



兰州大学

本科毕业论文

论文题目 (中文) 基于联邦学习的交通流量
预测方法研究

论文题目 (英文) Research on Method of Traffic Flow Prediction
Based on Federated Learning

学生姓名 苗洪磊

指导教师 焦栋斌 讲师

学 院 信息科学与工程学院

专 业 人工智能

年 级 2019 级

兰州大学教务处

诚信责任书

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人、集体已经发表或未发表的论文。

本声明的法律责任由本人承担。

论文作者签名：_____ 日 期：_____

关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文（设计）的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文（设计）。本人离校后发表、使用毕业论文（设计）或与该毕业论文（设计）直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文（设计）研究内容：

☒ 可以公开

☐ 不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名：_____ 导师签名：_____

日 期：_____ 日 期：_____

基于联邦学习的交通流量预测方法研究

中文摘要

随着人民生活水平的提高，车辆保有量不断增加，这对交通基础设施提出了更高的要求。为了缓解城市交通压力，促进智能交通系统的发展，对交通状况进行准确预测已经成为交通管理中亟待解决的问题。这有助于为出行者提供及时准确的交通信息，从而缓解交通拥堵，提高燃油效率。然而，交通预测在处理由交通站点和移动设备收集的交通数据时，会遇到数据收集和隐私保护问题。为了解决上述问题，采用联邦学习架构进行交通站点之间的数据共享。在此背景下，针对客户端在计算能力具有不同的资源约束的问题，本文研究改进的联邦学习参数更新方式，提出了基于异构联邦学习的交通流量预测方法，主要的研究内容总结如下：

本文首先通过道路网络和特征矩阵定义了交通流量预测问题，然后，使用基于谱的图卷积神经网络如 ChebNet 网络和 GCN 网络对交通道路网络的空间特征进行提取，考虑到交通流量预测本质上是一个时序预测问题，还采用 GRU 模型来提取交通流量数据的时间特征。最后，采用联邦学习架构进行交通预测模型的训练，并考虑客户端计算能力的异构，采用改进的联邦学习模型聚合算法 FedProx，在客户端异质性模拟实验中验证了其相对于传统的联邦学习算法 FedAvg 的优越性。

关键词：交通流量预测；图神经网络；门控循环单元；异构；联邦学习。

Research on Method of Traffic Flow Prediction Based on Federated Learning

Abstract

With the improvement of people's living standards and the continuous increase of vehicle ownership, higher requirements have been put forward for transportation infrastructure. In order to alleviate urban traffic pressure and promote the development of intelligent transportation systems, accurate prediction of traffic conditions has become an urgent problem in traffic management. This helps provide timely and accurate traffic information for travelers, thereby alleviating traffic congestion and improving fuel efficiency. However, when dealing with traffic data collected by traffic stations and mobile devices, traffic prediction will encounter data collection and privacy protection issues. To solve these problems, a federated learning architecture is used for data sharing between traffic stations. The use of a federated learning architecture for data sharing between traffic stations can solve the above problems well. Against this background, this paper studies an improved parameter updating method for federated learning and proposes a traffic flow prediction method based on heterogeneous federated learning. The main research contents are summarized as follows:

First, this paper defines the traffic flow prediction problem through road network and feature matrix, and then uses spectrum-based graph convolutional neural networks such as ChebNet and GCN to model the spatial features of the traffic road network. Considering that traffic flow prediction is essentially a time-series prediction problem, GRU model is also used to extract the time features of traffic flow data. Finally, the federated learning architecture is used for training the traffic prediction model, and considering the heterogeneity of client computing power, the improved federated learning model aggregation algorithm FedProx is used. Its superiority over the traditional federated learning algorithm FedAvg is verified in client heterogeneity simulation experiments.

Keywords: Traffic flow prediction, graph neural network, gated recurrent unit, heterogeneity, federated learning.

目 录

中文摘要	I
英文摘要	II
第一章 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 基于深度学习的交通流量预测	2
1.2.2 异构联邦学习	2
1.3 研究内容	3
1.4 组织结构	3
第二章 基于联邦学习架构的交通流量预测模型	5
2.1 交通流量预测问题	5
2.1.1 道路网络	5
2.1.2 特征矩阵	5
2.2 空间特征提取	6
2.2.1 ChebNet 网络	6
2.2.2 图卷积神经网络 (GCN)	7
2.3 时序特征提取	8
2.4 联邦学习架构	9
2.5 本章小结	9
第三章 异构客户端下联邦学习模型聚合改进算法	10
3.1 传统的联邦学习模型聚合算法	10
3.2 改进的联邦学习模型聚合算法	11
3.3 本章小结	13
第四章 实验结果分析	14

4.1 实验环境.....	14
4.2 实验数据与处理	14
4.3 评价指标.....	15
4.4 实验结果与分析	15
4.5 本章小结.....	18
第五章 总结与展望.....	19
5.1 论文总结.....	19
5.2 研究展望.....	20
参考文献	21
致 谢	23

图 目 录

图 1.1	主要内容	3
图 2.1	交通流量预测问题	5
图 2.2	GCN 模型	7
图 2.3	GRU 网络	8
图 2.4	联邦学习架构	9
图 4.1	不同联邦算法在 PeMS08 上 10%、50%、90%异质性模拟的损失收敛	16

表 目 录

表 4.1 实验环境 14

表 4.2 数据集介绍 14

表 4.3 联邦算法在交通数据集上的性能评估 17

表 4.4 联邦算法在不同数据集不同比例异质性模拟下的平均性能评估 18

第一章 绪 论

1.1 研究背景与意义

随着人民生活水平的提高,车辆保有量不断增加,这对交通基础设施提出了更高的要求。目前,我国各大中小城市面临车辆保有量越来越多,道路交通越来越拥堵以及交通事故频发等问题。为此,政府投入了大量人力物力,以便充分利用现有资源,缓解城市交通压力。然而,传统的交通管理方式中或多或少存在着一些弊端,难以解决现有交通管理中的矛盾。

2019年,政府工作报告提出了“新基建”的发展规划。智能交通系统(Intelligent Transportation System, ITS)^[1]作为新基建的重要组成部分,是一种融合基础设施的技术。这类基础设施与传统的基础设施具有很大区别,其深度应用互联网、大数据、人工智能等现代先进技术,是由传统的基础设施面向现代化需求转型升级而形成,服务于国家长远发展,也是国内交通管理系统的未来重点发展方向^[2]。

为了促进智能交通系统的发展,必须对交通状况进行准确预测,如交通流量和速度。这是因为,这些数据可以帮助司机做出有效的出行决定,从而缓解交通拥堵,提高燃油效率,并减轻空气污染。例如,在车联网中,准确的交通流量预测可以让驾驶员了解周围的道路条件并有效地提高驾驶体验^[3];在导航系统中,通过准确实时的旅行时间预测,驾驶员可以获得更智能的导航决策,以避免进入可能繁忙的道路^[4]。这些可以使交通预测在先进的旅客信息系统、交通管理系统和商业车辆运营中发挥主要作用,而这正是ITSs的目标^[5]。

为了达到这个目标,交通预测必须处理由交通站点和移动设备收集的实时和历史交通数据。例如,电感回路可以通过读取电感随时间的变化来测量行驶速度,这些数据可以用于交通速度预测。此外,移动设备如车载全球定位系统和电话的广泛使用使得普通公众的移动数据可以被获取,进一步促进了前所未有的交通数据收集。这种新兴的大数据可以在覆盖面和保真度方面大大增加数据的可用性,并大大促进对交通预测的研究^[6-10]。

由于交通站点可用的本地存储器有限,交通站点在其上游和下游站点获取数据时,直接共享交通数据会引起隐私泄露和通信成本等问题。谷歌在2016年提出的联邦学习^[11]在假设交通站点资源完全相同以及所存储数据是独立同分布的情况下可以很好地解决上述问题。然而,实际上不同交通站点数据大小和计算资源都存在差异^[12],交通数据也往往是非独立同分布的^[13],传统联邦学习算法难以有效地解决这些不同客户端的差异。如何更有效地在联邦学习架构下处理分布式下异构的交通数据,这是当前交通流量预测研究中亟待解决的重要问题。本文提出的基于异构联邦学习的交通流量预测方法,在保证普通公众数据隐私的前提下,重点研究分布式下异构的交通流量预测问题。

1.2 国内外研究现状

本文针对交通流量预测时遇到的实际问题,采用联邦学习架构进行交通站点之间的数据共享,并考虑客户端计算能力异构时,采用改进的联邦学习参数更新方式,提出了基于联邦学习的交通流量预测方法。本文的研究侧重于基于深度学习的交通流量预测和异构的联邦学习,下面将从这两方面对国内外的研究现状进行阐述。

