

中国科学技术大学
专业硕士学位论文



基于图卷积神经网络的
交通流预测方法研究

作者姓名： 丁子杰
专业类别： 电子信息
专业领域： 计算机技术
校内导师： 赵云波教授
实践导师： 尹航
完成时间： 二〇二四年五月二十四日

University of Science and Technology of China
A dissertation for master's degree



Research on Traffic Flow Prediction Method Based on Graph Convolutional Neural Networks

Author: Zijie Ding

Speciality: Electronic Information

Major: Computer Technology

Supervisor: Prof. Yunbo Zhao

Instructor: hang Yin

Finished time: May 24, 2024

中国科学技术大学学位论文原创性声明

本人声明所提交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：_____

签字日期：_____

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

控阅的学位论文在解密后也遵守此规定。

公开 控阅（____年）

作者签名：_____

导师签名：_____

签字日期：_____

签字日期：_____

摘 要

近年来,我国城市现代化进程迅速推进,城市公共设施和居民生活水平显著提升。然而,随着人口急速增长和交通规模不断扩大,交通拥堵成为城市居民生活的主要困扰。《2023年第3季度中国城市交通报告》报告指出,我国主要城市通勤出行总体呈拥挤趋势。城市路网的承载力未能与城市发展速度和车辆增长率同步提升,导致了許多城市在高峰时段经常出现严重的交通拥堵。在这种背景下,城市交通流预测的发展显得尤为重要。交通流预测能为出行者提供拥堵预警,同时也为交通管理部门实施交通管制提供了一个重要工具。

然而,一方面由于交通数据集时空性较复杂,常用的模型没有充分地挖掘出路网交通节点和邻接矩阵中的远程时空特征和隐藏动态空间特性,使得长期预测准确度欠佳,存在较大的预测误差。另一方面,先前方法在做出行需求预测时存在将起始地和目的地视为相同的语义实体和起始地-目的地(OD)数据复杂的空间依赖性问题,此外,还有需求分布的不均衡导致了OD数据的稀疏性问题。这一系列因素为基于图卷积神经网络的交通流预测方法研究带来了挑战。

针对上述挑战,本文开展了基于图卷积神经网络的交通流预测方法研究,主要研究内容包括:

1. 针对常用的方法没有充分地挖掘出路网交通节点和邻接矩阵中的远程时空特征和隐藏动态空间特性问题,提出了一种基于时空融合动态图卷积方法,实现了较为精准的长期交通流量的预测。本文将概率稀疏自注意力机制与时空卷积块结合在一起,使得模型具有建模长程时空相关性能的同时保持更高的计算效率。采用了动态图卷积网络,充分利用先验知识生成动态图来捕捉交通流的隐藏空间特征。通过在两个真实数据集上的对比实验,本文提出的长期交通流量预测方法与现有基线方法相比,展现了最佳的预测性能。

2. 针对先前方法在做出行需求预测时存在将起始地和目的地视为相同的语义实体、OD数据复杂的空间依赖性和OD数据稀疏性的问题,提出了一种融合渐进图卷积与构造网络更新的方法,实现了较为精准的OD需求预测。本文将渐进图卷积网络、图构造网络与分层记忆更新模块结合,使得模型更好地捕捉不同时间尺度上的相关特征。引入随机游走采样模块,区分原点和目的地为不同的语义实体,以避免混淆两种类型的信息。此外,还设计了一个损失函数,优化了OD数据稀疏性的问题。在真实数据集上的实验验证,本文所提出OD需求预测方法,相比于现有基线方法,展现了更优的预测准确性。

关键词: 时空相关性 深度学习 图卷积

ABSTRACT

In recent years, the process of urban modernization in China has been advancing rapidly, with significant improvements in public facilities and residents' living standards. However, with the rapid growth of the population and the continuous expansion of traffic scale, traffic congestion has become a major concern for urban residents. The "China Urban Traffic Report for the Third Quarter of 2023" points out that the overall commuting traffic in major cities in China shows a trend of congestion. The carrying capacity of urban road networks has not been able to keep pace with the speed of urban development and the rate of vehicle growth, leading to severe traffic congestion in many cities during peak periods. Against this background, the development of urban traffic flow prediction becomes particularly important. Traffic flow prediction can provide congestion warnings for travelers and also serve as an important tool for traffic management departments to implement traffic control.

However, on the one hand, due to the complex spatiotemporal nature of traffic datasets, commonly used models have not fully exploited the long-range spatiotemporal features and hidden dynamic spatial characteristics in the traffic network nodes and adjacency matrices, resulting in poor long-term prediction accuracy and significant prediction errors. On the other hand, previous methods have issues in predicting travel demand, such as treating the origin and destination as the same semantic entity and the complex spatial dependencies in origin-destination (OD) data. Additionally, the imbalance in demand distribution leads to the sparsity of OD data. These factors pose challenges to the research of traffic flow prediction methods based on graph convolutional neural networks.

To address the above challenges, this thesis conducts research on traffic flow prediction methods based on graph convolutional neural networks, with the main research contents including:

1. To address the issue that commonly used methods have not fully exploited the long-range spatiotemporal features and hidden dynamic spatial characteristics in the traffic network nodes and adjacency matrices, a spatiotemporal fusion dynamic graph convolution method is proposed to achieve more accurate long-term traffic flow prediction. This thesis combines the probabilistic sparse self-attention mechanism with the spatiotemporal convolution block, allowing the model to maintain higher computational efficiency while having the ability to model long-range spatiotemporal correlations. A

dynamic graph convolutional network is used to fully utilize prior knowledge to capture the hidden spatial features of traffic flow. Comparative experiments on two real datasets show that the long-term traffic flow prediction method proposed in this thesis has the best prediction performance compared to existing baseline methods.

2.To address the problems in previous methods of predicting travel demand, such as treating the origin and destination as the same semantic entity, the complex spatial dependencies in OD data, and the sparsity of OD data, a method combining progressive graph convolution with network construction updates is proposed to achieve more accurate OD demand prediction. This thesis combines the progressive graph convolutional network, graph construction network, and hierarchical memory update module to better capture relevant features at different time scales. A random walk sampling module is introduced to distinguish the origin and destination as different semantic entities, avoiding confusion between the two types of information. In addition, a loss function is designed to optimize the problem of OD data sparsity. Experiments on real datasets validate that the OD demand prediction method proposed in this thesis shows superior prediction accuracy compared to existing baseline methods

Key Words: spatial-temporal correlativity, deep learning, graph convolution

目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.2.1 基于统计学的方法	2
1.2.2 基于传统机器学习的方法	3
1.2.3 基于深度学习的方法	3
1.2.4 研究难点总结	7
1.3 本文主要工作	7
1.4 本文组织结构	8
第 2 章 基本概念与原理	10
2.1 交通数据的基本概念	10
2.1.1 交通数据	10
2.1.2 时空相关性	10
2.2 循环神经网络	11
2.2.1 RNN	11
2.2.2 LSTM 长短记忆网络	13
2.2.3 GRU 门控循环单元	14
2.3 因果卷积神经网络	15
2.4 图神经网络	16
2.4.1 图卷积神经网络	17
2.5 注意力机制	19
2.6 本章小结	20
第 3 章 基于时空融合动态图卷积的交通流预测方法	21
3.1 引言	21
3.2 相关定义与数据预处理	23
3.2.1 相关定义	23
3.2.2 数据预处理	23
3.3 基于时空融合动态图卷积方法	24
3.3.1 交互式学习模块	24
3.3.2 动态图卷积	26
3.3.3 时空卷积模块	28

3.3.4	稀疏自注意力机制模块	29
3.3.5	损失函数	30
3.4	实验结果与分析	30
3.4.1	数据集描述	31
3.4.2	实验环境与实验设置	31
3.4.3	超参数实验	33
3.4.4	模型对比实验	33
3.4.5	模型消融实验	36
3.4.6	预测值与真实值对比实验	38
3.4.7	空间依赖性的可视化分析	39
3.4.8	可视化分析	39
3.5	本章小结	40
第 4 章	融合渐进图卷积与构造网络更新的 OD 需求预测方法	43
4.1	引言	43
4.2	相关定义与数据预处理	45
4.2.1	相关定义	45
4.2.2	数据预处理	46
4.3	融合渐进图卷积与构造网络更新方法	46
4.3.1	随机游走采样模块	47
4.3.2	渐进图卷积构造网络模块	48
4.3.3	分层记忆更新模块	49
4.3.4	损失函数	52
4.4	实验结果与分析	52
4.4.1	数据集介绍	52
4.4.2	实验环境与实验设置	54
4.4.3	超参数实验	54
4.4.4	模型对比实验	55
4.4.5	模型复杂度和训练速度实验	58
4.4.6	模型消融实验	59
4.4.7	预测值与真实值对比实验	60
4.5	本章小结	62
第 5 章	总结与展望	63
5.1	本文工作总结	63
5.2	未来展望	64

目 录

参考文献	65
致谢	71
在读期间发表的学术论文与取得的研究成果	72

插图清单

图 1.1	交通数据的周期性和趋势性	2
图 1.2	本文主要研究内容	8
图 2.1	时间相关性	11
图 2.2	空间相关性	11
图 2.3	RNN 结构图 ^[53]	12
图 2.4	LSTM 结构图和输入门 ^[54]	13
图 2.5	LSTM 记忆门和输出门 ^[54]	14
图 2.6	GRU 结构图 ^[55]	14
图 2.7	因果卷积神经网络的基本结构 ^[56]	15
图 2.8	加入空洞卷积后的因果卷积神经网络 ^[56]	16
图 2.9	图神经网络的信息聚合	17
图 2.10	图卷积神经网络 ^[64]	18
图 2.11	注意力机制计算过程	19
图 3.1	动态交通网络和隐藏动态空间特性	22
图 3.2	交通流的时空相关性	22
图 3.3	IDG-PSAtt 框架图	25
图 3.4	交互式学习模块	26
图 3.5	动态图卷积模块	26
图 3.6	时空卷积模块	28
图 3.7	概率稀疏自注意力机制	30
图 3.8	不同区域的流量曲线	32
图 3.9	不同的隐藏单元数量对预测效果的影响	33
图 3.10	不同的 epoch 对预测效果的影响	34
图 3.11	在两个数据集上的 MAE 指标比较	37
图 3.12	在两个数据集上的 MAPE 指标比较	37
图 3.13	在两个数据集上的 RMSE 指标比较	38
图 3.14	METR-LA 的 69 号节点	38
图 3.15	PEMS-BAY 的 162 号节点	39
图 3.16	IDG-PSAtt 在 METR-LA 数据集上的空间依赖	40
图 3.17	在两个数据集上的 MAE 曲线	41
图 3.18	在两个数据集上的 MAPE 曲线	41

图 3.19	在两个数据集上的 RMSE 曲线	41
图 4.1	OD 矩阵的空间划分和 OD 矩阵的序列	44
图 4.2	城市不同区域的 OD 关联和空间依赖性	45
图 4.3	PGC-ODMUP 整体框架图	46
图 4.4	随机游走采样	47
图 4.5	渐进图	48
图 4.6	分层记忆更新	50
图 4.7	两个区域不同方向的出行需求	53
图 4.8	北京地铁和纽约出租车的迭代轮次实验	55
图 4.9	北京地铁和纽约出租车的 HMU 隐藏单元实验	55
图 4.10	北京地铁和纽约出租车的 PGC 隐藏单元实验	56
图 4.11	在 Beijing Subway 上不同变体的 RMSE 和 PCC	59
图 4.12	在 New York Taxi 上不同变体的 RMSE 和 PCC	60
图 4.13	区域 162 至区域 102 预测与真实曲线对比	61
图 4.14	区域 68 至区域 44 预测与真实曲线对比	61

表格清单

表 3.1	实验数据集的描述	31
表 3.2	实验环境	32
表 3.3	IDG-PSAtt 模型与基线模型在 METR-LA 和 PEMS-BAY 上的比较 ..	35
表 3.4	在 METR-LA 和 PEMS-BAY 数据集上对 IDG-PSAtt 变体的比较	36
表 4.1	实验数据集的描述	53
表 4.2	北京地铁的比较结果	57
表 4.3	纽约出租车的比较结果	57
表 4.4	模型参数量和速度对比	58
表 4.5	对比不同模型在北京地铁和纽约出租车数据集上的表现	59

第1章 绪 论

本章首先简要概述了交通流预测的背景知识与研究意义，随后对国内外相关研究现状进行难点总结和讨论，然后介绍了本文主要研究内容，最后介绍了本文的章节安排和组织结构。

1.1 研究背景与意义

近年来，我国城市现代化进程迅速推进，城市公共设施和居民生活水平显著提升。然而，随着人口急速增长和交通规模不断扩大，交通拥堵成为城市居民生活的主要困扰。根据国家统计局最新统计截至 2023 年 9 月底^[1]，全国机动车的总数已经上升到 4.3 亿辆。《2023 年第 3 季度中国城市交通报告》^[2] 报告也指出，我国主要城市通勤出行总体呈拥挤趋势。城市路网的承载力未能与城市发展速度和车辆增长率同步提升，导致了许多城市在高峰时段经常出现严重的交通拥堵。此外，交通管理和规划不足也是造成拥堵的重要因素之一。这一状况使得科学解决交通路段拥堵成为我国城市现代化进程中亟需解决的难题。

智能交通系统（Intelligent Transportation Systems, ITS）^[3-4] 的发展为缓解这一问题提供了途径。ITS 利用先进的信息技术、通信技术、控制技术和传感技术，实现对交通系统的实时监控、动态管理和智能服务，从而提高交通系统的效率和安全性。交通流预测作为 ITS 的一个重要组成部分，能够为出行者提供更准确的出行时间预估，帮助他们规划出行时间和选择最佳出行路线。还可以为他们提供拥堵预警，帮助他们提前做出调整，避开拥堵路段。同时也为城市交通管理部门实施交通管制提供了一个重要工具。

交通流预测旨在利用过往交通信息，结合交通网络构成、时间属性、周期性如图 1.1(a) 及趋势性如图 1.1(b) 等因素，以预估未来交通网络内路段的交通情况。研究人员将其分为短期预测（小于 30min）和长期预测（大于或等于 30min），本文主要以研究长期预测为主。长期预测会随着时间的增长和道路交通出现的随机因素增多，误差逐渐增大。除此之外，实现长期交通流预测仍然面临许多挑战，比如，如何准确捕捉和建模交通流的复杂远程时空特性、如何有效处理和解析交通流的隐藏动态空间特性等。

起始地-目的地（Origin-Destination, OD）需求预测作为交通流预测的一个重要分支同样不可忽略。OD 需求预测关注于预测从一个地点（起点）到另一个地点（终点）的出行需求量。这种预测对于城市规划、交通管理和公共交通系统的优化至关重要。OD 需求预测能够提供出行模式、乘客流量和交通流量分布的

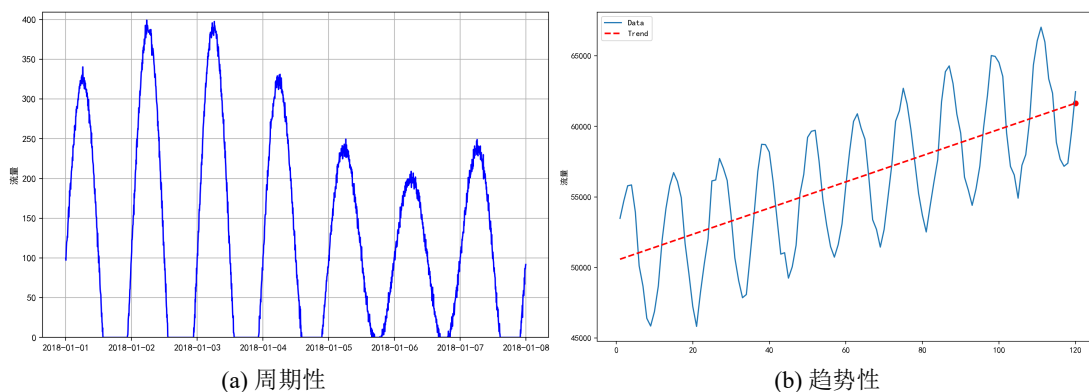


图 1.1 交通数据的周期性和趋势性

全面信息，帮助城市规划和交通工程师制定更有效的交通策略和基础设施规划。同时，OD 需求预测面临着多方面的挑战，比如在实际应用中，大量的起始地和目的地组合之间可能没有足够的出行记录、交通网络中的空间关系和相互作用对于出行需求有着显著的影响等。

综上所述，本文探讨的交通流预测对智慧城市建设和交通管理具有重要的研究意义。通过准确预测城市交通流量，本研究能够为缓解交通拥堵、减少交通事故发生率、以及优化交通规划和基础设施建设提供重要的科学依据。

1.2 国内外研究现状

近年来，交通流预测领域吸引了众多研究者的关注，他们致力于通过不同的方法提高预测的准确性和效率。研究交通流预测的方法主要分为三类：统计学方法、传统机器学习方法和深度学习方法。为了全面理解交通流预测的研究现状，本节将对这三种方法在国内外的应用和发展进行详细分析。

1.2.1 基于统计学的方法

早期的交通预测方法是基于固定理论假设，通过分析历史时段的数据变化规律，对未来时段的交通流量分布进行合理推断。基于统计学方法的预测技术，包括历史平均法（History Average, HA）^[5-6]、自回归移动平均（Autoregressive Integrated Moving Average Model, ARIMA）^[7-9]模型和向量自回归（Vector Autoregression, VAR）^[10]模型，被用于处理交通流量的预测问题，视之为单一或复合时间序列分析的范畴。HA 采用简明直接的方法，通过平均过去一定时间范围内的交通数据来预测未来值。虽然此方法简便易行、计算迅速，但由于不适应复杂的交通场景，其预测效果受限。ARIMA 模型通过差分化处理非平稳序列，利用移动平均和自回归方法来捕捉序列的时间相关性，适合处理具有稳定周期性的交通数据。VAR 模型通过分析多个序列间的线性依赖性来预测其动态关系，但

在面对复杂的非线性交通数据时效果不佳，随着结合变量数目的增加，模型复杂度显著上升。

虽然这些初期的统计方法因其模型简洁性而广泛应用，它们基于特定的数据前提假设进行预测。例如，ARIMA 假定需将非稳定序列转换为稳定序列进行分析，而 VAR 假设各变量间存在线性联系，并要求对序列的稳定性进行验证。交通数据的复杂性，受限于历史条件、环境和天气因素等多变量影响，往往不满足这些基本假设，导致在现实场景下应用效果受限。此外，忽略了序列间的空间关联性，这限制了其在满足现代交通系统对预测准确性高要求方面的应用。

1.2.2 基于传统机器学习的方法

随着技术发展，基于传统的机器学习算法逐渐被引入解决交通预测的问题，以克服传统统计方法在分析复杂交通场景时的限制。方法如 K 近邻 (K-Nearest Neighbor, KNN)^[11-12]、支持向量回归 (Support Vector Regression, SVR)^[13-15]、贝叶斯网络 (Bayesian Network, BN)^[16-18] 及高斯过程 (Gaussian Process, GP)^[19-20]，各自有各自的优势。

KNN 凭借其识别类似交通模式的能力，提供了快速且易于部署的预测解决方案。SVR 利用核函数技术处理低维数据的非线性问题，优化了小规模数据集的预测精度，尽管在大数据背景下其计算成本较高。BN 利用精心构建的特征进行交通预测，其过程既耗时又需要大量人力。这种方法通过增强数据来预测时间，属于一种无需参数设定的基于距离测算的学习技术，同时模型局限性大、泛化能力差，难以应用到真实场景。GP 利用时空核函数深入挖掘交通数据的潜在特征，但其在面对大数据集时的高计算与存储要求限制了其广泛应用。

虽然传统机器学习方法在处理非线性和时空数据方面取得了显著进步，但在面对高复杂度的交通数据时，它们仍然面临着诸多挑战，例如特征选择的复杂性、高计算需求和有限的模型适应性。这些限制因素影响了算法的效能。因此，尽管这些方法在早期应用中显示出了优势，但为了有效预测具有复杂时空特性的交通流量，还需要对这些方法进行进一步的优化和发展。

1.2.3 基于深度学习的方法

传统机器学习技术在预测交通流方面由于过度依赖特征工程而受到严重制约。相比之下，基于深度学习的方法在交通流预测中显示出了显著的优势，并因此被广泛采用。深度学习方法，特别是那些构建于深度神经网络上的模型，能够自动地从大量数据中提取复杂特征，无需进行繁琐的特征提取工程。

1. 基于卷积神经网络的方法

卷积神经网络 (Convolutional Neural Networks, CNN) 的提出和应用帮助人们解决了很多领域的难题, 研究人员也尝试过利用 CNN 进行路网的交通流量预测。例如, Qiao 等^[21]考虑了与预测路段相交路段的交通流量, 提出了基于 1D 卷积的网络实现对交通流的短时预测, 该方法通过 1DCNN 获取交通数据中的空间信息, 并将空间时间特征用于回归预测。Zhao 等^[22]提出了一个基于时间卷积神经网络 (Temporal Convolutional Network, TCN) 模型的深度学习框架, TCN 也是一种 1D 卷积的网络, 用于城市短期范围内的交通流量预测, 以准确捕捉交通流量的时间演化。由于 1D 卷积核尺寸的限制, 1D 卷积在捕获较长期的时间依赖性方面表现不佳, 同时也无法获取空间上的联系。

由此 2D 卷积捕获交通流中的空间关系尤为重要, Na 等^[23]提出了一种基于多变量分析的短期交通流量预测的混合神经网络算法, 该算法结合了 CNN 和长短期记忆网络 (Long Short-Term Memory, LSTM), 形成了一个 CNN-LSTM 模块, 它可以提取交通流信息的时空信息。Zhang 等^[24]使用残差神经网络结构来捕捉人群交通的时间接近性、周期性和趋势特征。对每个特征, 设计了一个残差卷积单元的分支, 每个分支都对人群交通的空间属性进行了模拟。Men 等^[25]提出了一种结合 CNN 和双向长短期记忆网络的混合模型, 用于长期城市路线的交通流量预测。该模型利用 CNN 提取输入数据中的隐藏特征, 并通过双向长短期记忆网络理解时间上下文。Gu 等^[26]结合 LSTM 和门控循环单元神经网络 (Gated Recurrent Unit, GRU), 建立了一个两层深度学习框架, 用于预测车道级交通速度。

基于卷积神经网络的方法在提取交通流预测的时空特征方面已取得显著成果。然而, 这种方法存在一些局限性。首先, CNN 的固定卷积核尺寸限制了它捕捉长距离空间依赖性的能力, 特别是在非相邻城市区域的交通流动变化上, 这是 CNN 难以有效处理的。其次, CNN 高度依赖于训练数据的质量和数量; 如果历史数据量不足或质量差, 会严重影响预测性能。因此, 尽管基于 CNN 的技术在处理特定交通流预测任务时表现出色, 其在处理长范围空间依赖性方面的限制不容忽视。

2. 基于循环神经网络的方法

基于 CNN 的方法主要捕捉局部空间特征, 在处理整体上下文信息, 特别是在需要理解交通流在广阔区域内的移动性或其他空间属性的场景中, 可能效果有限。而循环神经网络展现出了在处理序列化数据方面的独特优势, 其能力不仅限于捕捉时间序列中的动态变化, 也在理解空间关联性方面表现出色。这种对时间和空间依赖性的深刻理解, 让循环神经网络成为交通流预测这一应用场景中的强有力工具。Lin 等^[27]提出了一种结合自注意力机制和 LSTM 的新型的自

注意力记忆网络，通过关注数据中最重要的特征以提高模型的预测准确性。Yu 等^[28]提出一种基于深度学习的方法。该方法以长短时记忆单元为基础构建深度神经网络，结合深度 LSTM 和混合深度 LSTM 模型，不仅对高峰时段的交通流量进行准确预测，还能有效识别交通数据中的特异性，包括事故后的交通状况。Bai 等^[29]提出了一种针对数据的自适应图生成模块以及一种针对节点的自适应参数学习模块，结合 GRU，实现了对交通流数据细粒度时空相关性的自动学习。

Luo 等^[30]通过 K 近邻算法筛选出与目标站点密切相关的周边站点，从而把握交通流的空间属性。利用 LSTM 探索交通流量随时间的波动特性，并运用双层 LSTM 对选中的站点进行交通流量预测。最后，采取秩指数加权方法融合预测结果，以产生最终的预测值。Chu 等^[31]提出了一种名为多尺度卷积长短期记忆网络 (MultiConvLSTM) 的新型深度学习方法，并且提出了一种新的数据结构，称为 OD 张量，用以表示 OD 流量，以处理 OD 矩阵中的高维特征。MultiConvLSTM 考虑了时间和空间的相关性，以预测未来的出行需求和 OD 流量。Shu 等^[32]提出了一种融合 CNN 和 LSTM 的空间-时间跨区域出租车需求预测模型。该方法利用 CNN 提取特征间的相关性和时间上的紧密依赖，再通过 LSTM 将它们融合在时间序列中。

综合上述讨论，基于循环神经网络的模型已经在多个研究中得到应用，为交通流预测提供了更灵活、强大的方法。然而，由于交通数据的复杂性，包括非线性、多维性、时变性和周期性，精准的交通流预测仍然面临挑战。为了更好地捕捉这些复杂特性，并提高模型对不同交通情境的适应性，图神经网络提供了一种有前景的解决方案。图神经网络能够有效处理交通网络的图结构数据，捕捉交通流中的空间依赖性，从而为实现更精准和可靠的交通流预测提供了新的途径。

3. 基于图神经网络的方法

图神经网络 (Graph Neural Networks, GNN) 模型正逐步成为交通研究中的一种先进的深度时空分析工具。Zhao 等^[33]提出了一种时间图卷积网络模型，用于捕获交通网络的空间依赖性和时间动态性，通过使用图卷积网络来分析交通网络的空间布局，它识别了不同地区之间的相互依存性。同时，LSTM 用于追踪时间序列数据，如交通流或速度，捕捉其时间动态。Zhang 等^[34]提出了注意力图卷积序列到序列模型，它采用了带有注意力机制的图卷积网络，具有较高的预测精度，尽管混合模型结构相对较为复杂且具有大量待标定的参数。Li 等^[35]提出了一种数据驱动的“时间图”生成方法，以弥补空间图可能无法反映的真实相关性。该方法通过对不同时间段的各种空间图和时间图进行并行处理的新颖融合操作，能够有效地学习隐藏的时空依赖关系。Zeng 等^[36]提出了一个因果解耦表示模块来识别属性中的因果关系，并采用了一个变分样本重新加权模块来减少混淆偏差，解决了 OD 需求预测中只有部分因素表明有因果信息和因果建模

过程带来了混淆偏差。

Li 等^[37]设计了一个多头图注意力模块,旨在捕捉路段之间的空间依赖关系。建立了一个组件融合块,用于整合速度、流量和天气信息,并且构建了基于注意力的长短期记忆模块,用于学习时间依赖性和进行基于路段的速度预测。Wu 等^[38]采用时空卷积网络和自适应图卷积来捕捉时空相关性,用于检查交通流节点之间的时空属性。Zhang 等^[39]利用动态图卷积和一维卷积神经网络来探索交通流节点之间的时空特征,并且将卷积神经网络扩展到图域。Chen 等^[40]引入了注意力机制在嵌入式结构中,通过路网距离建立节点图,同时根据各种边缘交互模式建立边缘图。然后,采用双分量图卷积来实现节点和边缘之间的有效交互。对于 OD 需求预测,Wang 等^[41]提出了一个基于网格嵌入的多任务学习,分别关注空间信息和时间信息。网格嵌入部分旨在模拟乘客的空间移动模式和不同区域之间的邻近关系,其预加权聚合器的目的是感知数据的稀疏性和范围,多任务学习框架专注于建模时间属性。

此外,当前研究中许多方法通过定义各种图神经网络的新变体,以揭示交通流的深层结构^[42]。举例来说,Pan 等^[43]提出了一种专为多变量时间序列数据设计的通用图神经网络框架。该方法通过图学习模块自动提取变量之间的单向关系,并能轻松整合变量属性等外部知识。进一步引入了一个新颖的混合跳跃传播层和一个扩张萌芽层,以有效捕捉长时序列中的空间和时间依赖关系。Xu 等^[44]提出了一种图神经网络的新变体,被称为空间变换器。该方法通过自注意力机制以动态建模有向空间依赖关系,使模型能够更加灵活地识别和利用交通流的方向和实时状况,让该方法提高长期交通预测的准确性。Shi 等^[45]提出了一种名为多视角图卷积神经网络的新变体,用于捕捉 OD 流量预测中的复杂依赖关系。该方法利用长短期记忆网络提取每对 OD 对的时间特征,然后通过二维图卷积神经网络学习起点和终点的空间依赖性。此外,还设计了一个动态图和两个静态图来捕捉复杂的空间依赖关系,并使用平均策略获得最终预测的 OD 流量。

