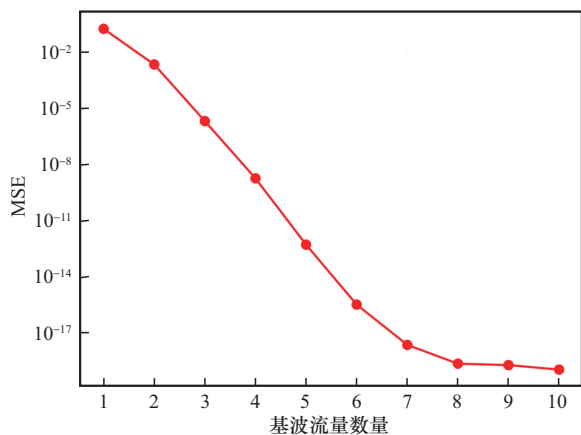


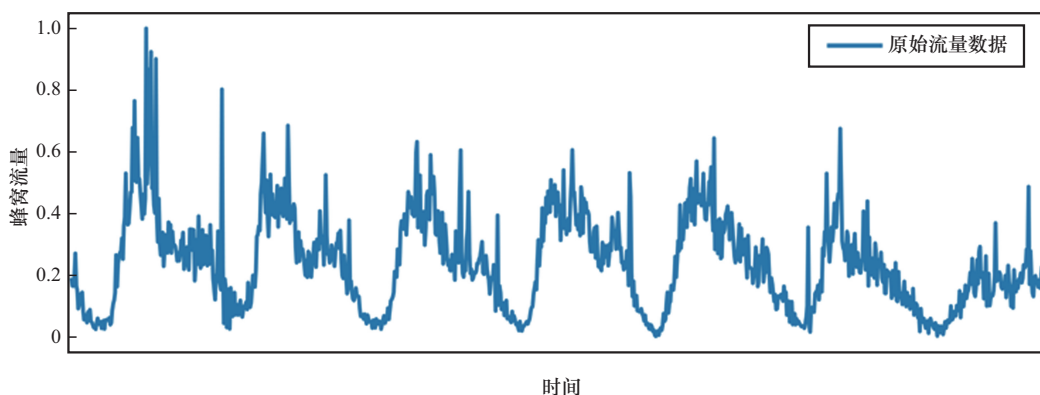
图2 不同数量基波流量拟合的精确度

的拟合效果如图 3 所示，从上到下依次展示了该地区的真实流量、基波拟合流量。通过这些图像可以明显看出，城市蜂窝流量具有显著的日周期性。这种周期性特征反映了该地区蜂窝流量在整个历史序列中的变化情况，有助于更好地捕捉自身流量的相关性。基波向量代表了蜂窝流量的日常周期模式，而真实数据则涵盖了为期 7 天的时间序列。基站自身的基波流量变化趋势包含了蜂窝流量的时间相关性特征，因为基波流量适用于整条流量数据，所以每天的系数矩阵的变化情况则体现了该地区蜂窝流量在一天中的变化过程特点。

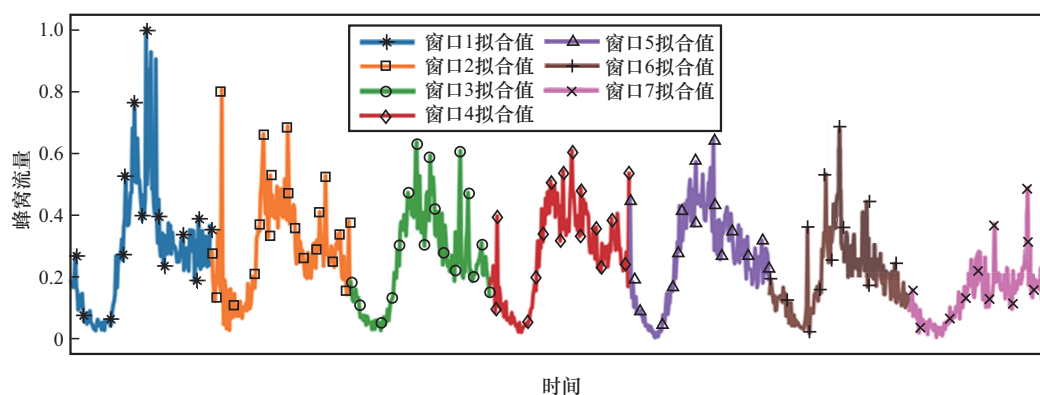
为了评估基波向量与真实数据的匹配程度，将基波系数与每一天的真实数据进行逐日对比，并计算了每一天的皮尔逊相关系数^[27]。皮尔逊相关系数反映了基波系数与实际流量数据之间的线性相关性，其值范围从 -1（完全负相关）到 1（完全正相关），0 表示无相关性。高皮尔逊相关系数表明基波系数能够有效地捕捉到真实数据的周期性特征，进而体现出蜂窝流量自身的稳定性和周期性。皮尔逊相关系数的计算式如下。

$$r = \frac{n(\sum AB) - (\sum A)(\sum B)}{\sqrt{[n \sum A^2 - (\sum A)^2][n \sum B^2 - (\sum B)^2]}} \quad (5)$$