1.2.1 基于深度学习的交通流量预测

关于交通预测的现有技术主要可以分为两类。第一类侧重于使用参数方法,如自回归综合移动平均模型 (ARIMA)^[6] 和卡尔曼滤波模型^[7]。当处理只呈现规律性变化的交通时(例如,在早晚高峰期经常发生的交通拥堵),参数化方法可以取得比较好的预测结果。然而,由于道路交通的随机性和非线性,使用参数化方法的交通预测可能会偏离实际值,特别是在突发性交通事件中^[8]。因此,与其像参数方法那样使用数学模型处理交通数据,不如使用非参数方法,其中基于深度学习 (Deep Learning, DL) 的方法最受欢迎^[8-10]。例如,文献[8]中提出了一个堆叠的自动编码器模型来学习通用的交通流特征进行预测。作者在文献[9]中研究了使用长短期记忆 (long short-term memory, LSTM) 和循环神经网络 (Recurrent Neural Network, RNN) 基于车站及其上下游交通站点收集的数据来预测交通流量、速度和占用率。

随着 RNN 的使用,文献[10]中的工作还利用卷积神经网络 (Convolutional Neural Network, CNN) 来捕捉底层道路网络中潜在的交通演变模式。由于复杂的耦合关系和动态变化的时空特征变化,交通流量预测比其他时间序列预测问题更加困难。图神经网络 (Graph Neural Network, GNN) 模型在交通预测问题上很受欢迎,因为它考虑了交通数据的时空特征和自然图的结构属性^[14,15]。

1.2.2 异构联邦学习

然而,当全局 GNN 模型在实践中面对大流量数据时有许多问题,如耗时的训练、高的通信成本。在合作和共享数据时,不同的数据提供者在合作和共享过程中交换数据时,还存在个人隐私泄露的潜在风险,因为拓扑结构信息也可能包含敏感数据,比如传感器的数量^[16]。联邦学习 (Federated Learning, FL) 是谷歌在 2016 年提出的一种分布式优化算法^[11,17,18]。本地参与者首先用全局共享模型进行训练,然后再向中央服务器提交更新的参数。由于 FL 算法是在本地训练的,只有模型参数被上传,原始数据的隐私可以得到保护,同时降低通信成本,分布式优化也降低了训练的时间成本。

尽管文献[8-10,14-16] 中的工作侧重于使用先进的 DL 模型进行交通预测,由于交通站点可用的本地存储器有限,本地训练数据可能极其不足,无法保证比较好的预测性能。再者,所有这些工作都假定交通站点收集的数据可以与其他站点单元共享,如文献[9]中交通站点从其上游和下游站点获取数据。在现实中,收集的数据可能包含个人信息,如摄像头拍摄的驾驶执照和移动电话用户的历史轨迹。在这种情况下,车站之间直接共享交通

数据会引起隐私问题。同时,通信成本是另一个主要问题。受 DL 和联邦学习^[19]结合的启发,FL 和 GNN 的结合可以很好地解决上述问题^[20]。

在联邦学习中,异构性包括两方面,一是每个客户端的系统特征存在显著的差异,即系统异构性;二是各交通站点中分布的数据不完全相同,即统计异构性^[13]。由于没有考虑到每个客户端的异构数据大小,例如,如果一个客户比其他客户有更多的数据,除非它有更好的计算资源,否则该客户将需要更长的时间来更新模型,这将延迟后续上传新模型参数的通信^[12]。同时由于原始的交通数据是非独立同分布的,在进行参数训练时,也必须考虑如何处理本地训练过程中的异构性,以保证多个客户端上传到服务器时数据都是一致的。所有这些关于异构数据的问题都将成为 FL 训练过程中的瓶颈。如何更有效地在联邦学习架构下处理分布式下异构的交通数据,这也是当前交通流量预测研究中亟待解决的重要问题。在本文中,主要考虑的是系统异构性。

1.3 研究内容

本文的主要研究内容如图 1.1 所示,本文考虑了交通流量预测问题及其在联邦学习架构下可能遇到的异构问题。针对该问题,本文采用了一种联邦学习算法 FedProx,该算法通过改进联邦学习全局模型聚合的方法来实现。具体来讲,本文首先实现了在集中式数据集下基于 GNN 的交通流量预测,再继续考虑分布式数据集,通过联邦学习架构完成模型的训练,最后针对联邦学习架构下异构问题,采用了改进的联邦学习模型聚合方法。本文基于 Python 编程语言进行算法实现,使用 Pytorch 框架来构建并训练交通流量预测模型,最终通过交通流量的公开数据集来验证模型与算法的有效性。

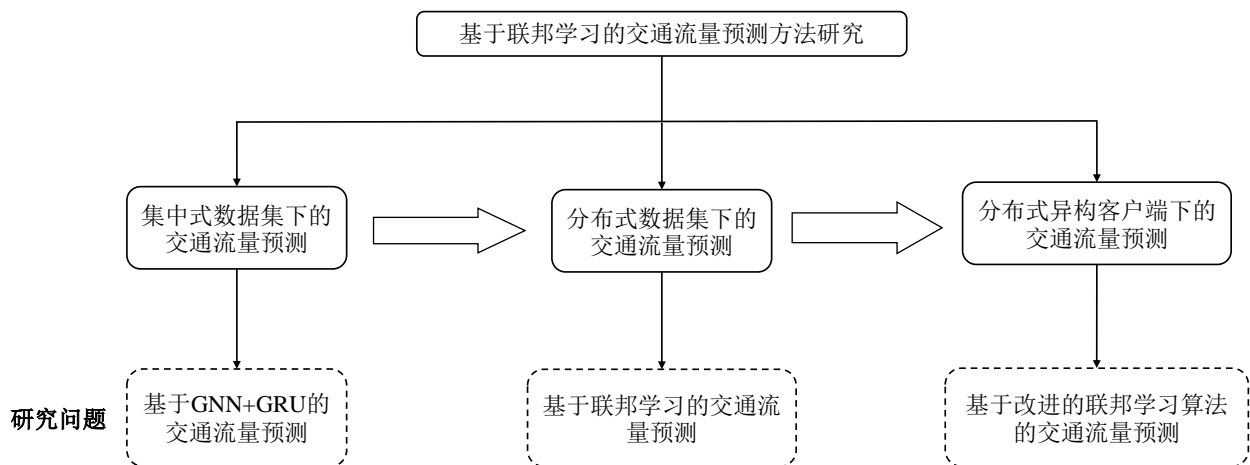


图 1.1 主要研究内容

1.4 组织结构

本文共有 5 个章节,具体结构如下:

第一章为绪论部分。首先说明了交通流量预测在智能交通系统中的重要性；其次综述了基于深度学习的交通流量预测的研究进展和存在的局限性，并简单阐述了基于联邦学习架构的交通流量预测的研究现状，引出了本文的研究内容是异构数据下基于联邦学习的交通流量预测；最后介绍了本文的组织结构。