综上所述,随着神经网络的出现和成熟,基于图神经网络的交通流预测研究更适合模拟道路结构的时空特征,图神经网络上进行时空特征捕获已经成为主流方向^[46-49]。图神经网络方法通过时空卷积网络、自适应图卷积等技术,能够高效、自动地提取复杂时空关系,从而提高了交通流预测的准确性。模型如多头图注意力模块、空间变换器、时间图生成方法等^[50-52]不仅消除了对手动特征设计的需求,还能够有效捕捉隐藏的时空依赖关系。因此,本文提出的相关算法均以图神经网络为基础进行研究,以更好地学习数据的多种特性,使预测结果更加准确。

1.2.4 研究难点总结

在回顾了交通预测领域的相关文献后，可以看到这个领域的研究非常活跃，并且有很多进步。总的来说，研究人员们一直在寻找更好的方法来进行交通预测。然而，现有研究还存在一些不足之处：

(1) 现有方法往往未能充分挖掘路网中交通节点以及邻接矩阵内蕴含的远程时空特征和隐藏的动态空间属性。这些特征和属性对于理解和预测交通流的长期趋势至关重要，因为它们反映了交通系统中各个组成部分之间的复杂相互作用和依赖性。远程时空特征指的是在时间和空间上距离较远的节点之间的关系和影响，这种影响可能不会立即显现，但对长期的交通流量有着深远的影响。同样，隐藏的动态空间属性涉及到路网内部随时间变化的空间结构和流动模式，这些模式可能由于各种外部因素而发生变化。

(2) 现有的出行需求预测方法面临着几个核心问题，未能充分解决这些问题，显著影响了预测的准确性和实用性。首先，许多方法将起始地和目的地视为同一种语义实体，忽略了它们之间在地理位置、功能定位以及与出行者需求相关的本质区别。其次，OD数据的复杂空间依赖性问题也往往未得到充分解决。城市的每个区域和地点都嵌入在一个复杂的空间网络中，这个网络中的空间关系和相互作用对于出行需求有着显著的影响。最后，OD数据的稀疏性问题是另一个重要挑战，大量的起始地和目的地组合之间可能没有足够的出行记录，导致OD矩阵中存在大量的空白或零值。

1.3 本文主要工作

由于交通预测问题的实质是时间序列预测问题，其数据序列的变化在空间方面有着明显的流动性，在时间方面有着明显的周期性。因此本文针对如何更好地学习交通路网数据的时空特性等一系列问题展开研究，提出两个基于图神经网络的方法，并在主流数据集上进行实验验证。具体研究结构如图1.2所示：

(1) 本文将概率稀疏自注意力机制与交互式动态时空图卷积结合在一起，引入到长期交通流量预测任务中。通过融合概率稀疏自注意力机制与交互式动态时空图卷积，以不同方式整合到预测模型中，充分利用额外信息提升预测精度。算法通过概率稀疏注意力机制优化图卷积网络结构，使得模型具有建模长程时空相关性能力的同时保持更高的计算效率。并在这些间隔内同步、交互地共享捕获到的动态时空特征，同时利用编解码器进一步提高预测数据准确度，这有助于更好地理解交通流的动态关联和时空演变，提高模型对交通流变化的敏感性。实验证明了算法的有效性。

(2) 本文将渐进图卷积网络、图构造网络与分层记忆更新模块结合在一起，引

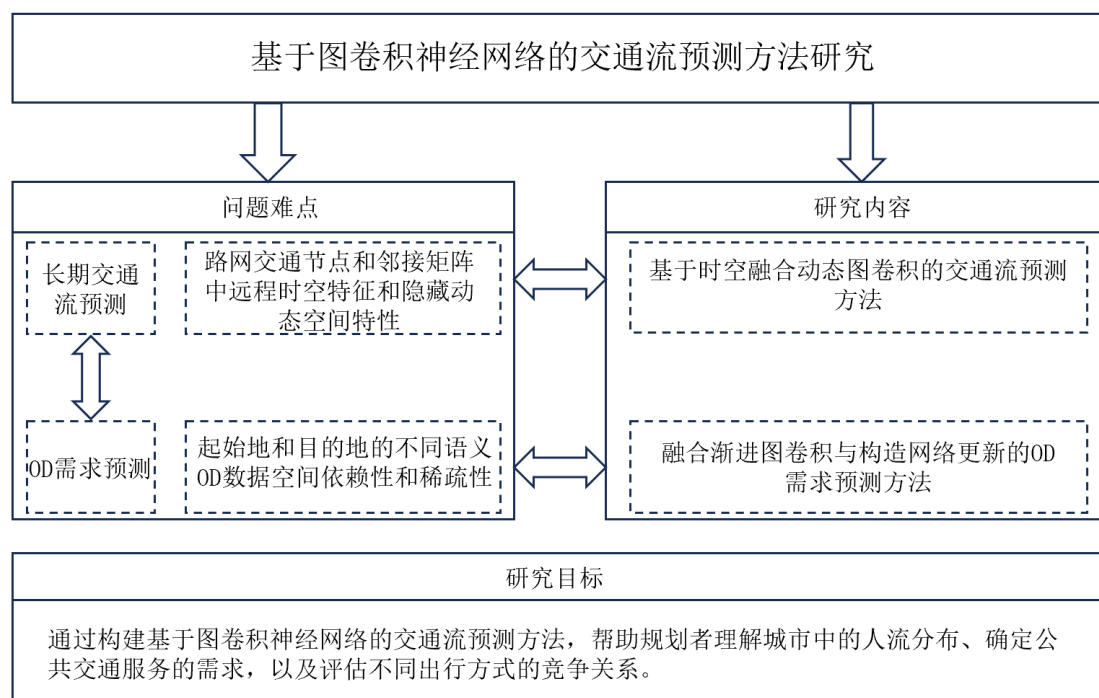


图 1.2 本文主要研究内容

入到出行需求预测任务中。先前方法在预测 OD 矩阵时存在将起始地和目的地视为相同的语义实体、OD 数据复杂的空间依赖性和 OD 数据稀疏性的问题。首先算法通过引入渐进图卷积网络和图构造网络来预测 OD 矩阵，从而更好地反映交通网络的空间动态特性，提高了 OD 矩阵预测的准确性和可靠性，随后，使用随机游走采样将原点和目的地视为不同的语义实体，以此区分它们，以避免混淆两种类型的信息。接着，基于分层记忆更新模块将节点的时间感知表示分层处理，使得模型更好地捕捉不同时间尺度上的相关特征，最后，使用基于 MSE 的变体损失函数，优化了 OD 数据稀疏性的问题。实验证明了算法的有效性。

1.4 本文组织结构

论文共分为五个章节，具体安排如下：

第一章：绪论。本章首先简述了交通流预测的研究背景和意义，强调了交通流预测对于现代城市交通管理系统、交通规划和道路网络设计的重要性。接着，分析了国内外在交通流预测领域的研究现状和成果，揭示了各种方法的优势与局限。例如：指出传统的统计学方法和机器学习技术已经不足以满足预测任务对精度的高要求，同时强调了图卷积神经网络已经成为了交通流预测的新兴研究方向，它能够处理更复杂的数据和模式。最后，介绍了文章的主要研究内容和各个章节的安排。

第二章：基本概念和原理。首先，探究交通流数据的时空相关性。然后，对

循环神经网络、图神经网络和注意力机制的基本原理进行介绍，为后续的模型搭建提供理论基础。最后，对本章进行小结。

第三章：基于时空融合动态图卷积的交通流预测方法。首先，对研究背景、遇到的问题、选取的数据集和预处理方法进行介绍。接着，将交互式学习策略、动态图卷积、时空卷积模块和概率稀疏自注意力块进行完整介绍。最后，进行基线实验、对比实验，通过消融实验分析单个模块对评价指标的影响。

第四章：融合渐进图卷积与构造网络更新的 OD 需求预测的方法。首先，对研究背景、遇到的问题、选取的数据集和预处理方法进行介绍。接着对渐进图卷积构造网络、分层记忆更新和随机游走采样等方法进行介绍。其次，在数据集上进行基线实验对比分析，最后通过消融实验分析单个模块对模型的影响。

第五章：总结与展望。本章对本文所有的内容进行总结，分析了本文目前还存在的问题，也进一步说明了未来的工作方向。

第 2 章 基本概念与原理

本章全面介绍了涉及交通数据的基本概念和原理、注意力机制、图卷积神经网络和循环神经网络，为后续基于交通流预测的方案和 OD 需求预测的方案奠定了理论基础。

2.1 交通数据的基本概念

2.1.1 交通数据

交通数据是现代城市交通系统中生成的重要时空信息资源。一般来说，这些交通数据由分布在城市的各个道路上的传感器通过预先设定时间频率进行采集。因为传感器广泛部署，交通数据可被视为带有时空信息的图数据结构。在实际的交通监测中，这些传感器记录了道路上多个重要信号值，包括但不限于交通流量、速度和密度等指标的时空变化信息。城市交通网络中车辆、行人和其他交通参与者的活动情况可以通过这些信号值进行反映。因此，交通数据也被称为图信号。假设有一个图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ ，其中 \mathcal{V} 表示图的节点集， \mathcal{E} 表示图的边集。每个节点 $v_i \in \mathcal{V}$ 表示一个传感器，而每条边 $e_{ij} \in \mathcal{E}$ 表示节点 v_i 和 v_j 之间的连接。现在，如果我们有一个图信号 $f(\cdot)$ 它在每个节点 v_i 处都有一个对应的数值，我们可以形式化地表示 \mathcal{G} 上的图信号如公式 (2.1)：

$$f(\mathcal{G}) = \{f(v_1), f(v_2), \dots, f(v_N)\} \quad (2.1)$$

其中， N 是传感器的总数。图信号的具体形式取决于具体的应用场景和信号类型

2.1.2 时空相关性

时空相关性涉及时间推移和空间位置间的关联。在交通数据和预测中，空间相关性考虑不同节点之间相互影响的关系，时间相关性考虑不同时刻之间的关系。同时，在交通数据中存在时空相关性，即不同地点和不同时刻的交通状态之间的关联程度。因此，时空相关性分析需要考虑时间和空间维度上的关系，以更全面理解交通现象的变化趋势和规律。

时间相关性涉及到交通现象随着时间变化的关系，如图 2.1，交通流量具有很强的周期性，举例来说，在 $t-1$ 时刻节点 1 和 2 还有没有很强的关联性，但在 t 时刻节点 1 和 2 关联性就变强了，此外，特殊天气条件或特殊日期也会对交通流产生很大的影响，时间相关性分析有助于理解交通的日变化、周变化，以及可能存在的季节性变化等。

空间相关性关注不同位置之间的交通流关联，如图 2.2，在早上的 8:30，从住宅区 D 到办公区 C 这段路上车流量非常密集，而在下午 6:00，从办公区 C 到购物区 A 的车流量也非常密集，同时住宅区域 D 对办公区 C 的影响显著降低。

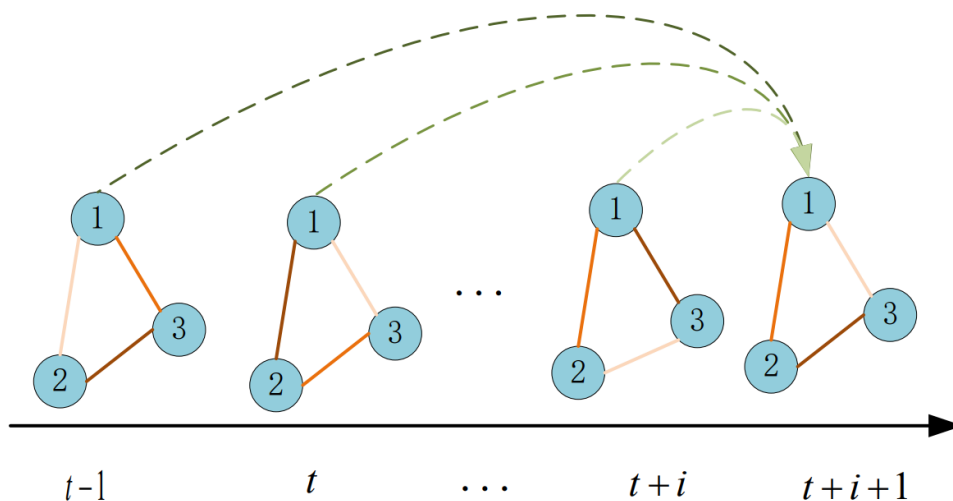


图 2.1 时间相关性

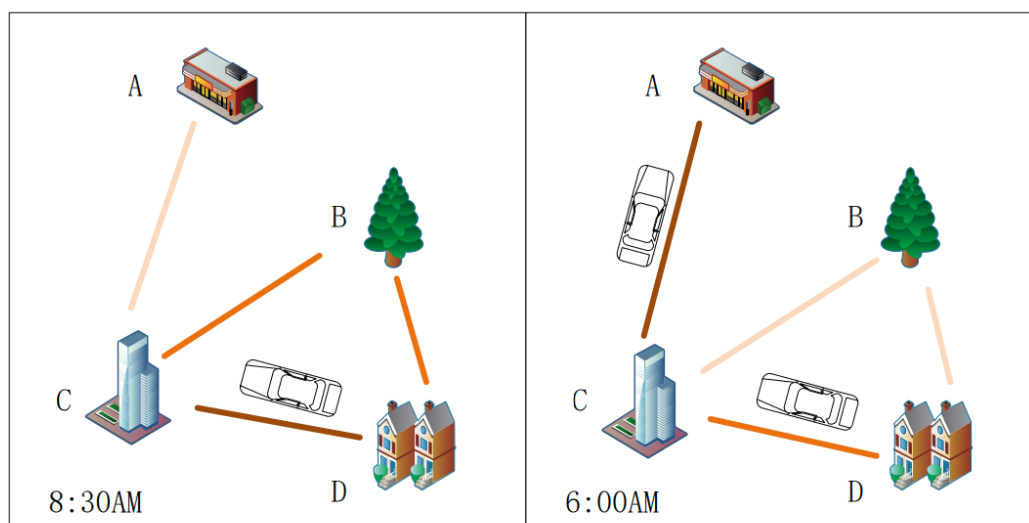


图 2.2 空间相关性

2.2 循环神经网络

2.2.1 RNN

RNN (Recurrent Neural Network, RNN) 是一种用于处理序列数据的神经网络。与传统的前馈神经网络不同，RNN 具有循环连接，当前循环单元的输出和之前的循环单元的输出以及当前的输入相关联。这种关联使得网络能够记忆历史

输出并将其用于计算当前输出，这使得 RNN 在处理序列时保持一种记忆状态。

在 RNN 中，每个时间步都有一个隐藏状态（hidden state），它接收当前时间步的输入和上一个时间步的隐藏状态作为输入。隐藏状态的输出不仅取决于当前时间步的输入，还取决于之前所有时间步的输入。这种循环连接使得 RNN 能够处理变长序列，并捕捉到序列中的时序信息。

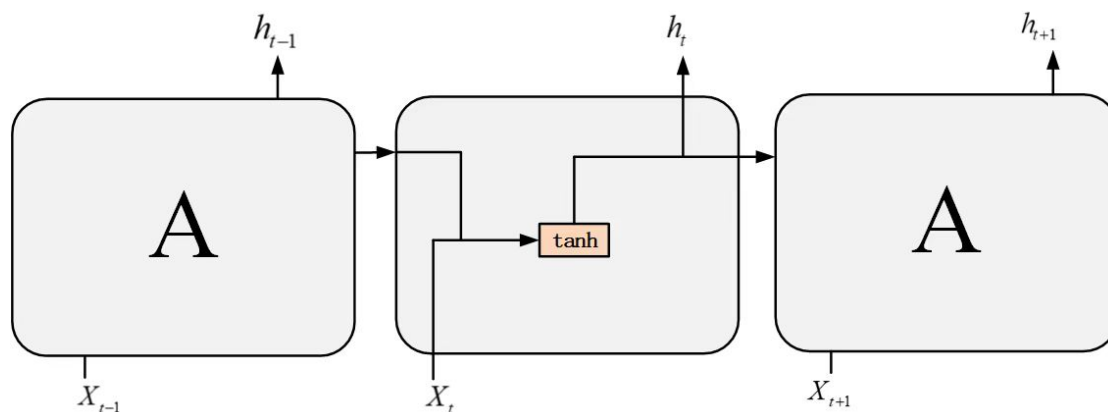


图 2.3 RNN 结构图^[53]

如图 2.3 左图所示，在 RNN 中，每个时间点 t 的输入由两部分组成：当前时间点的输入值 X_t 和前一时间点的隐藏状态 h_{t-1} 。这两部分各自乘以其权重，加和之后通过一个非线性的激活函数处理，以产生该时间点的输出 o_t 及新的隐藏状态 h_t 。然后，这个新的隐藏状态 h_t 会被用作紧接着的下一时间点的输入之一。这一过程可以用公式 (2.2) 来表示：

$$h_t = f(W_{hh}h_{t-1} + W_{xh}X_t) \quad (2.2)$$

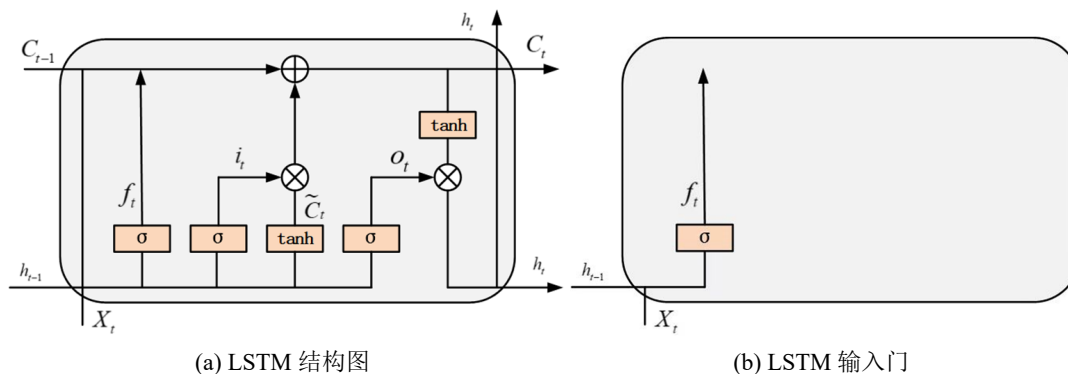
其中 f 是非线性激活函数。

RNN 在自然语言处理、时间序列预测等领域有广泛的应用，例如语言建模、天气预测、流量预测等任务。由于 RNN 能够处理变长序列，并且可以保持记忆状态，它在处理自然语言和时间序列预测时可以考虑上下文的信息，捕捉到词语和序列之间的依赖关系和语义信息。

然而，在传统的 RNN 中，当处理长序列数据时，反向传播过程中通过时间的梯度计算常常遭遇梯度消失或梯度爆炸问题。这是因为在计算时间点 t 的状态梯度时，使用链式法则会导致早期时刻的梯度通过多次乘法累积，影响模型捕捉长期依赖的能力。为了克服这些难题，研究人员开发了如长短期记忆网络和门控循环单元等 RNN 的先进变体。这些改进的算法通过引入门控机制来有效管理信息流，从而缓解梯度问题并增强模型对长序列数据的处理能力。它们引入了门控机制来控制记忆状态的更新，改善了对长期依赖关系的建模能力。

2.2.2 LSTM 长短记忆网络

LSTM 是对 RNN 的一种优化，专门设计来克服 RNN 面临的梯度消失和梯度爆炸的问题如图 2.4(a)。LSTM 引入了一个记忆单元，该单元可以存储和访问信息，并通过采用门控制机制实现信息流的有效管理。其核心组成部分涵盖了输入门、忘记门以及输出门。

图 2.4 LSTM 结构图和输入门^[54]

LSTM 的第一步涉及决定在主要记忆单元中丢弃哪些信息，这一决策由一个被称为“忘记门”的 sigmoid 层来执行如图 2.4(b)。该层接收输入 h_{t-1} 和 X_t ，然后输出一个介于 0 到 1 之间的值。其中，1 表示“完全保留这个值”，而 0 表示“完全丢弃这个值”。举例来说，对于基于上下文预测最后一个词的语言模型而言，主要记忆单元可能包含了当前主题的信息，以便预测下一个正确的词。然而，在遇到新的语言话题时，应该淡化对旧话题记忆的依赖，转而采纳新话题信息，以便更精确地预测接下来的词汇。

忘记门的计算方法如公式 (2.3)：

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (2.3)$$

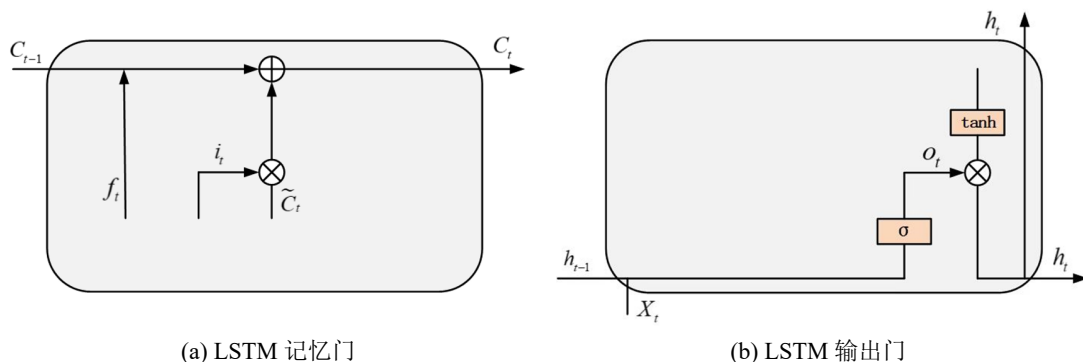
其中 f_t 是时间步 t 处的忘记门的输出， σ 是 sigmoid 函数， W_f 是忘记门的权重矩阵， b_f 是忘记门的偏置项， X_t 是输入。

LSTM 第二步是决定在主要记忆单元里存储的信息如图 2.5(a)。这个问题有两个部分。首先 sigmoid 层调用“输入门”以决定哪些数据是需要更新如公式 (2.4)。然后，一个 tanh 层为新的候选值创建一个向量 \tilde{C}_t ，这些值能够加入主要记忆单元中如公式 (2.5)。下一步，要将这两个部分合并以创建对主要记忆单元的更新。

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (2.4)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \quad (2.5)$$

在决定需要遗忘和需要加入的记忆之后，就可以更新 C_t 了。更新 C_t 操作是



(a) LSTM 记忆门 (b) LSTM 输出门

图 2.5 LSTM 记忆门和输出门^[54]

将 C_{t-1} 与 f_t 相乘如公式 (2.6)，遗忘之前所学习的东西，形成新的记忆信息。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{2.6}$$

最后的输出基于主要记忆单元。首先，运行一个 sigmoid 层，也就是输出门如公式 (2.7)，决定从主记忆单元中哪些部分需要被输出。接着，对这个主记忆单元应用 tanh 激活函数，以便将其数值标准化到-1 至 1 的范围内。最终，通过将 tanh 的结果与 sigmoid 门的输出进行相乘操作，完成了输出的生成。最后结果如图 2.5(b)。

$$o_t = \sigma (W_o [h_{t-1}, X_t] + b_o) \tag{2.7}$$

$$h_t = o_t * \tanh (C_t) \tag{2.8}$$

2.2.3 GRU 门控循环单元

GRU 是 Cho 等人在 2014 年提出的一种对 LSTM 稍作改进的循环神经网络结构如图 2.6 所示。与 LSTM 相比，GRU 将遗忘门和输入门合并为单个“更新门”，同时将主要记忆单元和隐藏状态合并，并进行了其他一些调整。相对于标准的 LSTM 模型，GRU 模型更加简单，因此越来越受到广泛关注和应用。

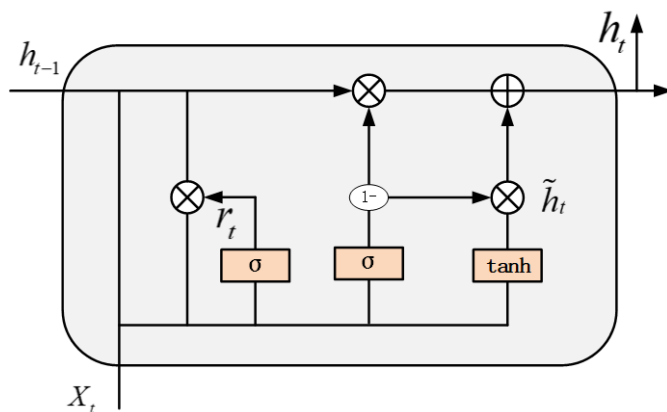


图 2.6 GRU 结构图^[55]

相比于 LSTM，GRU 在门控机制的设计上有所不同。GRU 只使用了一个更新门来控制信息流动和状态更新，从而减少了参数量和计算量。因此，相对于 LSTM，GRU 模型更加简洁，更容易训练，并且在某些任务上取得了相当好的性能。

GRU 由两个门构成，分别是重置门和更新门如公式 (2.9) (2.10)

$$z_t = \sigma(W_z [h_{t-1}, X_t]) \quad (2.9)$$

$$r_t = \sigma(W_r [h_{t-1}, X_t]) \quad (2.10)$$

$$\tilde{h}_t = \tanh(W [r_t * h_{t-1}, X_t]) \quad (2.11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (2.12)$$

这种机制有助于克服循环神经网络遇到的梯度消失难题，同时也能更有效地把握跨越较长时间段的序列依赖性。举例来说，更新门赋予较低的值时，意味着它决定“忘记”昨天的天气情况，因为这对于理解“今天的天气”并不是特别重要。最初的隐藏状态都能较为轻松地保持并在后续的各个时间步骤中传递。

2.3 因果卷积神经网络

因果卷积神经网络是一种改进的卷积神经网络模型，通常用于处理时序数据。时间序列预测问题是机器学习与深度学习领域中的常见任务之一，比如给出一段话，让机器预测其可能会出现的句子或者单词，或者给出过去几天的空气质量，预测明天或者后面几天的空气质量等，都属于时间序列建模要解决的问题，而因果卷积神经网络在处理具有时间关联性的数据上非常有优势。

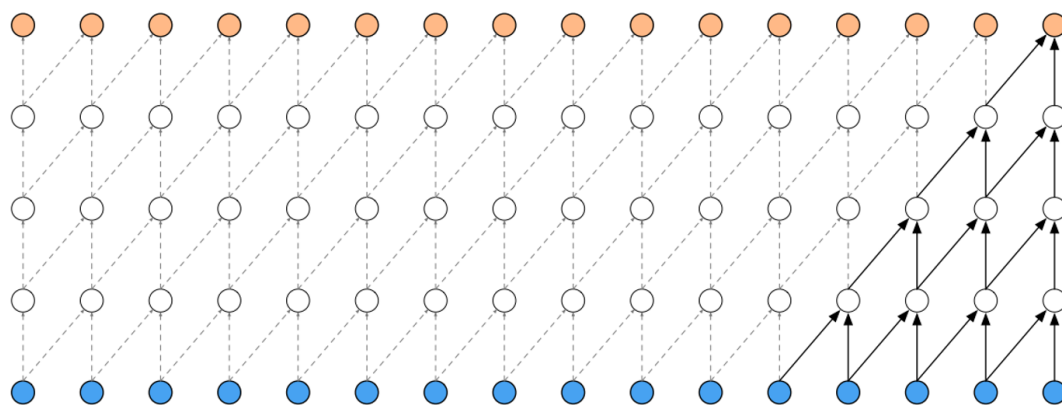


图 2.7 因果卷积神经网络的基本结构^[56]

在传统神经网络中，每层神经元都与下一层的所有神经元相连接，这导致了时间顺序上的约束问题。全连接违反了时间上的因果关系，因为输出的神经元（前一时刻）与输入的神经元（后一时刻）存在连接。可以通过使用掩膜技术，我

们可以部分去除层与层之间的连接，仅保留从前向后的连接，从而使网络满足时间上的因果依赖原则，如图 2.7 所示。

但是，在上述的网络中也会存在这样一个问题：因果卷积捕获历史信息的能力不是很强，比如最后的输出只依赖第一层的最后 5 个神经元，这样导致前面的信息无法获取。因此增强因果卷积捕获历史信息的能力是非常有必要的。为了解决这个问题，有人提出了采用空洞卷积的方法。这个技术通过在卷积过程中加入一个称作膨胀率的间隔参数，允许卷积核在移动过程中跳过固定数量的输入单位，如图 2.8 所示。