式 (5) 表示 A 、 B 两个观测值之间的相关性，具体比较了天内系数矩阵与真实流量的皮尔逊相关系数，这些结果揭示了基波向量在描述蜂窝流量日常波动方面的有效性和一致性。皮尔逊相关系数的阈值 $|r| > 0.7$ 表示强相关， $0.5 \leq |r| \leq 0.7$ 表示中等相关， $|r| < 0.5$ 表示弱相关。基波流量与每日流量的皮尔逊相关系数如图 4 所示，基波系数与每一天的真实数据之间的相关系数均维持在较高水平，说明在同一个基站下，基波系数变化能够较好地反映真实蜂窝流量的特征变化。



(a) 归一化时间序列流量数据



(b) 滑动窗口拟合数据效果

图3 基波流量对原始流量的拟合效果

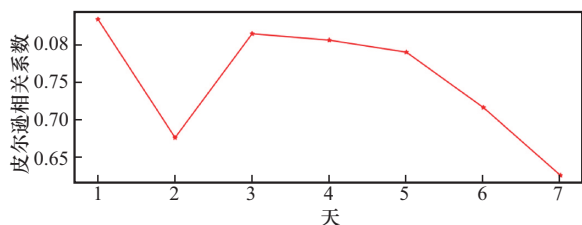


图4 基波流量与每日流量的皮尔逊相关系数

观测3 对不同地区基站之间的流量进行分析。不同的基站区域的基波流量的成分不同，可以将这个特征放到空间相关性中，地区之间地理位置及流量流动方向都包含着空间特征。有些地区的基波流量成分相似或变化趋势相同，这表示这些地区在地理位置上有着强相关性。这对可以用图神经网络去学习和捕捉蜂窝地区之间的流量相关性提供了重要信息。这里仍然用皮尔逊系数去衡量地区之间流量的相关性，随机挑选了10个地区的基波流量，并对它们之间皮尔相关系数

做了热图。不同地区之间皮尔逊相关系数热图如图5所示，可以看出有些不同的地区蜂窝基站之间在空间上有强相关性。

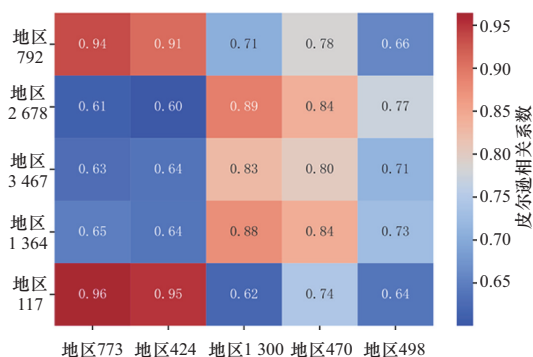


图5 不同地区之间皮尔逊相关系数热图

3 方法

3.1 模型结构

分解可解释的时空图卷积神经网络模型架构



如图6所示, 本文分解可解释的时空图卷积神经网络模型由3个重要的子模块组成, 分别是生成基波流量和基波系数的分解模块、生成自适应瞬时图的空间模块、时序建模的核心循环单元组成的时间模块。在空间模块中主要融入了分解模块特征, 并利用记忆模块来捕捉节点嵌入间的全局空间关系, 以捕捉更多的空间依赖性。在时间模块中, 利用注意力机制, 并结合上游的空间依赖信息与有门控机制的循环单元层。最后由分解模块的输出和时间模块的输出拟合, 实现对时序序列高效预测。本节将详细介绍模块内容。

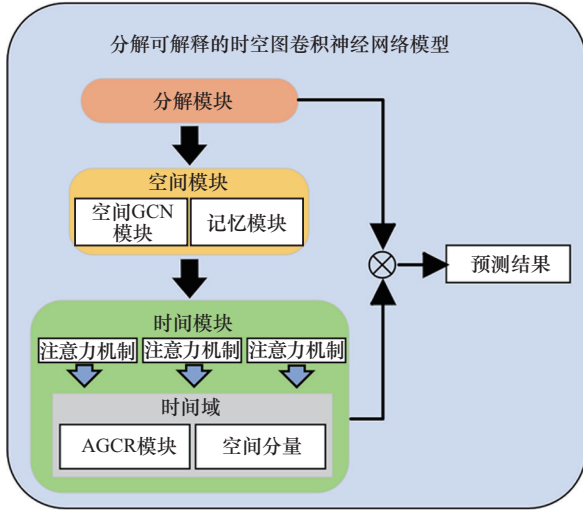


图6 分解可解释的时空图卷积神经网络模型架构

整个模型采用编码器—解码器架构, 其中编码器部分接收一段历史序列, 通过多层循环单元对序列进行编码, 输出隐藏状态序列。解码器则在未来时刻逐步解码, 利用编码器生成的最后状态作为初始输入, 并结合目标序列中的协变量进行逐步预测。

与传统的深度学习预测方法不同, DISTGCN首先从蜂窝流量的时间序列中和处理得到的基波流量特征中, 提取目标节点的空间网络拓扑结构, 并将空间信息存入节点组中, 再将节点嵌入的节点相关连接信息和时间域基波流量系数矩阵 w