第二章为模型建立部分。首先描述了交通流量预测问题，建立了进行交通流量预测的图神经网络模型和所采用的联邦学习架构。

第三章为训练算法部分。首先对模型训练的关键细节进行了说明；随后介绍了一种改进的联邦学习模型聚合算法，并给出了算法的伪代码，对算法流程进行了详细解释。

第四章为实验结果部分。介绍了交通流量数据集和实验的环境及设置，并对相应的实验结果进行了分析。

第五章为总结与展望部分。对全文进行了总结，并对未来的研究方向作出展望。

第二章 基于联邦学习架构的交通流量预测模型

2.1 交通流量预测问题

在本文中，使用历史交通流量信息预测未来交通流量是至关重要的。我们通过定义道路网络和特征矩阵来描述一个具体的交通流量预测问题，如图 2.1。

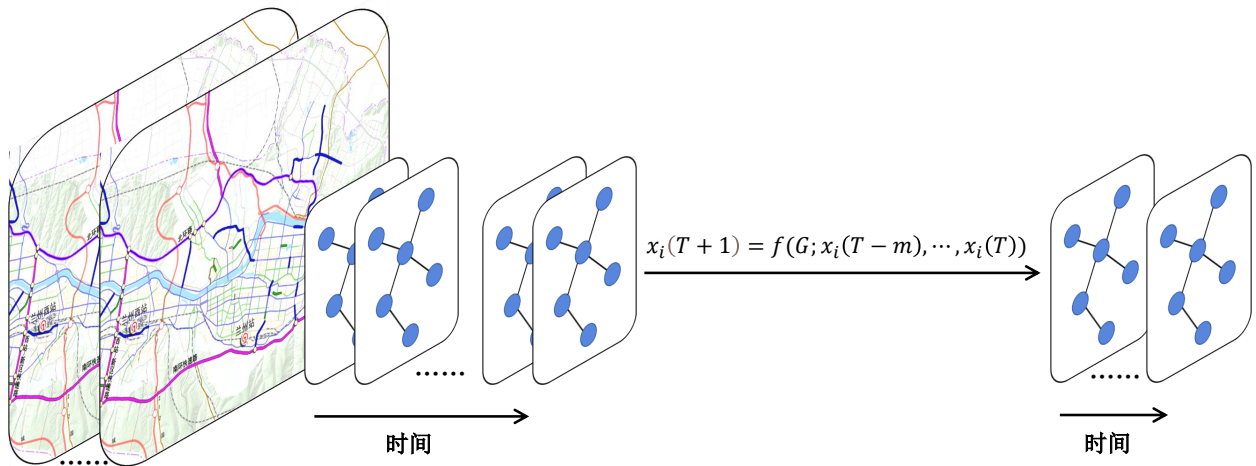


图 2.1 交通流量预测问题

2.1.1 道路网络

为了将道路网络中每个路段的空间拓扑信息纳入以提高预测的效果，本文将每条道路视为一个节点，并将道路网络定义为一个无向加权图 $G = (V, E)$ ，其中 $V = [v_1, v_2, \dots, v_N]$ 是道路节点集合， N 是节点数， E 是边集。此外，邻接矩阵 $A^{N \times N}$ 用于表示每个边的权重。当连接节点 i 和节点 j 之间的连接边上存在权重 $W(i, j)$ 时，我们设置 $A_{ij} = A_{ji} = W(i, j)$ ；否则，我们将邻接矩阵设置为零。

2.1.2 特征矩阵

我们将道路网络上的交通流信息视为网络中节点的特征，并表示为 $X^{N \times F}$ ，其中 F 表示节点特征数， $X_i(t)$ 表示时间 t 时刻道路 i 上的交通流量。因此，交通流量预测问题被视为基于道路网络拓扑 G 和特征矩阵 X 学习映射函数 f ，然后计算下一时刻的交通流量信息，如公式 2.1 所示：

$$X_i(T+1) = f(G; X_i(T-m), \dots, X_i(T)), \quad (2.1)$$

其中 m 是滑动窗口参数，这表示历史交通流量的间隔大小， T 是以前时间戳的数量。

2.2 空间特征提取

图是一种数据格式,通过点和边来描述个体以及个体之间的关系。近年来,图神经网络在非欧几里得空间中用图对空间特征提取方面显示出了巨大的潜力,包括图像分类^[21-24],节点分类^[25],流量预测^[26,27]。在本文的交通流量预测模型中,主要用图神经网络模型对交通道路网络的空间特征进行提取,下面是本文中使用的图神经网络模型。

2.2.1 ChebNet 网络

图卷积方法主要分为两类:基于谱的方法和基于空间的方法^[28]。从图信号处理的角度来看,基于谱的方法通过引入滤波器来定义图卷积。因此,基于频谱的图卷积可以理解为从图信号中去除噪声^[29]。基于空间的图卷积方法通过收集邻居节点的信息来构建图卷积。当图卷积在节点级别运行时,图池化模块和图卷积可以交错重叠;因此,该图可以粗略地处理成高级子图^[30]。目前,基于谱的图卷积网络主要依赖拉普拉斯矩阵的特征分解。本文所采用的图神经网络模型都是基于谱的。

对于无向图,设 A 为图的邻接矩阵, D 为节点度的对角矩阵, $D_{ii} = \sum_j (A_{i,j})$ 。图拉普拉斯矩阵定义为 $L = D - A$, 拉普拉斯矩阵的规范化定义为 $L = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I_n - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, 它是图论中图的矩阵表示。 L 是对称的和半正定的,归一化拉普拉斯矩阵 L 可以被分解为 $L = U \Lambda U^T$, 其中 $U \in \mathbb{R}^{N \times N}$ 是按特征值排序的特征向量矩阵, Λ 是特征值的对角矩阵。

谱图卷积运算是在傅里叶域中通过计算图拉普拉斯矩阵的特征分解来定义的。 $x \in \mathbb{R}^N$ 是图中节点的特征向量。图的傅里叶变换到 x 的定义为 $F(x) = U^T x$ 。傅里叶变换将输入图的 x 投影到正交空间中,这相当于将图上定义的任意特征向量 x 表示为拉普拉斯矩阵的特征向量的线性组合。图的反傅里叶变换定义为 $F^{-1}(\hat{x}) = U \hat{x}$, 其中 \hat{x} 是 x 通过图的傅里叶变换得到的输出。将卷积定理应用于图的傅里叶变换,图上的谱卷积被定义为 x 与滤波器 $g \in \mathbb{R}^N$ 在频域的乘法,可以表示为式 2.2:

$$x * g = F^{-1}(F(x) \odot F(g)) = U \left(U^T x \odot U^T g \right), \quad (2.2)$$

其中 $*$ 表示卷积运算, \odot 表示 Hadamard 积。对于维数为 $m \times n$ 的两个矩阵 A 和 B , Hadamard 积 $(A \odot B)$ 是一个与操作数相同维数的矩阵,其元素由 $(A \odot B)_{ij} = (A)_{ij} (B)_{ij}$ 给出。将滤波器 g 定义为 $g_\theta = \text{diag}(U^T g)$, 式 2.2 可简化为式 2.3, 这里我们可以将 g_θ 理解为 L 特征值的函数,即 $g_\theta(\Lambda)$ 。

$$x * g_\theta = U g_\theta U^T x, \quad (2.3)$$

式 2.3 的计算代价比较大,为了解决这个问题,Defferrard 等人提出了 ChebNet^[23], 它使用特征值对角矩阵的 Chebyshev 多项式来近似 g_θ , 将 g_θ 参数化为 Λ 的 K 阶多项式 2.4:

$$g_\theta = \sum_{i=0}^K \theta_i T_i(\tilde{\Lambda}), \quad (2.4)$$

其中 $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - I_N$, λ_{\max} 表示 L 的最大特征值。Chebyshev 多项式递归定义为 $T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x)$ 其中 $T_0(x) = 1$ $T_1(x) = x$ 。现在 x 与滤波器 g_θ 的卷积可以定义为式 2.5:

$$x * g_\theta = \sum_{i=0}^K \theta_i T_i(\tilde{L})x, \quad (2.5)$$

式中 $\tilde{L} = 2L/\lambda_{\max} - I_N$ 。因为 Chebyshev 多项式在 $[-1, 1]$ 上形成正交基, \tilde{L} 表示特征值从 $[0, \Lambda_{\max}]$ 映射到 $[-1, 1]$, θ_i 。

对于图 $G(V, E)$, ChebNet 网络可以表示为式 2.6:

$$\text{ChebNet}(x, A) = \sigma\left(\sum_{i=0}^K \theta_i T_i(\tilde{L})x\right), \quad (2.6)$$

$\sigma(\cdot)$ 是激活函数。

2.2.2 图卷积神经网络 (GCN)

Kipf 等人提出的图卷积网络 (GCN) 是在 ChebNet 基础上的进一步改进, 将 ChebNet 网络中的切比雪夫多项式 K 设置为 1, 限制了感受野的大小。作为 ChebNet 的一阶近似, GCN 不仅提高了大规模神经网络的性能, 而且提高了可扩展性 [25]。

图卷积操作有三个步骤 [31], 具体如图 2.2 所示。首先, 每个节点将其特征信息传递给其邻居节点; 然后, 每个节点收集其邻居节点和自身的特征信息以融合局部结构; 最后, 在图卷积中加入激活函数, 对节点信息进行非线性变换, 增强模型的表达能力。

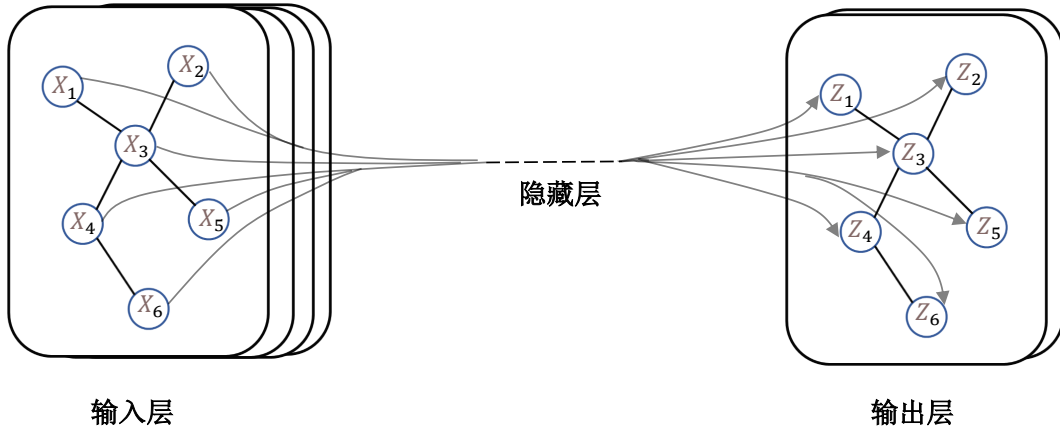


图 2.2 GCN 模型

因此, GCN 的关键是学习一个函数, 该函数可以聚合当前节点与其邻居节点的特征信息。节点之间的连接及其长度也会影响相关性。两层 GCN 模型可表示为公式 2.7:

$$f(X, A) = \sigma\left(\hat{A}\sigma\left(\hat{A}XW^{(0)}\right)W^{(1)}\right), \quad (2.7)$$