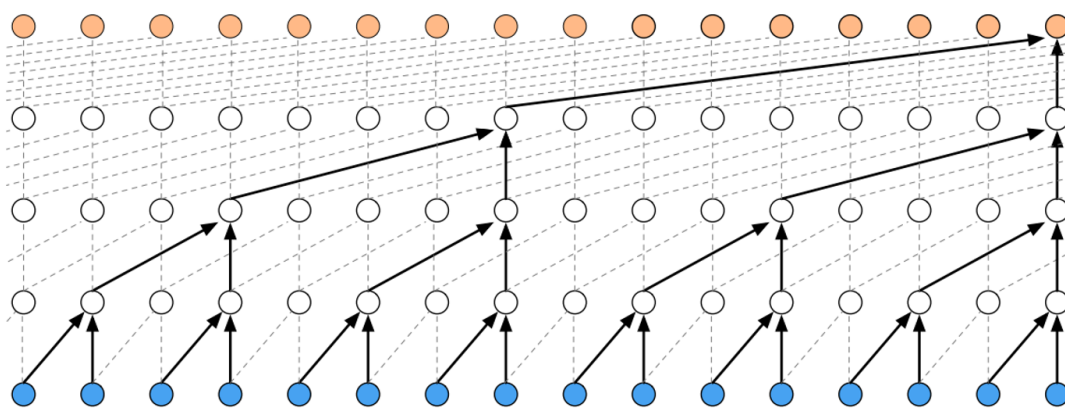


图 2.8 加入空洞卷积后的因果卷积神经网络^[56]

通过空洞卷积扩展后，模型能够更有效地处理和理解图像或序列数据中的长范围信息。空洞卷积通过引入空洞（即在卷积核的元素之间添加空格）来增加卷积核的感受野，而不增加参数数量或计算负担。具体而言，对于输入为 \mathbf{x} ，卷积核大小为 k 的空洞卷积公式如 (2.13) 所示。

$$F(s) = (\mathbf{x} *_d f)(s) = \sum_i^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (2.13)$$

这里， s 代表时间序列的输入数据， $\mathbf{x} *_d f$ 用以指代空洞卷积过程，而 d 是空洞率，表示在执行卷积时元素间的间隔， $s - d \cdot i$ 用于确定历史信息的具体位置。

通过引入空洞机制，增强了模型在处理需要长期记忆的时间序列预测任务中的建模能力，使其不仅能够捕捉到广泛的上下文信息，还能确保信息流的方向性，从而提高了时间序列预测的精度和效率。

2.4 图神经网络

近些年来，GNN 成为了一个热门的研究主题。GNN 专注于处理图形状的非欧式空间数据，与传统的卷积神经网络不同，对于处理欧几里得空间中的网状数据，后者展现了更好的适用性。在现实世界中，众多复杂的网络结构，例如

物流网络、铁路网络、公路网络和信息网络等，都能以图的形式进行抽象描述。基于图中节点和边的种类数目，图神经网络可以被分为同构图网络与异构图网络。其中，同构图网络指的是那些图仅涵盖单一种类的节点与边，典型例子有 DCRNN^[57]、GCN^[58]、GIN^[59]、GAT^[60]。异构图网络则指节点和边的类型数量之和大于二的图，代表性的异构网络有 R-GCN^[61]、HIN2Vec^[62]、HAN^[63]等。

首先引入图的一些概念和符号。将一个图表示为 $G = (V, E)$ ，其中 G 代表整个图， V 表示图的节点列表，大小为 n ； E 表示图的边集合；度矩阵 D 表示节点除了自身之外的连接数，如公式 (2.14)。

$$D \in \mathbb{R}^{n \times n}, D_{ii} = \sum_j W_{ij} \quad (2.14)$$

其中 $W \in \mathbb{R}^{n \times n}$ 表示邻接矩阵，用于描述图中各节点之间的连通状态。

GNN 通过对图进行可优化的转换来保持图对称性（排列不变性），包括节点、边以及全局上下文等所有属性。对称性指的是对图中的节点进行重新排序后，整个结果保持不变。核心思想是消息传递。在信息聚合过程中，节点更新其信息通过汇聚周围邻居的特征进行。如图 2.9 所示，GNN 里的每个节点都负责传递与接收信息，其表征受到相邻以及更远节点影响，直到系统稳定。因此，确立邻接关系和制定信息汇聚策略对于 GNN 的有效学习至关重要。

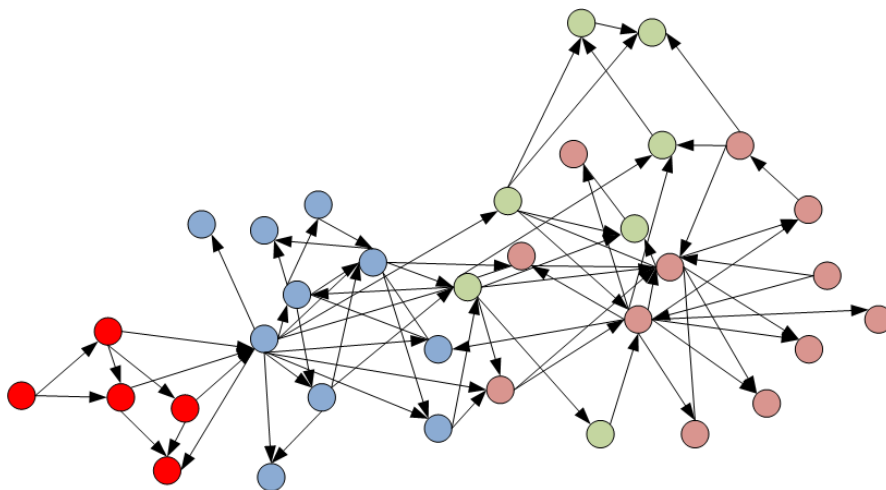


图 2.9 图神经网络的信息聚合

2.4.1 图卷积神经网络

图卷积神经网络（Graph Convolutional Networks, GCN）首次于 2017 年国际学习表征会议中提出。神经网络在传统任务中取得了显著的成就，例如卷积神经网络在图像领域（如图像识别）和循环神经网络在序列数据（如语言处理）中的成功应用。然而，传统卷积神经网络主要处理规则的数据结构，如图像或语言，这些属于欧式空间的数据，具有规则结构和维度的概念。在现实世界中，我们常

见到各种非线性的数据结构，图或网络结构就是其中的典范，包括但不限于交通网络、生态系统网络、以及社交网络平台等。语言结构本身，以其内在的复杂树状组织形式，也可以视为图结构的一种表现。如图 2.10 所示，展示了 GCN 的基本结构。

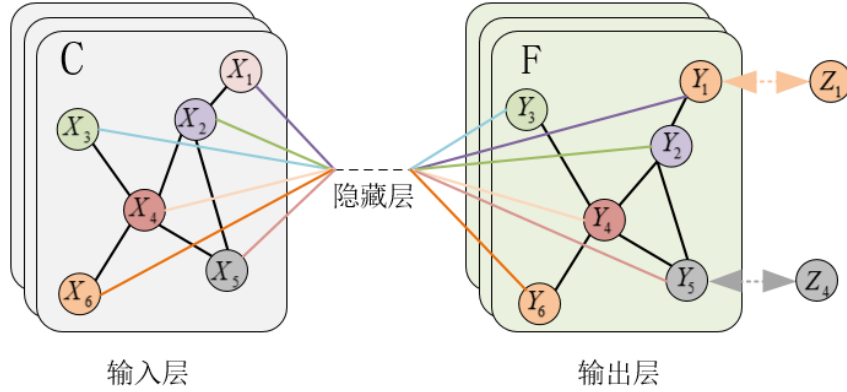


图 2.10 图卷积神经网络^[64]

由于图数据展现出结构不规则性，它们可被理解为存在于无限维空间中，因而缺乏常规的平移不变特性。在图中，各个节点周围的环境可能各不相同，这种独特的不规则性使得传统的卷积神经网络和循环神经网络在处理这类数据时面临挑战。尽管图数据丰富，但却缺乏相应的直接可用模型，于是 GCN 应运而生。GCN 是一种图神经网络模型，起源于谱图卷积的 ChebNet^[64]。这一模型旨在处理图形数据，通常接受图的拓扑结构通过邻接矩阵 A 和节点特征通过特征矩阵 X 表示的输入。它的前向传播机制可通过下述公式描述 (2.15) 所示：

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \quad (2.15)$$

其中， σ 为非线性激活函数， $\tilde{A} = A + I_N$ 为有自连接的邻接矩阵， I_N 表示单位矩阵，而 $D_{ii} = \sum \tilde{A}_{ij}$ 用于计算图中每个节点的度数。在这里， W 代表模型的权重参数， H 则是表示图中各节点特征的矩阵。针对特定的目标节点 i ，GCN 中的信息汇集机制可以被进一步简化为使用目标节点 i 的度与每个邻居节点的度相乘的倒数作为权重，表示为公式 (2.16)。

$$h^{(l+1)} = \sigma\left(\sum_{j \in N(i)} \frac{\tilde{A}_{ij}}{\sqrt{d_i} \sqrt{d_j}} h_j^{(l)} W^{(l)}\right) \quad (2.16)$$

其中， d_i 和 d_j 分别表示目标节点 i 与其邻居节点 j 的度， \tilde{A}_{ij} 的值表示节点间是否相连， W 为卷积核参数。节点 i 在更新后的特征是 $h_i^{(l+1)}$ ，其邻居节点 j 的特征由 $h_j^{(l)}$ 表示。

GCN 使用节点度的无参的邻域加权方法，以有效地捕捉图结构中的节点关系和特征分布，但并没有考虑节点本身的特征。其次，GCN 的设计本质上是基

于整图训练方式，这使得它在处理大规模图数据时面临着挑战。作为由谱图卷积理论推导而来的模型，GCN 对图的处理依赖于拉普拉斯矩阵的谱分解，这一过程要求拉普拉斯矩阵必须是对称的。这些限制凸显了 GCN 在处理大规模或有向图数据时的局限性。

2.5 注意力机制

在深度学习领域，注意力机制已经变成一个基础的理念，它被普遍应用于图像识别、语言理解等众多领域中。该机制可借鉴人类视觉系统的直观机理来理解，而人类视觉往往集中于图像中对决策有帮助的信息，而忽视那些无关紧要的内容。类似地，在阅读理解过程中，人们通常更关注帮助消除疑虑的文字信息，而忽略一些不相关的文字。因此，注意力机制在自然语言处理等领域具有重要应用，有助于计算机更有效地掌握人类交流的语境，即学习与遗忘上下文语境。

在预测任务中，注意力机制用于提高模型对输入数据的关注程度，使其更专注于对预测任务有贡献的部分。具体而言，注意力机制允许模型在每个时间步或每个输入元素上分配不同的权重，这些权重反映了模型对输入的关注程度。在自然语言处理或机器翻译等任务中，通过注意力机制，模型可以在生成每个输出元素时对输入序列的不同部分进行加权，使模型更灵活地关注输入序列中与当前输出相关的信息，从而提高性能。在计算机视觉领域，注意力机制被广泛应用于视觉信息的筛选与处理，它本质上是一个方法学，而非有着严格数学定义的概念。注意力机制的计算过程如图 2.11 所示。

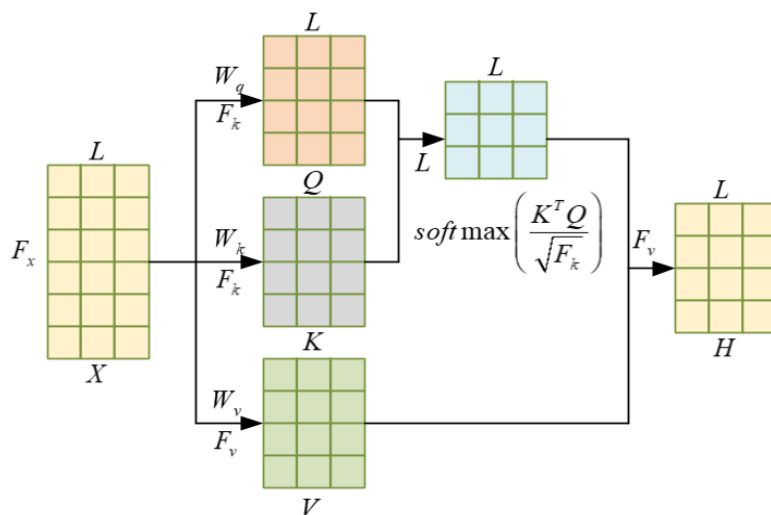


图 2.11 注意力机制计算过程

注意力机制包含三个核心向量：查询（Query）、键（Key）和值（Value）。通过对输入执行三个不同的线性变换得到这些向量。在计算过程中，给定查询键 Query 和一组键值对（Key, Value），首先计算 Query 和 Key 的向量点积以得到注

注意力分数，表示它们之间的关联强度。然后通过 softmax 函数计算归一化值，得到权重系数。最后，将权重系数与对应的 Value 向量进行加权求和，作为注意力机制的输出，其公式如 (2.17) 所示。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{K^T Q}{\sqrt{d}}\right)V \quad (2.17)$$

其中， Q 、 K 、 V 分别对应的 Query、Key、Value 矩阵。 d 是 Query 和 Key 矩阵中向量的维度。

注意力机制又可分为自注意力机制和通用注意力机制两种，自注意力机制，又被称为缩放点积注意力，是一种用于计算序列中元素之间相互作用的注意力机制。这种机制最早应用于 Transformer 模型，用于处理序列数据，例如在自然语言处理中处理语句。自注意力机制专注于数据内部的相互关系，而不依赖于外部数据信息。通用注意力机制通过神经网络实现了对输入的动态关注，从而显著提升了模型对输入信息的建模能力。

近年来，注意力机制在交通预测领域得到了广泛的应用，比如 ASTGCN^[65]、STANN^[66] 和 GMAN^[67]，这些方法在交通预测任务中取得了显著的成果，进一步证明了注意力机制在处理时空数据、提升模型性能方面的有效性。随着领域的不断发展，注意力机制仍然是研究的热点之一，未来可能会有更多创新性的方法涌现。

2.6 本章小结

本章详细探讨了与本研究密切相关的理论方法。首先，对时空数据的相关特性进行了介绍，包括时间相关性和空间相关性。接着，深入研究了循环神经网络、图神经网络和图卷积神经网络在预测领域的应用。最后，阐述了注意力机制的基本概念和在预测领域的研究。这些理论方法为后续第三章、第四章和第五章的研究提供了关键的基础，对于交通流预测起到了至关重要的支撑作用。

第3章 基于时空融合动态图卷积的交通流预测方法

本章研究城市路网中各节点的长期交通流量预测。首先对交通流量预测所面临的一些问题进行了全面描述，并对问题进行了定义。随后，详细介绍了本章提出基于时空融合动态图卷积方法总体框架。接着分别介绍基于交互式学习的策略、基于动态图卷积的特征提取、时空卷积块和基于概率稀疏注意力机制这四个不同的模块。最后，对本章方法分别做了模型对比实验、模型消融实验和预测值与真实值对比实验，以验证本章提出的方法的有效性。

3.1 引言

交通流预测是城市交通管理和规划中的一个关键任务，其目的是通过预测未来一段时间内的交通流量和模式，以便更有效地管理城市交通网络，减少拥堵，提高道路使用效率，从而改善城市居民的出行体验和生活质量。随着城市化进程的加速，城市路网的承载力未能与城市发展速度和车辆增长率同步提升，导致了許多城市在高峰时段经常出现严重的交通拥堵，此外，交通管理和规划不足也是造成拥堵的重要因素之一。因此，开发高效、准确的交通流预测模型变得尤为重要。

现代城市的交通流预测面临着多方面的挑战。如图 3.1(a) 示意了空间交通条件可以相互影响并随时间变化的情况。例如，交通事故可能影响周围道路的交通流，而在同一条道路上不同方向的交通流也可能不同。图 3.1(b) 展示了交通流的隐藏动态空间特性，如空间异质性、动态关联性和不确定性。空间异质性指的是具有不同交通模式和时空特性的区域，例如住宅区和工业区。在给定历史交通数据和道路网络拓扑的情况下，动态关联性指的是随时间推移道路网络节点之间的关联。不确定性涉及到交通流如何受事故、假期、恶劣天气等影响。

最新研究通过图神经网络模拟城市交通传感器之间的非线性空间关系，利用图卷积和时空卷积技术预测交通流量。虽然前期研究已经取得了重要成就，但这些研究主要侧重于从时间和空间相关性的匹配度来预测交通流量，而没有深入探讨产生这种时空相关性的根本原因。为了更全面地考虑时空关联性，引入注意力机制，可以使模型更加灵活地捕捉数据中的重要时空特征，从而提高交通流量预测的准确性。此外，还存在以下问题：

(1) 许多当前的研究还通过定义各种邻接矩阵来表示交通流的深层结构，以捕捉隐藏的动态空间特征，自适应邻接矩阵可以探索道路网络节点之间的隐藏关系，提高模型对交通流空间异质性的学习，但随着模型训练的停止，自适应邻

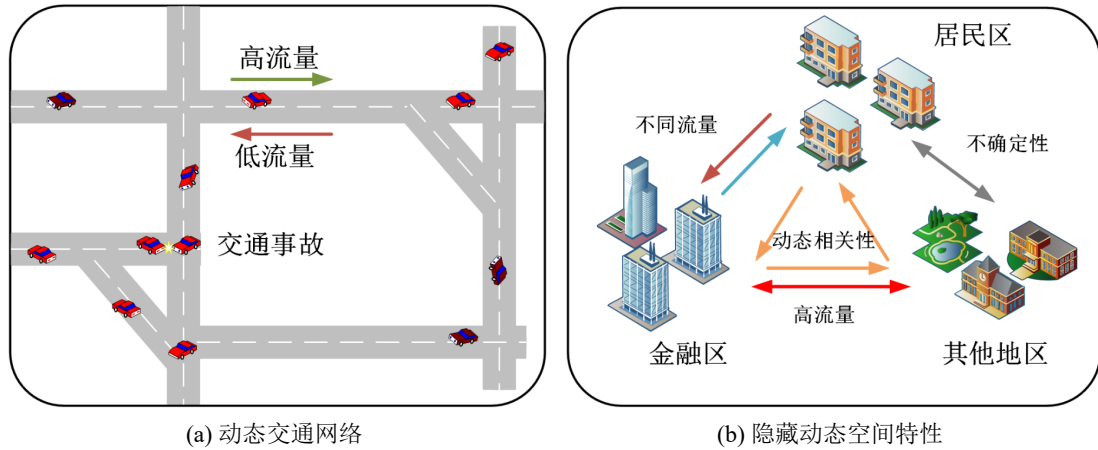


图 3.1 动态交通网络和隐藏动态空间特性

接矩阵无法学习图节点之间的动态关联，导致模型无法充分利用历史交通流。

(2) 在以往的研究中，许多采用图神经网络的策略通常通过静态邻接矩阵来表征节点间的联系，但这种做法未能充分考虑节点间的互动性及其动态的相关性如图 3.2。若空间和时间模块在提取动态时空特征时相互作用能力较弱，会影响交通流预测模型对时间序列周期性和变化趋势的感知，也无法充分捕捉交通流的动态时空特征。因此，当前方法在处理空间相关性时存在一些限制，需要更有效的机制来捕获节点间的交互和动态关系，以提高交通流预测模型对时空特征的准确性。

(3) 对于交通流的长期预测来说，涉及到对未来较远时段的预测，这是一个具有挑战性的问题。在长期预测中，时空相关性的影响可能更加复杂，因为存在更多的不确定性和外部因素的干扰。解决这个问题需要结合多种数据源和建模方法，以提高长期交通流预测的准确性和可靠性。

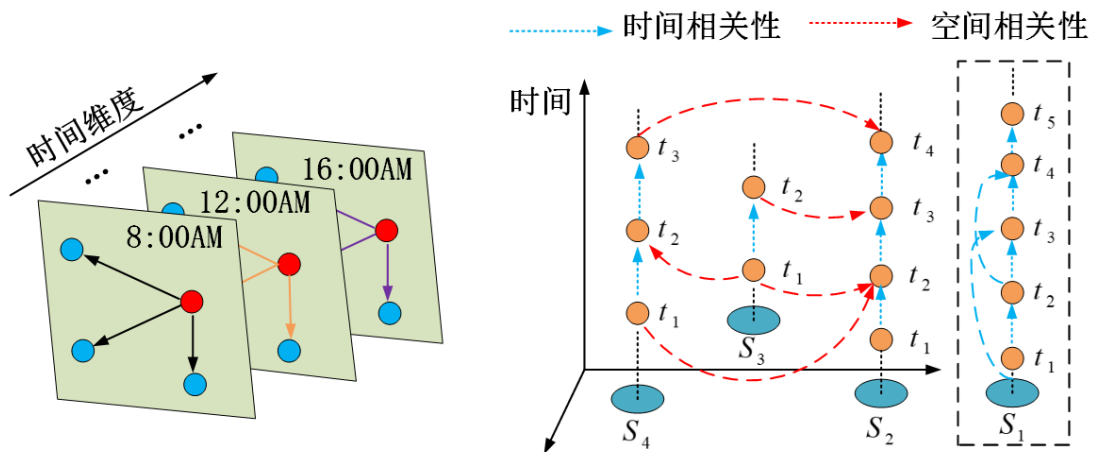


图 3.2 交通流的时空相关性

为解决上述研究问题，提出了一种基于时空融合动态图卷积的交通流预测方

法 (IDG-PSAtt), 实现了长期交通流量的预测。采用了动态图卷积网络 (DGCN), 可以充分利用先验知识生成动态图来捕捉交通流的隐藏空间特征。将 DGCN 嵌入到交互式学习结构中, 形成交互式动态图卷积网络 (IL-DGCN), 该网络分析交通流的周期性, 然后通过间隔划分序列, 通过划分后的子序列之间的交互学习捕捉深层次的动态时空特征。IDG-PSAtt 模型通过交互学习策略将多个 IL-DGCN 模块相结合, 充分提取交通流的动态时空特征。此外, 引入了自适应邻接矩阵和动态邻接矩阵, 以进一步探索节点之间随时间变化的动态关联。最后, 通过时空卷积块 (ST-Conv Block) 捕捉同一位置的复杂时间依赖性和邻近位置同一时间步的交通流动态空间相关性, 利用概率稀疏自注意力机制 (ProbSSAtt) 融合动态时空特征, 建模长程时空相关性, 降低模型的计算复杂性。

3.2 相关定义与数据预处理

3.2.1 相关定义

在下文中, 对本章主要使用的符号以及问题进行定义, 并且这些定义参考了文献^[68]。

定义 3.1 (交通图): $G = (V, E, A)$, 其中 V 是节点集合, $|V| = N$, E 表示边集合, 其权重表示节点之间的距离; $A \in \mathbb{R}^{N \times N}$ 表示由图 G 生成的初始邻接矩阵, 如果 i 和 j 之间有边, 则 $a_{i,j}$ 为 1, 否则为 0。交通流预测的任务旨在基于给定的历史信息预测未来的交通流。由原始交通网络获得的邻接矩阵 A 被用作先验知识, 通过历史时间序列 X 来预测未来的交通流 Y , 其中 X_t 表示在时刻 t 的图 G 的观察值, C 表示特征通道的数量, T 表示给定历史时间序列的长度, T' 表示预测未来交通流的长度。

定义 3.2 (问题定义): 交通流预测问题的映射关系可以表示如公式 (3.1)。

$$\left[X_G^{(t-T+1)}, X_G^{(t-T+2)}, \dots, X_G^{(t)} \right] \xrightarrow{f} \left[X_G^{(t+1)}, X_G^{(t+2)}, \dots, X_G^{(t+T)} \right] \quad (3.1)$$

3.2.2 数据预处理

这一部分将介绍了数据的预处理过程。

(1) 输入数据是历史交通流量 $X \in \mathbb{R}^{C \times N \times T}$, 其中 T 代表时间步数, N 为节点个数, C 为特征数量。本章按照时间先后的顺序排列这些交通流量, 构建了一个连续的时间序列数据集, 通过对这些时间序列数据的深入分析, 可以更准确地预测未来的交通流量变化趋势。

接下来, 鉴于不同城市区域间交通流量的规模差异可能使得模型倾向于调整那些预测误差较大区域的参数, 为避免这一潜在偏差, 本章采用了数据标准化方法。这一处理步骤通过缩放各个特征值来实现, 同时确保保持数据原有的分布

特性。标准化的具体实现如公式 (3.2) 所示。

$$X = \frac{X - X_{mean}}{X_{std}} \quad (3.2)$$

其中, $X \in \mathbb{R}^{T \times N \times C}$ 为交通流量数据, 训练集数据的均值表示为 X_{mean} , 标准差表示为 X_{std} 。标准化的目的就是消除不同特征的量纲差异, 有助于加速模型的收敛。

(2) 由于数据在未进行归一化之前处理起来过于繁琐, 所以在将数据输入预测模型之前, 需要对数据进行了最小-最大归一化, 将数据限制在 $[0, 1]$ 范围内。归一化公式为 (3.3)。

$$X_i^{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (3.3)$$

其中, X_i^{norm} 表示第 i 个原始数据的归一化后的数据, x_{min} 和 x_{max} 分别表示原始数据的最小值和最大值。

3.3 基于时空融合动态图卷积方法

如图 3.3 所示, 本章提出的 IDG-PSAtt 预测方法包括四个部分, 分别是交互式学习模块、动态图卷积模块、时空卷积模块及概率稀疏自注意力模块组成。首先将原始数据馈送至 SC 卷积以得到数据的高维空间表示进而捕获更深层次的依赖关系。模型的输入通过历史时间序列 $X_G^{(t-T+1)}, X_G^{(t-T+2)}, \dots, X_G^{(t)}$, 在动态图卷积的基础上以交互学习策略对开始卷积层提取到的特征进行处理。交互式动态图卷积的输入为 $X \in \mathbb{R}^{C \times N \times T}$, X 通过交错采样之后, 得到两个子序列, 分别为奇序列 $X_{odd} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$, 偶序列 $X_{even} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$ 。第一次交互式学习的输出为 X'_{odd}, X'_{even} 。 X'_{odd} 和 X'_{even} 通过进一步的交互式学习得到最终输出序列 $X_{odd_out}, X_{even_out}$ 。

再通过串联融合, 所有的输出子序列按时间索引顺序重组, 然后馈送至扩散图卷积和时空卷积以提取交通流的全局动态时空特征, 每个时空卷积块将前一个时空注意力块的输出作为输入 $X_N^{(l)} \in \mathbb{R}^{C^{(l)} \times N \times T_h}$ 。最后, 将捕获到的动态时空特征馈送至概率稀疏自注意力 $A(Q, K, V)$ 和多层感知机输出预测序列 $X_G^{(t+1)}, X_G^{(t+2)}, \dots, X_G^{(t+T)}$ 。

3.3.1 交互式学习模块

交互式学习模块由 CNN 和 GCN 实现, 具有处理非线性数据的能力。相较于基于 CNN 和 TCN 的方法, 它更有效地学习交通流的时空依赖关系, 并更充分地捕捉复杂的时序特征和动态的空间特征。鉴于交通流具有周期性、趋势性

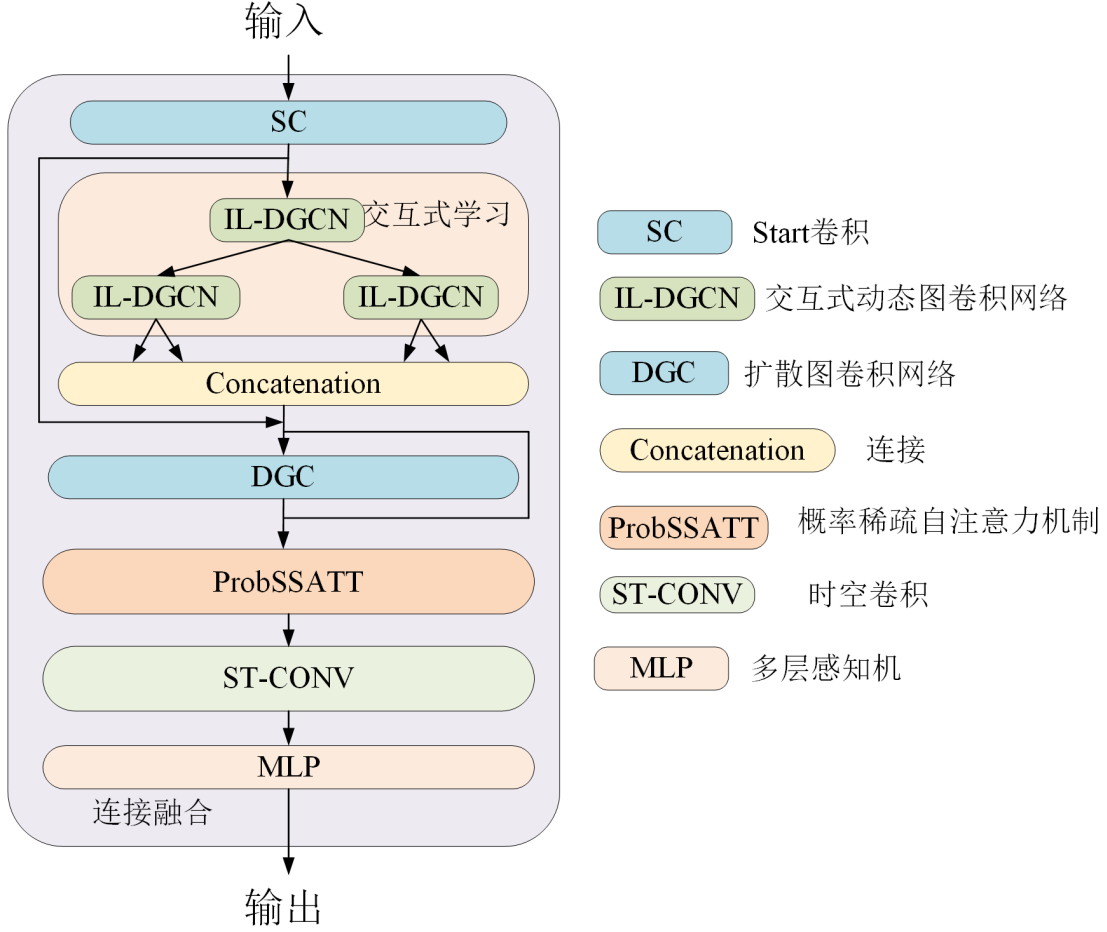


图 3.3 IDG-PSAtt 框架图

和相似性，采用交错采样方法生成的子序列仍然保留了原始序列的大部分信息。因此，本文利用交错采样对原始数据进行多分辨率分析，从而扩展感知领域。

整个交互式学习框架由三个相同的 IL-DGCN 组成，其中的核心部分是 IL-DGCN 模块。如图 3.4，两个子序列通过交互学习分别获取它们的动态时空特征。每个子序列通过卷积进行特征预处理，以扩展感受域。同时，这两个子序列在 DGCN 中共享参数权重，相互捕捉动态时空特征。