传入递归神经网络中学习, 这里得到输出预测的基波流量系数 w' , 最后再将最后的输出与最开始得到的基波流量 b 进行拟合, 拟合的结果就是最后的预测流量。

3.2 分解模块

为进一步实验对数据时空特征的提取, 在第2.3节所提出的网络流量分解基础上, 引入了分解模块, 用于提取区域的基波向量和流量系数。该模块通过对时间流量序列进行滑动窗口分割, 并基于随机梯度下降 (SGD) 算法迭代更新多个基向量, 从而有效提取能够代表区域流量特征的基向量和基波系数。这种分解方式可以有效捕捉时间序列数据中的多尺度特征, 能够更精准地提取数据的局部特征和长期趋势。此外, 它能够有效去除复杂数据中的噪声成分, 保留主要信号, 同时保证较低的计算复杂度和较高的效率。具体的实现流程如下: 为了方便对数据的处理, 首先对原始数据进行归一化处理, 接着定义时间窗口长度为 T 的子序列, 代表特定时间段内的数据片段。通过滑动窗口, 得到多个时间窗口, 便于模型学习不同时间片段特征。这一过程与第2.3节中提出的时间窗口的划分方法一致, 确保了整体算法框架的一致性和连贯性。

在每个窗口中, 利用多个基波向量和基波系数来表征时间序列的潜在周期模式。基波向量的初始化为随机值, 并在后续迭代中通过随机梯度下降算法进行更新。基波向量的更新策略基于误差最小化原则, 具体规则如下。

对于第 k 次迭代中, 第一个基向量的更新计算式为:

$$b_1^{(k+1)} = b_1^{(k)} + \eta(y_i - b_1^{(k)}) \quad (6)$$

其中, η 为学习率, 用于控制基波向量的更新幅度。通过迭代, 基波向量逐渐逼近当前窗口的目标数据 y_i 。基波向量更新后, 计算残差 $r = y_i - b_1$, 以此继续更新其他基波向量, 一直更新到最后一个

向量 (第 N 个)。

$$\mathbf{b}_j^{(k+1)} = \mathbf{b}_j^{(k)} + \eta(\mathbf{r}_{j-1} - \mathbf{b}_j^{(k)}), j=2, 3, \dots, N \quad (7)$$

在基波向量结束更新后, 使用一组权重矩阵 $\mathbf{w}_j$ 进行组合, 以生成最终的拟合值。每一个基波向量 $\mathbf{b}_j$ 对应一个权重 $\mathbf{w}_j$, 更新同样基于随机梯度下降, 最小化与真实值之间的误差。最终, 通过模型对基波系数进行预测, 并通过矩阵加权组合, 得到第 i 个窗口的预测值:

$$\mathbf{w}_j^{(k+1)} = \mathbf{w}_j^{(k)} + \eta(\mathbf{y}_i - \hat{\mathbf{y}}_i) \mathbf{b}_j^T \quad (8)$$

$$\hat{\mathbf{y}}_i = \sum_{j=1}^M \mathbf{w}_j' \cdot \mathbf{b}_j \quad (9)$$

3.3 时空图学习图层

(1) 图卷积单元

本文受图神经网络的发展影响, 特别是最近几年比较火热的 GCN, 可以利用图结构产生空间数据, 如蜂窝塔之间的连接, 将 GCN 捕捉到的数据作为空间信息输入。图神经网络卷积操作被加入门控的循环单元, 由此产生了图卷积循环单元 (GCRU), 该单元可以同时捕捉由输入图拓扑表示的空间依赖性和顺序表示的时间依赖性, 不仅表达了蜂窝塔之间的物理连接, 而且捕捉到在城市网络中蜂窝流量在蜂窝塔之间的流动特征。采用图卷积运算下使用门控循环单元作为空间建模的基本单元。

$$\mathbf{H} = \sigma(\mathbf{X} \cdot \zeta \Theta) = \sigma\left(\sum_{k=0}^K \tilde{\mathbf{P}}^k \mathbf{X} \mathbf{W}_k\right) \quad (10)$$

$$\begin{cases} u_t = \text{sigmoid}([X_t, H_{t-1}] \cdot \zeta \Theta_u + b_u) \\ r_t = \text{sigmoid}([X_t, H_{t-1}] \cdot \zeta \Theta_r + b_r) \\ C_t = \tanh([X_t, (r_t \odot H_{t-1})] \cdot \zeta \Theta_c + b_c) \\ H_t = u_t \odot H_{t-1} + (1 - u_t) \odot C_t \end{cases} \quad (11)$$

其中, $\mathbf{X} \in \mathbf{R}^{N \times C}$ 和 $\mathbf{H} \in \mathbf{R}^{N \times h}$ 分别表示图卷积处理的输入和输出, Θ 和 $\mathbf{W}_k \in \mathbf{R}^{K \times C \times h}$ 是 K 阶切比雪夫多项式的核参数, σ 是激活函数。 u 、 r 、 C 分别表示循环门控单元的更新门、复位门和候选状态, 用来控制想记住过去状态的数量和控制状态