其中 $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ 表示对自连接矩阵 A 和度矩阵 D 的邻接矩阵进行预处理, $W^{(0)}$ 表示输入层与隐藏层之间的权重矩阵, $W^{(1)}$ 为可训练权值, $\sigma(\cdot)$ 是激活函数。

2.3 时序特征提取

由于交通流量预测本质上是一个时序预测问题,因此循环神经网络(RNN)、长短期记忆网络(LSTM)^[32]和门控循环单元(GRU)^[33]等模型,广泛用于交通流量预测研究。近年来,这些模型也与图神经网络模型相结合^[34,35],共同提取交通流量数据的时空特征。为了有效地利用空间和时间因素之间的相互影响,文献[34]利用时空块从空间和时间维度进行数据分析,并提出了时空图卷积网络(STGCN)模型。在STGCN模型中,时间卷积层由CNN实现,图卷积层由GCN实现,试图更好地分析附近时刻和位置之间的相关性。为了更灵活地处理动态变化的交通网络,文献[35]提出了一种图注意力长短期记忆网络(GAT-LSTM)来预测交通流量,它使用GAT更新空间信息,使用LSTM概括隐藏在历史交通数据中的时间模式,以进行动态时空分析。

最常用的提取自然序列时间特征的模型是RNN。然而,随着时间序列变得太长,输入信息变得更大,训练RNN收敛到全局最优成为一项具有挑战性的任务。GRU和LSTM是特殊的循环神经网络,它们改变了RNN的隐藏层,以更好地捕获深度连接,并有效改善由于长序列而导致的梯度消失问题。由于GRU的内部结构比LSTM更简单,训练时间也更短。因此,本文采用GRU网络来提取交通流量数据的时间特征。

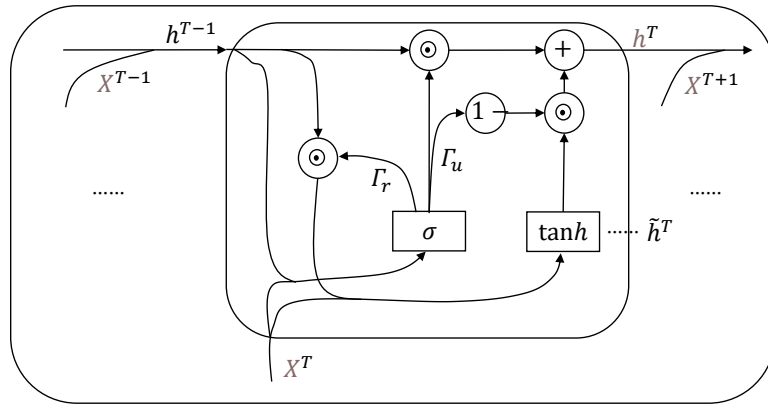


图 2.3 GRU 网络

本文所考虑的GRU的结构如图2.3所示,其中 h^{T-1} 表示 $T-1$ 时刻的隐藏状态, X^T 表示 T 时刻的交通流量特征。 T 时刻的隐藏状态 h^T 通过更新门 Γ_u 确定是保存为前一时刻的隐藏状态 h^{T-1} 还是更新为当前的候选隐藏状态 \tilde{h}^T ,其中 σ_u 是一个使自身等于0或1的值的函数。另外, Γ_r 是一个重置门,用于控制先前状态 h^{T-1} 是否被忽略。GRU的结构能够捕获远程依赖关系,适合提取长相关序列的时间特征,这对于预测随时间变化的交通流量非常适合。

2.4 联邦学习架构

联邦学习^[36]是一种可以避免隐私泄露的分布式机器学习方法。在联邦学习中，不同的组织可以参与全局模型的训练，而不需要将自己唯一的数据上传到云端。使用联邦学习，我们可以有效地减少网络波动的影响，避免模型训练过程中大量的数据传输，这样也节省了训练时间。在本文中，具体的联邦学习架构如图 2.4 所示。

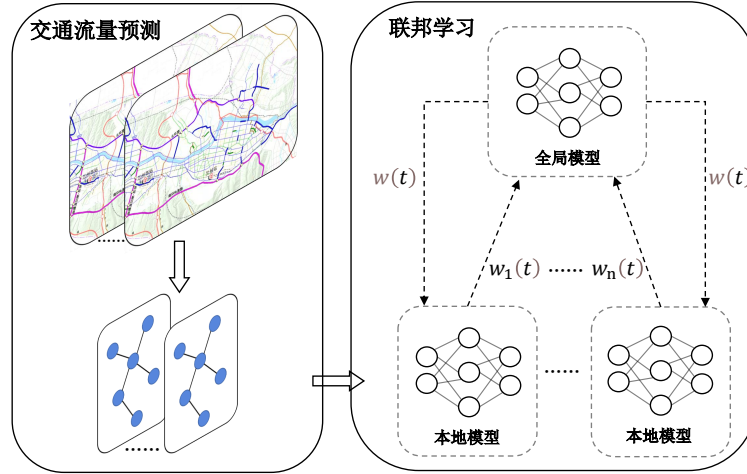


图 2.4 联邦学习架构

在传统的联邦学习方法中，一般假设有一组内存大小为 k 的计算设备 E ，每个设备存储其本地数据 L_k ，大小为 D_k 。因此总数据量为 $D = \sum_{k=1}^n D_k$ 。对于某一计算设备 k ，用 $\{x_i, y_i\}_{i=1}^{D_k}$ 表示子模型的输入和标签。那么子模型在计算设备 k 上的损失可以表示为 $f_i(w, x_i, y_i)$ 。该装置上子模型的损失函数可表示为公式 2.8，

$$J_k(\omega) = \frac{1}{D_k} \sum_{i=1}^{D_k} f_i(\omega) + \lambda h(\omega). \quad (2.8)$$

其中， $\omega \in \mathbb{R}^d$ 为子模型参数， $\lambda h(\omega)$ 为正则化项。云通过聚合每个设备上的子模型参数来更新全局模型。

全局模型的损失函数如公式 2.9，

$$\arg \min_{\omega \in \mathbb{R}^d} J(\omega) = \sum_{k=1}^n \frac{\sum_{i=1}^{D_k} f_i(\omega) + \lambda h(\omega)}{D}. \quad (2.9)$$

2.5 本章小结

本章首先通过道路网络和特征矩阵定义了交通流量预测问题，在本地的交通预测模型训练中，使用基于谱的图卷积神经网络如 ChebNet 网络和 GCN 网络对交通道路网络的空间特征进行提取，考虑到交通流量预测本质上是一个时序预测问题，还采用 GRU 模型来提取交通流量数据的时间特征，这样本文的交通预测模型就可以通过提取交通流量的时空特征进行预测。最后，采用联邦学习架构进行交通预测模型的训练。

第三章 异构客户端下联邦学习模型聚合改进算法

3.1 传统的联邦学习模型聚合算法

联邦学习^[36]是为了引入一种新的集中式分布式训练方法而提出的，该方法允许多个客户端在其共同的数据上协同训练深度神经网络模型，而任何参与者都不必向中央服务器或其他参与者透露其本地数据。

算法 1: FedAvg 算法

输入: N (客户端总数), η (每个客户端的学习率), S (包含所有客户端的集合), E (通信间隔), C (客户端的比例), w_0 (初始全局模型参数)。

输出: w_n (经过 n 次更新的全局模型参数)。

// 中心服务器执行

```

1 初始化: 全局模型  $w_0$ ;
2 for 每次全局迭代  $t \in 0, \dots, n-1$  do
    // 确定参与的客户端数量。
3      $m \leftarrow \max(C \cdot N, 1)$ ;
    // 随机选择参与的客户端。
4      $S_p = \text{random.choice}(S, m)$ ;
5     forall 客户端  $k \in S_p$  并行 do
        // 获得客户端更新的模型。
6          $w_{t+1}^k \leftarrow \text{TrainLocally}(k, w_t)$ 
7     end
    // 更新全局模型。
8      $w_{t+1} \leftarrow \sum_{k=0}^N \frac{D_k}{D} w_{t+1}^k$ 
9 end
10 Function  $\text{TrainLocally}(k, w_t)$  is
11     for 第  $k$  个客户端迭代  $e \in 0, \dots, E-1$  do
        // 进行本地模型训练
12          $w_{e+1} \leftarrow w_e - \eta \nabla J_k(w_e)$ 
13     end
14     return  $w_{e+1}$ 
15 end