$X \in \mathbb{R}^{C \times N \times T}$ 表示 IL-DGCN 的输入，经过交错采样后获得两个子序列，分别是奇数序列 $X_{\text{odd}} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$ 和偶数序列 $X_{\text{even}} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$ 同时，IL-DGCN 中的 Conv1、Conv2、Conv4 和 Conv4 分别表示一维卷积操作。IL-DGCN 第一次交互学习的输出为 $X'_{\text{odd}} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$ 和 $X'_{\text{even}} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$ 。通过额外的交互学习， X'_{odd} 和 X'_{even} 得到最终的输出序列 $X'_{\text{odd_out}} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$ 和 $X'_{\text{even_out}} \in \mathbb{R}^{C \times N \times \frac{T}{2}}$

交互式动态图卷积中的具体操作如公式 (3.4)，(3.5)，(3.6)，(3.7)，(3.8)。

$$X_{\text{even}}, X_{\text{odd}} = \text{Con}(\text{Split}(X)) \quad (3.4)$$

$$X'_{\text{odd}} = \sigma(DG(\text{Con1}(X_{\text{even}}))) \odot X_{\text{odd}} \quad (3.5)$$

$$X'_{\text{even}} = \sigma(DG(\text{Con2}(X_{\text{odd}}))) \odot X_{\text{even}} \quad (3.6)$$

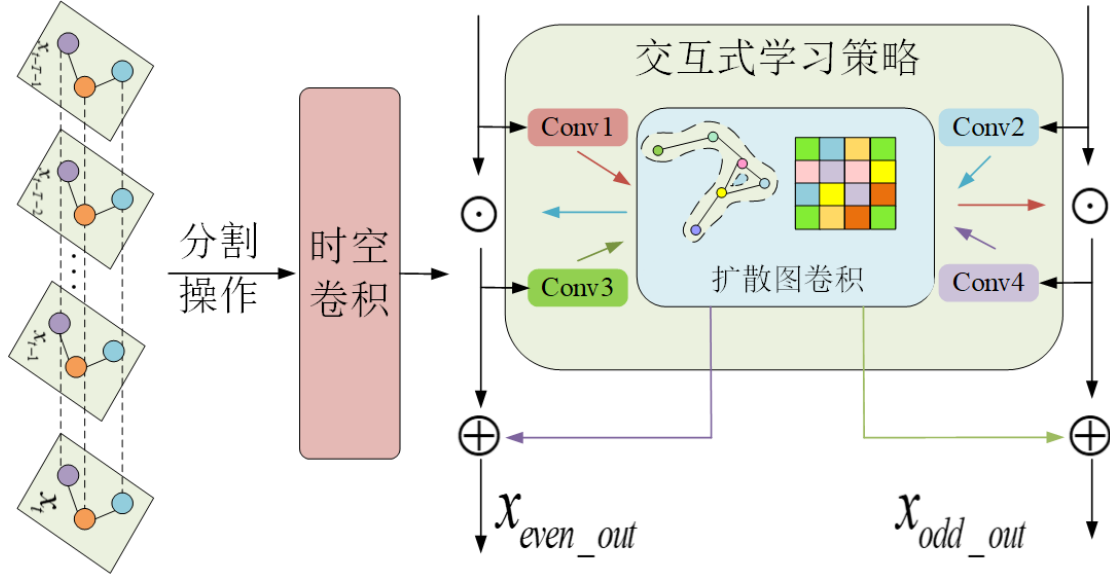


图 3.4 交互式学习模块

$$X_{odd_out} = X'_{odd} + \sigma\left(DG\left(Con3(X'_{even})\right)\right) \quad (3.7)$$

$$X_{even_out} = X'_{even} + \sigma\left(DG\left(Con4(X'_{odd})\right)\right) \quad (3.8)$$

其中，符号 \odot 表示 Hadamard 乘积， σ 表示激活函数。

3.3.2 动态图卷积

动态图卷积 (DGCN) 主要由扩散图卷积网络和图生成模块组成，以更好地学习深度动态空间特征，提高 IDG-PSAtt 方法在捕捉空间异质性方面的性能。其结构如图 3.5 所示。

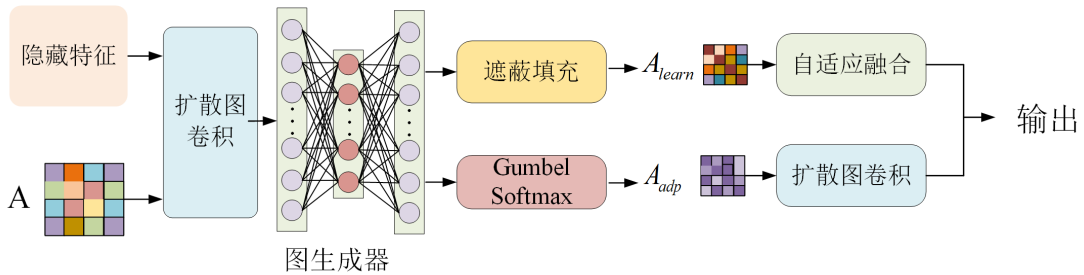


图 3.5 动态图卷积模块

DGCN 将隐藏特征 $H \in \mathbb{R}^{C \times N \times T}$ 和预定义的初始邻接矩阵 $A \in \mathbb{R}^{N \times N}$ 作为输入传递给扩散图卷积网络，然后输入到图生成器和 MLP 层，生成包含时空信息的离散矩阵 $A' \in \mathbb{R}^{N \times N}$ 。 A' 的表示如公式 (3.9)。

$$A' = \sigma\left(MLP\left(G(H, A)\right)\right) \quad (3.9)$$

其中， G 表示扩散卷积和图生成操作，MLP 表示多层感知机。

该模型使用 Gumbel 重参数化，因为需要确保在训练过程中采样过程可行，其公式 (3.10)。

$$A_{learn} = \text{GumbelSoftmax}(A') = \sigma((\log(A') - \log(-\log(g)))/\tau) \quad (3.10)$$

其中， $g \sim \text{Gumbel}(0,1)$ 表示一个随机变量， τ 是 softmax，其值为 0.5， A_{learn} 表示由图生成器生成的邻接矩阵，可以模拟节点之间的动态依赖关系。

与此同时，该模型构建了一个自适应邻接矩阵 $A_{apt} \in \mathbb{R}^{N \times N}$ ，表示如公式 (3.11)。

$$A_{apt} = \sigma(\text{Relu}(E_1 E_2^T)) \quad (3.11)$$

其中， $E_1 \in \mathbb{R}^{N \times c}$ 和 $E_2^T \in \mathbb{R}^{N \times c}$ 表示可学习的参数， A_{apt} 的初始值是基于原始图数据的预定义邻接矩阵 $A \in \mathbb{R}^{N \times N}$ 。

在该模型中，通过自适应融合模块融合 A_{learn} 和 A_{apt} ，提取交通道路中的隐藏动态时空关联性，然后将得到的动态邻接矩阵 $A_{dyn} \in \mathbb{R}^{N \times N}$ 输入到扩散图卷积网络中。该融合模块的具体操作如公式 (3.12)。

$$A_{dyn} = \alpha A_{apt} + (1 - \alpha) A_{learn} \quad (3.12)$$

其中， α 表示可学习的自适应参数。

动态图卷积中使用扩散图卷积、融合图卷积，并统一定义扩散图卷积的输入为 $X_{in} \in \mathbb{R}^{C \times N \times T}$ 。

扩散图卷积的定义如公式 (3.13)。

$$G(X_{in}, A_{apt}) = \sum_{k=0}^K A_{apt}^k X_{in} W \quad (3.13)$$

其中， k 是扩散步长， K 描述了最大扩散步数， W 表示参数矩阵。

融合图卷积的输入邻接矩阵，在融合图卷积模块中表示为符号 A_{dyn} ，具体表示如公式 (3.14)。

$$G(X_{in}, A_{dyn}) = \sum_{k=0}^K A_{dyn}^k X_{in} W \quad (3.14)$$

IDG-PSAtt 模型通过在连接模块中按时间索引顺序重新组合从交互学习结构中提取的动态时空特征，将其输入到扩散图卷积中，实现对整个时间序列特征的捕捉和校正。

在 DGCN 中同时使用了预定义的初始邻接矩阵 $A \in \mathbb{R}^{N \times N}$ 和通过交互学习结构获得的动态邻接矩阵 $A_{dyn} \in \mathbb{R}^{N \times N}$ 。对于初始邻接矩阵 A ，本文采用有向图表示，并定义前向传递矩阵 $P_f = A/\text{rowsum}(A)$ 和后向传递矩阵 $P_b =$

$A^T/\text{rowsum}(A^T)$ 。在这种情况下，连接融合模块中的扩散图卷积的表示如下。

$$G\left(X_{in}, A, A_{dyn}\right) = \sum_{k=0}^K \left(A_f^k X_{in} W_1 + A_b^k X_{in} W_2 + A_{dyn}^k X_{in} W_3 \right) \quad (3.15)$$

DGCN 模块不仅能够通过探索交通网络中节点之间的不可见依赖关系来提取深层次的隐藏空间特征，还能够基于对输入交通流时间序列的模拟生成数据之间的动态相关性。此外，通过将 DGCN 嵌入到交互式学习框架中，可以充分利用 DGCN 捕获的动态空间信息，在训练过程中增强对交通流复杂时空依赖关系的捕捉。

3.3.3 时空卷积模块

在交通道路网络中，每个传感器检测到的数据表现出一定的周期性。例如，在工作日的早晚高峰阶段，交通流量显著增加，车速普遍较低。交通流的隐藏空间特征与各传感器间的距离相关，这一空间属性不会受到时间因素的影响。

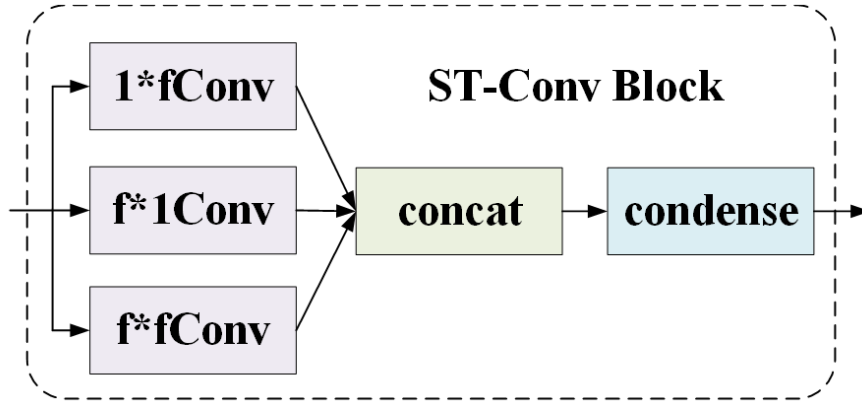


图 3.6 时空卷积模块

为了更好的获取交通流量的时空属性，设计了一个具有三个核心组件的时空卷积模块，这些核心分别针对时间域、空间域和时空域的不同维度进行分析，如图 3.6 所示。这些核旨在捕获从扩散图卷积模块中提取的时空特征，并在交通流拓扑图结构中考虑多个节点特征对单个节点特征的影响。具体而言，时域核用于捕捉相同位置在不同时间的交通流依赖关系，而空域核则用于捕捉相邻位置在同一时间步的空间关系上。此前的时空注意力块产生的输出成为随后每个时空卷积块的输入，即 $X_N^{(l)} \in \mathbb{R}^{C^{(l)} \times |V_N| \times T_h}$ 。通过公式 (3.16) 和 (3.17) 可以计算出输出 $X_N^{(l+1)}$ 。

$$H = \sigma \left[\omega_{st}^{[l+1]} * X_N^{(l)}; \omega_t^{[l+1]} * X_N^{(l)}; \omega_s^{[l+1]} * X_N^{(l)}; \right] \quad (3.16)$$

$$X_N^{(l+1)} = \sigma \left(\omega_o^{[l+1]} * H \right) \quad (3.17)$$

其中， $\omega_t^{(l+1)}$ 表示 $f \times 1$ 的时域核， $\omega_s^{(l+1)}$ 表示 $1 \times f$ 的空域核， $\omega_{st}^{(l+1)}$ 表示 $f \times f$

的时空核。 $*$ 表示卷积操作。最终，三个卷积核的输出被连接在一起，然后使用 1×1 卷积 $\omega_o^{(l+1)}$ 进行特征压缩和通道数量限制。

3.3.4 稀疏自注意力机制模块

典型的自注意力机制的输入形式为 (Q, K, V) ，点积计算如公式(3.18)。

$$A(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (3.18)$$

其中， $Q \in \mathbb{R}^{L_Q \times d}$ ， $K \in \mathbb{R}^{L_K \times d}$ ， $V \in \mathbb{R}^{L_V \times d}$ ， d 分别表示输入的查询(query)、键(key)、值(value)和维度。第 i 个查询的注意力因子 $A(q_i, K, V)$ 计算如公式(3.19)。

$$A(q_i, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = E_{p(k_j|q_i)}[v_j] \quad (3.19)$$

其中， q_i 、 k_i 和 v_i 分别是 Q 、 K 和 V 中的第 i 行。 $p(k_j|q_i) = \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)}$ ，而 $k(q_i, k_l)$ 使用了不对称的指数核 $\exp\left(\frac{q_i k_l^T}{\sqrt{d}}\right)$ 。

自注意力机制计算点积 $p(k_j|q_i)$ 的空间复杂度是 $O(L_Q L_K)$ 。然而，在稀疏自注意力机制模块的计算中，查询和关键字的输入长度通常是相等的，即 $L_Q = L_K = L$ ，使得总的时间复杂度和空间复杂度为 $O(L \ln L)$ 。此外，稀疏自注意力机制模块如图3.7所示，通过在自注意力机制之上调整注意力系数，结合了概率稀疏性自注意力机制，使得对于每个查询，只有一些关键字对其很重要，即一些关键字的点积提供主要的注意力，其余的点积被忽略。这种方法可以间接地结合复杂的时空特征，以节省计算资源，而不影响模型的准确性。

ProbSSAtt使用 $M(q_i, K)$ 表示第 i 个查询的稀疏度，使用KL散度来衡量查询的稀疏度如公式(3.20)。

$$M(q_i, K) = \ln \sum_{j=1}^{L_K} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \quad (3.20)$$

其中，所有关键字的算术平均值是第二项，第一项是 q_i 对所有关键字的对数和指数。可以使用这个思想创建概率稀疏自注意力机制如公式(3.21)。

$$A(Q, K, V) = \sigma\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \quad (3.21)$$

常数采样因子 c 控制 $u = c \cdot \ln(L_Q)$ ，其中 \bar{Q} 表示与 q 维度相同的稀疏矩阵。 \bar{Q} 只包括在由 $M(q, K)$ 评估的稀疏性下的前 u 个查询。因此，概率系数自注意力机制的查询只具有 $O(\ln(L_Q))$ 复杂性。为防止主要信息丢失，本研究采用了多头概率稀疏自注意力技术。该机制可以提供各种稀疏查询-关键字对。

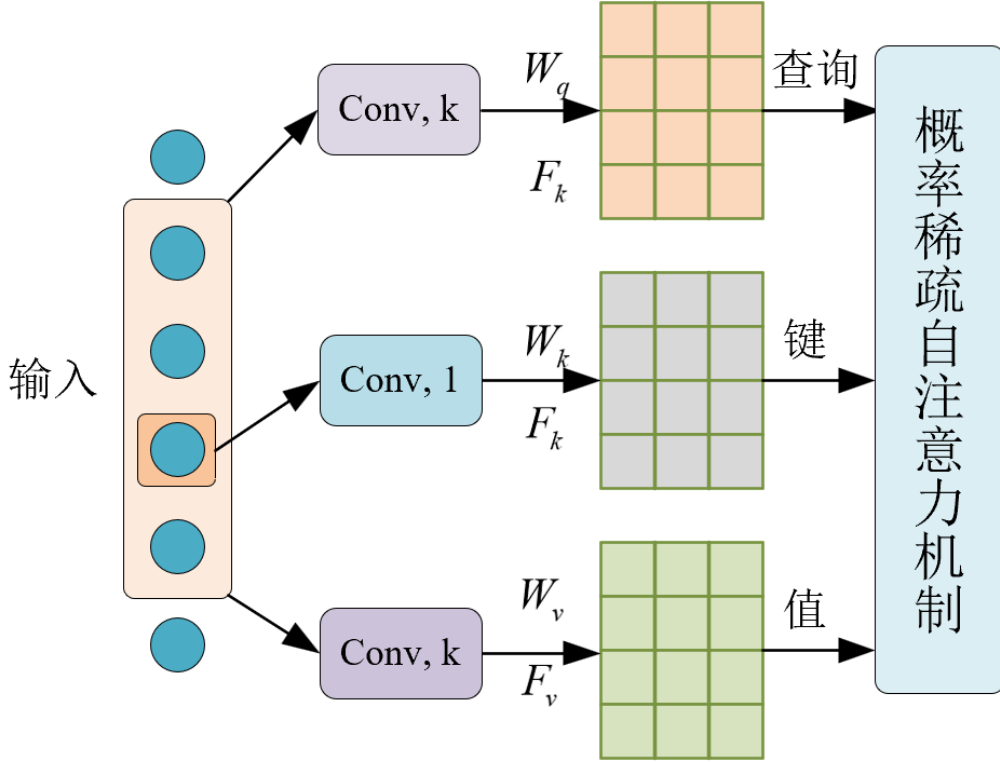


图 3.7 概率稀疏自注意力机制

3.3.5 损失函数

本章节中使用的损失函数是常用于回归任务的均方误差（MSE）如公式 (3.22)，其定义为预测值与真实值之差的平方和的平均值，利用此误差通过反向传播算法来调整并优化模型的参数。

$$L(W_\theta) = MSE(X_G^{(t+T)}, \hat{X}_G^{(t+T)}) = \frac{1}{N_l * N_l} \sum_{i=1}^N \sum_{j=1}^N (m_{i,j} - \hat{m}_{i,j})^2 \quad (3.22)$$

其中， W_θ 为整个模型的参数， $X_G^{(t+T)}, \hat{X}_G^{(t+T)} \in \mathbb{R}^{N_l \times N_l}$ 分别为 $t+T$ 时刻真实值和预测值。 N 为节点数量。

3.4 实验结果与分析

本节旨在通过实验来验证提出方法的效果。实验基于两个真实世界的交通流数据集进行，目的是评估所提方法的性能。具体实验内容包括：

实验一：超参数实验。

实验二：比较提出方法与现有基准方法在预测性能上的差异。

实验三：评估模型内部各组件对预测结果的影响。

实验四：分析模型在处理实际交通流量时的预测表现。

3.4.1 数据集描述

本研究验证了 IDG-PSAtt 模型在 METR-LA^[69] 和 PEMS-BAY^[70] 公共交通数据集上的预测性能。METR-LA 是由洛杉矶县沿高速公路放置的 207 个传感器在四个月内收集的交通速度统计数据的集合。PEMS-BAY 由在旧金山湾区交通动脉上放置的 325 个传感器在六个月内收集的交通数据组成。METR-LA 和 PEMS-BAY 都记录了检测站点、检测日期和数据类型。表 3.1 显示了实验数据集的具体信息。

表 3.1 实验数据集的描述

数据	METR-LA	PEMS-BAY
类型	连续	连续
城市	洛杉矶高速公路	旧金山
节点数量	207	325
边	1515	2369
时间步数	34272	52116

1. 数据集优势说明

大规模的交通流量数据：METR-LA 和 PEMS-BAY 数据集涵盖了广泛交通网络，包括高速公路、城市道路和主要交叉口。这使得研究人员能够获得大量的交通流量观测数据，从而进行交通流量预测和交通状态估计的研究。

高时空分辨率：数据集提供的交通流量观测数据具有高时空分辨率，可以以分钟级别的时间粒度和交通网络中的具体位置来表示。这种高时空分辨率使得研究人员能够更精确地捕捉交通流量的变化和交通拥堵的动态情况。

多源数据整合：METR-LA 和 PEMS-BAY 数据集整合了多个数据源，包括密度传感器、速度传感器等。通过整合这些多源数据，可以提供更全面的交通和环境信息，有助于深入了解交通流量与其他因素之间的关联性。

2. 数据分析

挑选 METR-LA 中的一部分数据集，如图 3.8 所示，A、B、C 分别代表商业区、工业区、住宅区，该图显示了三个区域不同的车流量曲线随时间的变化。其中 A、B 区域都有早晚高峰的特点，C 区域有晚高峰的特点，这说明预测流量与不同的区域有较大关联。而且数据呈现周期性，通过高峰和非高峰时段的差异，来揭示交通流量的规律和模式。这些时间特征可以帮助研究人员预测未来的交通流量并制定相应的交通管理策略。

3.4.2 实验环境与实验设置

1. 实验环境

IDG-PSAtt 模型使用的实验环境如表 3.2 所示。

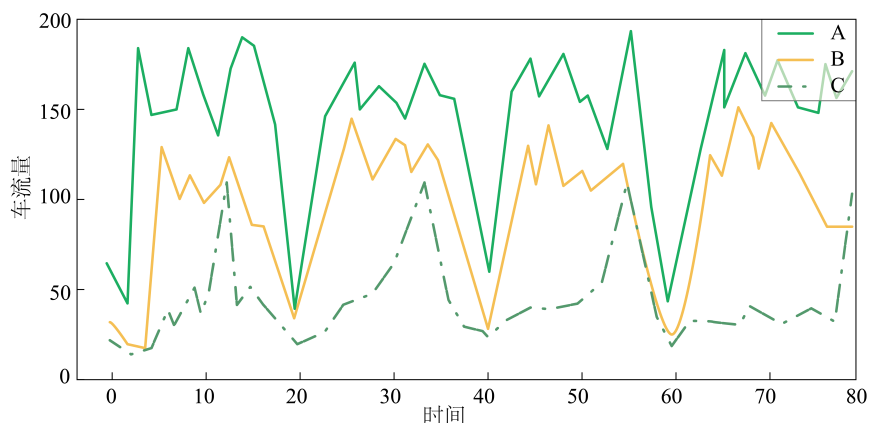


图 3.8 不同区域的流量曲线

表 3.2 实验环境

实验环境	详细信息
操作系统	windows 系统
集成开发环境	PyCharm
编程语言	python
CPU 配置	Intel(R) i7-13700kf
GPU 配置	RTX4090
机器记忆	64G

2. 评价指标

本章通过使用过去一小时内连续的 12 个时间步长的交通流量数据，来预测接下来一小时内连续的 12 个时间步长的交通流量。在实验中，数据集按照 7:2:1 的比例划分为训练集、测试集和验证集，用于预测未来 15 分钟、30 分钟和 60 分钟的交通流预测。同时，本文使用三个标准指标，平均绝对误差（MAE）、均方根误差（RMSE）和平均绝对百分比误差（MAPE）来评估所有方法的预测性能，其定义如公式 (3.23)，(3.24) 和 (3.25)：

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3.23)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.24)$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3.25)$$

其中 N 为观测次数， y_i 和 \hat{y}_i 分别表示实际和预测的交通速度。MAE、RMSE 和 MAPE 的预测值越小，说明 IDG-PSAtt 模型的预测效果越好。

3.4.3 超参数实验

在本研究采用了 Adam 优化器来调整神经网络的权重。实验中设定的学习率为 0.0001，这个值较小，有助于模型在训练过程中稳定地逼近最优解。此外，实验的隐藏层的单元数统一设置为 64，epoch 训练轮次选择的是 400。在这些基础上，本次实验均在 METR-LA 数据集上进行超参数实验。

图 3.9展示了两个不同性能指标的折线图，图 3.9(a) 中，RMSE 随着预测时间的增加误差在增大。隐藏层的单元大小为 128 时的性能最差，而大小为 64 时的性能最好。图 3.9(b) 也类似，这表明对于本研究，较小的模型复杂度足以捕捉到所需的模式，而更大的模型可能导致过拟合或不必要的复杂度。

图 3.10展示了两个不同性能指标的折线图，图 3.10(a) 中，RMSE 随着预测时间的增加误差在增大。epoch 为 400 时的性能最好，epoch 为 300 时性能最差。这表明当训练周期为 300 时，模型可能还没有充分学习到数据中的模式，导致预测性能不佳，而 epoch 为 500 时性能相较而言也较差，这表明随着训练周期的增加，模型开始过度拟合训练数据，其泛化能力会下降。图 3.9(b) 中也类似。

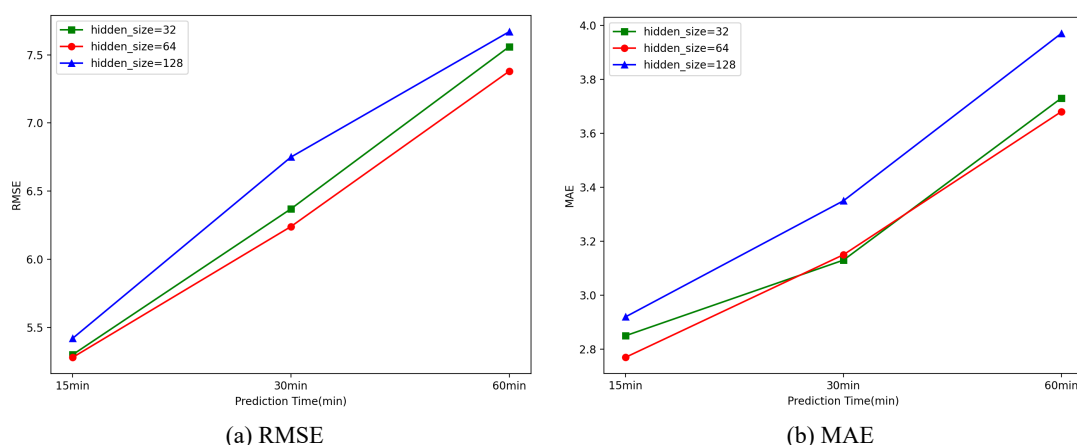


图 3.9 不同的隐藏单元数量对预测效果的影响

3.4.4 模型对比实验

1. 基线实验

为了深入评估 IDG-PSAtt 模型的性能，本文将其与几个基准模型进行了对比分析，包括当前在交通流预测任务的最优模型（Graph WaveNet）。

(1) HA^[5]: 将历史和当前交通流的平均值作为下一步的预测值。

(2) VAR^[10]: 是一种统计模型，用于描述多个时间序列变量之间相互依赖的动态关系。

(3) SVR^[14]: 在回归问题中使用支持向量机分类的扩展。

(4) FNN^[71]: 是一种基本的神经网络结构，数据处理过程中，数据仅单向流

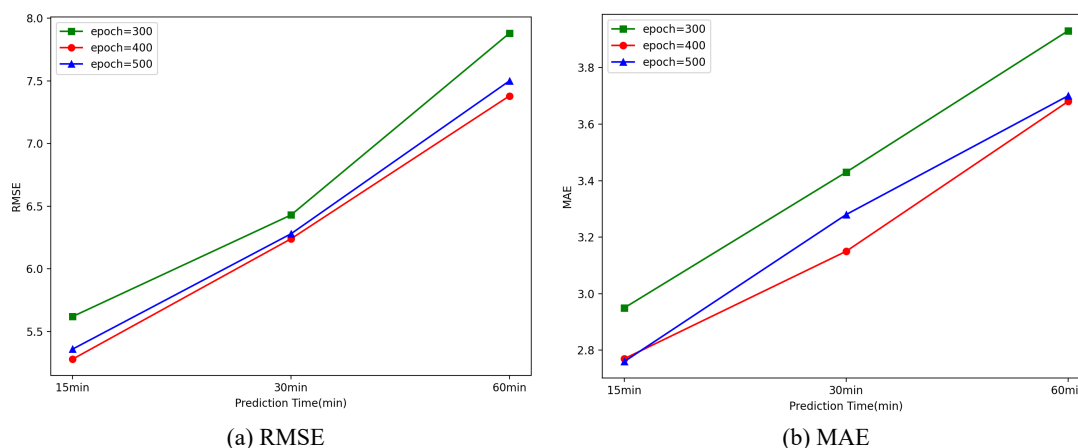


图 3.10 不同的 epoch 对预测效果的影响

动。自输入层向输出层传递，过程中可能会经过一个或数个隐藏层。

(5) ARIMA^[8]: 是时间序列分析中常见的一种工具，其核心由自回归阶数、差分次数和移动平均阶数这三个参数构成。

(6) FC-LSTM^[72]: 这是一个经典的循环神经网络，通过全连接神经网络学习时间序列并进行预测。

(7) WaveNet^[56]: 用于预测序列数据的卷积神经网络。

(8) Graph WaveNet^[38]: 由图卷积网络和门控时间卷积层构建而成，采用图卷积网络和因果卷积网络来分别挖掘时空数据的空间和时间依赖性。

(9) STGCN^[73]: 采用图卷积层和卷积序列层。

(10) ASTGCN^[65]: 采用注意机制捕捉时空动态相关性。

(11) STSGCN^[52]: 分别捕捉局部空间和时间相关性。

2. 模型对比实验与分析

本文将 IDG-PSAtt 模型与 11 个常见的基线模型进行了 15 分钟、30 分钟和 60 分钟的预测性能比较。在 METR-LA 和 PEMS-BAY 的两个数据集上，IDG-PSAtt 的评价指标均实现了最优的预测表现。

