

基于锚点注意力机制的联邦网络流量预测^①

李文峰, 薛林东, 师少伟, 李彦沛, 冯晨琛

(西安科技大学 通信与信息工程学院, 西安 710699)

通信作者: 薛林东, E-mail: 761772964@qq.com



摘要: 流量预测有利于提高网络服务质量, 现有的流量预测方法涉及大规模的流量数据传输, 会导致用户隐私泄露等安全问题. 采用联邦学习的方法能够保护用户隐私并有效减少数据计算的工作量. 然而在实际情景中, 不同基站流量的数据存在异构现象, 且数据量有限, 导致训练得到的全局模型泛化能力较差. 针对这些问题, 在联邦学习的基础上提出一种基于注意力机制的深度神经网络模型 (CALs). 在 3 个非相同分布的公开数据集上进行训练, 通过训练该模型可以更好地捕捉基站流量的变化特性, 最后与常用的深度学习算法如 GRU、LSTM 等模型进行比较, 平均绝对误差性能分别提升 9.42%、11.1%.

关键词: 流量预测; 深度学习; 联邦学习; 注意力机制

引用格式: 李文峰, 薛林东, 师少伟, 李彦沛, 冯晨琛. 基于锚点注意力机制的联邦网络流量预测. 计算机系统应用, 2025, 34(9): 192-199. <http://www.c-s-a.org.cn/1003-3254/9963.html>

Federated Network Traffic Prediction via Anchor-based Attention Mechanism

LI Wen-Feng, XUE Lin-Dong, SHI Shao-Wei, LI Yan-Pei, FENG Chen-Chen

(College of Communication and Information Technology, Xi'an University of Science and Technology, Xi'an 710699, China)

Abstract: Traffic prediction enhances the quality of network services, yet existing methods typically require large-scale traffic data transmission, raising concerns about user privacy leakage. To address this, federated learning is adopted to preserve user privacy and reduce computational workload. However, in practical applications, traffic data from different base stations are often heterogeneous and limited in volume, which hinders the generalization ability of the resulting global model. To overcome these issues, a deep learning model, termed CALs, based on an attention mechanism within the framework of federated learning is proposed. The proposed model is trained on three publicly available datasets with non-identical distributions, enabling the model to better capture the dynamic characteristics of base station traffic. Compared to standard deep learning models such as GRU and LSTM, the proposed model reduces the mean absolute error by 9.42% and 11.1%, respectively.

Key words: traffic prediction; deep learning; federated learning (FL); attention mechanism

随着 5G 网络覆盖范围的扩大, 用户数量不断增加、用户使用的网络设备多样以及物联网设备的大规模使用, 这使得网络流量快速增长. 网络流量的快速增长带来了一些问题, 尤其是网络拥塞和延迟, 这些问题直接影响到网络的服务质量和用户体验. 此外虚拟/增强现实 (VR/AR)、智能健康及智慧城市等技术也驱动

网络需求爆炸式增长. 这些需求依赖超大带宽和超低延迟, 来保障数据的实时性、安全性和大容量传输能力. 对 VR/AR 带来的瞬时数 Gbps 级流量洪峰以及智慧城市中交通监控、环境感知等异构系统产生的多模态数据流, 运营商需分析用户行为数据与边缘节点负载状态, 提前预判数据流向^[1]. 为了有效应对这些挑战,

① 基金项目: 陕西煤业化工集团有限责任公司科技研发类投资项目 (2023SMHKJ-B-51)

收稿时间: 2025-02-19; 修改时间: 2025-03-12; 采用时间: 2025-04-07; csa 在线出版时间: 2025-07-25

CNKI 网络首发时间: 2025-07-28

对无线网络流量进行准确预测显得尤为重要。网络流量预测不仅可以帮助运营商和网络管理者提前了解网络负载情况^[2],还能合理规划网络资源,优化网络结构,避免或减轻网络拥塞^[3],降低延迟,提升用户体验。

网络流量的本质上是时间序列数据,其预测问题可以转化为时间序列建模预测问题。在网络流量预测技术的早期阶段,大部分技术都引入了结构建模来进行预测,如线性时间序列模型,其方法基于传统的时间序列预测技术,这种方法的基本思想是提供具有足够参数的广泛模型类别来拟合各种数据集。网络流量预测的经典线性预测模型有自回归(AR)、移动平均(MA)、自回归移动平均(ARMA)^[4]和自回归综合移动平均(ARIMA)^[5]等。所有这些模型都是线性的,表现为短时记忆,无法解释模型以及捕捉自相似性。在基于时间序列的流量预测背景下,Trinh等人^[6]将LSTM与MLP和ARIMA进行了比较,结果表明LSTM显著优于其他模型,这是研究者们将深度学习算法应用于网络流量预测的重要探索,验证了深度学习算法预测时间序列的能力。Kulin等人^[7]研究了有关使用深度学习技术增强网络性能的现有研究,描述了如何在各种网络场景中应用深度学习技术。Lin等人^[8]提出了一个模型采用多变量的方法,利用扩展的图注意力网络以及注意力机制来提高捕捉空间依赖性的效率,从而对网络流量之间的相关性进行探索。随着隐私保护意识的增强,从大量不同的数据所有者手中收集数据变得不可行^[9],这限制了深度学习在数据分散场景中的应用。