```

虽然在机器学习背景下有大量关于分布式优化的工作，但将联邦学习与传统分布式优化区分出来的两个关键挑战是：高度的系统异构和数据异构。为了处理异构性和解决高通信成本，允许本地更新和低参与的优化方法是联邦学习的流行方法。特别是 Federated Averaging(即 FedAvg) 的联邦学习模型聚合方法已成为联邦学习设置中最常用的优化方法，算法 1 总结了 FedAvg 的细节。第 1 行到第 9 行是中心服务器的执行，首先初始化全局模型，每次全局迭代都确定参与的客户端数量，并随机选择参与的客户端，把所有接收到的本地模型参数的均值作为全局模型的参数。第 10 到第 15 行是本地更新的步骤，每个本地客户端将接收到的全局模型按照本地数据进行梯度下降更新，并将训练完的本地模型参数发给中心服务器。在 FedAvg 中，设备 k 处全局目标函数的本地代理为 $J_k(\cdot)$ ，本地求解器为随机梯度下降 (SGD)，在每个设备上使用相同的学习率和本地 epoch 数。在每一轮中，选择总设备的一个子集 $K \ll N$ ，并在本地运行 E 个 epoch，然后用所有本地模型参数的均值对全局模型更新。

文献 [17] 的研究表明，适当调整 FedAvg 的优化超参数至关重要。特别是 FedAvg 的本地 epoch 数对收敛性起着重要的作用。一方面，执行更多的本地 epoch 允许更多的本地计算并可能减少通信，这可以大大提高通信受限网络的整体收敛速度。另一方面，对于不同的本地目标函数 $J_k(w)$ ，更大的本地 epoch 数可能会导致每个设备走向其局部的最优且与全局目标相反，这可能会损害收敛性，甚至导致方法发散。此外，在具有异构系统资源的联邦网络中，设置较高的本地 epoch 数可能会增加设备在给定的通信轮内未完成训练的风险。

虽然 FedAvg 在异构环境中取得了成功，但它并没有完全解决与异构相关的潜在挑战。在系统异构的背景下，FedAvg 不允许参与设备根据其底层系统约束执行可变数量的本地工作；相反，通常会简单地丢弃无法在指定时间窗口内计算 E 个 epoch 的设备 [37]。然而这可能会对收敛产生负面影响，因为它限制了有助于训练的有效设备的数量，并且如果丢弃的设备具有特定的数据特征，可能会在设备采样过程中引起偏差。从统计角度来看，FedAvg 在数据跨设备非均匀分布的情况下会出现经验分歧 [17]。

3.2 改进的联邦学习模型聚合算法

FedProx 算法 [38] 类似于 FedAvg 算法，因为在每轮中选择设备的子集，执行本地更新，然后将所有本地更新的模型参数的均值进行全局更新。然而，FedProx 做了以下简单但关键的修改，这导致了比较显著的性能提升，也为方法提供收敛保证。

联邦网络中的不同设备在计算硬件、网络连接和电池水平方面通常具有不同的资源约束。因此，强制每个设备运行相同的本地 epoch 数是不现实的。在 FedProx 中，通过允许根据可用的系统资源在本地跨设备执行灵活的本地训练次数来推广 FedAvg 算法，然后，相比于在 FedAvg 算法中放弃这些客户端，FedProx 考虑把这些未训练完的本地模型参数直接发

送给服务器。在整个训练过程中, 通过使用 γ 来度量每一轮局部求解器的局部计算量, 其中 $\gamma \in [0, 1]$, FedProx 没有对所有设备假设一个统一的 γ , 而是隐含地适应不同设备和不同迭代的变量 γ , 通过对本地训练的非精确求解, 动态调整本地迭代次数, 保证对异构系统的容忍度。设备 k 在迭代 t 时的 γ_k^t 非精确解为 $w^* = \min_w h_k(w; w_t)$, 其中函数 h_k 为 $(w; w_t) = J_k(w) + \frac{\mu}{2} \|w - w_t\|^2$, $\|\nabla h_k(w^*; w_t)\| \leq \gamma_k^t \|\nabla h_k(w_t; w_t)\|$ 且 $\nabla h_k(w; w_t) = \nabla J_k(w) + \mu(w - w_t)$ 。 γ_k^t 衡量在第 t 轮中为解决设备 k 上的本地子问题执行的本地计算量, γ_k^t 越小, 所得到的解越精确, 但同时本地计算量也会越大。FedProx 算法的具体步骤被总结在算法 2 中。

算法 2: FedProx 算法

输入: N (客户端总数), γ (每个客户端解的精确度), μ (近端项系数), S (包含所有客户端的集合), C (客户端的比例), w_0 (初始全局模型参数)。

输出: w_n (经过 n 次更新的全局模型参数)。

// 中心服务器执行

1 初始化: 全局模型 w_0 ;

2 **for** 每次全局迭代 $t \in 0, \dots, n-1$ **do**

 // 确定参与的客户端数量。

3 $m \leftarrow \max(C \cdot N, 1)$;

 // 随机选择参与的客户。

4 $S_p = \text{random.choice}(S, m)$;

5 **forall** 客户端 $k \in S_p$ **并行 do**

 // 获得客户端更新的模型。

6 $w_{t+1}^k \leftarrow \text{TrainLocally}(k, w_t)$

7 **end**

 // 更新全局模型。

8 $w_{t+1} \leftarrow \sum_{k=0}^N \frac{D_k}{D} w_{t+1}^k$

9 **end**

10 **Function** $\text{TrainLocally}(k, w_t)$ **is**

11 **while** 第 k 个客户端迭代直至得到 γ_k^t 非精确解 **do**

 // 进行本地模型训练

12 $w_{e+1} \leftarrow \min_w h_k(w_e; w_t)$, 其中函数 h_k 为 $(w; w_t) = J_k(w) + \frac{\mu}{2} \|w - w_t\|^2$,

$\|\nabla h_k(w^*; w_t)\| \leq \gamma_k^t \|\nabla h_k(w_t; w_t)\|$ 且 $\nabla h_k(w; w_t) = \nabla J_k(w) + \mu(w - w_t)$

13 **end**

14 **return** w_{e+1}

15 **end**

改进后的算法 FedProx 中心服务器第 1 行到第 9 行执行步骤和 FedAvg 一致, 主要在

第 10 行到第 15 行对本地更新的步骤进行了改进。 γ_k^t 衡量了本地模型 k 在第 t 个回合的计算量, 它越小意味着本地模型的训练完成度越高。FedProx 算法允许 γ_k^t 随着本地模型的不同和训练数的不同产生变化, 从而允许出现没有完成训练的本地模型。同时, 在损失函数中增加一个近端项, 以有效地限制模型本地更新与全局模型相差太大。

虽然容忍跨设备执行的工作量不一致有助于减轻系统异构性的负面影响, 但太多的本地更新仍然可能导致由于底层异构数据而导致的分歧。因此考虑在本地子问题中增加一个近端项, 以有效地限制模型本地训练的影响。这样的话, 设备 k 不是最小化本地函数 $J_k(w)$, 而是使用其选择的本地求解器来近似地最小化以下目标函数 h_k 如式 3.1,

$$\min_w h_k(w; w^t) = J_k(w) + \frac{\mu}{2} \|w - w^t\|^2. \quad (3.1)$$

近端项的优势包括两个方面: 一, 它通过限制本地更新更接近初始的全局模型来解决统计异构性的问题, 这样的话就无需手动设置本地训练 epoch 的数量。二, 它可以合理地整合由于系统异构而产生的相同时间内不同客户端内不同迭代次数的本地训练。

3.3 本章小结

本章对前文所提到的联邦学习中采用的传统模型聚合算法 FedAvg 与在此基础上改进的模型聚合算法 FedProx 进行了详尽的描述, 并给出了两种算法的伪代码。首先介绍了联邦学习中的异构问题, 并分析了在异构客户端中 Fedavg 算法难以有效解决异构问题的原因。然后, 针对这些原因, 详细阐述了 FedProx 算法相比于 Fedavg 所做的改进及其理论基础。

第四章 实验结果分析

4.1 实验环境

本文所有实验的运行配置如表 4.1 所示。

表 4.1 实验环境

软硬件配置	说明
CPU	AMD Ryzen Threadripper 3970X 32-Core Processor
GPU	NVIDIA GeForce RTX3090Ti-24GB
RAM	128GB-DDR4
OS	Ubuntu Server 20.04 LTS
CUDA	11.6
cuDNN	8.2.0
Python	3.7.13
Pytorch	1.11.0