中有多少个旧状态的副本。此外 GCRU 还需要添加邻接矩阵 $\mathbf{P} \in \mathbf{R}^{N \times N}$ 对图 ζ 拓扑的辅助性输入。图卷积的操作可以表示为:

$$\mathbf{X}' = \sum_{k=0}^{K-1} T_k(\tilde{\mathbf{L}}) \mathbf{X} \mathbf{W}_k \quad (12)$$

$\mathbf{X}$ 表示输入矩阵, $T_k(\tilde{\mathbf{L}})$ 是第 k 阶切比雪夫多项式基, 应用于 $\tilde{\mathbf{L}}$, $\mathbf{W}_k$ 是切比雪夫多项式对应的权重矩阵, 形状为 $(d_{\text{in}}, d_{\text{out}})$ 。切比雪夫多项式在卷积网络中能够高效地聚合不同尺度的节点特征, 第 k 阶切比雪夫多项式基 T_k 能够捕捉到节点到 k 阶邻居的特征。使用切比雪夫多项式避免了计算拉普拉斯矩阵的幂次, 大大降低了图结构的计算复杂度。

(2) 图结构学习

要使用图卷积神经网络对时间序列进行分析, 则必须要提供一个图结构。在实践中, 有启发式和从数据中学习两种方式被用来从数据中去生成缺少的图结构。在上述定义中, 矩阵 $\mathbf{P}$ 通常是通过一些规定好的度量方式去获得的, 如空间邻近性、成对连通性和成对相似度等。这里将采用两个节点之间的相似度来产生自适应图进行时空图学习, 用于定义节点之间的连接关系。这样可以将图结构学习集成到时空建模的过程中进行同步优化, 这里的相似度方法定义为:

$$\tilde{\mathbf{P}} = \text{softmax}(\text{relu}(\mathbf{E} \mathbf{E}^T)) \quad (13)$$

矩阵 $\tilde{\mathbf{P}}$ 是通过随时间变化的通过随机将嵌入的 $\mathbf{E} \in \mathbf{R}^{N \times e}$ 的可训练节点矩阵积的非负部分归一化进行转置得到, 也可以用输入信号 $\mathbf{X}_t$ 或者隐藏状态 $\mathbf{H}_t$ 以相似的方式进行定义。

$$\tilde{\mathbf{P}}_t = \text{softmax}(\text{relu}((\mathbf{H}_t * \mathbf{W})^T)) \quad (14)$$

其中, 参数矩阵 $\mathbf{W}$ 将隐藏状态 $\mathbf{H}_t$ 映射到另一个嵌入空间, 瞬时图的产生都类似于式 (14)。通过相似度计算得到的瞬时图能够使模型更加灵



活, 配合基波流量的特征能够高效捕捉蜂窝流量之间的关系, 将节点上更多下文信息考虑进来, 从而提高模型的预测性能和泛化能力。

这里将介绍本文设计的全新的时空学习模块。利用流量分解模块获得分解特征, 以帮助构建空间图结构。这个模块的启发来自记忆网络的技术的研究^[28]。记忆网络是记忆所见样本中不同时间的典型特征的异构信息, 以进一步的模式匹配。在图卷积模块中图表示图结构学习的节点嵌入生成, 其中加入了基波流量成分在图结构学习过程中的偏重影响。所以, 图结构的生成的瞬时图决定了最终学习任务中的空间信息。为获得更多空间信息, 将无结构学习的结果保存在建立的节点信息组 $\Phi \in \mathbf{R}^{\phi \times d}$ 中, 组中的 ϕ 和 d 分别代表了存储点的个数和每个项的维度。这个节点组的主要功能如下。

$$\mathbf{Q}_t^{(i)} = \mathbf{H}_t^i * \mathbf{W}_Q + \mathbf{b}_Q \quad (15)$$

$$\begin{cases} a_j^{(i)} = \frac{\exp(\mathbf{Q}_t^{(i)} * \Phi^T[j])}{\sum_{j=1}^{\phi} \exp(\mathbf{Q}_t^{(i)} * \Phi^T[j])} \\ \mathbf{M}_t^{(i)} = \sum_{j=1}^{\phi} a_j^{(i)} * \Phi[j] \end{cases} \quad (16)$$

其中, (i) 表示第 i 个节点的向量式 (15) 是一个 FC 的全连接层, 将隐藏状态 $\mathbf{H}_t^{(i)}$ 映射到局部查询中。

式 (16) 表示通过 $\mathbf{Q}_t^{(i)}$ 进行内存读取操作与每个 $\Phi[j]$ 匹配去计算得到 a_j , a_j 在表示 $\mathbf{Q}_t^{(i)}$ 与 $\Phi[j]$ 的相似度。节点向量 $\mathbf{M}_t \in \mathbf{R}^{N \times d}$ 表示进一步的内存向组合。从节点组中学习每个节点的空间连接信息, 根据蜂窝流量变化情况自适应地更新图结构。