传统的集中式学习要求将手机等本地设备上收集的所有数据都集中存储在数据中心或云服务器上。这一要求不仅存在数据泄露的问题,而且海量的数据对服务器的存储和计算能力提出了很高的要求。联邦学习(federated learning, FL)则可以在保护数据隐私的同时对分布在不同设备上的数据进行训练。联邦学习中最常用的算法是Fedavg,由McMahan等人^[10]提出。在联邦学习中最重要的一步是模型聚合。模型聚合是通过计算基于数据量的客户模型的加权平均值来执行的,从而使客户具有更大的影响力,在此阶段,中央服务器收集和聚合来自参与者的模型,以更新全局模型的状态。这个过程存在一定的困难,尤其在处理客户端之间的Non-IID数据^[11]和异构数据时联邦学习的性能将大大减弱。Mai等人^[12]提出了一种模型可以在服务器和节点端进行学习与模型蒸馏,从而得到训练模型,缓解了数据异构性。但额外的模型蒸馏操作同样

引入了额外的时间和计算开销。为缓解这个问题,更好地捕获数据异构性,采用另一种算法FedNova^[13],它在平均过程中对局部模型更新并进行归一化,用标准化的局部梯度除以每个客户端单独执行的局部步骤数,从而求平均值。

联邦学习的一个关键特性是数据本身永远不会在设备和服务器之间传输,这不仅减少了通信需求,而且能解决与数据共享相关的隐私问题。与集中式学习相比FL也有一定局限,包括本地设备所需的计算能力和能耗,以及学习过程中可能存在的大量本地设备的异质性和随机性。这些问题导致FL采用简单求平均的聚合方法进行训练的准确性下降。针对以上问题,本文提出了CALs模型。该模型对主流深度学习算法进行改进并增添注意力机制,可以实现较为准确的网络流量预测。本文贡献如下:(1)设计了一种结合注意力机制和长短时记忆网络的网络流量预测算法,能够对复杂且动态的网络流量数据建模分析。(2)将联邦学习引入到网络流量算法中,并在聚合部分规范化和加权聚合模型参数,均衡不同传感器对全局模型的贡献,确保模型的稳定性和有效性。(3)在真实的网络流量数据集上对设计的算法进行实验验证。在保证隐私的安全条件下,实现了数据的有效聚合,增强对时序数据异常模式的识别。

1 模型设计

1.1 CALs模型

模型CALs(CAA-LSTM-MLP)是一种融合卷积注意力机制与长短期记忆网络的混合算法,具有空间-时间双阶段特征提取,灵活的结构适应性支持任意二维空间构造以及端到端时空联合优化,先提取单时间步的空间结构,再建模跨时间步的动态演化的特点,专为同时捕捉时间序列数据的空间和时序特征而设计。模型采用上下文锚点注意力提取输入信息的通道特征,并分别对水平和垂直方向进行卷积操作,从而提取局部上下文信息,将提取的局部上下文信息进行全局平均池化得到特征图信息,然后归一化输出到LSTM模型中,进行隐藏状态向量的计算,最后通过一个全连接的多层感知机进行最终的预测输出。在联邦学习中使用CALs模型进行预测,可以有较好的准确率,并且因为模型添加了上下文锚点注意力对不同特征有不同的计算因子从而使结果更加准确,同时通过联邦模型FedNova可以减少Non-IID数据对于联邦学习的影响。图1为CALs模型的结构图。