4.2 实验数据与处理

在本文中,使用三个交通流量相关的公开数据集进行算法测试,包括 PeMS04、PeMS07 和 PeMS08,它们都是加利福尼亚运输机构 (CalTrans) 性能测量系统 (PeMS) 数据集的一部分,分别包括 307 个传感器、883 个传感器和 170 个传感器。PeMS 是一个从加州交通局性能测量系统获得的数据源,它从近 40,000 个单独的探测器实时收集数据,这些探测器横跨加州所有主要城市地区的高速公路系统,并收集了来自加州交通局和其他地方机构系统的各种信息,包括交通检测器、事故、车道关闭、收费标签、交通计数、车辆分类等等。

表 4.2 数据集介绍

名称	数据源	空间区域	时间范围	时间窗口
PeMS04		307 个传感器	2018/1/1~ 2018/2/28	
PeMS07	加利福利亚州交通传感器	883 个传感器	2017/5/1~ 2017/8/31	5 分钟
PeMS08		170 个传感器	2016/7/1~ 2016/8/31	

PeMS 提供超过十年的数据用于历史分析,数据集的采样频率为 5 分钟/次,即一天包

含 288 个采样点，通过这些数据对高速公路网络进行全面的评估，基于高速公路的状态分析堵塞以管理高速公路。具体统计信息如表 4.2 所示。

每个数据集的前 80% 作为训练集，训练集中 20% 的数据作为初始训练的验证集。保存最好的模型后，使用完整的训练集进行进一步的训练，最后 20% 的数据作为测试集。在预测之前，使用 MinMaxScaler 函数对样本数据进行归一化。这使得数据标准化到 [0,1]，并在输出结果之前进行反向归一化。

4.3 评价指标

为了比较不同方法的性能，我们遵循文献 [9, 16] 采用三个著名的指标：平均绝对误差 (MAE)，均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE)。指标的定义如式 4.1，其中 \hat{y}_i 为预测值， y_i 为真实值。

$$\begin{aligned} MAE &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ MAPE &= \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \end{aligned} \quad (4.1)$$

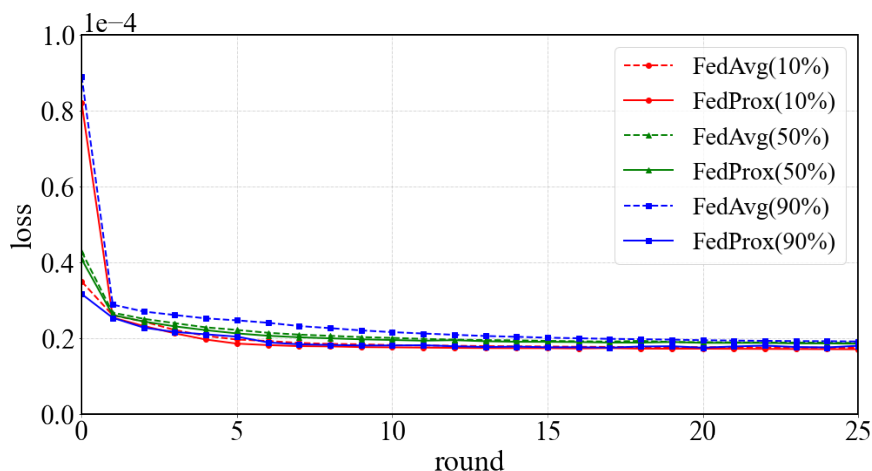
4.4 实验结果与分析

本文的实验在 PeMS04、PeMS07 和 PeMS08 三个数据集上进行。在实验中，客户端总数均为 10 个，人为设置一定比例的特殊客户端为在 E 个 epoch 内无法完成本地训练的客户端，以此作为不同联邦学习模型聚合算法的客户端异质性模拟，在本文的实验中，将这个比例分别设置为 10%，50%，90%。在本实验中，选择 $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 作为损失函数，选择 SGD 作为优化器，学习率设置为 0.001，如果选用了 FedProx 算法，算法参数 μ 设置为 0.001，本地模型训练的 epoch 为 2，batchsize 设置为 4，时间窗口设置为每 6 个时间戳的交通流量预测未来 1 个时间戳的交通流量，也即用过去 30 分钟的数据预测未来 5 分钟的数据。

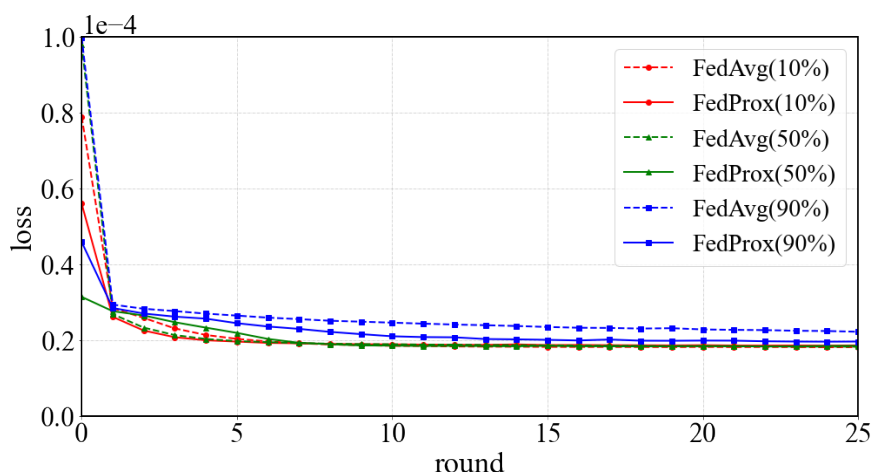
交通流量预测模型分为 ChebNet+GRU 模型和 GCN+GRU 模型，其中 ChebNet+GRU 模型分为两层图卷积层和一层 GRU 层，模型参数分别为 114，57 和 18，模型的总参数数量为 189；GCN+GRU 模型分为两层图卷积层和一层 GRU 层，模型参数分别为 42，21 和 18，模型的总参数数量为 81。

FedAvg 算法和在 PeMS08 上 10%、50%、90% 异质性模拟下的的损失收敛情况如图 4.1 所示。其中虚线和实线分别表示 FedAvg 算法和 FedProx 算法的损失随全局模型聚合轮次

的变化情况,红圆、绿上三角、蓝正方形分别为 10%, 50%, 90% 异质性客户端模拟的情况。可以看到, FedAvg 算法和 FedProx 算法在 ChebNet+GRU 模型和 GCN+GRU 模型下最终的损失都能达到收敛。



(a) ChebNet+GRU 模型



(b) GCN+GRU 模型

图 4.1 不同联邦算法在 PeMS08 上 10%、50%、90% 异质性模拟的损失收敛

在图4.1(a)中,除了10%异质性客户端模拟下 FedProx 算法的收敛更慢,其他的异质性客户端模拟下都是 FedProx 算法能更快地达到损失的收敛,这可能是由于在低比例的异质性客户端情况下, FedProx 算法相比于 FedAvg 算法并没有明显的优势。在图4.1(b)中,在所有的异质性客户端模拟下都是 FedProx 算法能更快地达到损失的收敛,并且 90% 异质性客户端模拟下 FedAvg 算法在达到收敛后的损失值明显比 FedProx 算法要更高,这说明在高比例的异质性客户端情况下, FedProx 算法相比于 FedAvg 算法的优势可能会更明显。

联邦学习模型聚合算法 FedAvg 算法和 FedProx 算法在三个交通数据集 PeMS04、PeMS07 和 PeMS08 上的性能评估如表 4.3 所示,性能评估包括三个指标:平均绝对误差 (MAE),均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE),这三个评价指标都是数值越小说明模

型预测的越准确。表中标为粗体的数据为某一数据集在某一比例异质性客户端模拟中的最好表现。

表 4.3 联邦算法在交通数据集上的性能评估

数据集	异质性模拟 特殊客户端占比	交通预测模型	FedAvg 算法			FedProx 算法		
			MAE	RMSE	MAPE	MAE	RMSE	MAPE
PeMS04	10%	ChebNet+GRU	18.26	29.02	21.46%	17.28	28.04	20.45%
		GCN+GRU	18.97	29.92	19.57%	18.94	29.90	19.73%
	50%	ChebNet+GRU	18.28	29.05	22.43%	17.30	28.06	21.35%
		GCN+GRU	18.93	29.89	20.53%	17.95	28.91	19.95%
	90%	ChebNet+GRU	18.33	29.12	22.44%	17.36	28.16	21.06%
		GCN+GRU	19.44	30.40	20.39%	19.03	29.95	20.36%
PeMS07	10%	ChebNet+GRU	19.18	29.43	9.85%	18.78	29.09	9.31%
		GCN+GRU	23.61	32.56	11.85%	22.85	31.52	11.22%
	50%	ChebNet+GRU	20.20	31.23	10.40%	19.37	29.61	9.78%
		GCN+GRU	22.98	32.73	10.95%	23.64	30.59	10.89%
	90%	ChebNet+GRU	20.19	30.35	9.60%	19.39	29.70	9.37%
		GCN+GRU	23.10	39.85	11.70%	20.20	32.76	11.19%
PeMS08	10%	ChebNet+GRU	14.18	20.81	11.75%	13.34	20.94	11.10%
		GCN+GRU	15.02	22.86	11.86%	14.84	22.70	11.66%
	50%	ChebNet+GRU	14.89	22.50	11.93%	14.28	22.08	11.34%
		GCN+GRU	14.96	22.85	12.56%	14.86	22.72	11.80%
	90%	ChebNet+GRU	14.34	21.99	12.88%	13.83	20.41	12.18%
		GCN+GRU	15.63	23.43	13.83%	15.60	22.43	13.67%