如表 3.3 中的实验结果表明，统计方法（HA、VAR、ARIMA）、传统机器学习方法（SVR 和 FC-LSTM）表现较差，因为这些模型只考虑时间依赖性，忽视了交通流的动态空间特征。基于 GCN 的模型能够更有效地处理非欧几里得交通数据，并更好地捕捉道路网络节点之间的隐藏关系，因此基于 GCN 的 STGCN 和 STSGCN 模型表现较好。虽然 STSGCN 能够同时捕捉空间和时间数据，但由于它只强调捕捉时间依赖性，并使用简单的滑动窗口来捕捉时间相关性，因此性能不佳。另外，由于注意机制捕捉了序列的时间依赖性，基于注意机制的模型 ASTGCN 也表现不错。Graph WaveNet 将 GCN 嵌入到 TCN 中，使其性能优于 ASTGCN，但 Graph WaveNet 没有引入自注意力机制以进一步捕捉隐藏的时

空特征。

表 3.3 IDG-PSAtt 模型与基线模型在 METR-LA 和 PEMS-BAY 上的比较

Data	Models	15min			30min			60min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA	HA	4.17	7.81	13.05%	4.17	7.81	13.05%	4.15	7.81	13.05%
	VAR	4.41	7.89	10.19%	5.41	9.11	12.84%	6.49	10.09	15.79%
	SVR	3.97	8.42	9.29%	4.99	10.85	12.10%	6.76	13.71	16.72%
	FNN	3.96	7.94	9.89%	4.19	8.17	12.91%	4.49	8.65	14.05%
	ARIMA	3.91	8.23	9.62%	5.16	10.46	12.71%	6.92	13.24	17.41%
	FC-LSTM	3.45	6.33	9.61%	3.78	7.22	10.91%	4.36	8.68	13.21%
	WaveNet	3.00	5.88	8.05%	3.60	7.27	10.26%	4.46	8.94	13.67%
	GWN	2.97	5.91	7.93%	3.60	7.28	10.25%	4.45	8.96	13.65%
	STGCN	2.87	5.73	7.61%	3.46	7.23	9.56%	4.58	9.41	12.71%
	ASTGCN	4.86	9.28	9.19%	5.39	10.68	10.09%	6.53	12.52	11.61%
	STSGCN	3.32	7.63	8.05%	4.16	9.76	10.30%	5.04	11.65	12.90%
	IDG-PSAtt	2.77	5.28	7.24%	3.15	6.24	8.73%	3.62	7.38	10.52%
PEMS-BAY	HA	2.87	5.60	6.81%	2.87	5.60	6.81%	2.87	5.60	6.81%
	VAR	1.75	3.15	3.61%	2.31	4.26	5.05%	2.94	5.45	6.55%
	SVR	1.84	3.60	3.81%	2.47	5.19	5.51%	3.24	7.05	8.05%
	FNN	2.21	4.41	5.20%	2.34	4.62	5.48%	2.48	4.99	5.86%
	ARIMA	1.61	3.35	3.53%	2.35	4.77	5.42%	3.39	6.52	8.35%
	FC-LSTM	2.04	4.18	4.81%	2.25	4.50	5.21%	2.38	4.95	5.71%
	WaveNet	1.40	3.00	2.90%	1.82	4.24	4.15%	2.34	5.42	5.86%
	GWN	1.40	3.00	2.88%	1.82	4.23	4.10%	2.34	5.42	5.79%
	STGCN	1.35	2.94	2.91%	1.80	4.28	4.14%	2.48	5.68	5.76%
	ASTGCN	1.55	3.09	3.26%	2.03	4.25	4.50%	2.59	5.45	6.59%
	STSGCN	1.43	3.04	3.05%	1.82	4.17	4.16%	2.25	5.22	5.39%
	IDG-PSAtt	0.96	1.72	1.96%	1.64	3.68	3.77%	1.91	4.36	4.53%

IDG-PSAtt 模型在 METR-LA 和 PEMS-Bay 的 15 分钟、30 分钟和 60 分钟预测中相较于最先进模型有显著的提升，这是因为 IDG-PSAtt 模型通过交互式学习结构、DGCN 以及 ST-Conv 充分捕捉了交通流的动态时空特征，并利用 ProbSSAtt 模块实现了有效的长程预测。例如，在 PEMS-Bay 的 30 分钟和 60 分钟预测中，IDG-PSAtt 模型在 MAE、RMSE 和 MAPE 方面分别超过最先进方法 9.8%、11.7%、9.2% 和 8.3%、11.5%、15.9%。验证了 IDG-PSAtt 模型能够充分捕捉交通流的动态时空特性。

IDG-PSAtt 模型将交互式学习策略与 ST Conv 和 ProbSSAtt 模块相结合，以有效地捕捉动态的时空相关性。因此，与基线模型相比，IDG-PSAtt 模型可以更好地捕捉交通流在每个时间段的动态时空相关性，并在 15 分钟、30 分钟和 60 分钟时实现了最佳的预测结果。研究结果表明，IDG-PSAtt 模型能够探索道路网络节点中的不可见动态关系，捕捉隐藏的空间相关性。随着预测周期的增加，预测难度增加，IDG-PSAtt 模型在长期预测中的预测效果仍然很好，进一步验证了 IDG-PSAtt 模型交互式学习策略的有效性。

3.4.5 模型消融实验

为了进一步调查本文提出的 IDG-PSAtt 模型各个模块的性能, 本文设计了 IDG-PSAtt 模型的八个变体, 以验证每个模块对 IDG-PSAtt 模型的影响, 这八种变体和 IDG-PSAtt 模型分别在 METR-LA 和 PEMS-BAY 数据集上进行了比较分析。上的 MAE、RMSE 和 MAPE 的均值方面进行比较, 实验结果如表3.4所示。

这八个模型变体与 IDG-PSAtt 模型之间的区别是:

- (1) GCN w/o: 在 IDG-PSAtt 模型基础上移除 GCN 网络。
- (2) DGCN w/o: 在 IDG-PSAtt 模型基础上移除 DGCN 网络。
- (3) Conv w/o: 在基于 IDG-PSAtt 建模的交互式学习结构中移除一维卷积模块。
- (4) Interaction w/o: 在 IDG-PSAtt 模型基础上移除交互式学习结构。
- (5) Apt Adj w/o: 在 IDG-PSAtt 模型基础上移除自适应邻接矩阵。
- (6) Learned Adj w/o: 在 IDG-PSAtt 模型基础上移除图生成器结构。
- (7) ProbSSAtt w/o: 在 IDG-PSAtt 模型基础上移除 ProbSSAtt 块模块。
- (8) ST-Conv Block w/o: 在 IDG-PSAtt 模型基础上移除时空卷积模块。

表 3.4 在 METR-LA 和 PEMS-BAY 数据集上对 IDG-PSAtt 变体的比较

Dataset	Model	MAE	RMSE	MAPE
METR-LA	w/o GCN	5.46	8.42	11.32%
	w/o DGCN	4.89	7.98	10.67%
	w/o Conv	4.37	7.47	10.13%
	w/o Interaction	3.92	7.23	9.89%
	w/o Apt Adj	3.71	6.89	9.56%
	w/o Learned Adj	3.44	6.72	9.24%
	w/o ProbSSAtt	3.63	6.53	9.03%
	w/o ST-Conv	3.54	6.65	9.12%
	IDG-PSAtt	3.12	6.14	8.62%
PEMS-BAY	w/o GCN	3.89	6.12	6.27%
	w/o DGCN	3.43	5.41	5.43%
	w/o Conv	2.92	4.87	4.98%
	w/o Interaction	2.73	4.66	4.66%
	w/o Apt Adj	2.47	4.17	4.18%
	w/o Learned Adj	2.03	3.89	4.09%
	w/o ProbSSAtt	2.20	4.03	4.37%
	w/o ST-Conv	2.28	3.96	4.23%
	IDG-PSAtt	1.57	3.47	3.59%

在 METR-LA 数据集上, 本文设计了八个 IDG-PSAtt 模型的变体, 通过移除不同的模块以验证每个模块对 IDG-PSAtt 模型的影响。表格显示了这些变体与 IDG-PSAtt 模型在 MAE、RMSE 和 MAPE 均值方面的比较结果。具体而言, 移除 GCN 网络导致 MAE 从 3.12 上升到 5.46, RMSE 从 6.14 上升到 8.42, MAPE 从 8.62% 上升到 11.32%。这表明 GCN 网络在模型性能中发挥着关键作用。

在 PEMS-BAY 数据集上的实验结果显示，对 IDG-PSAtt 模型进行不同模块的移除同样对预测性能产生影响。例如，去除交互学习结构导致 MAE 从 1.57 上升到 2.73，RMSE 从 3.47 上升到 4.66，MAPE 从 3.59% 上升到 4.66%。这说明交互学习结构在模型中的重要性，尤其是在处理 PEMS-BAY 数据集时。

总体而言，通过对比 IDG-PSAtt 模型及其变体在不同数据集上的性能，可以得出结论：IDG-PSAtt 模型中的 GCN、DGCN、交互学习结构以及 ProbSSAtt 块等模块都对模型的性能起到了关键作用。这些模块的移除往往导致性能指标的显著下降，证明了它们在捕获交通流动态时空特征方面的有效性。因此，IDG-PSAtt 模型的整体性能优越性得以验证。

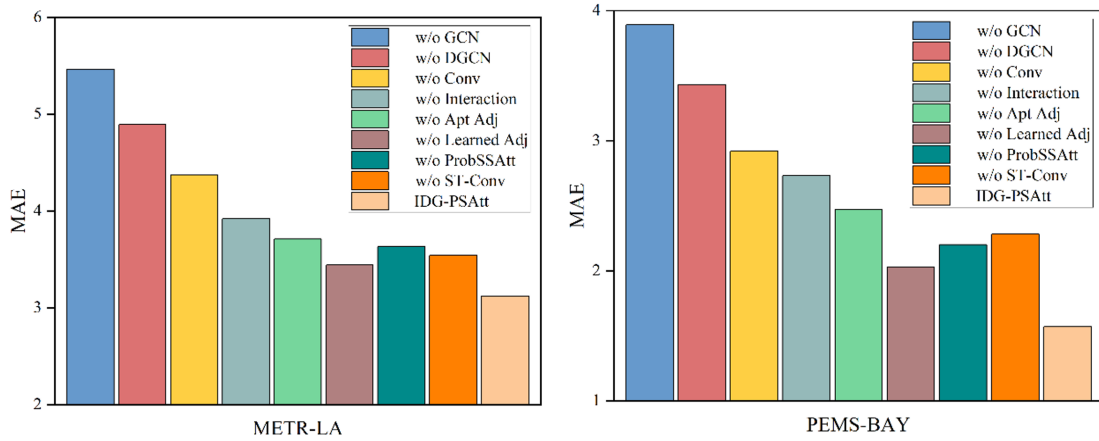


图 3.11 在两个数据集上的 MAE 指标比较

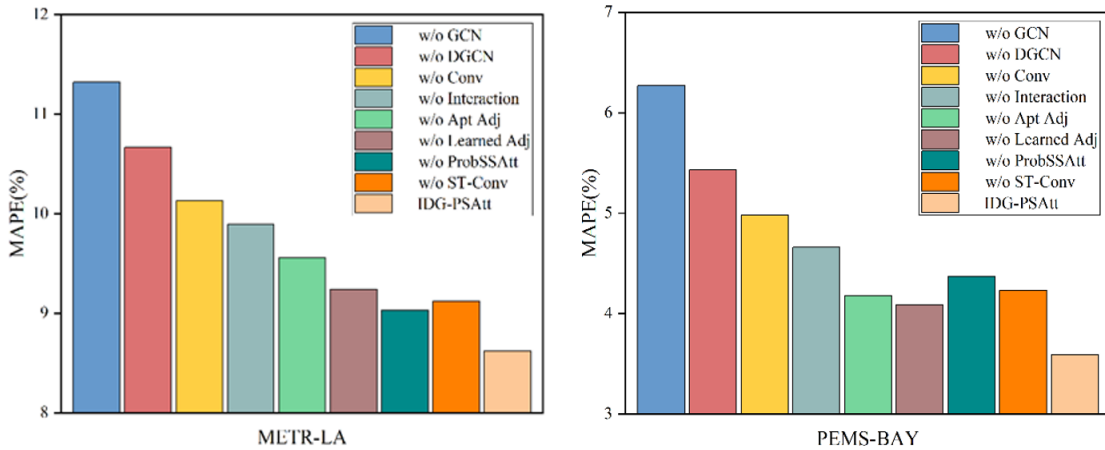


图 3.12 在两个数据集上的 MAPE 指标比较

ProbSSAtt 和 IL-DGCN 模块是提高模型性能的关键部分。交互学习框架的重要组成部分之一是使用一维卷积来扩展感受野，消融实验证明一维卷积能够极大地提升模型性能。在 DGCN 模块中定义的两个邻接矩阵被消融的同时，IDG-PSAtt 模型还对模型的自适应邻接矩阵的有效性进行了研究，如图3.13所示。动态邻接矩阵是通过将可学习邻接矩阵与自适应邻接矩阵相结合而创建的。动态邻接矩阵使得图卷积能够更准确地描述交通数据中的隐藏空间相关性，如图3.11、

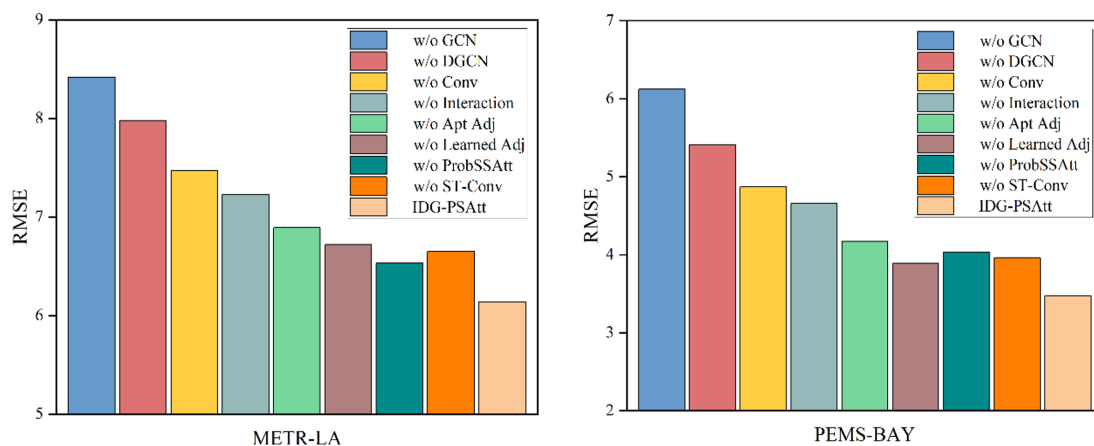


图 3.13 在两个数据集上的 RMSE 指标比较

图3.12和图3.13所示，展示了本文提出的三个重要结构，即交互学习、动态图卷积和稀疏自注意力机制的有效性。

3.4.6 预测值与真实值对比实验

为了更直观地观察预测效果，说明时空数据融合的有效性，本节在 METR-LA 和 PEMS-BAY 数据集上进行了预测结果可视化实验。如图 3.14和图 3.15所示，在两个数据集上分别选取了 69 号节点和 162 号节点，并分别预测它们未来 1750 个时间步和 1200 个时间步的交通流量。可以观察到，IDG-PSAtt 在高峰期的预测中表现出了和真实值较为吻合的曲线，这说明 IDG-PSAtt 有能力对突发事况进行有效的预测。

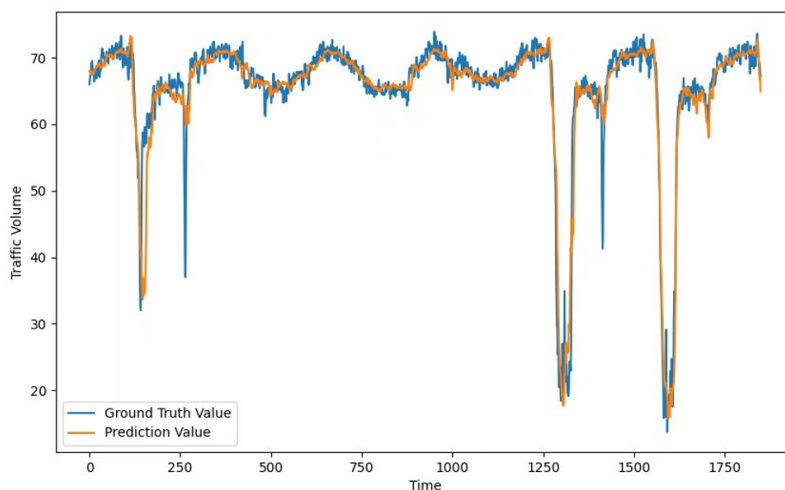


图 3.14 METR-LA 的 69 号节点

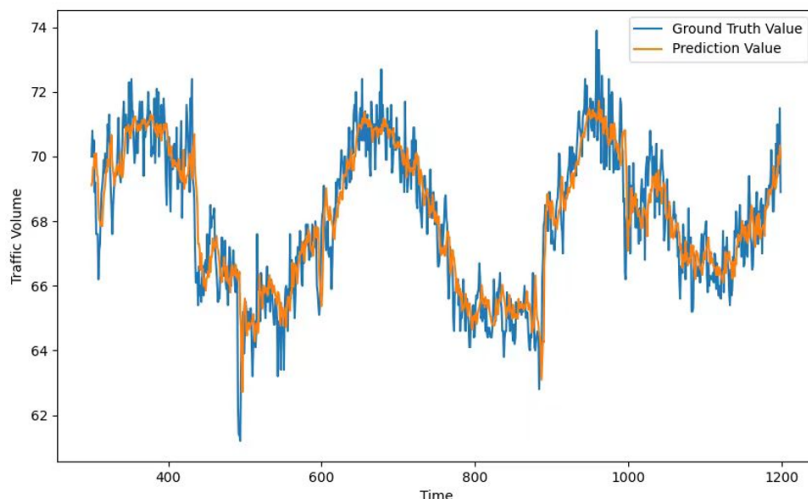


图 3.15 PEMS-BAY 的 162 号节点

3.4.7 空间依赖性的可视化分析

为了展示本章提出的动态图的细节，本小节通过可视化前 10 个节点的空间依赖性来阐明其效果。如图 3.16(a) 和图 3.16(b) 所示，IDG-PSAtt 不仅关注一定距离范围内的所有节点，而且还能够在不同的时间段动态地分配权重。这一特性使得 IDG-PSAtt 能够灵活地适应交通状况的动态变化，从而有效地识别不同时间段内的交通流量和拥堵情况。

相比之下，传统的方法通常采用预定义的空间依赖图来表示节点之间的关系，如图 3.16(c) 所示。这种静态的空间依赖图缺乏灵活性，无法适应交通状况的动态变化，从而限制了模型在预测动态交通流量方面的性能。

3.4.8 可视化分析

为了深入阐述本文提出的 IDG-PSAtt 模型，本章将其与 FNN、FC-LSTM、Graph WaveNet、STGCN、STSGCN 和 GMAN 等模型在 PEMS-BAY 和 METR-LA 两个数据集上进行了比较研究。通过观察图表（图 3.17、图 3.18 和图 3.19），本章发现在短期预测（5 分钟或 10 分钟）环节，STGCN 和 Graph WaveNet 展现出较强的性能。然而，随着预测时间的延长，FNN、FC-LSTM 和 STSGCN 的表现迅速下降，而 Graph WaveNet 和 GMAN 能够维持较好的性能。特别是在未来一小时的预测中，尽管 GMAN 是表现最佳的基准模型，IDG-PSAtt 模型仍然显示出更加卓越的预测能力。总体而言，IDG-PSAtt 在两个数据集上的多步预测性能超越了所有基线模型。这一结果证明了 IDG-PSAtt 在提取交通流动态时空特征方面的高效性。此外，随着预测时间的增加，该模型的预测误差增长较慢，尤其是在超过 15 分钟的长期预测中，IDG-PSAtt 的误差显著低于其他模型，突显其在长期预测方面的强大优势。

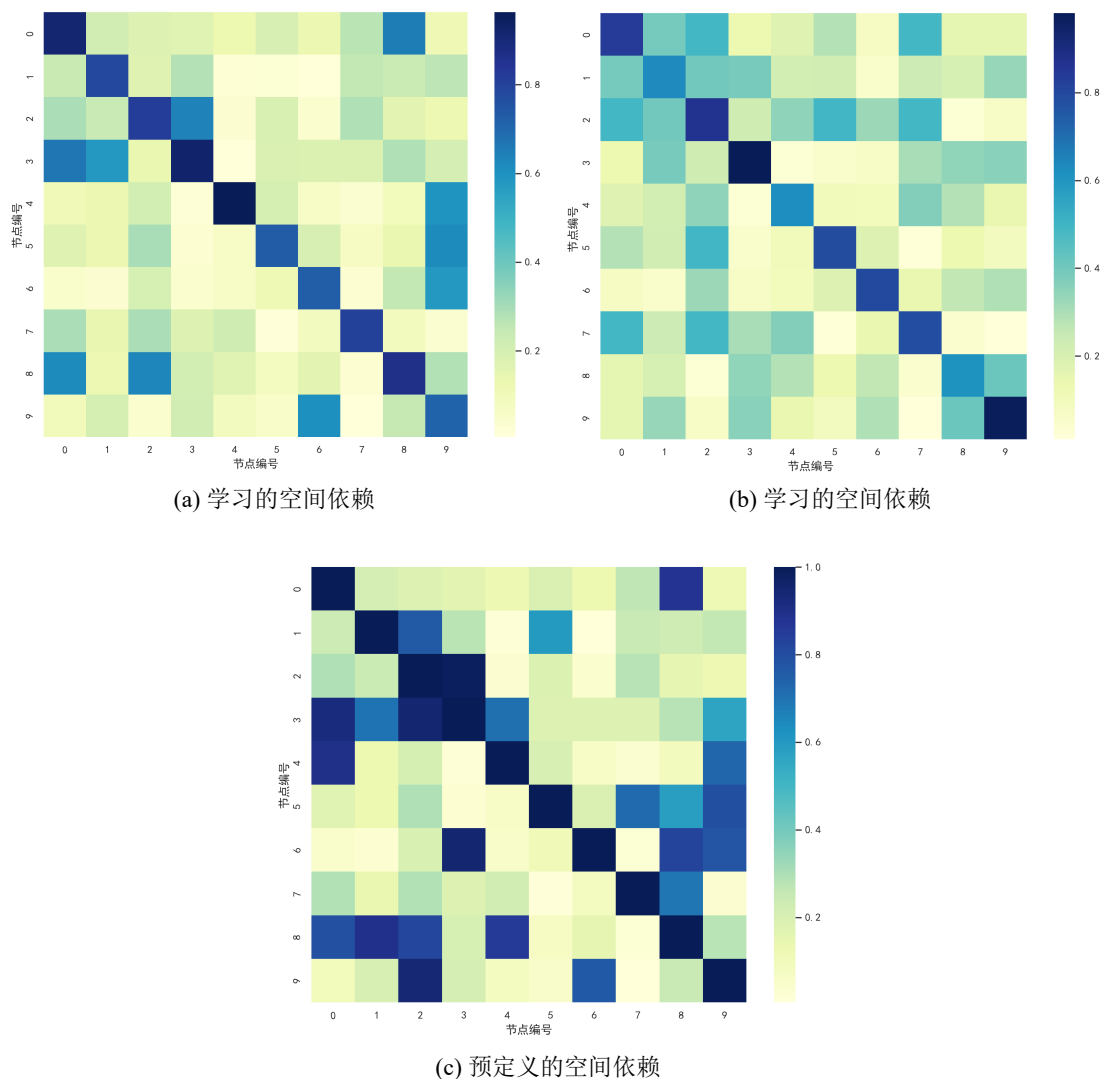


图 3.16 IDG-PSAtt 在 METR-LA 数据集上的空间依赖

以上研究表明，IDG-PSAtt 模型在长期预测取得了极为优秀的预测结果。IDG-PSAtt 模型准确预测交通拥堵，捕捉交通流的趋势，识别高峰交通流期的起始时间和结束时间，证明了该模型在交通流长期预测任务中的出色预测性能以及在处理交通数据方面的高效性能。

3.5 本章小结

本章针对在做交通流预测时没有充分得挖掘出路网交通节点和邻接矩阵中的远程时空特征和隐藏动态空间特性问题，本章提出了一种高效而准确的交通流长期预测模型 IDG-PSAtt。该模型不仅考虑了非线性交通流，还将交互式学习策略、动态图卷积、时空卷积模块和概率稀疏自注意力块结合起来，以充分捕捉交通流的动态时空特征。这解决了先前模型存在的非交互式、隐藏空间特性、未能充分捕捉时空特征且难以进行长期预测的问题。具体而言，IDG-PSAtt 模型通