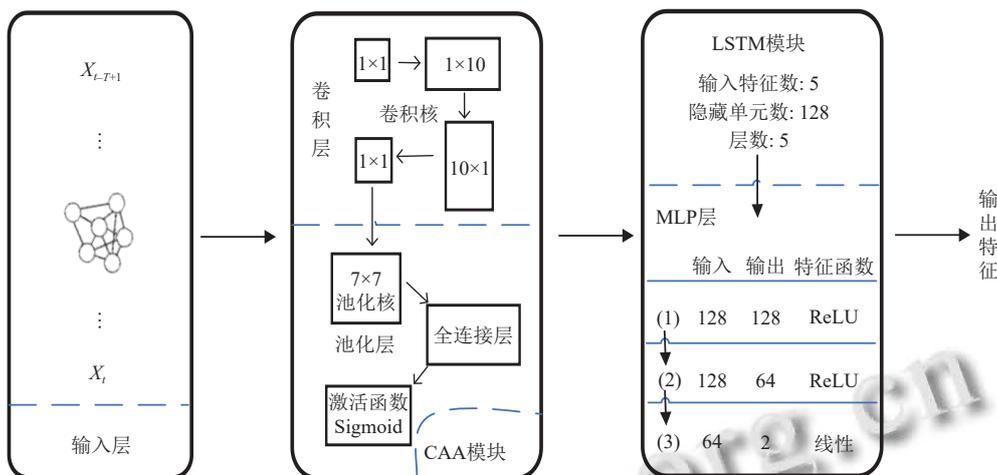


图1 CALS 模型

该模型处理时间序列数据的流程为:首先将原始时间序列 (批次 B , 时间步 T , 特征数 F) 的每个时间步特征构造为二维空间结构 (批次 B , 通道数 C, H, W), 随后输入到 CAA 模块中, 利用 1×1 卷积调整通道, 接着利用水平和垂直方向的分组卷积 ($\text{kernel_size}=(1, 10)$ 和 $(10, 1)$) 提取空间特征得到特征 $(B, C, 1, 1)$, 结合全局平均池化和全连接层将通道维度从 10 压缩到 5, 得到每个时间步特征 $(B, 5)$, 所有时间步处理后的输出特征为 $(B, T, 5)$, 作为 LSTM 的输入序列, 5 层 LSTM 堆叠, 每层隐藏状态维度为 128, 前一层的输出作为下一层的输入, 在中间层保留所有时间步的输出, 以便下一层能接收完整的时序输入, 最终层设置为仅返回最后时间步的隐藏状态得到特征 $(B, 128)$, 然后输入到多层感知机 (MLP) 中, 在多层感知机的第 1 层将输入从 128 维映射到 128 维, 应用 ReLU 引入非线性. 第 2 层将 128 维映射到 64 维, 应用 ReLU. 第 3 层将 64 维映射到 2 维. 最终输出特征 $(B, 2)$, 用于回归任务.

1.2 联邦学习模型

联邦学习是一种分布式学习模型, 在训练模型时不损害用户的隐私. 每位用户可以在本地设备端进行模型的训练, 同时将训练数据保存至本地. 首先是关于个体学习问题, 设 $X_t = \{x_{t,1}, x_{t,2}, \dots, x_{t,v}\}$ 表示时间步长 t 的测量值, 其中 v 为变量数. 对于给定的 t , 需要考虑过去观测的窗口 $T \in [t-T+1, t]$ 和 $X_t^* = \{X_{t-T+1}, \dots, X_t\}$. 目的是利用观测到的 X_t^* 来预测下一步的测量结果. 然后利用整个测量数据建立一个可以预测未来未知序列的模型. 在集中式学习中, 测量数据被传送给中央服务器, 由第三方使用整个测量数据来建立一个可以预测

未来序列的模型.

联邦学习介于个体学习和集中学习之间, 主要由参与者合作建立一个可以预测网络流量的全局模型 θ , 这样可以最小化聚合局部损失函数 $f_k(\theta^k)$, 以及最优全局模型 $f(\theta)$, 其中 n_k 是本地数据大小, l 为损失函数, x 为数据特征, y 为数据标签, k 为客户端指数, M 为用户参与数量, n 为样本总数.

$$f_k(\theta^k) = \frac{1}{n_k} \sum_i^{n_k} l(x_i, y_i, \theta^k) \quad (1)$$

$$f(\theta) = \sum_{k=1}^M \frac{n_k}{n} f_k(\theta^k) \quad (2)$$

每位参与者都拥有自己的时间序列, 需要在有限的时间内对学习模型做局部训练. 并将训练完成的数据提供给中心服务器进行聚合生成平均模型. 这个过程一直持续到全局模型推广到 n 个参与者的观察结果为止. 因此, 在联邦学习中, 每个参与实体基于本地数据优化模型, 类似于个体学习. 联邦聚合通过接收的本地模型来更新全局模型, 并多次进行该操作直到联邦聚合生成的全局模型具有泛化到 n 个实体的能力. 图 2 为联邦学习的模型图.