3.4 时间模块

为了有效处理上游得到的蜂窝流量的特征, 本文设计了一个带有注意力机制的循环神经网络结构用于时间序列数据的预测任务。该模块以节点为中心, 结合注意力机制增强时序特征的建模能力, 利用 GCN 的空间特征学习能力与注意力机制的动态加权能力。在每一个时间步, 输入与前一时刻的隐藏状态拼接为输入矩阵。

$$\mathbf{X}^{(t)} = [\mathbf{X}^{(t)}, \mathbf{h}^{(t-1)}] \quad (17)$$

拼接之后的矩阵同时包含了当前时刻输入特征和过去的状态信息, 作为后续的 GCN 操作输入。为了捕获节点间的复杂空间依赖性, 模型采用了时间学习的门控机制, 模型通过图卷积计算门控值与重置值。

$$[z^{(t)}, r^{(t)}] = \sigma(\text{GCN}_{\text{gate}}(\mathbf{X}^{(t)}, \mathbf{A})) \quad (18)$$

其中, σ 为 Sigmoid 函数, GCN 为图卷积运算, $\mathbf{A}$ 为节点之间的邻接矩阵, 用于表达空间之间

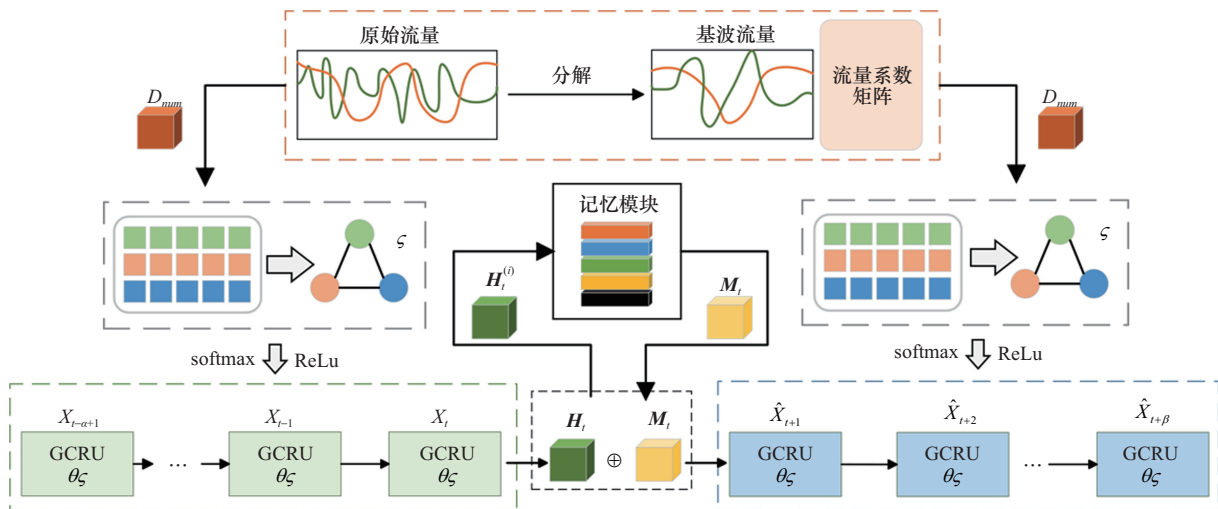


图7 时空图学习图层

的依赖性。通过此方法去更新隐藏状态的选择和重置。最终，利用门控机制计算新的隐藏状态。

$$h^{(t)} = r^{(t)} \cdot h^{(t-1)} + (1 - r^{(t)}) \hat{h}^{(t)} \quad (19)$$

4 实验结果

4.1 实验设置

在本文的实验中，选择了数据集中常用的米兰中心区域（30×30 网格）中一个月的蜂窝流量数据。首先本文采用归一化的方法，将流量数据缩放至[0,1]，然后将数据集的训练集、验证集、测试集划分，比例为2:0:1。为了便于实验结果比较，数据集的数据分割比、邻接矩阵、历史和预测窗口均与文献[34]一致。模型的编码器和解码器由1个神经网络层组成，每个时间预测模块有32个单元，在内存中保存了20个元节点，每一个都是64维的可学习向量，分解的基波流量数量为8。为了观测2 h的预测情况，时间步长设置为6。使用Adam优化器训练网络，其中优化器中的自适应学习率 α 的初始值为0.01，分别在50%和75%的训练周期后降低到0.001和0.000 1，批量大小设置为64。实验使用了一张 GeForce RTX 3090 GUPs 显卡，框架为Pytorch 1.9.1。

本文将DISTGCN与典型蜂窝流量预测方法GRU^[29]、LSTM^[8]以及通用的最新的时空序列预测模型STGCN^[30]、AGCRN^[31]、AHSTGNN^[32]进行了预测性能比较。

4.2 评价指标

本次采用现在研究所认可的蜂窝流量预测指标，均方根误差（RMSE）、平均绝对误差（MAE）和平均绝对百分比误差（MAPE），来评价该模型。

评价指标中的 N 表示样本数量， $\hat{y}$ 是预测值， y 表示真实目标值。3种评价指标都是值越小，模

型的预测结果就更加精确。

- MAE表示预测值与实际值的绝对差值的平均值

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (20)$$

- RMSE表示预测值与实际值的差值

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (21)$$

- MAPE是绝对百分比错误的平均值

$$\text{MAPE} = \frac{100\%}{N} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (22)$$