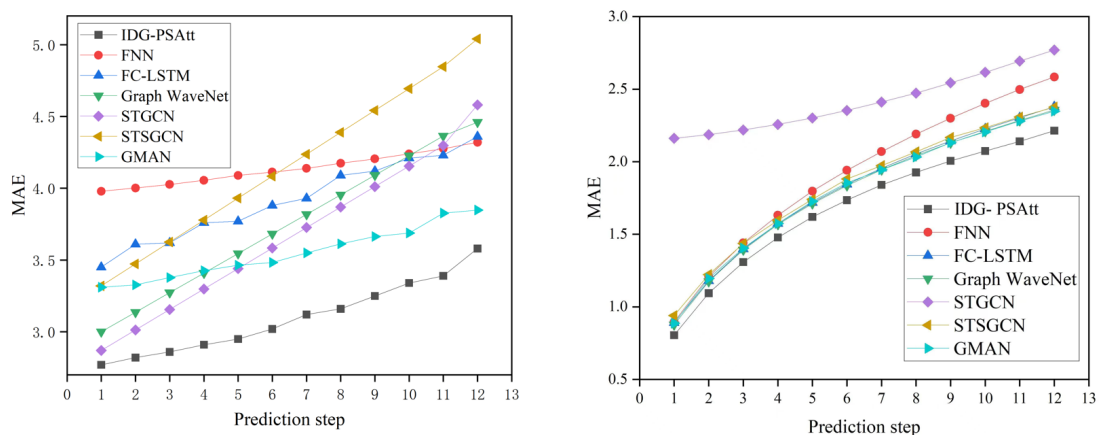


图 3.17 在两个数据集上的 MAE 曲线

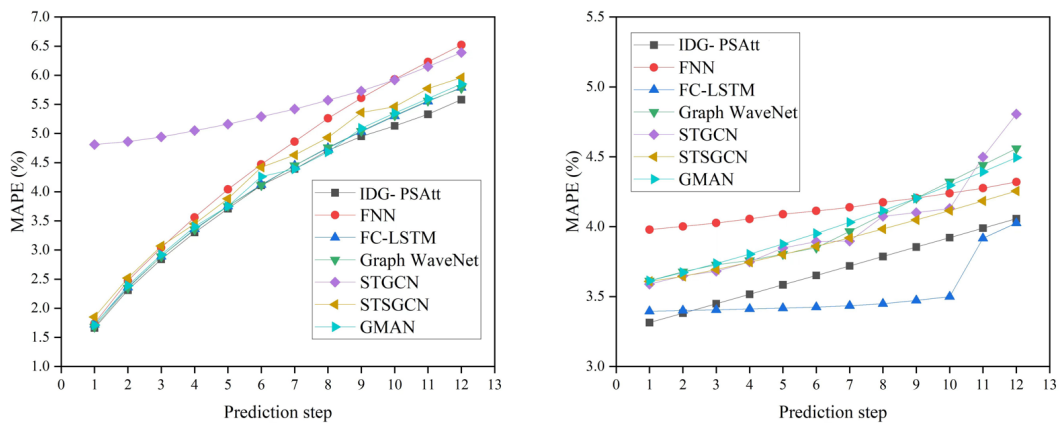


图 3.18 在两个数据集上的 MAPE 曲线

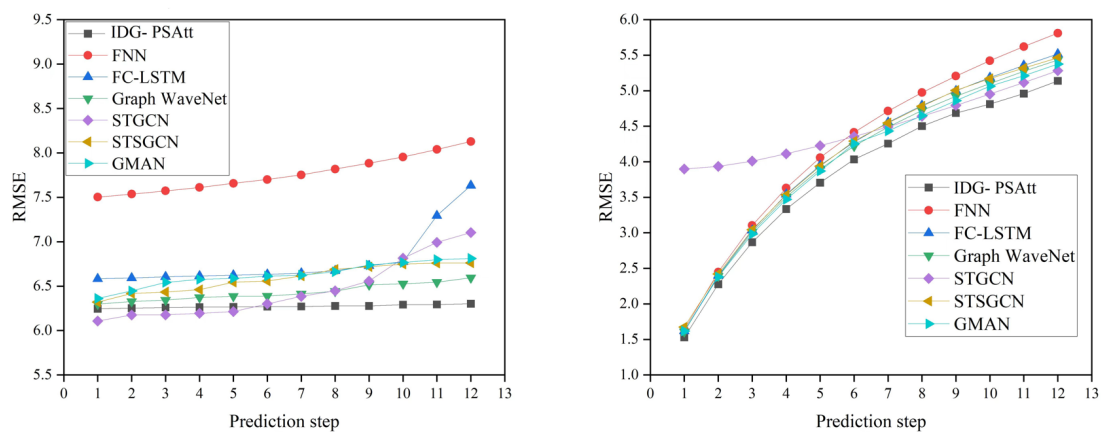


图 3.19 在两个数据集上的 RMSE 曲线

过调整输入时空信息和预设初始邻接矩阵来创建动态图结构，模拟节点之间的动态关系，从而探索交通网络中不可见节点之间的动态关联并捕捉隐藏的时空相关性。同时，IL-DGCN 通过将 DGCN 块嵌入交互式学习框架中来构建，以同时学习交通流的周期特性和趋势，并捕捉时空依赖性。通过在公开的交通流量预测数据集上的实验验证，本文所提出的方法在竞争性能上显著优于基线方法。还通过消融实验，验证了各个模块对 IDG-PSAtt 模型的影响。

然而这种交通流预测方法无法解决从起点到终点的出行需求量问题，OD 需求预测则专注于这个问题。它考虑了起点和终点之间的复杂空间依赖性，并通过分析历史数据和实时数据来预测未来的出行需求。这种方法能够更好地理解城市交通网络中的流动模式，为交通管理和规划提供更准确的信息。

第4章 融合渐进图卷积与构造网络更新的OD需求预测方法

在第3章，本文提出了高效且精确的交通流预测方法 IDG-PSAtt，并通过结合交互式学习策略、动态图卷积、时空卷积模块以及概率稀疏自注意力块，有效捕获了交通流的远程时空特征和隐藏动态空间特性，对于有效的路径规划、缓解拥堵以及人流管理，起着重要的作用。然而，起始目的地流量的预测仍然是一个具有挑战性的问题，尤其是在城市规划、公共交通系统设计和交通流量管理等方面。本章提出融合渐进图卷积与构造网络更新的OD矩阵预测模型总体框架。接着分别介绍随机游走采样、渐进图卷积构造网络和分层记忆更新模块。最后，本章对模型分别做了模型对比实验和模型消融实验，以验证本章提出的方法的有效性。

4.1 引言

随着城市化进程的加速，城市交通需求日益增长，带来了交通拥堵、能源消耗和环境污染等一系列问题。为了有效应对这些挑战，准确预测城市的出行需求变得尤为关键。OD需求预测，即预测从起点到终点的出行需求量，是城市交通规划和管理中的一个核心环节。通过准确的OD需求预测，可以为交通网络的优化、公共交通系统的规划和交通政策的制定提供科学依据，从而提高交通系统的效率，减少交通拥堵，并指导城市的可持续发展。

OD矩阵是一种用于描述交通流量的工具。它是一个二维矩阵，其中的行表示出发地，列表示目的地。在城市交通领域中，可以将城市区域视为一个大矩形，并将其划分为 $m \times n$ 个小区域，如图4.1(a)所示。这样的划分方式形成了一个格网结构，每个小区域可以表示为一个单元格。针对每个时间片，可以构建一个OD矩阵如图4.1(b)所示，其维度为 $N = m \times n$ 。这个OD矩阵描述了在该时间片内从一个区域到另一个区域的出行次数或者交通需求。

为了进行出发地与目的地交通流预测，可以形成一个按时间顺序排列的OD矩阵序列。这个序列由历史上多个时刻的OD矩阵组成，用于预测未来一个时刻的OD矩阵。因此，OD矩阵预测的目标是基于历史上多个时刻的OD矩阵，来预测未来一个时刻的OD矩阵。通过分析和建模OD矩阵序列之间的关系，可以利用时间序列分析、机器学习或其他预测方法来预测未来交通流量的分布和变化趋势。此外，还存在以下问题：

- (1) OD矩阵中出发地与目的地语义关系混淆：城市的各个区域因其点位兴

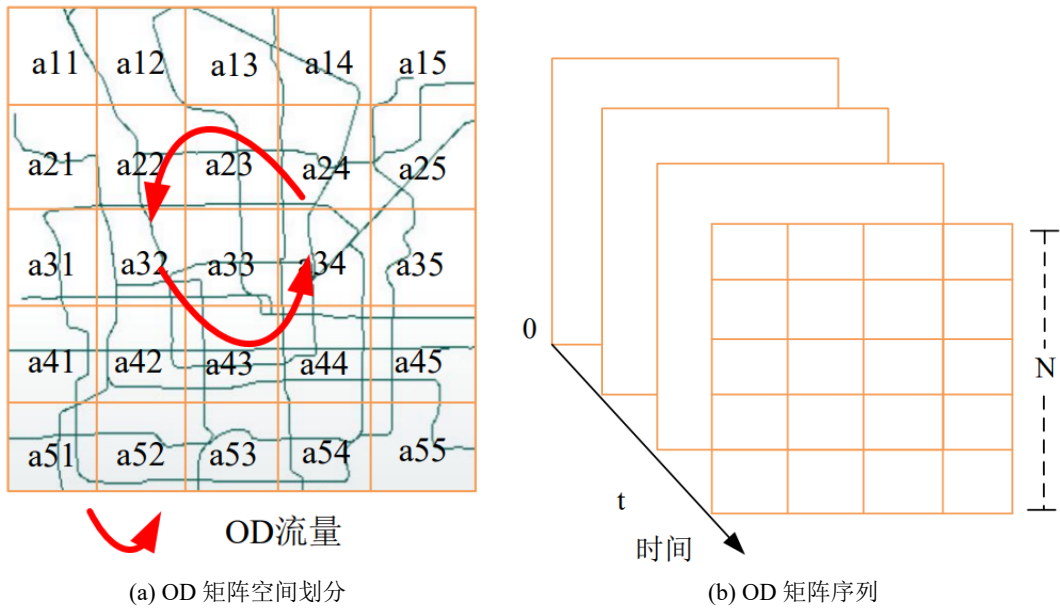


图 4.1 OD 矩阵的空间划分和 OD 矩阵的序列

趣 (Point of Interest, POI) 的分布差异展现出独特的功能特性, 进而影响着出行订单的地理空间分布格局。区域功能的不同, 在图 4.2(a) 中商业集中的地区或是办公楼密集的地带, 往往吸引大量寻求上下车服务的人群, 从而成为打车需求的高发区域。相对而言, 城市的郊区、公共绿地、以及农村等区域由于人流相对稀少, 这些地方作为打车服务的起点或终点的订单数量远远低于城市中心。这种基于区域功能不同而产生的出行需求分布不均衡, 不仅体现了城市空间的多样性, 同时也揭示了在进行交通流量预测时面临的重要挑战。

(2) 在 OD 需求预测中还会有需求分布的不均衡导致 OD 数据的稀疏性。在具体的时间片段内, OD 矩阵可能包含大量的零值, 反映了某些区域对之间的出行需求极低或根本不存在。这种稀疏性不仅凸显了城市交通需求的非均质性, 还对交通流量预测模型的性能提出了挑战。

(3) OD 需求中的空间依赖性指的是不同空间位置之间存在相互影响相互制约的关系。在 OD 需求中, 空间依赖性意味着不同地理位置的交通状况存在相互影响的关系。例如, 在图 4.2(b) 中, 区域 s_1 的流入会影响周边区域 s_2 的交通流出。同样, 如果区域 s_2 发生交通拥堵, 会影响相邻区域 s_3 , 使得车辆无法进入该区域, 导致周边区域也发生交通拥堵。这种相邻区域之间的交通流关联性, 使得交通拥堵现象在城市中产生连锁反应。

为了应对上述挑战, 本章提出一种融合渐进图卷积与构造网络更新的 OD 需求预测模型 (PGC-ODMUP), 该模型主要包含随机游走采样、渐进图卷积构造网络和分层记忆更新模块。其中, 随机游走采样利用交替且递归采样的方法解决源节点和目标节点语义关系混淆的问题。渐进图卷积构造网络会逐步处理图中

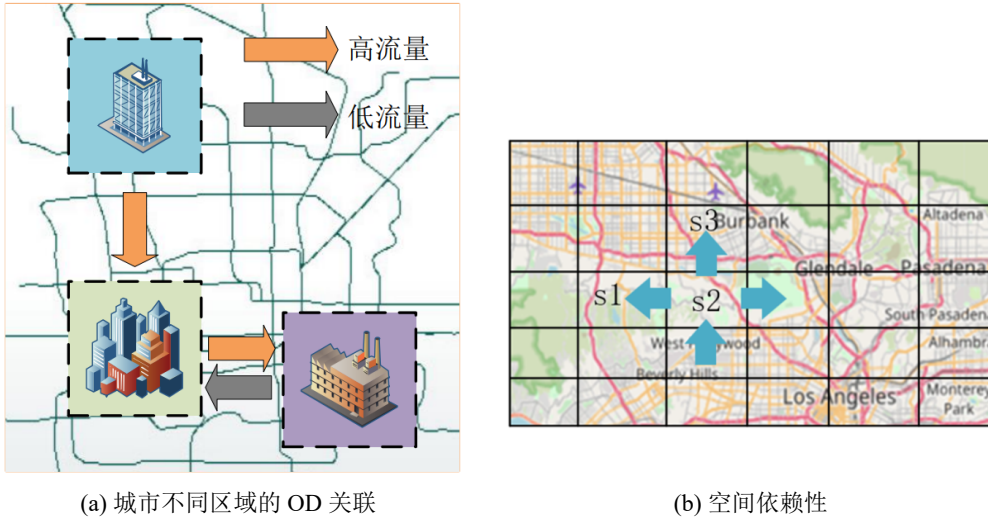


图 4.2 城市不同区域的 OD 关联和空间依赖性

的信息，以适应不同规模和复杂度的图形数据，解决 OD 需求中的空间依赖性。这样的网络能够更好地处理大规模图数据。分层记忆更新可以在不同层次上同时考虑时间信息和空间信息，实现不同时间粒度信息的集成，解决连续消息和离散消息在捕捉全局时空信息方面的局限性问题。最后，设计了一个损失函数，优化了 OD 数据稀疏性的问题。实验结果表明，PGC-ODMUP 方法可以有效地完成 OD 需求的预测，与流行的基线方法相比具有最佳的预测性能。

4.2 相关定义与数据预处理

4.2.1 相关定义

定义 4.1 (动态行程图): 在本研究中, 动态行程图定义为 $G = (V, E)$, 其中 $V = \{v_1, v_2, \dots, v_N\}$ 是 N 个节点的集合, $E = \{e_1, e_2, \dots, e_M\}$ 是 M 条边的边集。这些节点代表现实世界中的站点或区域。每条边 $e_k = \{u_k, v_k, t_k, f_k\}$ 表示在时间 t_k 从 u_k 出发到 v_k 的乘客, 并具有相关特征向量 $f_k \in \mathbb{R}^{d_F}$ 。特征向量包含行程长度、订单编号、结算等信息。在特定时间点 t , 动态图可以表示为 $G_t = (V, \{e_k | t_k < t\})$, 其中包含 t 之前发生的所有行程。

定义 4.2 (问题定义): 从时间 t 到 $t + \tau$ 的时间间隔内的 OD 矩阵表示为 $Y_{t,t+\tau} \in \mathbb{R}^{N \times N}$ 。OD 预测的目的是在给定历史行程数据作为动态图 G_t 的情况下, 预测未来的 OD 矩阵 $\hat{Y}_{t,t+\tau}$ 。给定时间戳 t 处的动态图 G_t 和时间 t 到 $t + \tau$ 的真实 OD 矩阵 $Y_{t,t+\tau}$, 期望学习一个函数 Ψ 来预测 OD 矩阵如公式 (4.1), 即 $\hat{Y}_{t,t+\tau} = \Psi(G_t)$ 。

$$\Psi^* = \arg \min_{\Psi} \text{ODLoss}(Y_{t,t+\tau}, \Psi(G_t)) \quad (4.1)$$

4.2.2 数据预处理

在本章中，对数据进行预处理，具体的预处理步骤如下：

(1) 首先，将数据加载并转换为 NumPy 数组，其中包含了源节点、目标节点和时间戳等信息。在处理数据时，采用线性插值方法来填充缺失值。接下来，对数据进行最小-最大归一化处理，以将数据范围限制在 $[0, 1]$ 之间。归一化的公式表示为公式 (4.2)。

$$g_i^{norm} = \frac{\vartheta_i - \vartheta_{min}}{\vartheta_{max} - \vartheta_{min}} \quad (4.2)$$

其中， ϑ_i 表示第 i 个原始数据， ϑ_{min} 和 ϑ_{max} 分别表示原始数据的最小值和最大值。而 g_i^{norm} 表示归一化后的输入数据。

(2) 根据加载的数据，创建 Data 类的实例，该类包含了数据的各种属性和相关方法，为后续的任务提供了基础。

4.3 融合渐进图卷积与构造网络更新方法

如图 4.3 所示，本章提出一种融合渐进图卷积与构造网络更新的 OD 需求预测方法。该方法主要包括四个部分，分别是随机游走模块、渐进图卷积构造网络模块、分层记忆更新模块和输出层。首先，输入动态图 G_t 到随机游走模块，保持交替且递归地采样前向边缘和反向边缘。再通过渐进图卷积构造网络，让不同节点之间的相关性随着时间的推移而更新。接着通过分层记忆更新模块，维护节点表示并基于消息传递动态更新它们，每个节点上维护 $D + 1$ 个记忆 $H \in \mathbb{R}^{(D+1) \times d_H} = \{H^0, H^1, \dots, H^D\}$ 。最后为输出层和损失函数，对模型的最后一个隐藏状态 H' ，利用 MLP 进行预测，输出时间 t 到 $t + \tau$ 的 OD 矩阵 $\hat{Y}_{t,t+\tau}$ ，用于进行最终的预测和训练。

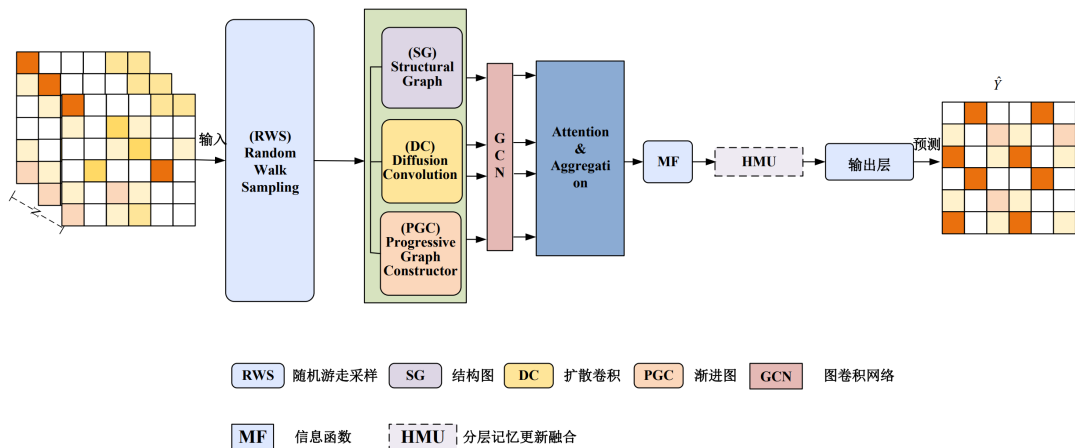


图 4.3 PGC-ODMUP 整体框架图

4.3.1 随机游走采样模块

基于随机游走的方法由于其在图嵌入技术中图结构感知的能力，为理解图中的复杂关系提供了强有力的支持。然而，传统的随机游走方法在网络表示学习中得到了广泛应用，但在 OD 需求预测中的应用存在一定的挑战。由于 OD 需求预测关注起点和终点之间的关系，直接使用传统的随机游走方法可能无法准确地区分源节点和目标节点，从而导致语义关系的混淆。为了解决这一问题，本章在 OD 需求预测中进行了深入考虑，并提出了一个方法。这个方法通过对随机游走的改进，有效地区分了源节点和目标节点之间的语义关系。具体而言，在随机游走过程中引入了额外的约束条件和信息，以保持源节点和目标节点的语义一致性。通过在游走路径中加入源节点和目标节点的标识信息，方法能够明确指示节点的角色，并避免语义关系的混淆。这个方法可以被解释为以下两个步骤的循环执行：

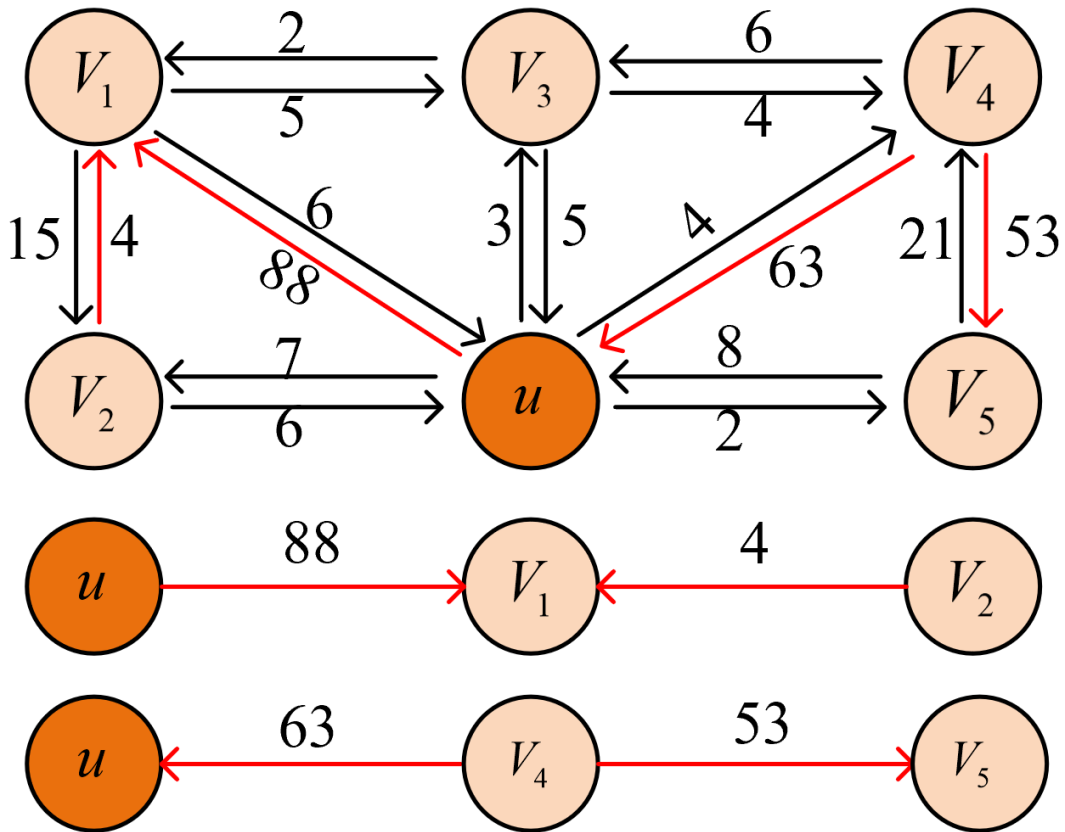


图 4.4 随机游走采样

(1) 保持交替且递归地采样前向边缘和反向边缘：在这一步骤中，首先从源节点开始，通过采样前向边缘（从源节点指向其他节点的边）来扩展路径如图 4.4。通过递归地采样前向边缘，可以沿着路径不断向前延伸，同时保持源节点到目标节点的方向性。然后，在每一步采样后，根据一定的策略选择下一个节点进行采样，直到达到预定的路径长度或满足其他终止条件。

(2) 保持交替且递归地采样反向边缘和前向边缘：在这一步骤中，从目标节点开始，通过采样反向边缘（从其他节点指向目标节点的边）来扩展路径。通过递归地采样反向边缘，路径可以向目标节点的方向不断延伸。类似地，在每一步采样后，根据一定的策略选择下一个节点进行采样，直到达到预定的路径长度或满足其他终止条件。

通过交替执行这两个步骤，源节点和目标节点之间的语义关系得以保持，并且能够在采样过程中逐步扩展路径。这样，最终得到的路径能够准确地反映源节点和目标节点之间的关联性，并用于后续的语义分析和预测任务。这种交替且递归采样的方法在OD需求预测中取得了显著的效果。它能够有效地区分源节点和目标节点的语义关系，避免混淆，并生成准确的路径表示。在实验中，这个方法表现出了良好的性能和鲁棒性，为OD需求预测提供了一种可靠的解决方案。对于每一步，随机采样概率公式为(4.3)。

$$P = \begin{cases} \Phi(\exp(\hat{t} - t^-)), & d = 0 \\ \Phi(t - 2^{d-1}\Delta T, t + \epsilon), & d \neq 0 \end{cases} \quad (4.3)$$

其中， Φ 表示按行归一化，使概率之和等于 1， $d = 0$ 表示根据连续时间边缘生成随机游走， $d \neq 0$ 表示根据第 d 个离散时间记忆生成随机游走。

4.3.2 渐进图卷积构造网络模块

在本小节中，使用了可以根据节点信号相似性逐步适应流量变化的渐进图。Han 等^[74]采用类似的想法来自适应学习邻接矩阵。

渐进图随着不同节点之间的相关性随着时间的推移而演变，使用在线数据逐步更新节点关系是直观的，如图 4.5 所示。给定四个节点 $\{v_1, v_2, v_3, v_4\}$ ，矩阵中的每一行说明了在最后 5 个时间步中观察到的节点信号。提出两个时间矩阵 t 和 $t+h$ 。在 t 处，可以观察到 v_1 与 v_4 之间有很强的相似性，而 v_1 与 v_2 在时间 $t+h$ 处最相似，边缘可能会随着时间的推移而出现或者消失。

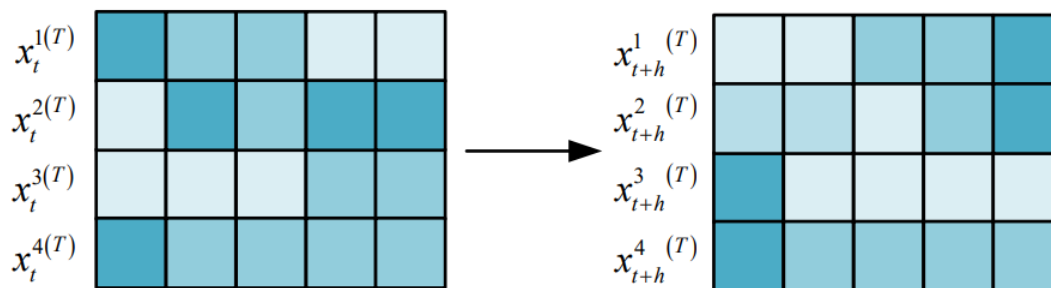


图 4.5 渐进图

渐进图 G^t 是一组图，其中 $G^t = (V, A_p^t)$ 。这里， A_p^t 是时间 t 的渐进邻接矩

阵，其中包含从节点信号相似性中学习到的成对权重。目标是在具有相似信号的节点之间施加更高的权重，无论它们的空间接近程度如何。节点相似度是使用其信号的余弦相似度来测量的。两个节点 v_i 和 v_j 之间的余弦相似度 s_{ij} 定义如 (4.4) 所示。

$$s_{ij}^t = \tilde{x}_i^{t(T)} \cdot \tilde{x}_j^{t(T)} \quad (4.4)$$

其中， $\tilde{x}_i^{t(T)}$ 为单位向量。为了引入相似度学习的随机性，本章实现了一个可学习的调整矩阵 $W_{adj} \in \mathbb{R}^{T \times T}$ ，并且渐进邻接矩阵 $A_{P_{ij}}^t$ 的每个元素定义如 (4.5) 所示。

$$A_{P_{ij}}^t = \text{softmax} \left(\text{ReLU} \left(\tilde{x}_i^{t(T)} W_{adj} \tilde{x}_j^{t(T)} \right) \right) \quad (4.5)$$

应用 softmax 函数对渐进邻接矩阵进行归一化，并通过 ReLU 激活消除负连接。任何图卷积模块的核心思想都是聚合邻居节点的信息来提取目标节点的空间特征。图卷积模块的基本形式是将图信号与可学习参数乘以按定义方法处理的邻接矩阵。在交通预测中，最流行的图卷积模块形式之一是扩散卷积，其中交通网络上的交通流被视为扩散过程。利用转移矩阵 $P = \frac{A}{\text{rowsum}(A)}$ ，带有滤波器 f_w 的 K 步扩散过程的有向图上的扩散卷积可以定义如 (4.6) 所示。

$$Z_t = X_t * Gf_W = \sum_{k=0}^{K-1} P^k X_t W_{k,1} + P^{T^k} X_t W_{k,2} \quad (4.6)$$

其中， $*Gf_W$ 是与滤波器 f_W 的图卷积运算， $W_{k,1}$ 和 $W_{k,2} \in \mathbb{R}^{C \times D}$ 是可学习参数。 P 和 P^T 用于反映前向和后向扩散过程。如果给定的邻接矩阵是无向的，则仅使用方程的第一项。

以扩散卷积作为渐进图卷积的基本图卷积模块，向扩散卷积添加了渐进邻接矩阵、图信号矩阵和附加权重参数的乘法如 (4.7) 所示。

$$Z_t = X_t * Gf_W = \sum_{k=0}^{K-1} P^k X_t W_{k,1} + P^{T^k} X_t W_{k,2} + A_P^t X_t W_{k,3} \quad (4.7)$$

4.3.3 分层记忆更新模块

本节设计了一种解决历史行程中时间信息丢失的问题的方法。传统方法在对边进行计数时通常使用固定的时间窗口，并将时间离散化处理，这样做会导致实际的时间信息丢失。特别是当时间窗口的时间粒度较粗时，更容易丢失大量的时间相关信息。为了解决这个问题，本章节出了一种分层记忆更新模块来维护节点的分层时间记忆如图 4.6。