联邦训练完成后, 中央服务器将最终的全局模型传输给参与者. 每个参与者在得到全局模型后应进行预处理操作, 并将转换后的数据提供给训练过的模型. 然后将预测结果转换回原始特征范围, 以获得最终的预测结果. 此外, 参与者还可以进行局部微调步骤, 使模型更接近他们的本地数据, 从而提供更高质量的预测.

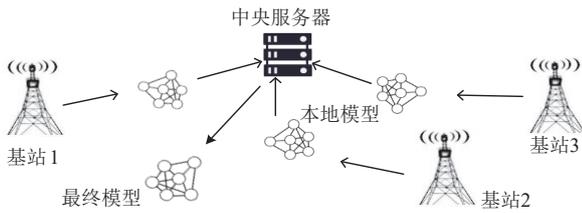


图2 联邦学习模型

1.3 上下文锚点注意力模块

上下文锚点注意力模块 (context anchor attention, CAA)^[14,15] 模块是一种新型的注意力机制,旨在增强特征提取过程中的上下文感知能力,处理输入特征之间的相互依赖关系,聚焦于不同尺度的局部和全局信息,从而在流量预测任务中实现更精确的注意力分配。

CAA 模块的结构主要包括若干层卷积操作、批归一化、非线性激活、平均池化层以及全连接层等部分。首先,输入特征图为 $X \in \mathbb{R}^{C \times H \times W}$,其中 C 、 H 、 W 分别是通道数、高度和宽度,接下来进行横向与纵向卷积操作,并同时全局平均池化提取上下文的特征信息,

$$X_{\text{hor}} = \text{Conv}_{1 \times k}(X) \quad (3)$$

$$X_{\text{ver}} = \text{Conv}_{k \times 1}(X) \quad (4)$$

$$X_{\text{avg}} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j) \quad (5)$$

其中, $\text{Conv}_{1 \times k}$ 、 $\text{Conv}_{k \times 1}$ 分别是 $1 \times k$ 和 $k \times 1$ 的深度可分离卷积操作, $X_{\text{avg}} \in \mathbb{R}^C$ 。

$$X_{\text{all}} = X_{\text{hor}} + X_{\text{ver}} + X_{\text{avg}} \quad (6)$$

接下来通过 1×1 卷积和 Sigmoid 激活函数生成通道注意力权重,最后通过对输入特征图按通道权重进行加权得到输出:

$$A = \sigma(\text{Conv}_{1 \times 1}(X_{\text{all}})) \quad (7)$$

$$X_{\text{out}} = X \odot A \quad (8)$$

其中, σ 表示 Sigmoid 激活函数,输出 $A \in \mathbb{R}^C$, \odot 表示通道维度上的逐元素相乘。

1.4 长短时记忆网络

长短时记忆网络 (long short-term memory, LSTM) 是一种特殊的循环神经网络,它能够学习长期依赖关系。LSTM 通过引入门机制来控制信息的流动,从而解决了传统 RNN 中的梯度消失问题,如图3为 LSTM^[16] 结构图。

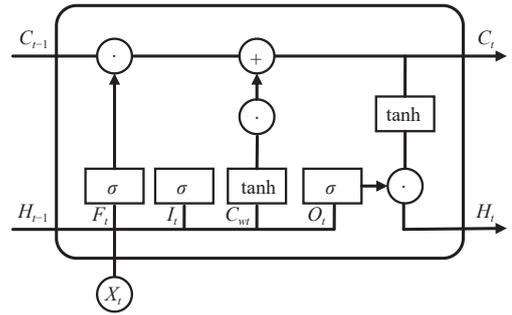


图3 LSTM 结构

组合计算是 LSTM 单元的核心部分,包括输入门、遗忘门、输出门和内存单元。其中输入门、遗忘门、输出门使用 Sigmoid 激活函数,接收当前输入和前一时间步的隐藏状态作为输入,分别决定哪些信息需要被更新到内存单元中、哪些信息需要从内存单元中遗忘、哪些信息需要被输出。内存单元更新使用 tanh 激活函数来处理输入信息,并结合输入门和遗忘门的输出,通过元素逐位相乘和加法操作来更新内存单元的状态。LSTM 的 3 个门控单元,分别是输入门 I_t 、遗忘门 F_t 、输出门 O_t ,计算公式如下:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \quad (9)$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \quad (10)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \quad (11)$$