4.3 实验结果

DISTGCN和基线模型在米兰数据集上的性能比较见表1，展示了几种深度学习模型在米兰数据集预测的总体性能，该模型DISTGCN在此数据集上性能最好，和传统神经网络GRU和LSTM对比，性能在RMSE的评价指标上提升了30.37%和25.98%，这是由于传统神经网络中只对时序部分进行预测，忽略了空间特征对蜂窝流量预测的影响。在和经典的图神经网络STGCN对比中，RMSE的性能提升了10.17%，这是因为STGCN在利用GCN的方法对时空依赖关系建模时，缺少异质性。对比AGCRN和AHSTGNN两种模型，RMSE的性能分别提升了5.45%和1.92%，这两个模型对比，该模型泛化能力较弱，难以达到最优的结果。

表1 DISTGCN和基线模型在米兰数据集上的性能比较

模型	RMSE	MAE	MAPE
GRU	80.53	45.28	18.76%
LSTM	75.77	43.08	18.53%
STGCN	62.43	34.65	18.21%
AGCRN	59.31	30.27	17.29%
AHSTGNN	57.18	27.96	16.92%
DISTGCN（本文所提方案）	56.08	24.08	16.88%

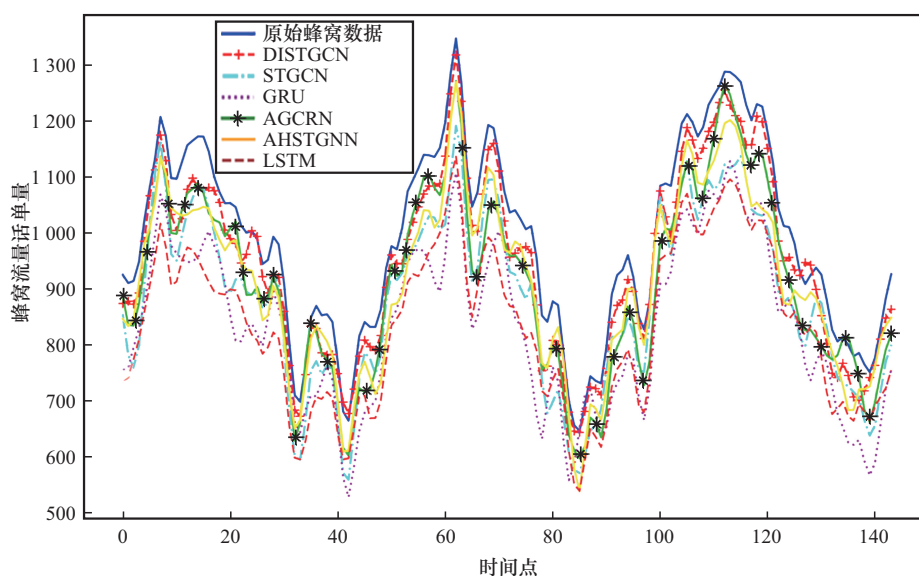


图8 不同模型预测流量与真实流量的对比

不同模型预测流量与真实流量的对比如图8所示,展现了不同模型预测值和真实值的对比情况。通过对5种模型的预测结果比较分析,验证了DISTGCN在蜂窝流量预测中的可靠性。GCN对空间依赖性的捕捉弥补了对以往时间序列模型对空间特征的忽略,通过将蜂窝流量分解特征加入学习循环网络中,有助于探索时空信息下更多的有效信息,从而提高预测性能。

不同模型在多个时间步下预测的RMSE值变化如图9所示,不同模型在多个时间步下预测的MAE值变化如图10所示,展示了在1 h内多个时间步的预测性能。结果表明,从多个时间步的预测结果来看,图神经网络模型在总体的预测性能上要优于传统的时间序列模型。这一现象表明,空间信息的引入在预测性能上带来了显著的提升。此外,还观察到,DISTGCN在短期和长期预测中均能实现最佳性能,即便在1 h的时间步下,预测误差也能保持相对稳定。这进一步验证了DISTGCN在处理复杂时空依赖性蜂窝流量数据方面的有效性。

对DISTGCN、STGCN、AGCRN和AHSTGNN这些运用图神经模型的计算复杂度进行了比较,

米兰数据集上的计算成本见表2。本次实验都是在了一块GeForce RTX 3090 GUP显卡的服务器上进行的。在确保最优预测性能下,本文提出的模型在训练时间相较于采用全卷积结构的STGCN略有延长。该差异的根本原因在于,STGCN作为一种基础的图神经网络,模型复杂度较低,因此训练时间较短,但是其预测精度与近些年更复杂的图神经网络相比存在一定差距。与其他的图神经网络相比,该模型通过流量分解减少了所需学习参数数量,有效降低了训练时间。