如表 4.3 所示，在 PeMS04 数据集中，除了在 10% 异质客户端模拟中，FedAvg 算法和 GCN+GRU 交通预测模型的 MAPE 评价指标表现更好，这可能由于异质客户端比例比较低，FedProx 算法相比于 FedAvg 算法优势相对而言不明显，在其它实验设置中，都是 FedProx 算法的三项评价指标都表现最好。在大部分实验设置中，也是 ChebNet+GRU 交通预测模型三项评价指标都表现更好，由于 ChebNet+GRU 模型和 GCN+GRU 模型都是两层图卷积层加一层 GRU 层，考虑到前者的参数更多，可能在一定范围内模型的参数越多效果更好。

在 PeMS07 数据集中，在所有比例的异质客户端模拟中，都是 FedProx 算法和 ChebNet+GRU 模型的三项评价指标都表现更好。同时 FedProx 算法的三项评价指标大部分也呈现出随着异质客户端的比例增加而表现逐渐变差，这说明异质客户端确实会对联邦学习架构下的交通预测模型性能产生影响。

在 PeMS08 数据集中，除了在 10% 异质客户端模拟中，FedAvg 算法和 ChebNet+GRU 交通预测模型的 RMSE 评价指标表现更好，这可能与 PeMS04 数据集中的原因一致，即异质客户端比例比较低，在其它实验设置中，都是 FedProx 算法的三项评价指标都表现最好。

将表 4.3 的数据进行整理，对不同数据集不同比例异质性模拟下的三项评价指标分别求平均值，可以得到表 4.4，以此对联邦学习模型聚合算法和交通预测模型进行更普遍直

观的性能评估。

表 4.4 联邦算法在不同数据集不同比例异质性模拟下的平均性能评估

交通预测模型	FedAvg 算法			FedProx 算法		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ChebNet+GRU	17.54	27.06	14.75%	16.77	26.23	13.99%
GCN+GRU	19.18	29.39	14.80%	18.66	27.94	14.50%

如表 4.4 所示, FedProx 算法和 ChebNet+GRU 模型的组合三项评价指标都表现更好。在 ChebNet+GRU 模型的 MAE 指标上, 采用了 FedProx 算法比采用 FedAvg 算法的性能提高了 4.4%, 在 GCN+GRU 模型的 MAE 指标上, 性能提高了 2.7%; 在 ChebNet+GRU 模型的 RMSE 指标上, 采用了 FedProx 算法比采用 FedAvg 算法的性能提高了 3.1%, 在 GCN+GRU 模型的 RMSE 指标上, 性能提高了 4.9%; 在 ChebNet+GRU 模型的 MAPE 指标上, 采用了 FedProx 算法比采用 FedAvg 算法的性能提高了 5.2%, 在 GCN+GRU 模型的 MAPE 指标上, 性能提高了 2.0%。

综合上述实验数据, 可以认为 FedProx 联邦学习模型聚合算法相比于传统的 FedAvg 算法在不同数据集不同比例异质性模拟下在三项评价指标上均有比较明显的性能提升。

4.5 本章小结

本章对前文所提到的改进的联邦学习模型聚合算法 FedProx 与传统的联邦学习模型聚合算法 FedAvg 的实验结果进行性能评估与对比。首先, 介绍了实验的环境与设置, 然后详细介绍了数据集的具体情况和性能评估的相关指标, 最后设计了在多个数据集上的多组实验对模型的性能进行了分析。

第五章 总结与展望

本章对全文进行总结和展望，首先对本文的主要研究内容和贡献做一个简短的总结，然后对未来的工作进行展望。

5.1 论文总结

本文针对交通流量预测时遇到的数据收集和隐私保护问题，采用联邦学习架构进行交通站点之间的数据共享，并考虑客户端计算能力异构时，采用改进的联邦学习参数更新方式，提出了基于联邦学习的交通流量预测方法，主要的研究内容总结如下：

本文首先通过道路网络和特征矩阵定义了交通流量预测问题，然后使用交通预测模型通过提取交通流量的时空特征进行预测。最后，由于交通流量预测时会遇到的数据收集和隐私保护问题，采用联邦学习架构进行交通预测模型的训练，并考虑客户端计算能力的异构，采用改进的联邦学习模型聚合算法，在客户端异质性模拟实验中验证了其相对于传统的联邦学习算法的优越性。

在本地的交通预测模型训练中，使用基于谱的图卷积神经网络如 ChebNet 网络和 GCN 网络对交通道路网络的空间特征进行提取，考虑到交通流量预测本质上是一个时序预测问题，还采用 GRU 模型来提取交通流量数据的时间特征，这样本文的交通预测模型就可以通过提取交通流量的时空特征进行预测。

在联邦学习架构下，第一步，中央服务器初始化全局模型，并将全局模型发送到每个参与的客户端。第二步，客户端根据本地数据迭代训练接收到的全局模型。之后，它们将有利于全局模型更新的模型权重或梯度返回给服务器。第三步，中央服务器利用所有客户端的信息来更新全局模型。重复第二步和第三步，直到收敛。其中，改进的联邦学习模型聚合算法 FedProx 与传统的联邦学习算法 FedAvg 区别主要在第三步。在系统异构的背景下，FedAvg 通常会简单地丢弃无法在指定时间内完成本地训练的设备，然而这可能会对收敛产生负面影响。相比于在 FedAvg 算法中直接放弃这些客户端，FedProx 考虑把这些未训练完的本地模型参数直接发送给服务器，并在本地子问题中增加一个近端项，以有效地限制模型本地更新与全局模型相差太大。

最后，通过在三个数据集的客户端异质性模拟实验验证 FedProx 算法的优越性。实验结果表明，FedProx 联邦学习模型聚合算法相比于传统的 FedAvg 算法在不同数据集不同比例异质性模拟下在三项评价指标上均有比较明显的性能提升。

5.2 研究展望

本文针对交通流量预测时遇到的数据收集和隐私保护问题,采用联邦学习架构进行交通站点之间的数据共享,并考虑客户端计算能力异构时,采用改进的联邦学习参数更新方式,提出了基于联邦学习的交通流量预测方法,这对于准确预测交通状况,促进智能交通系统的发展具有重要的意义。但是,本文的工作还可以从以下几个方面进一步扩展。

首先,由于研究重点放在验证改进的联邦学习模型聚合算法 FedProx,在交通预测模型上并没有进行大量实验以调整模型的层数和超参数,提高预测模型的准确率,在这方面还可以继续改进。

其次,本文的交通预测模型只考虑了交通流量这一特征,实际上还可以考虑天气、交通事件等额外因素对预测的影响,以及各个交通站点之间的交通流量变化的相关性,通过特征融合提高预测的准确率。

最后,在本文中只考虑了系统异构的情况,并只通过客户端异质性模拟的实验验证 FedProx 算法的优越性,然而在实际情况下,不同交通站点的交通流量数据由于地理位置等因素也是异构的,因此接下来还可以考虑数据异构性的实验模拟,并进一步改进联邦学习模型聚合算法,提高联邦学习架构下交通预测的准确率。