分层记忆更新模块由两部分组成，每一部分都有不同的功能和目的。首先，一部分将时间视为连续特征，并在边到达时进行更新，以保留连续的时间信息。

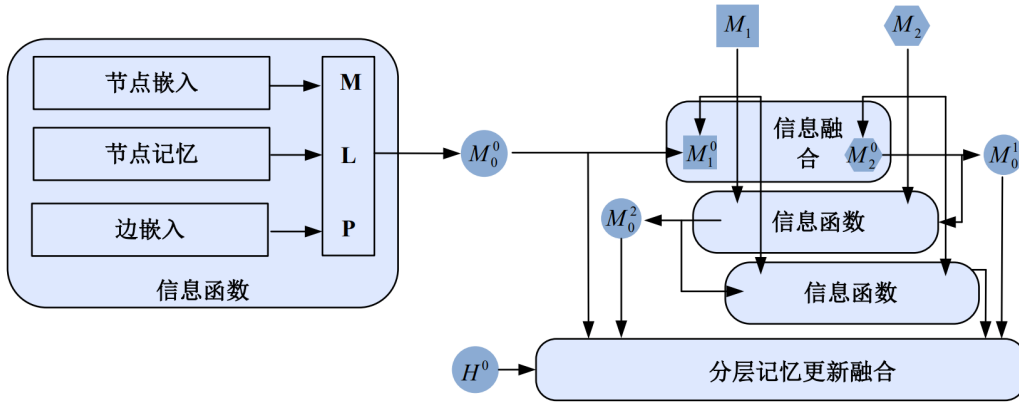


图 4.6 分层记忆更新

这意味着每当有新的边到达时，分层记忆更新模块会更新节点的时间记忆，而不是简单地在固定时间窗口内进行计数。通过保留连续的时间信息，分层记忆更新模块能够更好地捕捉节点在不同时间点上的活动模式和演化趋势。

另一部分的分层记忆更新模块则维护多个离散时间存储器，用于揭示节点的宏观状态。这些离散时间存储器可以看作是对节点在不同时间层次上的状态进行编码。通过维护多个离散时间存储器，分层记忆更新模块能够捕捉到节点在不同时间尺度上的特征和模式，从而提供更全面的时间相关信息。

此外，本小节还设计了记忆融合机制，根据相应级别的消息进行主要接收。这样可以综合利用不同级别的时间记忆，从而更好地捕捉节点的时间特征。通过引入分层记忆更新模块和记忆融合机制，本小节的方法能够更有效地维护节点的时间信息，并在时间上更全面地描述节点的状态。

形式上，每个节点上维护 $D+1$ 个记忆 $H \in \mathbb{R}^{(D+1) \times d_H} = \{H^0, H^1, \dots, H^D\}$ ，其中 H^0 代表连续时间记忆， H^d ($d \geq 1$) 代表覆盖 ΔT_d 时间单位的宏离散时间记忆如公式 (4.8)、(4.9)、(4.10)。

$$M' = \begin{cases} \left[H^d \parallel \Theta(H^d, G_t) \parallel \Omega_d(H^d, G_t) \right], & d = 0 \\ \left[H^d \parallel Y_{t-2^{d-1}\Delta T, t}^i \parallel \Omega_d(H^d, G_t) \right], & d \neq 0 \end{cases} \quad (4.8)$$

$$M_d^0 = W^{m_1} \sigma(W^{m_2} M' + b^{m_2}) + b^{m_1} \quad (4.9)$$

$$\Theta(H^d, G_t) = (1 + W^e(t - t_i^-) + b^e) \odot H^d \quad (4.10)$$

其中， $M_d^0 \in \mathbb{R}^{d_M}$ 表示上述消息， t_i^- 表示节点 i 的最后更新时间， $Y_{t-2^{d-1}\Delta T, t}^i$ 表示 OD 矩阵的第 i 行。 W^* 和 b^* 是可学习的参数。这代表时间编码，表示节点嵌入方法。

为了解决连续消息和离散消息在捕捉全局时空信息方面的局限性，并实现不同时间粒度信息的集成，该方法引入了一种消息融合机制。该机制能够将特定

轨道消息与其他类型的消息进行融合，以整合多种时空信息。通过消息融合机制，连续消息和离散消息可以保留各自的特定轨道信息，同时与其他时间粒度的信息进行集成。这样就可以获得更全面且更丰富的时空特征。

形式上，最新消息 $M \in \mathbb{R}^{(L+1) \times d_M} = \{M^0, M^1, \dots, M^L\}$ 存储在相应的存储器中，其中 M^0 是消息函数的输出， M^l ($l \geq 0$) 的计算如下：在计算 M^l 时，首先聚合来自其他时间级别的其他消息，然后与 M^{l-1} 连接。然后将连接传递到全连接层来计算结果。 M^0 表示节点的特征， M^l 将 M^{l-1} 与其 l 跳邻居聚合。该机制可以表述为公式 (4.11)、(4.12)。

$$M_d^l = W_{mess}[M_d^{l-1} || AGG(\{M_k^{l-1}, \forall k \neq d\})] \quad (4.11)$$

$$AGG(\{M_t\}) = \max(\{\sigma(W^{pool} M_t + b^{pool})\}) \quad (4.12)$$

其中， M_d^l 表示第 l 层第 d 个记忆的消息， σ 是 ReLU 函数， W^{pool} 和 b^{pool} 是可学习参数。

因此，由于能够以减少的参数来记忆信息，因此采用 GRU 来更新记忆。 M_d^L 直接用作 GRU 的输入，可以表示为公式 (4.13)。

$$H^d = GRU(H^d, M_d^L) \quad (4.13)$$

为了聚合这些信息，还设计了信息聚合器，该聚合器能够减少训练时间并处理大量的边缘。具体而言，该信息聚合器不仅简单地对多个边缘的信息进行平均，而是引入了另一个项目来扩展最后更新的记忆。该项目将节点状态和批次中的时间影响视为指数衰减。这意味着在信息聚合过程中，较早的边缘会受到较新的边缘的影响更多，而较早的边缘的影响会随着时间的推移逐渐减弱。通过使用指数衰减来调整边缘的权重，可以更好地平衡边缘之间的时间影响，从而提高消息的聚合效果。

该设计的目的是在保留重要时间信息的同时减少训练时间。通过将边缘批次的连续时间消息进行聚合，模型可以更高效地处理大量的边缘，并在训练过程中降低时间复杂度。 H'' 的更新如公式 (4.15)。形式上，当 $d=0$ 时，对于特定节点 i ，可以重写如公式 (4.15)

$$H'' = \text{mean}_{b \in B}([H^0 || H^0 * \exp(t - t_b^-)]) \quad (4.14)$$

$$M' = \begin{cases} [H'' || \Theta(H^d, G) || \Omega_d(H^d, G)], & d = 0 \\ [H^d || Y_{t-2d-1}^i \Delta T, t || \Omega_d(H^d, G_t)], & d \neq 0 \end{cases} \quad (4.15)$$

其中， b 表示批次中从 i 出发的边。

4.3.4 损失函数

由于 OD 矩阵的稀疏性，有许多零 OD 流会显著影响预测性能。为了应对这一挑战，本章引入了一种掩码损失函数，如公式 (4.16) 所示。该函数通过构建一个掩码矩阵来实现，该矩阵根据 ODAD (Origin-Destination Attraction Degree, ODAD)^[75] 等级来遮挡 OD 流。

$$\mathcal{L} = \frac{1}{(n \times n)_{no_mask}} \sum_{i,j} mask \times (Y_{i,j} - \hat{Y}_{i,j})^2 \quad (4.16)$$

其中 $mask$ 是由 0 和 1 组成的矩阵，用于表示相应的 OD 流是否被遮挡。由于特定 OD 对的 ODAD 值随时间变化，因此掩码矩阵将相应随时间更新如公式 (4.17)、(4.18)。值得注意的是，如果值被屏蔽，则不会进行反向传播， $w_{new} = w$ 。如所描述的那样，只有重要流量的误差才会在这里反向传播。

$$MSE = \frac{1}{(n \times n)_{no_mask}} \sum_{i,j} mask \times (Y_{i,j} - w \times \hat{Y}_{i,j})^2 \quad (4.17)$$

$$w_{new} = w - lr \times \frac{\partial MSE}{\partial w} = w + lr \times \frac{2 \times Y_{i,j}}{(n \times n)_{no_mask}} \sum_{i,j} mask \times (Y_{i,j} - w \times \hat{Y}_{i,j}) \quad (4.18)$$

其中 $Y_{i,j}$ 和 $\hat{Y}_{i,j}$ 分别是真实值和预测值， lr 代表学习率。

4.4 实验结果与分析

本节旨在通过实验来验证提出方法的效果。实验基于两个真实世界的交通流数据集进行，目的是评估所提方法的性能。具体实验内容包括：

实验一：超参数实验。

实验二：比较提出方法与现有基准方法在预测性能上的差异。

实验三：评估模型内部各组件对预测结果的影响。

实验四：分析模型在处理实际 OD 需求时的预测表现。

4.4.1 数据集介绍

本章使用北京地铁 (Beijing Subway) 数据集和纽约出租车 (New York Taxi) 数据集来验证本章所提出的 PGC-ODMUP 模型的预测性能。北京地铁数据集是一个综合性的数据集，用于研究和分析北京市地铁系统的运行情况和乘客出行行为。纽约出租车数据集是一个包含了纽约市出租车行程记录的综合性数据集，提供了丰富的信息用于分析城市交通和乘客出行行为。两组实验数据集的详细

信息如表 4.1。数据集依照时间顺序划分，训练集、验证集、测试集的比例为 7:1.5:1.5。

表 4.1 实验数据集的描述

数据集	北京地铁	纽约出租车
类型	铁路出行数据	出租车订单
城市	北京	纽约
时间段	2018/6/1-2018/7/1	2019/1/1-2019/6/30
单位时间间隔	30 分钟	1 小时
网格图尺寸	(32,16)	(16,16)

1. 数据集优势说明

大规模的交通流量数据：北京地铁和纽约出租车数据集涵盖了广泛交通网络数据，包括经纬度坐标、行程距离和时间等。这些数据集提供详尽的时间戳和地理位置信息，使得研究者能够准确地追踪交通流动的时空分布。

高覆盖度：对于北京地铁数据集来说，该数据集覆盖全市的主要地铁网络，包括重要线路和站点，涵盖城市的主要区域和交通枢纽。对于纽约出租车数据集来说，数据反映了从住宅区到商业区的全面交通流量。

丰富的行为洞察：北京地铁的乘客流量数据和纽约出租车的行程数据可以揭示城市居民的出行习惯、偏好以及对特定事件（如节假日、大型活动）的响应。

2. 数据分析

挑选纽约出租车中的一部分数据集，如图 4.7所示，A、B 分别代表住宅区、工业区，该图显示了两个区域不同方向 9 天的出行需求曲线随时间的变化。其中可以看到有大量的以 B 为终点的出行需求，但是以 A 为终点的出行需求较少。通过高峰和非高峰时段的差异，来揭示出行需求的规律和模式。

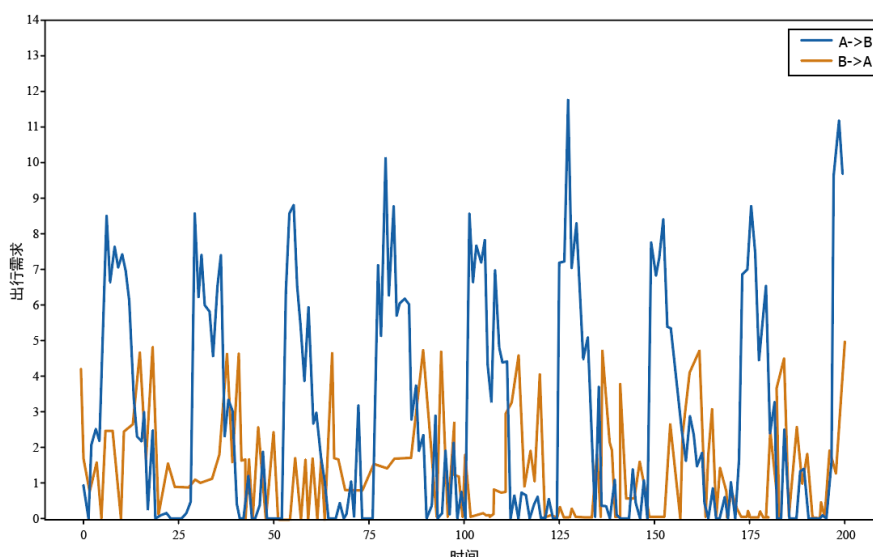


图 4.7 两个区域不同方向的出行需求

4.4.2 实验环境与实验设置

1. 实验环境

本章采用的软件及硬件环境与第3章 3.3.2 节所述的实验设置相同。

2. 评价指标

为了更好的评估模型的预测性能，本章使用均方根误差（RMSE）如公式(4.19)、皮尔逊相关系数（Pearson Correlation Coefficient, PCC）如公式(4.20)以及平均绝对误差（MAE）如公式(4.21)作为 PGC-ODMUP 模型的评价指标。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.19)$$

$$PCC = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma_X} \right) \left(\frac{Y_i - \bar{Y}}{\sigma_Y} \right) \quad (4.20)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.21)$$

其中， $\frac{X_i - \bar{X}}{\sigma_X}$ 、 \bar{X} 和 σ_X 分别表示对 X_i 样本的标准分数、样本平均值以及样本标准差，同理等同于样本 Y。

4.4.3 超参数实验

在本研究采用了 Adam 优化器来调整神经网络的权重。实验中设定的学习率为 0.0001，这个值较小，有助于模型在训练过程中稳定地逼近最优解。此外，实验中还确定了 PGC 和 HMU 的维度，分别为 64 和 128。这意味着，在构建模型的相关部分时，PGC 层使用了 64 个隐藏单元，而 HMU 层则使用了 128 个隐藏单元。epoch 训练轮次选择的是 500。在这些基础上，进行超参数实验。

如图 4.8 显示了两组训练轮次对模型性能的影响，其中模型性能通过皮尔逊相关系数来衡量。两图的趋势大致相同，都是随着训练轮次的增加，模型性能呈现出上升趋势，这可能表明随着模型在更多数据上训练，其对数据的拟合能力增强。

如图 4.9 显示了 HMU 维度数对模型性能的影响，从图 4.9(a) 可以看出，在北京地铁数据集中，当 HMU 维度从 64 增加到 128 时，模型性能有所提升，PCC 值略有上升。然而，当继续增加 HMU 维度到 256 时，模型性能反而下降，PCC 值有明显的下降。这可能意味着在 128 维度时模型已经达到了性能的顶峰，而更多的维度没有带来性能提升，反而因为模型过于复杂导致过拟合或其他问题。图 4.9(b) 虽然在 HMU 维度为 64 时模型可能已经达到了性能的顶峰，但整体考虑模型的复杂性时，选择 128 维的 HMU 维度是更为合适的选择。

如图 4.10 展示了 PGC 维度数量变化对模型性能的影响，图 4.10(a) 显示，当 PGC 维度为 64 时，PCC 值最高，随后随着维度的增加性能逐渐降低。这表明，对于北京地铁数据集，较小的 PGC 维度能够得到较好的性能。图 4.10(b) 中也呈现了相似的下降趋势。综合分析两个图表，可以推测出增加 PGC 维度可能导致模型过度复杂导致的过拟合的迹象。

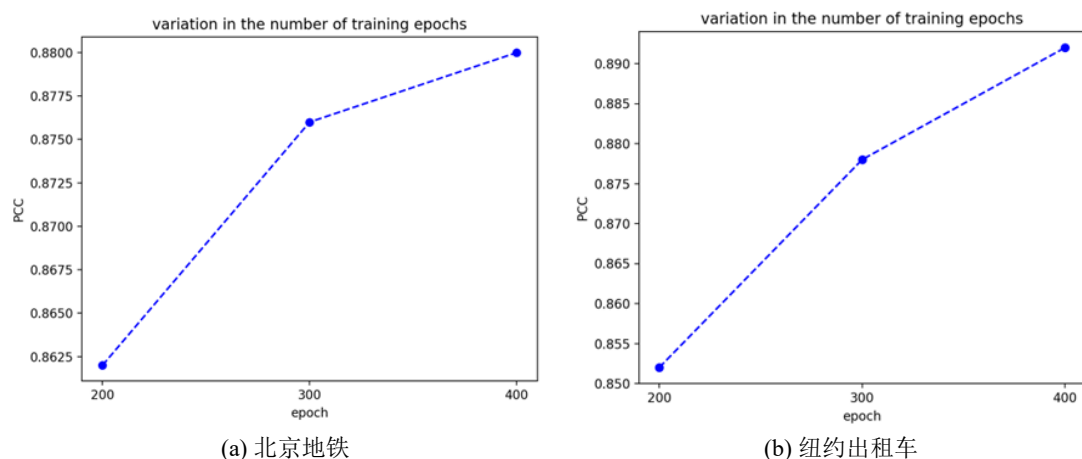


图 4.8 北京地铁和纽约出租车的迭代轮次实验

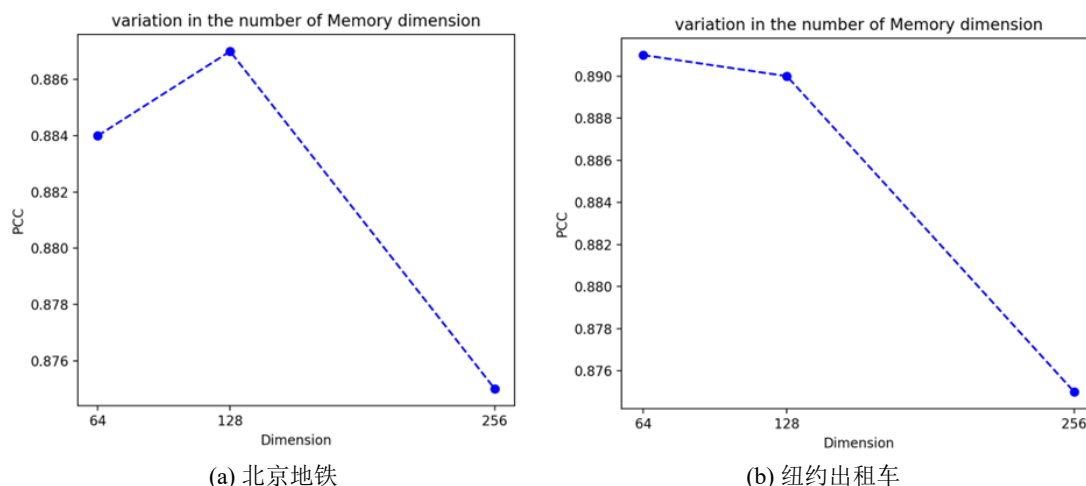


图 4.9 北京地铁和纽约出租车的 HMMU 隐藏单元实验

4.4.4 模型对比实验

1. 基线方法

为了进一步评估 PGC-ODMUP 模型的有效性，本章将其性能与几个基线模型进行比较，包括当前在 OD 矩阵预测任务的最优模型（GEML）。

(1) HA^[5]: 将历史 OD 需求矩阵的平均值作为预测值。

(2) LR (Linear Regression, LR)^[76]: 是一种利用输入和输出之间线性相关性的回归模型。

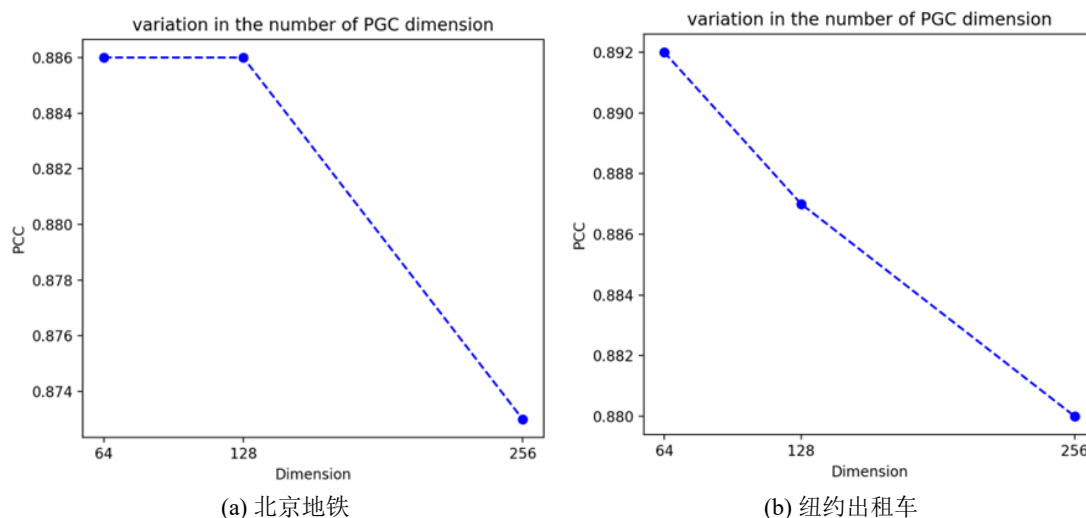


图 4.10 北京地铁和纽约出租车的 PGC 隐藏单元实验

(3) MDL^[77]: 提出了一种多任务深度学习框架, 通过耦合中间层的潜在表示进行连接并一起训练, 用于同时预测时空网络中的节点流量和边流量。

(4) XGBoost^[78]: 是一种基于梯度提升树的模型, 能够从先前的数据模式中进行学习和提升。

(5) GEML^[41]: 是一种基于多任务学习框建模时间属性和使用图卷积进行节点嵌入的 OD 矩阵预测模型。

(6) DNEAT^[79]: 是另一种基于多任务学习框建和节点边缘嵌入的 OD 需求预测模型。

(7) TCN^[80]: 是一种连续时间动态图表示学习框架。

通过与这些基线模型进行比较, 可以评估 PGC-ODMUP 模型在预测 OD 需求方面的性能。这些模型比较将提供关于 PGC-ODMUP 模型相对于其他方法的优势和效果的信息。

2. 实验结果对比与分析

如表 4.2 和 4.3 所示, 实验对比北京地铁和纽约出租车中的评价指标结果。根据实验结果显示, 本文提出的 PGC-ODMUP 模型在 MAE、RMSE 和 PCC 指标上优于所有基线模型, 并且在两个数据集的所有情况下表现出色。在北京地铁数据集中, TCN 模型相对于其他 OD 预测方法的性能稍逊。这可能是因为北京地铁具有大量的节点和边, 这增加了 TCN 模型找到重要的邻居节点的难度。

尽管在北京地铁数据集中的性能稍逊, 但在纽约出租车数据集中, TCN 模型的性能优于其他基线模型, 这证明了在这种情况下连续时间信息的有效性。纽约出租车数据集包含的是纽约市的出租车出行数据, 而这些数据具有连续的时间信息。TCN 模型能够利用这些连续时间信息来更好地捕捉出行模式和交通流量的变化, 从而提高了预测性能。由于 TCN 模型采用了分层时间节点表示, 它

能够同时捕捉细粒度和宏观信息，并在各种情况下都表现良好。这表明所提出的方法在处理 OD 需求预测问题时具有优势，并能够有效地利用时间信息来改善预测性能。

表 4.2 北京地铁的比较结果

Method	MAE	RMSE	PCC
HA	1.4638	4.8672	0.8235
LR	1.5326	5.3556	0.7521
MDL	1.4356	4.7496	0.8231
XGBoost	1.6138	5.7716	0.7039
GEML	1.2746	4.6143	0.8568
DNEAT	1.6432	5.6378	0.7734
TCN	1.7346	6.1313	0.6453
PGC-ODMUP	1.2464	3.7225	0.8872

表 4.3 纽约出租车的比较结果

Method	MAE	RMSE	PCC
HA	0.7151	1.4501	0.8481
LR	0.8098	1.3631	0.8586
MDL	0.7636	1.3836	0.8521
XGBoost	0.7847	1.3575	0.8599
GEML	0.6847	1.3188	0.8718
DNEAT	0.7973	1.5249	0.8242
TCN	0.7207	1.2947	0.8747
PGC-ODMUP	0.6506	1.2021	0.9013

本文对 GEML 和 PGC-ODMUP 都使用了 Adam 优化器进行训练，并将初始学习率设置为 0.001。可以观察到，传统模型 HA 和 LR 的预测效果较差，这可能是因为纽约出租车数据集中部分区域间的 OD 出行需求具有随机性。HA 方法使用历史平均值进行预测，简单但效果不佳。而 LR 模型适用于平稳的时间序列数据，对于这种具有非平稳性质的数据集来说，LR 模型的表现最差。

相比之下，深度学习模型在预测纽约出租车数据集上表现更好。这表明深度学习模型能够更好地捕捉非线性和复杂的模式，从而提高预测性能。因此，在处理具有随机性和非平稳性质的 OD 需求预测问题时，深度学习模型可能是更好的选择。

MDL 利用二维卷积来获取空间关联性，但没有明确地获取数据中的时序相关性，故表现效果并不好。GEML 是一种基于图神经网络的 OD 需求预测模型。它利用图卷积进行节点嵌入，以聚合 OD 对的语义邻居和地理邻居的信息，并预先加权考虑邻居节点的重要性。此外，它还使用 LSTM 来捕捉时间上的关联。然而，GEML 模型存在一些限制。首先，它未充分考虑到 OD 需求预测中起点和终点的有向语义。在 OD 矩阵预测任务中，起点和终点之间的有向语义对预测结果具有重要影响，但 GEML 模型未明确建模这些信息。这可能导致模型在捕捉起

点和终点之间的相关性时存在一定的不足。

其次，GEML模型并未明确区分邻接矩阵中边的方向。在实际的OD需求预测任务中，起点和终点之间的交通流通常是有方向性的，因此明确区分邻接矩阵中边的方向对于准确预测OD矩阵很重要。然而，GEML模型未对这一点进行明确处理，可能导致模型在考虑方向性信息时存在一定的局限性。

相比之下，本文提出的PGC-ODMUP模型考虑了连续时间细粒度信息的集成和时空依赖性。它通过节点信号相似性逐步适应流量变化的渐进图，从而更好地捕捉了时空关系。同时，PGC-ODMUP模型使用随机游走采样的OD节点嵌入方法，以聚合邻居的语义信息，提高了预测性能。总之，PGC-ODMUP模型在考虑起点和终点的语义信息、处理稀疏矩阵以及捕捉时空依赖性方面进行了改进，表现出良好的预测性能。

4.4.5 模型复杂度和训练速度实验

本节对比了PGC-ODMUP和MDL、TCN、DNEAT的参数数量和训练速度，模型复杂度和速度是模型部署的重要参考指标。在OD需求预测任务中，在本实验中，将连续15个时间片的OD矩阵作为输入特征，而紧随其后的2个时间片的OD矩阵被定义为预测目标，从而构成了样本数据。训练速度则定义为单位时间能处理的样本数量（samples/s）。

表 4.4 模型参数数量和速度对比

模型	可训练参数量	训练速度
MDL	7,694,106	73
TCN	6,202,579	79
DNEAT	3,056,256	87
PGC-ODMUP	1,602,456	92

如表4.4所示，MDL模型使用多任务学习，通过耦合中间层的潜在表示进行连接并一起训练，用于同时预测OD矩阵特征的节点流量和边流量。但其每项任务都需要使用深度神经网络，并且要从高维度的历史交通数据提取时空关系，这就使其网络复杂度很高。相较之下，TCN仅对OD矩阵进行特征提取，由于其单任务特性，模型的参数量更少，训练速度也更快。

本文提出的PGC-ODMUP模型，和具有多任务学习框建和节点边缘嵌入的DNEAT模型相比，具有更少的参数量。对于不同时间步的输入，采用分层记忆更新将节点的时间感知表示分层处理，以便更好地捕获不同时间尺度上的模式和趋势，极大减少了参数量。相比之下，PGC-ODMUP模型通过节点信号相似性逐步适应流量变化的渐进图卷积，加快了训练速度。还通过随机游走采样的OD节点嵌入方法，以聚合邻居的语义信息。这样的扩展使得模型能够更好地捕捉交通数据中的时空关系，同时降低了参数数量并加快了训练速度。

4.4.6 模型消融实验

为深入分析 PGC-ODMUP 模型各部分的效能，本部分提出了 PGC-ODMUP 模型的三种变体，研究了 PGC、RWS 和 HMU 对模型预测性能的影响，并在 Beijing Subway 和 New York Taxi 数据集上将这三个变体与 PGC-ODMUP 模型进行比较。GSTFCN 模型与这三种变体的差异为：

- (1) w/o PGC：该模型没有渐进图卷积构造网络模块。
- (2) w/o RWS：该模型没有随机游走采样模块中的交替与递归。
- (3) w/o HMU：该模型没有分层记忆更新模块。

表 4.5 对比不同模型在北京地铁和纽约出租车数据集上的表现

Dataset	Method	RMSE	PCC
Beijing Subway	w/o PGC	4.132	0.853
	w/o RWS	3.845	0.821
	w/o HMU	4.040	0.832
	PGC-ODMUP	3.750	0.892
New York Taxi	w/o PGC	1.660	0.852
	w/o RWS	1.775	0.830
	w/o HMU	1.556	0.843
	PGC-ODMUP	1.202	0.901