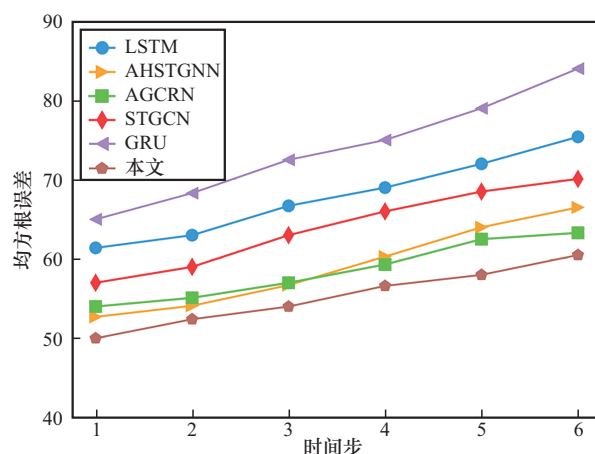


图9 不同模型在多个时间步下预测的RMSE值变化

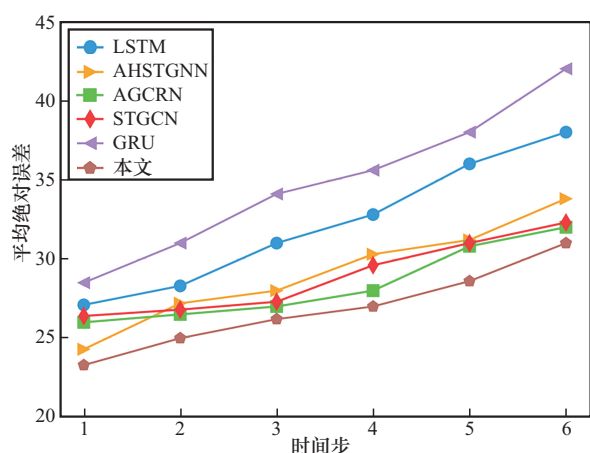


图 10 不同模型在多个时间步下预测的MAE值变化

表 2 米兰数据集上的计算成本

模型	参数	训练时间/(s·epoch ⁻¹)
STGCN	127 420	6.5
AGCRN	754 350	13.1
AHSTGNN	255 862	13.2
DISTGCN (本文所提方案)	240 257	12.9

5 结束语

本文研究对米兰市真实的蜂窝流量进行了时空分析,分析了基波流量对整体流量预测的实质性帮助。基于最新的图神经网络,提出了DISTGCN深度学习网络架构,该模型利用流量的分解性捕捉时间序列的空间与时间信息,提高了预测的准确性和可靠性。并且,该模型具有良好的可解释性,以基波流量和系数矩阵的特征来分析和调整训练过程中的预测问题。在对米兰数据集的预测性能比较中,本文提出的DISTGCN在预测结果上都优于传统的时间序列模型和主流的图神经网络模型。尽管DISTGCN在流量预测中展现了优异的性能,但仍发现模型在应对不同数据集的流量分解选择时存在一定局限性。对于不同的数据集,超参数需要重新进行选择和调整。未来的研究应重点探索不同数据集下的最佳流量分解策略,以进一步提高模型的普适性和适应性。

参考文献:

- [1] Cisco. Cisco annual internet report (2018 - 2023) white paper[R]. 2020.
- [2] SRINIVASAN R. IEEE 802.16m evaluation methodology document (EMD)[J]. IEEE 802.16m-08/004r3, 2008.
- [3] XU F L, LIN Y Y, HUANG J X, et al. Big data driven mobile traffic understanding and forecasting: a time series approach[J]. IEEE Transactions on Services Computing, 2016, 9(5): 796-805.
- [4] RIZWAN A, ARSHAD K, FIORANELLI F, et al. Mobile Internet activity estimation and analysis at high granularity: SVR model approach[C]//Proceedings of the 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2018: 1-7.
- [5] 程定国, 曾浩洋. 无线通信网络中流量分析技术综述[J]. 电讯技术, 2023, 63(3): 441-447.
CHENG D G, ZENG H Y. Overview of network traffic analysis techniques in wireless communication networks[J]. Telecommunication Engineering, 2023, 63(3): 441-447.
- [6] QIU C, ZHANG Y Y, FENG Z Y, et al. Spatio-temporal wireless traffic prediction with recurrent neural network[J]. IEEE Wireless Communications Letters, 2018, 7(4): 554-557.
- [7] KHAN A, FOU DA M M, DO D T, et al. Short-term traffic prediction using deep learning long short-term memory: taxonomy, applications, challenges, and future trends[J]. IEEE Access, 2023, 11: 94371-94391.
- [8] TRINH H D, GIUPPONI L, DINI P. Mobile traffic prediction from raw data using LSTM networks[C]//Proceedings of the 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2018: 1827-1832.
- [9] 高勇, 陆钱春, 李锋. 面向IP网络扩容应用的复杂网络流量预测方法[J]. 电信科学, 2023, 39(9): 21-31.
GAO Y, LU Q C, LI F. A complex network traffic prediction method for IP network expansion applications[J]. Telecommunications Science, 2023, 39(9): 21-31.
- [10] BHATTI U A, TANG H, WU G L, et al. Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence[J]. International Journal of Intelligent Systems, 2023(1): 8342104.
- [11] KOCHETKOVA I, KUSHCHAZLI A, BURTSEVA S, et al. Short-term mobile network traffic forecasting using seasonal ARIMA and holt-winters models[J]. Future Internet, 2023, 15(9):