参考文献

- [1] 康有余, 王伟华. 基于物联网技术的智慧地铁系统设计 [J]. 物联网技术, 2021, 11(103-104+107).
- [2] 王晓宇, 赵鹏. 新基建下的智能交通系统发展综述 [J]. 上海公路, 2022, No.165(91-95+167-168).
- [3] Sadiku M N, Tembely M, Musa S M. Internet of vehicles: An introduction[J]. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2018, 8(1):11.
- [4] Dai R, Xu S, Gu Q, et al. Hybrid spatio-temporal graph convolutional network: Improving traffic prediction with navigation data[C]. 2020. 3074--3082.
- [5] Zhang J, Wang F Y, Wang K, et al. Data-driven intelligent transportation systems: A survey[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2011, 12(4):1624--1639.
- [6] Ghosh B, Basu B, O'Mahony M. Multivariate short-term traffic flow forecasting using time-series analysis[J]. *IEEE transactions on intelligent transportation systems*, 2009, 10(2):246--254.
- [7] Yang F, Yin Z, Liu H, et al. Online recursive algorithm for short-term traffic prediction[J]. *Transportation Research Record*, 2004, 1879(1):1--8.
- [8] Lv Y, Duan Y, Kang W, et al. Traffic flow prediction with big data: A deep learning approach[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 16(2):865--873.
- [9] Kang D, Lv Y, Chen Y y. Short-term traffic flow prediction with LSTM recurrent neural network[C]. 2017. 1--6.
- [10] Lv Z, Xu J, Zheng K, et al. Lc-rnn: A deep learning model for traffic speed prediction.[C]. 2018. 27.
- [11] Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving communication efficiency[J]. *Conference on Neural Information Processing Systems*, 2016.
- [12] Yuan X, Li P. On convergence of FedProx: Local dissimilarity invariant bounds, non-smoothness and beyond[J]. *Advances in Neural Information Processing Systems*, 2022, 35:10752--10765.
- [13] Kim M, Yu S, Kim S, et al. DepthFL: Depthwise federated learning for heterogeneous clients[C]. 2023.
- [14] Putri A, Brahmana F, Joelianto E, et al. Traffic flow estimation using graph neural network with aggregation of traffic features[C]. 2022. 972--977.
- [15] Yuan L, Fang W, Xiao H, et al. Short-term traffic flow prediction by graph deep learning with spatial temporal modeling[C]. 2022. 172--177.
- [16] Zhang C, Zhang S, James J, et al. Fastgnn: A topological information protected federated learning approach for traffic speed forecasting[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(12):8464--8474.
- [17] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]. 2017. 1273--1282.
- [18] Konečný J, McMahan H B, Ramage D, et al. Federated optimization: Distributed machine learning for on-device intelligence[J]. *arXiv preprint arXiv:1610.02527*, 2016.
- [19] Huang W, Song G, Hong H, et al. Deep architecture for traffic flow prediction: Deep belief networks with multitask learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(5):2191--2201.

- [20] Yoon J S, Kang S M, Park S B, et al. Gdfed: Dynamic federated learning for heterogenous device using graph neural network[C]. 2023. 683--685.
- [21] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[J]. *International Conference on Learning Representations*, 2014.
- [22] Chen Z M, Wei X S, Wang P, et al. Multi-label image recognition with graph convolutional networks[C]. 2019. 5177--5186.
- [23] 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.
- [24] Monti F, Boscaini D, Masci J, et al. Geometric deep learning on graphs and manifolds using mixture model cnns[C]. 2017. 5115--5124.
- [25] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. *International Conference on Learning Representations*, 2017.
- [26] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[J]. *Conference on Neural Information Processing Systems*, 2017.
- [27] Diao Z, Wang X, Zhang D, et al. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting[C]. 2019. 890--897.
- [28] Tang J, Liang J, Liu F, et al. Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 124:102951.
- [29] Chen W, Jiang M, Zhang W G, et al. A novel graph convolutional feature based convolutional neural network for stock trend prediction[J]. *Information Sciences*, 2021, 556:67--94.
- [30] Ma T, Wang H, Zhang L, et al. Graph classification based on structural features of significant nodes and spatial convolutional neural networks[J]. *Neurocomputing*, 2021, 423:639--650.
- [31] Yang Z, Dong S. Hagerec: Hierarchical attention graph convolutional network incorporating knowledge graph for explainable recommendation[J]. *Knowledge-Based Systems*, 2020, 204:106194.
- [32] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural computation*, 1997, 9(8):1735--1780.
- [33] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. *Conference on Neural Information Processing Systems*, 2014.
- [34] Xiao G, Wang R, Zhang C, et al. Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks[J]. *Multimedia Tools and Applications*, 2021, 80:22907--22925.
- [35] Wu T, Chen F, Wan Y. Graph attention LSTM network: A new model for traffic flow forecasting[C]. 2018. 241--245.
- [36] Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving communication efficiency[J]. *Conference on Neural Information Processing Systems*, 2016.
- [37] Bonawitz K, Eichner H, Grieskamp W, et al. Towards federated learning at scale: System design[J]. *Proceedings of machine learning and systems*, 2019, 1:374--388.
- [38] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks[J]. *Proceedings of Machine learning and systems*, 2020, 2:429--450.

致 谢

终于快要毕业了，想写些大学期间的总结，心中千头万绪，下笔却不知从何说起，那就从一切的起点开始吧。回想起第一次到兰大的情景，仿佛就在眼前，那时的我大概是抱着上大学以后就轻松了的心态进学校的吧，从没有去过离家这么远的地方，还挺新奇的，懵懵懂懂，忽然之间大学就要毕业了，一切变得那么不一样了，我应该成长了许多吧，不知道当时的我是否满意呢。至少一路走来，面对曾向我传授专业知识和人生经验的各位老师，曾与我在校园一同度过难忘时光的各位同学和好朋友，曾一直给予我坚定支持和情感慰藉的家人，我想我可以自豪地对当时的我说，你是那么幸福！

首先感谢我的论文指导老师焦栋斌老师。我与焦老师结缘于大二下学期的《最优化方法》，焦老师在课上鼓励我们班的同学跟他进行课外的科研训练，我也主动加入了焦老师的课题组。在课题组里，我们亦师亦友，焦老师教会了我如何进行科研阅读，文献报告，引导我去思考和学习工具的使用，同时，在课题组里，我也结识了许多优秀的同学，我们相互学习，互相帮助，在他们身上，我学到了很多。在两年的相处中，焦老师的言传身教也一定程度上影响我选择继续研究生的学习，虽然不确定自己一定会在学术道路上走下去，但我相信这些会是我人生中的一笔宝贵的财富。

感谢绽琨老师，在《模式识别》和《人工智能综合实践》课上用通俗易懂的方式讲解了常用的机器学习算法，并通过卷积神经网络的 `numpy` 库实现和基于神经网络的互信息估计 (MINE) 两个实验帮助我深刻理解神经网络的实现与运用。在实验中，我和绽老师交流了许多代码中的细节，对于我不理解的地方，绽老师不厌其烦地讲解，这让我学到了很多。也因此，我后来在考虑兴趣的基础上选择了在兰大读研，并将有幸在研究生期间与绽老师结下师生缘。

感谢大学期间我的所有任课老师，在大学的学习和高中很不一样，我也改变了许多学习的方法，但我相信你们认真备课所传授给我的知识都会在未来的某一刻给予我帮助。

感谢我的好朋友裴雄和王玉强，在预防医学专业学习的一年里，我们一起早起，一起去图书馆背书，在兰大的第一年没有荒废学习，我们相互鼓励，互相学习，都找到了自己喜欢的专业，并成功通过转专业面试，我也即将开始在兰大的研究生学习，真的很怀念我们一起为了自己的理想奋斗的那段时光，那决定了我未来的道路。

感谢我的同学和好朋友吴晓鹏、郑泓凌、魏楚扬、孙锐、李亚霖和其他 19 级人工智能班的同学，人工智能专业的学习并不简单，有很多不同专业的课程，在学习中和这些同学一起交流作业、小组任务，这加深了我对这些专业知识的思考。他们中的一部分也在四年的时间里和我成为了相知的挚友，我和他们课下之余的交流、娱乐也是四年美好校园时光难忘的一部分。

感谢焦栋斌老师课题组的各位同学们，刚进入课题组时，从刘波、过祖煜等优秀学长身上我学到了很多以前课上没学到的知识，他们的优秀为当时的我提供了前进的目标和动力，在后来的学习中，他们也为我提供了很多帮助。在课题组的两年里，我也与同级的邹振涛、王凌宇、贾博文和李淼同学相互学习，我们虽然来自不同专业，但在课题组内，我们凝聚在一起，一起进行科研训练，与他们的相识是我在大学期间意外而又珍贵的财富。

感谢我大学四年的舍友王文保、孙哲和杨智超，四年的朝夕相伴，我们学习、打闹、相互帮助，融洽的宿舍氛围为我的大学生活增添了太多的色彩，我想我们不仅是四年的舍友，更是值得一生珍惜的好友。

感谢我的家人，在大学的记忆里，你们出现的很少，但在学校里的时光有太多想家的情绪只能通过电话和视频聊天缓解。我也深深地知道，我在学校学习的每一天都离不开你们默默地辛劳。面对你们的付出，我能做出的回报实在有限，只希望我有能力回报你们的那一天不要太迟。

最后，诚挚地感谢焦栋斌老师在我毕业论文形成过程中的指导、讨论与修改，也感谢在百忙之中参加论文评阅和答辩的各位老师！

毕业论文（设计）成绩表

导师评语

建议成绩 _____

指导教师（签字） _____

答辩委员会意见

答辩委员会负责人（签字） _____

成绩 _____

学院（盖章） _____

年 月 日