表 4.5 的消融实验通过逐步移除模型的特定组件来评估这些组件对模型性能的贡献。在分析 PGC-ODMUP 模型与其变体（w/o PGC、w/o RWS、w/o HMU）的消融实验结果时，可观察到 PGC-ODMUP 模型在所有评价指标和所有数据集上均展现出最佳性能。如图 4.11 和 4.12 这一结果表明，PGC-ODMUP 模型中的特定组件对于提升 OD 需求预测准确性至关重要。

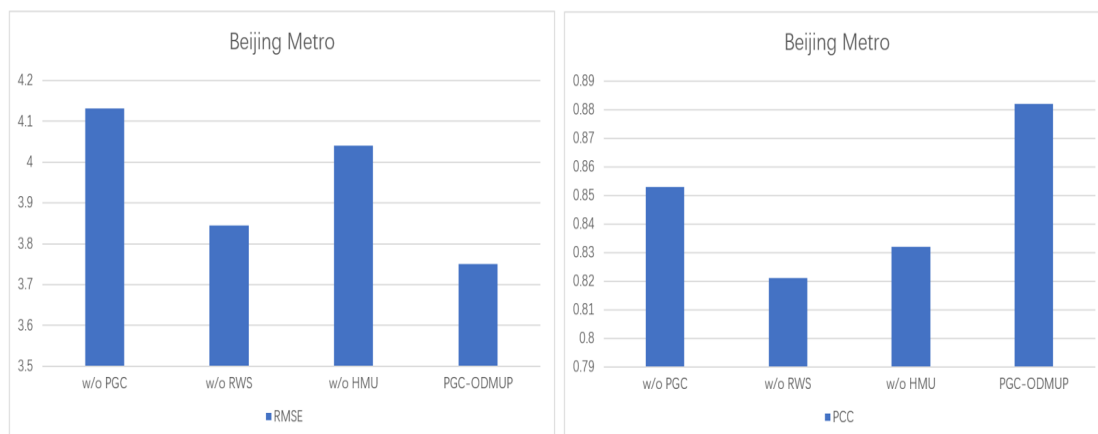


图 4.11 在 Beijing Subway 上不同变体的 RMSE 和 PCC

具体而言，PGC-ODMUP 相较于其变体模型展现出的优越性能，可能归因于更高效的时间细粒度信息的集成或是更先进的聚合邻居的语义信息。这些组件的存在使得 PGC-ODMUP 能够更精确地预测 OD 矩阵中的复杂模式和依赖关系，从而提高预测的准确度。

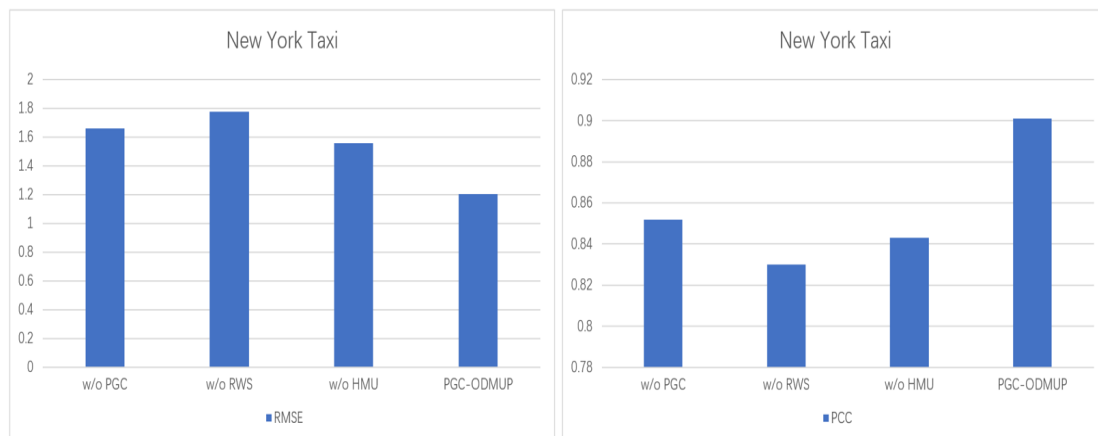


图 4.12 在 New York Taxi 上不同变体的 RMSE 和 PCC

而 PGC-ODMUP 的变体模型，通过消融实验的性能下降，揭示了它们各自缺少的关键特性对于整体预测性能的影响。这可能包括空间信息处理的不足、聚合邻居的语义信息的简化。

4.4.7 预测值与真实值对比实验

在本研究中，本小节通过选取纽约出租车的数据进行实验，分析了区域 162 至区域 102 和区域 68 到区域 44 的出行需求变化情况（从 2019 年 1 月 11 日到 2019 年 1 月 19 日共 9 天），旨在探索真实出行需求与预测值之间的差异及其随时间的动态变化特性。

从图 4.13 和 4.14 中可以看出，出行需求随时间展现出不规则的变化，整体上具有明显的周期性特征，并在特定时间段内显示出早晚高峰现象。这种趋势反映了人们日常出行模式的一般规律，如上下班高峰期间的出行需求增加。通过细致观察，还可以发现真实值与预测值在数值范围和波动趋势上存在差异，但总体上保持一定的一致性，这表明本章构建的模型能够在一定程度上捕捉到出行需求变化的基本模式。

尽管如此，模型在预测具体的高峰值时仍存在一定的偏差，尤其是在真实数据表现出的突然增长或下降时。这可能是由于模型未能完全捕获到影响出行需求变化的所有因素，例如特殊事件、天气变化或节假日等。此外，真实值中的轻微偏差反映了现实世界数据的不确定性，这对于提高预测准确性提出了更高的要求。

综上所述，通过对区域 162 至区域 102 和区域 68 到区域 44 的出行需求分析，本研究不仅验证了模型在捕捉出行需求变化趋势上的有效性，也发现了模型在细节预测上存在的局限性。未来的研究可以通过引入更多相关因素和改进算法来提高预测的准确度和鲁棒性，以更好地服务于城市交通规划和管理。

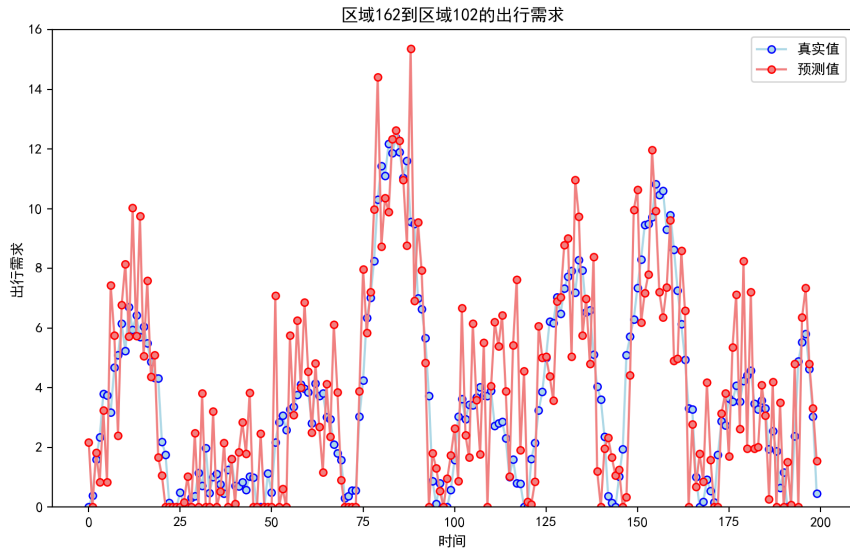


图 4.13 区域 162 至区域 102 预测与真实曲线对比

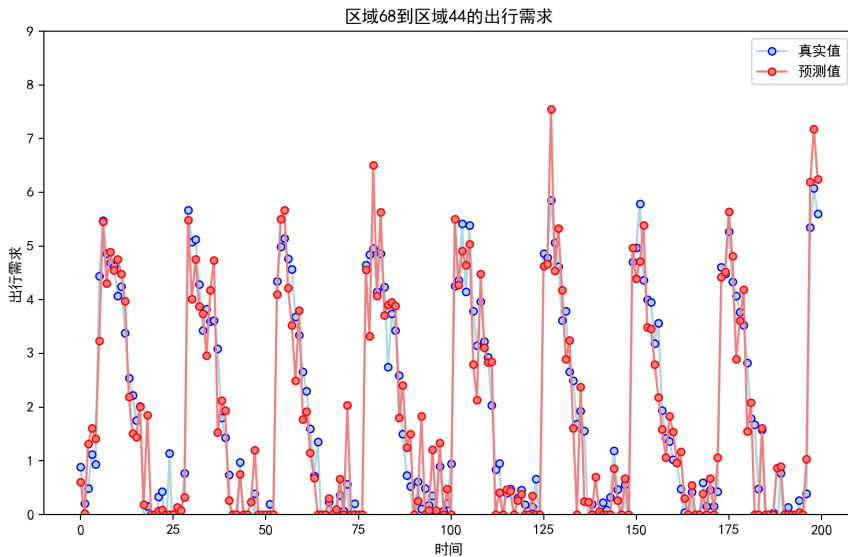


图 4.14 区域 68 至区域 44 预测与真实曲线对比

4.5 本章小结

本章针对在做出行需求预测时存在将起始地和目的地视为相同的语义实体、OD 数据复杂的空间依赖性问题和 OD 数据稀疏性的问题，提出了一种融合渐进图卷积与构造网络更新的 OD 需求预测模型。具体地，提出了一种由记忆融合机制支持的分层记忆更新模块，以集成连续时间细粒度时空依赖性和离散时间一般时空依赖性，这样的扩展使得模型能够更好地捕捉交通数据中的时空关系。通过图构造器和渐进图卷积网络用于测量简单空间邻接之外的节点相似性来学习隐藏于在线流量数据本身中的丰富语义。最后还设计了一个损失函数，优化了 OD 数据稀疏性的问题。通过对现实世界数据集的实验，本章证明了所提出的 PGC-ODMUP 模型具有较高的预测精度和鲁棒性。此外，通过消融实验，验证了 PGC-ODMUP 模型中各个模块的有效性。这意味着每个模块都对模型的整体性能做出了贡献，并且没有冗余的组件。通过逐步去除或修改模型中的各个模块，可以评估每个模块对预测精度的影响。

第 5 章 总结与展望

5.1 本文工作总结

近年来,随着我国经济快速发展和城市化水平的不断提高,城市建设和居民生活质量都有了显著的进步。公共交通设施的改善、住宅条件的优化以及城市服务功能的增强,共同推动了城市现代化进程的快速前进。不过,城市发展的同时也带来了一系列挑战,其中最为突出的就是交通拥堵问题。随着城市人口的急剧增长和机动车保有量的持续上升,城市交通系统面临前所未有的压力。这种状况不仅影响了市民的出行效率和生活质量,也给城市的可持续发展带来了挑战。在此背景下,提高城市交通流的预测准确性和实用性变得至关重要。通过对交通流量的精准预测,可以在一定程度上预见到交通拥堵的发生,从而为出行者提供实时的拥堵信息和绕行建议,帮助他们规切合理的出行计划,减少等待时间,提高出行效率。然而,一方面常见的交通流预测模型并没有充分挖掘路网中交通节点以及邻接矩阵的远程时空特征和隐藏的动态空间特性,导致了预测的准确度不够理想。特别是在长期预测上,这些模型往往无法准确捕捉未来交通流的变化,从而影响了预测结果的可靠性。另一方面,先前的方法在进行出行需求预测时,往往将起始地点和目的地视为同一种语义实体来处理,忽略了它们在空间分布和功能属性上的本质区别。此外,起始点与目的地数据之间复杂的空间依赖性,以及需求分布的不均衡性,进一步加剧了 OD 数据的稀疏性问题。这种稀疏性不仅给数据处理带来了难度,也极大地影响了模型学习和预测的准确性。这一系列因素为基于图卷积神经网络的交通流预测方法研究带来了挑战。

针对上述挑战,本文开展了基于图卷积神经网络的交通流预测方法研究,本文主要内容和创新点总结如下:

(1) 针对先前方法在做交通流预测时没有充分得挖掘出路网交通节点和邻接矩阵中的远程时空特征和隐藏动态空间特性问题,本文提出了一种交通流预测方法 (IDG-PSAtt), 该方法通过整合概率稀疏自注意力机制、动态图卷积和时空卷积技术,以及编解码器策略,优化了图卷积网络结构,实现了交通流动态时空特征的高效捕捉和交互式学习。具体而言, IDG-PSAtt 模型通过调整输入时空信息和预设初始邻接矩阵来创建动态图结构,模拟节点之间的动态关系,从而探索交通网络中不可见节点之间的动态关联并捕捉隐藏的时空相关性。同时, IL-DGCN 通过将 DGCN 块嵌入交互式学习框架中来构建,以同时学习交通流的周期特性和趋势,并捕捉时空依赖性。这种方法不仅提高了交通流量(速度)预测的精度,还增强了模型对交通流变化敏感性,解决了以往模型在长期预测和捕捉隐藏时空依赖性方面的不足。通过在公开的交通流量预测数据集上的实验验

证, 该方法在竞争性能上显著优于基线方法。

(2) 针对先前方法在做 OD 需求预测时存在将起始地和目的地视为相同的语义实体、OD 数据复杂的空间依赖性和 OD 数据稀疏性的问题, 本文提出了一种 OD 需求预测方法 (PGC-ODMUP)。该方法采用分层记忆更新将节点的时间感知表示分层处理, 以便更好地捕获不同时间尺度上的模式和趋势。还首次通过连续时间动态图视图对交通节点表示学习进行了扩展, 这样的扩展使得模型能够更好地捕捉交通数据中的时空关系。通过图构造器和渐进图卷积用于测量简单空间邻接之外的节点相似性来学习隐藏于在线流量数据本身中的丰富语义。引入随机游走采样模块, 区分原点和目的地为不同的语义实体, 以避免混淆两种类型的信息。最后, 还设计了一个损失函数, 优化了 OD 数据稀疏性的问题。实验结果显示, 在公开的交通数据集上, 该方法的性能超越了基准模型, 表现出卓越的竞争力。

5.2 未来展望

虽然本研究提出的交通流预测方法在预测性能上超越了现有的交通流预测模型, 但仍有一些局限性和改进空间需要进一步研究和优化:

(1) 交通流预测任务受到多种外部因素的影响, 例如天气、社会事件、道路施工等。这些因素的变化能够对交通流的预测产生直接或间接的影响, 从而增加预测任务的复杂度和挑战性。因此, 未来研究需深入探讨这些外部影响因素, 目的是为了增强模型的鲁棒性和预测准确性。

(2) 为了实现更广泛的应用和提升预测的整体性能, 有必要深入探索如何将现有模型有效扩展至更大规模的数据集。这不仅涉及提高模型处理大量数据时的时间效率, 确保快速响应和更新预测, 同时也要关注在处理庞大和可能更加复杂的数据时保持或提高预测精度的策略。

(3) 确保预测模型具有良好的泛化性和实际应用价值, 必须涉及对各类交通数据集的广泛分析和应用。这涵盖了来自多样化地理区域的数据, 包括但不限于城镇道路、快速干道和农村通路上的交通流数据, 及各种交通方式如小汽车、大众交通与步行等方式的数据。通过对这类多元化数据的深入训练与测试, 能够对模型进行全面的性能验证, 揭示其潜在缺陷, 并针对性地对模型进行调整和优化, 以增强其在多种场景下的鲁棒性和适用性。

(4) 需要系统的全面实施和布署。当前, 算法验证依托于静态数据。在接下来的研究中, 应当从整体系统的视角出发, 处理不同算法之间的集成问题, 以搭建和实施一个综合的交通流量预测系统。

参 考 文 献

- [1] 国家统计局. 中国统计年鉴[EB/OL]. 2023. <http://www.stats.gov.cn/sj/ndsj/2023/indexch.htm>.
- [2] 百度地图. 2023 年第 3 季度中国城市交通报告[EB/OL]. 2023. <https://jiaotong.baidu.com/>.
- [3] NJOKU J N, NWAKANMA C I, AMAIZU G C, et al. Prospects and challenges of metaverse application in data-driven intelligent transportation systems[J]. IET Intelligent Transport Systems, 2023, 17(1): 1-21.
- [4] ZULKARNAIN P T D. Intelligent transportation systems (ITS): A systematic review using a natural language processing approach[J]. Heliyon, 2021, 7(12): e08615.
- [5] SMITH B L, DEMETSKY M J. Traffic flow forecasting: comparison of modeling approaches [J]. Journal of Transportation Engineering, 1997, 123(4): 261-266.
- [6] 关为生, 肖建力. 联合时空特征的交通流参数预测综述[J]. 上海理工大学学报, 2022, 44(06): 592-602.
- [7] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results[J]. Journal of Transportation Engineering, 2003, 129(6): 664-672.
- [8] MIAO Y, BAI X, CAO Y, et al. A novel short-term traffic prediction model based on SVD and ARIMA with blockchain in industrial internet of things[J]. IEEE Internet of Things Journal, 2023, 10(24): 21217-21226.
- [9] KONTOPOULOU V I, PANAGOPOULOS A D, KAKKOS I, et al. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks[J]. Future Internet, 2023, 15(8): 255.
- [10] XU X, JIN X, XIAO D, et al. A hybrid autoregressive fractionally integrated moving average and nonlinear autoregressive neural network model for short-term traffic flow prediction[J]. Journal of Intelligent Transportation Systems, 2023, 27(1): 1-18.
- [11] HABTEMICHAEL F G, CETIN M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns[J]. Transportation Research Part C: Emerging Technologies, 2016, 66: 61-78.
- [12] ZHUANG W, CAO Y. Short-term traffic flow prediction based on a k-nearest neighbor and bidirectional long short-term memory model[J]. Applied Sciences, 2023, 13(4): 2681.
- [13] HU W, YAN L, LIU K, et al. A short-term traffic flow forecasting method based on the hybrid PSO-SVR[J]. Neural Processing Letters, 2016, 43: 155-172.
- [14] DAI G, TANG J, LUO W. Short-term traffic flow prediction: An ensemble machine learning

- approach[J]. Alexandria Engineering Journal, 2023, 74: 467-480.
- [15] OKWUASHI O, NDEHEDEHE C E. Deep support vector machine for hyperspectral image classification[J]. Pattern Recognition, 2020, 103: 107298.
- [16] SCANAGATTA M, SALMERÓN A, STELLA F. A survey on Bayesian network structure learning from data[J]. Progress in Artificial Intelligence, 2019, 8: 425-439.
- [17] MA D, ZHU J, SONG X B, et al. Traffic flow and speed forecasting through a Bayesian deep multi-linear relationship network[J]. Expert Systems with Applications, 2023, 213: 119161.
- [18] KUANG L, YAN H, ZHU Y, et al. Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor[J]. Journal of Intelligent Transportation Systems, 2019, 23(2): 161-174.
- [19] DERINGER V L, BARTÓK A P, BERNSTEIN N, et al. Gaussian process regression for materials and molecules[J]. Chemical Reviews, 2021, 121(16): 10073-10141.
- [20] GS V, VS H. Prediction of bus passenger traffic using gaussian process regression[J]. Journal of Signal Processing Systems, 2023, 95(2): 281-292.
- [21] QIAO Y, WANG Y, MA C, et al. Short-term traffic flow prediction based on 1DCNN-LSTM neural network structure[J]. Modern Physics Letters B, 2021, 35(02): 2150042.
- [22] ZHAO W, GAO Y, JI T, et al. Deep temporal convolutional networks for short-term traffic flow forecasting[J]. Ieee Access, 2019, 7: 114496-114507.
- [23] NARMADHA S, VIJAYAKUMAR V. Spatio-temporal vehicle traffic flow prediction using multivariate CNN and LSTM model[J]. Materials Today: Proceedings, 2023, 81: 826-833.
- [24] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 31. 2017: 1655-1661.
- [25] MÉNDEZ M, MERAYO M G, NÚÑEZ M. Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model[J]. Engineering Applications of Artificial Intelligence, 2023, 121: 106041.
- [26] GU Y, LU W, QIN L, et al. Short-term prediction of lane-level traffic speeds: A fusion deep learning model[J]. Transportation Research Part C: Emerging Technologies, 2019, 106: 1-16.
- [27] LIN Z, LI M, ZHENG Z, et al. Self-attention convlstm for spatiotemporal prediction[C]// Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 34. 2020: 11531-11538.
- [28] YU R, LI Y, SHAHABI C, et al. Deep learning: A generic approach for extreme condition traffic forecasting[C]//Proceedings of the 2017 SIAM International Conference on Data Mining. SIAM, 2017: 777-785.
- [29] BAI L, YAO L, LI C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17804-17815.

- [30] LUO X, LI D, YANG Y, et al. Spatiotemporal traffic flow prediction with KNN and LSTM [J]. *Journal of Advanced Transportation*, 2019, 2019: 1-10.
- [31] CHU K F, LAM A Y, LI V O. Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(8): 3219-3232.
- [32] SHU P, SUN Y, ZHAO Y, et al. Spatial-temporal taxi demand prediction using LSTM-CNN[C]//2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). IEEE, 2020: 1226-1230.
- [33] ZHAO L, SONG Y, ZHANG C, et al. T-gcn: A temporal graph convolutional network for traffic prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(9): 3848-3858.
- [34] ZHANG Z, LI M, LIN X, et al. Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies[J]. *Transportation Research Part C: Emerging Technologies*, 2019, 105: 297-322.
- [35] LI M, ZHU Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting[C]//*Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 35*. 2021: 4189-4196.
- [36] ZENG J, ZHANG G, RONG C, et al. Causal learning empowered OD prediction for urban planning[C]//*Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2022: 2455-2464.
- [37] LI D, LASENBY J. Spatiotemporal attention-based graph convolution network for segment-level traffic prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(7): 8337-8345.
- [38] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[M]//*Proceedings of the 28th International Joint Conference on Artificial Intelligence: Vol. 7*. 2019: 1907-1913.
- [39] ZHANG Q, CHANG J, MENG G, et al. Spatio-temporal graph structure learning for traffic forecasting[C]//*Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 34*. 2020: 1177-1185.
- [40] CHEN W, CHEN L, XIE Y, et al. Multi-range attentive bicomponent graph convolutional network for traffic forecasting[C]//*Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 34*. 2020: 3529-3536.
- [41] WANG Y, YIN H, CHEN H, et al. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019: 1227-1235.
- [42] SHI Z, ZHANG Y, WANG J, et al. DAGCRN: Graph convolutional recurrent network for

- traffic forecasting with dynamic adjacency matrix[J]. *Expert Systems with Applications*, 2023, 227: 120259.
- [43] WU Z, PAN S, LONG G, et al. Connecting the dots: Multivariate time series forecasting with graph neural networks[C]//*Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020: 753-763.
- [44] XU M, DAI W, LIU C, et al. Spatial-temporal transformer networks for traffic flow forecasting [A]. 2020.
- [45] SHI H, YAO Q, GUO Q, et al. Predicting origin-destination flow via multi-perspective graph convolutional network[C]//*2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020: 1818-1821.
- [46] ZHANG X, WEN S, YAN L, et al. A hybrid-convolution spatial-temporal recurrent network for traffic flow prediction[J]. *The Computer Journal*, 2024, 67(1): 236-252.
- [47] KE J, QIN X, YANG H, et al. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 122: 102858.
- [48] CHAI D, WANG L, YANG Q. Bike flow prediction with multi-graph convolutional networks [C]//*Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2018: 397-400.
- [49] LV M, HONG Z, CHEN L, et al. Temporal multi-graph convolutional network for traffic flow prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(6): 3337-3348.
- [50] PAN Z, LIANG Y, WANG W, et al. Urban traffic prediction from spatio-temporal data using deep meta learning[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019: 1720-1730.
- [51] JI J, WANG J, HUANG C, et al. Spatio-temporal self-supervised learning for traffic flow prediction[C]//*Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 37*. 2023: 4356-4364.
- [52] SONG C, LIN Y, GUO S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]//*Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 34*. 2020: 914-921.
- [53] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[A]. 2014.
- [54] YU Y, SI X, HU C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. *Neural Computation*, 2019, 31(7): 1235-1270.
- [55] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural

- networks on sequence modeling[A]. 2014.
- [56] OORD A V D, DIELEMAN S, ZEN H, et al. Wavenet: A generative model for raw audio[A]. 2016.
- [57] LI Y, YU R, SHAHABIC, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[A]. 2017.
- [58] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [A]. 2016.
- [59] XU K, HU W, LESKOVEC J, et al. How powerful are graph neural networks?[A]. 2018.
- [60] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[A]. 2017.
- [61] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15. Springer, 2018: 593-607.
- [62] FU T Y, LEE W C, LEI Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 1797-1806.
- [63] WANG X, JI H, SHI C, et al. Heterogeneous graph attention network[C]//The World Wide Web Conference. 2019: 2022-2032.
- [64] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in Neural Information Processing Systems, 2016, 29: 3844-3852.
- [65] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 33. 2019: 922-929.
- [66] DO L N, VU H L, VO B Q, et al. An effective spatial-temporal attention based neural network for traffic flow prediction[J]. Transportation Research Part C: Emerging Technologies, 2019, 108: 12-28.
- [67] ZHENG C, FAN X, WANG C, et al. Gman: A graph multi-attention network for traffic prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 34. 2020: 1234-1241.
- [68] SUN J, ZHANG J, LI Q, et al. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(5): 2348-2359.
- [69] CHEN C, LI K, TEO S G, et al. Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2020, 14(4): 1-23.

- [70] HUANG B, DOU H, LUO Y, et al. Adaptive spatiotemporal transformer graph network for traffic flow forecasting by IoT loop detectors[J]. *IEEE Internet of Things Journal*, 2022, 10(2): 1642-1653.
- [71] CHEN J, XU M, XU W, et al. A flow feedback traffic prediction based on visual quantified features[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(9): 10067-10075.
- [72] YAN H, MA X, PU Z. Learning dynamic and hierarchical traffic spatiotemporal features with transformer[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(11): 22386-22399.
- [73] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[M]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 2018: 3634-3640.
- [74] HAN L, DU B, SUN L, et al. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting[C]//*Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021: 547-555.
- [75] ZHANG J, CHEN F, WANG Z, et al. Short-term origin-destination forecasting in urban rail transit based on attraction degree[J]. *IEEE Access*, 2019, 7: 133452-133462.
- [76] SAYED S A, ABDEL-HAMID Y, HEFNY H A. Artificial intelligence-based traffic flow prediction: a comprehensive review[J]. *Journal of Electrical Systems and Information Technology*, 2023, 10(1): 13.
- [77] ZHANG J, ZHENG Y, SUN J, et al. Flow prediction in spatio-temporal networks based on multitask deep learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 32(3): 468-478.
- [78] CHEN T, GUESTRIN C. Xgboost: A scalable tree boosting system[C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016: 785-794.
- [79] ZHANG J, CHE H, CHEN F, et al. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 124: 102928.
- [80] ROSSI E, CHAMBERLAIN B, FRASCA F, et al. Temporal graph networks for deep learning on dynamic graphs[A]. 2020.

致 谢

在我这段漫长而又充满挑战的硕士旅程中，有太多的感激之情需要表达。每个帮助、支持和鼓励我的人都是我能够顺利完成这一重要学术任务的重要原因。

首先，我必须深深感谢我的导师们，他们的智慧如同灯塔，为我照亮了前进的道路。赵云波教授，您的深邃的学术造诣和严谨的学术态度，为我的研究指明了方向。尹航老师，您的细致指导和悉心关怀使我在学术探索中免于许多弯路。黄志辉老师，您对我的研究项目提出了宝贵的建议和见解，使我的论文更加深入和全面。何卓识博士，您在我遇到困难时给予的鼓励和支持，让我有勇气面对挑战。您们的共同努力，不仅让我的学术之路充满了光明，也让我学会了如何成为一名真正的学者。

其次，我要感谢我的室友们，你们是我在这段学术旅程中最宝贵的朋友。我们共同度过的日夜，你们的理解、支持和陪伴，使我在写作论文的压力之下仍感到温暖和力量。我们一起讨论问题，共同解决困难，这些宝贵的经历将成为我人生中不可磨灭的记忆。

再次，我要向我的父母表达最深的感谢。是你们无条件的爱和支持，给了我追求学术梦想的勇气和力量。在我遇到挑战和困难时，你们总是我的坚强后盾，用你们的鼓励和关怀为我加油。你们的理解和牺牲，让我能够专心于我的研究，从而顺利完成这篇论文。

最后，虽然目前的我还在学术旅程的起点，但我将不断努力，希望有一天能够站在更高的地方，回望过去，对所有支持和帮助过我的人表达我的最深的感激。

感谢所有帮助和支持过我的人，是你们让我的学术之旅充满了意义和价值。在未来的日子里，我将继续努力，不负众望。

在读期间发表的学术论文与取得的研究成果

已发表论文

1. Z Ding, Z He, H Yin, et al. Traffic Flow Prediction Research Based on an Interactive Dynamic Spatial-Temporal Graph Convolutional Probabilistic Sparse Attention Mechanism (IDG-PSAtt)[J]. Atmosphere, 2024, 15(4): 413.