- 290.
- [12] DABLAIN D, KRAWCZYK B, CHAWLA N V. DeepSMOTE: fusing deep learning and SMOTE for imbalanced data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(9): 6390-6404.
- [13] ROMAN I, SANTANA R, MENDIBURU A, et al. In-depth analysis of SVM kernel learning and its components[J]. Neural Computing and Applications, 2021, 33(12): 6575-6594.
- [14] HUANG C W, CHIANG C T, LI Q H. A study of deep learning networks on mobile traffic forecasting[C]//Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2017: 1-6.
- [15] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [16] DALGKITSIS A, LOUTA M, KARETSOS G T. Traffic forecasting in cellular networks using the LSTM RNN[C]//Proceedings of the 22nd Pan-Hellenic Conference on Informatics. New York: ACM Press, 2018: 28-33.
- [17] SHIRI F M, PERUMAL T, MUSTAPHA N, et al. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU[EB].2023: 2305.17473.
- [18] ZHANG C Y, PATRAS P. Long-term mobile traffic forecasting using deep spatio-temporal neural networks[C]//Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2018: 231-240.
- [19] 皇甫晓瑛, 钱惠敏, 黄敏. 结合注意力机制的深度神经网络综述[J]. 计算机与现代化, 2023(2): 40-49, 57.
- HUANGFU X Y, QIAN H M, HUANG M. A review of deep neural networks combined with attention mechanism[J]. Computer and Modernization, 2023(2): 40-49, 57.
- [20] ZHANG Z W, YAN L, GU Y T. ST2T: a spatio-temporal transformer for cellular traffic prediction in digital twin systems[C]//Proceedings of the 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT). Piscataway: IEEE Press, 2023: 1112-1117.
- [21] WANG X, ZHOU Z M, XIAO F, et al. Spatio-temporal analysis and prediction of cellular traffic in metropolis[J]. IEEE Transactions on Mobile Computing, 2019, 18(9): 2190-2202.
- [22] ZHANG H Q, LU G Q, ZHAN M M, et al. Semi-supervised classification of graph convolutional networks with Laplacian rank constraints[J]. Neural Processing Letters, 2022, 54(4): 2645-2656.
- [23] YU B, YIN H T, ZHU Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[EB]. 2017.
- [24] GUO S N, LIN Y F, WAN H Y, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415-5428.
- [25] YAO Y, GU B, SU Z, et al. MVSTGN: a multi-view spatial-temporal graph network for cellular traffic prediction[J]. IEEE Transactions on Mobile Computing, 2023, 22(5): 2837-2849.
- [26] BARLACCHI G, DE NADAI M, LARCHER R, et al. A multi-source dataset of urban life in the city of Milan and the Province of Trentino[J]. Scientific Data, 2015, 2: 150055.
- [27] SACCENTI E, HENDRIKS M H W B, SMILDE A K. Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models[J]. Scientific Reports, 2020, 10: 438.
- [28] PARK H, NOH J, HAM B. Learning memory-guided normality for anomaly detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 14372-14381.
- [29] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB]. 2014.
- [30] ZHAO N, YE Z Y, PEI Y Y, et al. Spatial-temporal attention-convolution network for citywide cellular traffic prediction[J]. IEEE Communications Letters, 2020, 24(11): 2532-2536.
- [31] BAI L, YAO L N, LI C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804-17815.
- [32] WANG X, YANG K X, WANG Z D, et al. Adaptive hybrid spatial-temporal graph neural network for cellular traffic prediction[C]//Proceedings of the ICC 2023 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2023: 4026-4032.
- [33] GAO L, GUAN L. Interpretability of machine learning: recent advances and future prospects[J]. IEEE MultiMedia, 2023, 30(4): 105-118.
- [34] WANG X, ZHAO J, ZHU L, et al. Adaptive multi-receptive field spatial-temporal graph convolutional network for traffic forecasting[C]//Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2021: 1-7.

[作者简介]



张子天 (1988-), 男, 博士, 浙江工商大学信息与电子工程学院副研究员, 主要研究方向为基于机器学习的网络流量预测与资源管理。



吕智豪 (1999-), 男, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为通信与网络。



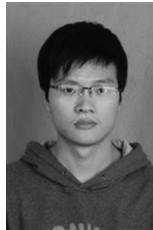
温之馨 (1998-), 男, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为时间序列网络流量预测。



董黎刚 (1972-), 男, 博士, 浙江工商大学信息与电子工程学院教授, 主要研究方向为智能网络、在线教育。



诸葛斌 (1976-), 男, 博士, 浙江工商大学信息与电子工程学院教授, 主要研究方向为网络和通信技术、互联网技术和网络安全。



蒋献 (1988-), 男, 浙江工商大学信息与电子工程学院讲师、实验员, 主要研究方向为在线教育。