其中, σ 为 Sigmoid 函数, W 分别为权重函数。输入门控制候选记忆状态,遗忘门控制上一步的记忆状态,当输入门 $I_t = 0$ 遗忘门 $F_t = 1$ 时将无法传递新的信息,记忆状态就会保存之前的状态,隐藏状态用于存储 LSTM 在处理序列数据时学习到的信息。其中候选记忆状态 C_{wt} 、记忆状态 C_t 、隐藏状态 H_t ,计算公式如下:

$$C_{wt} = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \quad (12)$$

$$C_t = F_t \cdot C_{t-1} + I_t \cdot C_{wt} \quad (13)$$

$$H_t = O_t \cdot \tanh(C_t) \quad (14)$$

2 实验设计

2.1 数据预处理

为了验证本文提出的联邦预测模型的性能,实验选用了来自联邦预测挑战赛上的公开数据集^[17],该数据集包含 3 个不同的基站的样本数据,该数据集包含 11 个与下行和上行流量相关的特性。在每个时间步,测量下行链路和上行链路流量,无线网络临时标识数量,

上下行分配的资源块数量及其方差,上下行调制和编码方案及其方差.每个时间步表示将这些测量值聚合为一个 3 min 的间隔.

时间序列数据通常通过传感器、监控系统或其他自动化记录机制实时或定期收集的.由于数据收集过程的自动化特性,数据集中包含冗余信息、错误或异常峰值是非常常见的现象.这些问题可能源于传感器故障、网络传输错误、记录设备的不稳定性,或其他外部干扰因素,因而对于数据的预处理显得尤为重要.如图 4 和图 5 所示的箱型图展示了每 24 h 内上行流量和下行流量的分布情况,从图中可以清晰地看出,下行流量的分布始终高于上行流量,因为大多数情况下,用户下载数据的需求多于上传数据的需求,这与一般常识相符.然而,图中也显示出存在较多的异常数据点,这些异常值可能是由于网络拥塞、设备故障或恶意攻击等原因引起的,因此需要对数据进行预处理,通过对于收集到的数据进行预处理,来改善原始数据的质量,以确保数据能够更好地适用于深度学习的模型.

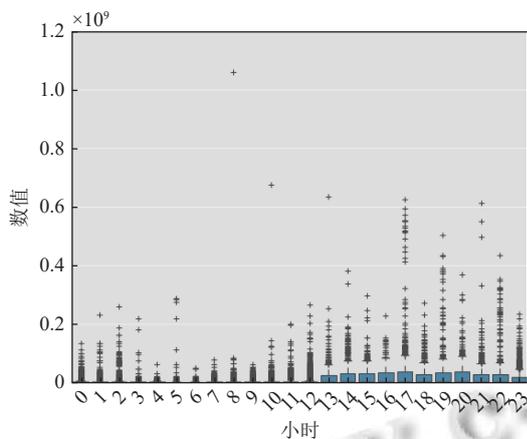


图 4 每小时上行流量分布图(24 h 内)

数据预处理主要包含数据清洗、数据转换、数据集划分等.数据清洗是数据预处理的第 1 步,主要进行缺失数据的处理,对缺失量较大的数据记录进行删除、检测异常值,对于明显异常的数据进行修正;数据转换将原始数据通过最大-最小归一化的操作转换成适合模型训练的格式,并保留了原始数据的分布特性,其计算方式如式(15)所示:

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

接下来进行数据集的划分,将处理好的数据划分成训练集、验证集、测试集,其中训练集占 60%,验证

集与测试集各占 20%,通过对数据集的划分,可以更好地进行深度学习的训练,还可以针对数据集进行多次划分,实现交叉验证从而减少由于划分数据集带来的误差.最后一步在数据上滑动一个固定大小为 T 的窗口来创建多个重叠的子序列,进行特征提取.

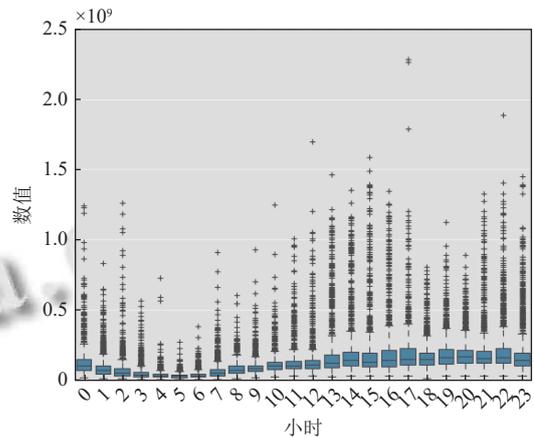


图 5 每小时下行流量分布图(24 h 内)

2.2 基准模型

前文分析了针对网络流量的时间序列进行预测的主流方法,比如经典的线性预测模型以及深度学习算法,通过与这些方法中最经典且性能优异的算法相比,来展示所提出模型在不同技术路线下的优势.

(1) CNN^[18,19]通过卷积操作能够有效地提取时间序列中的局部特征,尤其在具有周期性或重复模式的数据时,卷积核能够在不同时间步上捕捉到相似的模式.还能增强模型对输入数据位置变化的鲁棒性,使其能够处理时间序列中的平移不变性问题.

(2) LSTM^[20]通过输入门、遗忘门和输出门来控制信息流,并捕捉时间序列中的复杂模式和趋势,即使数据中存在较长的时间间隔,也能有效地传递相关信息.这种机制使其在处理具有长期依赖关系的时间序列数据时表现出色.

(3) ARIMA 包括自回归、差分和滑动平均 3 个部分,自回归部分利用历史数据预测未来值,差分部分通过消除非平稳性使数据更适合建模,而滑动平均部分则用于处理噪声和随机波动.模型能够通过调整参数灵活地适应不同类型的时间序列数据.

(4) GRU^[21,22]合并了 LSTM 的遗忘门和输入门为更新门,这种设计使 GRU 在处理时间序列数据时,能够在捕捉长期依赖关系的同时,减少计算复杂度和训练时间. GRU 的轻量级结构使其在资源有限的环境中

更具优势。

2.3 评估方法

本次模型仿真采用平均绝对误差 (mean absolute error, MAE) 来表示输出和预测输出之间的绝对误差、均方根误差 (root mean squared error, $RMSE$) 用于衡量预测值与实际值之间的差异以及决定系数 (R -square, R^2) 来评估模型的性能。 MAE 、 $RMSE$ 的值越小、 R^2 的值越大表示模型效果越好。各种指标的定义式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (16)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{avg})^2} \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (18)$$

在以上指标中 y_i 为第 i 次测量值, \hat{y}_i 为第 i 次对应的预测值, \hat{y}_{avg} 表示加权平均值, n 代表数据点的个数。

3 实验结果

为确保 CALS 模型的有效性, 分别在个体学习、集中学习、联邦学习模型架构上进行多次实验, 并与

其他多种深度学习模型相比较。本次实验在一台 CPU 为 Intel(R) Core(TM) i5-9300H CPU @ 2.40 GHz, GPU 为 NVIDIA GeForce GTX 1660 的计算机上进行, 编程语言采用 Python 3.9。

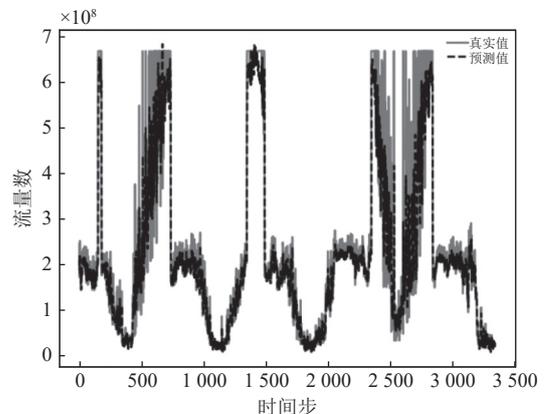
本次联邦学习进行 50 轮全局联邦训练, 采用联邦聚合算法 FedNova 用于聚合客户端的模型参数, 每个客户端在本地训练 3 轮, 总的训练周期为 150 (全局联邦训练轮数 \times 本地训练轮数)。使用 Adam 优化器, 来优化模型参数, 其中优化器的学习率为 0.001, 每次训练时使用 128 个样本。本地早停耐心值为 50, 表示如果验证集性能在 50 个训练周期内没有提升, 则停止训练。通过对不同模型的仿真实验得到如表 1 的结果。

从表 1 中可以看出个体学习效果不如集中学习与联邦学习, 这是因为在个体学习中数据仅来自单个设备, 模型的训练数据量有限, 无法利用全局信息进行改进。本次实验提出的模型 CALS 对比 LSTM 在个体学习、集中学习、联邦学习上对 MAE 性能分别提升了 5.27%、5.62%、11.1%, 对比 GRU 分别提升了 5.51%、10.8%、9.42%, 这表明了 CALS 模型能提高网络流量的预测结果且能在一定程度上解决联邦学习中的数据异构问题。图 6 分别给出了 CALS、ARIMA、LSTM 模型的流量预测图, 方便直观地看出不同模型的预测效果。

表 1 不同方法在个体学习、集中学习、联邦学习上的比较

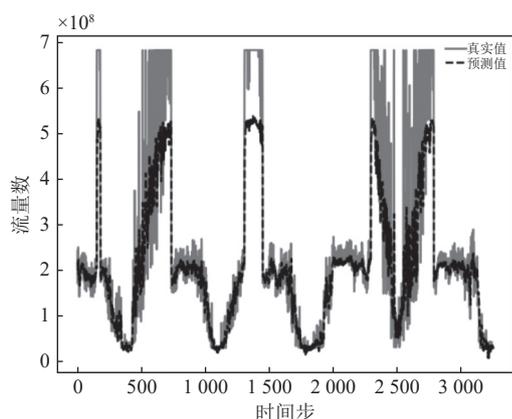
模型	个体学习			集中学习			联邦学习		
	MAE	R^2	$RMSE$	MAE	R^2	$RMSE$	MAE	R^2	$RMSE$
ARIMA	0.2481	0.7643	0.4119	0.1235	0.7883	0.3551	0.0331	0.8583	0.0312
LSTM	0.2387	0.7852	0.4061	0.1068	0.8015	0.2675	0.0314	0.8778	0.0288
GRU	0.2393	0.7819	0.3955	0.1130	0.7981	0.3162	0.0308	0.9311	0.0291
MLP	0.2524	0.7491	0.5051	0.1263	0.7962	0.3171	0.0452	0.7328	0.1172
CALS	0.2261	0.7914	0.3825	0.1008	0.8067	0.2086	0.0279	0.8757	0.0252

由图 6 可以看出 3 个模型对流量的趋势都有很好的预测效果, 但是 ARIMA 模型在具体数值上的预测结果较差一些, 对比 3 个模型可以看出 CALS 模型对网络流量预测的结果更加可靠, 具有更好的拟合效果。值得注意的是在 3 个模型的预测结果中, 在时间步为 500、2400 以及 2600 步时, 每个模型的预测结果都出现了不同程度的偏差, 其原因是噪声问题, 即数据中存在的异常值 (如突发的流量尖峰) 干扰了模型的学习预测, 尽管已经对数据集进行了预处理, 但仍无法完全消除噪声影响后续将深入研究该问题。

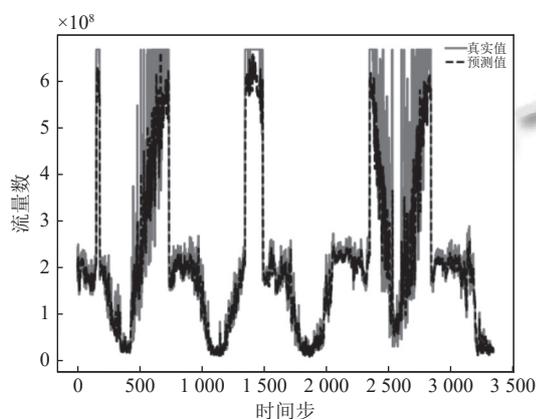


(a) CALS模型预测效果

图 6 预测结果



(b) ARIMA模型预测效果



(c) LSTM模型预测效果

图6 预测结果(续)

4 结束语

本文研究了网络流量的预测问题,与联邦学习相结合设计了 CALS 模型,该模型使用经典的深度学习方法如长短时记忆网络等,并在其中添加 CAA 注意力机制通过对特征数据的池化,量化出不同特征值的贡献,该模型较好地解决了传统深度学习方法泛化能力不足以及联邦学习中数据异构的问题,有效提升了基站网络流量的预测能力.通过在具体数据集上进行模拟,实验结果表明,CALS 模型能够对预测结果有一定的提升.得到流量预测的结果后可以在流量峰值接入 MESH 自组网来减轻基站的负担.需要注意的是本次数据集并没有区分流量数据的类型,如短信、视频等,无法有效分析出用户流量的具体类型,以及减少数据噪声的影响,这些将是接下来的努力方向.

参考文献

1 Panayiotou T, Michalopoulou M, Ellinas G. Survey on

machine learning for traffic-driven service provisioning in optical networks. *IEEE Communications Surveys & Tutorials*, 2023, 25(2): 1412–1443. [doi: [10.1109/COMST.2023.3247842](https://doi.org/10.1109/COMST.2023.3247842)]

2 Wang ZH, Wong VWS. Bayesian meta-learning for adaptive traffic prediction in wireless networks. *IEEE Transactions on Mobile Computing*, 2024, 23(6): 6620–6633. [doi: [10.1109/TMC.2023.3325301](https://doi.org/10.1109/TMC.2023.3325301)]

3 Jiang WW. Cellular traffic prediction with machine learning: A survey. *Expert Systems with Applications*, 2022, 201: 117163. [doi: [10.1016/j.eswa.2022.117163](https://doi.org/10.1016/j.eswa.2022.117163)]

4 Moon J, Hossain MB, Chon KH. AR and ARMA model order selection for time-series modeling with ImageNet classification. *Signal Processing*, 2021, 183: 108026. [doi: [10.1016/j.sigpro.2021.108026](https://doi.org/10.1016/j.sigpro.2021.108026)]

5 Kontopoulou VI, Panagopoulos AD, Kakkos I, *et al.* A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, 2023, 15(8): 255. [doi: [10.3390/fi15080255](https://doi.org/10.3390/fi15080255)]

6 Trinh HD, Gambin ÁF, Giupponi L, *et al.* Mobile traffic classification through physical control channel fingerprinting: A deep learning approach. *IEEE Transactions on Network and Service Management*, 2021, 18(2): 1946–1961. [doi: [10.1109/TNSM.2020.3028197](https://doi.org/10.1109/TNSM.2020.3028197)]

7 Kulin M, Kazaz T, De Poorter E, *et al.* A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer. *Electronics*, 2021, 10(3): 318. [doi: [10.3390/electronics10030318](https://doi.org/10.3390/electronics10030318)]

8 Lin CY, Su HT, Tung SL, *et al.* Multivariate and propagation graph attention network for spatial-temporal prediction with outdoor cellular traffic. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. ACM, 2021. 3248–3252. [doi: [10.1145/3459637.3482152](https://doi.org/10.1145/3459637.3482152)]

9 Han QD, Lu SQ, Wang WH, *et al.* Privacy preserving and secure robust federated learning: A survey. *Concurrency and Computation: Practice and Experience*, 2024, 36(13): e8084. [doi: [10.1002/cpe.8084](https://doi.org/10.1002/cpe.8084)]

10 McMahan HB, Moore E, Ramage D, *et al.* Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017*. Fort Lauderdale: PMLR, 2017. 1273–1282.

11 Zhao Y, Li M, Lai LZ, *et al.* Federated learning with non-IID data. *arXiv:1806.00582*, 2018.

12 Mai W, Yao JC, Chen G, *et al.* Server-client collaborative

- distillation for federated reinforcement learning. *ACM Transactions on Knowledge Discovery from Data*, 2023, 18(1): 9. [doi: [10.1145/3604939](https://doi.org/10.1145/3604939)]
- 13 Wang JY, Liu QH, Liang H, *et al.* Tackling the objective inconsistency problem in heterogeneous federated optimization. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 7611–7623. [doi: [10.5555/3495724.3496362](https://doi.org/10.5555/3495724.3496362)]
- 14 Li YX, Hou QB, Zheng ZH, *et al.* Large selective kernel network for remote sensing object detection. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023. 16748–16759. [doi: [10.1109/ICCV51070.2023.01540](https://doi.org/10.1109/ICCV51070.2023.01540)]
- 15 Tang YH, Han K, Guo JY, *et al.* GhostNetV2: Enhance cheap operation with long-range attention. *Proceedings of the 36th International Conference on Neural Information Processing Systems*. New Orleans: Curran Associates Inc., 2022. 724.
- 16 Wang L, Che LX, Lam KY, *et al.* Mobile traffic prediction with attention-based hybrid deep learning. *Physical Communication*, 2024, 66: 102420. [doi: [10.1016/j.phycom.2024.102420](https://doi.org/10.1016/j.phycom.2024.102420)]
- 17 Trinh HD, Giupponi L, Dini P. Mobile traffic prediction from raw data using LSTM networks. *Proceedings of the 29th IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Bologna: IEEE, 2018. 1827–1832. [doi: [10.1109/PIMRC.2018.8581000](https://doi.org/10.1109/PIMRC.2018.8581000)]
- 18 Yao GL, Lei T, Zhong JD. A review of convolutional-neural-network-based action recognition. *Pattern Recognition Letters*, 2019, 118: 14–22. [doi: [10.1016/j.patrec.2018.05.018](https://doi.org/10.1016/j.patrec.2018.05.018)]
- 19 Dhillon A, Verma GK. Convolutional neural network: A review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 2020, 9(2): 85–112. [doi: [10.1007/s13748-019-00203-0](https://doi.org/10.1007/s13748-019-00203-0)]
- 20 Zhang KX, Hong M, Zhang KX, *et al.* Forecasting crude oil price using LSTM neural networks. *Data Science in Finance and Economics*, 2022, 2(3): 163–180. [doi: [10.3934/DSFE.2022008](https://doi.org/10.3934/DSFE.2022008)]
- 21 Khorram A, Khalooei M, Rezghi M. End-to-end CNN+LSTM deep learning approach for bearing fault diagnosis. *Applied Intelligence*, 2021, 51(2): 736–751. [doi: [10.1007/s10489-020-01859-1](https://doi.org/10.1007/s10489-020-01859-1)]
- 22 Shiri FM, Perumal T, Mustapha N, *et al.* A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *arXiv:2305.17473*, 2023.

(校对责编: 张重毅)