



基于个性化联邦学习的无线通信流量预测

林尚静^{1*}, 马 冀², 李月颖¹, 庄 珩¹, 李 铁¹, 李子怡¹, 田 锦²

(1. 北京邮电大学安全生产智能监控北京市重点实验室 北京 海淀区 100876; 2. 金陵科技学院网络与通信工程学院 南京 211169)

【摘要】提出一种基于分布式的城市全域通信流量预测算法 Fed-DenseNet, 各个边缘计算服务器在中心服务器的协调下进行协同训练, 中心服务器利用 KL 散度挑选出流量分布相似的区域流量模型, 并采用联邦平均算法对具有相似流量分布的区域流量模型的参数进行融合, 以较低的复杂度和通信开销实现城市全域流量预测。此外, 城市范围内不同地区流量具有高度差异化的特征, 为此, 在 Fed-DenseNet 算法基础上, 提出基于合作博弈的个性化联邦学习算法 p-Fed-DenseNet, 将本地区的各个区域性数据特征作为合作博弈的参与者, 通过合作博弈的超可加性准则, 进行本地区特征的筛选, 从而达到既能提高模型的泛化性, 又能够保持对本地流量精准刻画的目的。

关键词 云边协同; 合作博弈; 联邦学习; 时空依赖性; 无线流量预测

中图分类号 TN92 **文献标志码** A **doi**:10.12178/1001-0548.2022102

Citywide Wireless Traffic Prediction Based on Personalized Federated Learning

LIN Shangjing^{1*}, MA Ji², LI Yueying¹, ZHUANG Bei¹, LI Tie¹, LI Ziyi¹, and TIAN Jin²

(1. Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications Haidian Beijing 100876;

2. School of Network and Communication Engineering, Jinling Institute of Technology Nanjing 211169)

Abstract Wireless communication network traffic prediction is of great significance to operators in network construction, base station wireless resource management and user experience improvement. However, the existing centralized algorithm models face the problems of complexity and timeliness, thus difficult to meet the traffic prediction of the whole city scale. A distributed urban global traffic prediction algorithm Fed-DenseNet is proposed in this paper. Each edge computing server of the algorithm performs collaborative training under the coordination of the central server, and the central server uses KL (Kullback-Leibler) divergence to select regional traffic models with similar traffic distribution and uses the federated average algorithm to fuse the parameters of these regional traffic models. In this way, the urban global traffic prediction can be realized with lower complexity and communication cost. In addition, the traffic in different areas within the city is highly differentiated, so how to improve the accuracy of model prediction is also facing challenges. Based on Fed-Densenet algorithm, a personalized federated learning algorithm p-Fed-DenseNet based on cooperative game is proposed. Each regional data feature in the region is taken as a participant of cooperative game, and local features are screened by the super-additivity criterion of cooperative game, so as to achieve the purpose of both improving the generalization of the model and maintaining the accurate description of local traffic.

Key words cloud-edge collaboration; coalitional game; federated learning; spatio-temporal dependence; wireless traffic prediction

精准预测城市全域尺度通信网络流量能够辅助运营商进行精细化运营、高效配备与部署基站资源, 从而满足涌现的各种业务需求。

然而, 城市全域尺度的通信流量预测面临 3 方

面困难。1) 预测模型的复杂性。现有的单个或局部范围的流量预测模型, 由于数据体量不大, 通常是集中式算法模型。如果直接扩展到城市全域范围, 必然导致在模型训练过程中的复杂度急剧扩张。

收稿日期: 2022-04-11; 修回日期: 2022-08-31

基金项目: 国家重点研发计划 (2019YFC1511400); 中央高校基本科研业务费 (2021RC07); 泛网无线通信教育部重点实验室开放基金 (KFKT-2020102)

作者简介: 林尚静 (1986-), 女, 博士生, 主要从事深度学习、分布式计算及边缘计算等方面的研究。

*通信作者: 林尚静, E-mail: linshangjing@bupt.edu.cn

2) 预测模型的时效性。在模型持续运行过程中, 分散在城市全域范围的各个基站的实时流量数据需要全部汇聚到部署了全域流量预测模型的中心节点上进行演算, 这不仅导致通信负载的增加, 还导致预测模型的时效性较低。3) 预测模型的准确性。城市尺度的流量特征是高度差异化, 与城市的规划、城市居民生活习惯、通信设施部署息息相关。城市尺度的流量预测模型需要准确性以刻画本城市的流量特征。

本文的主要贡献体现在以下两个方面。

1) 提出了基于联邦学习的城市全域流量预测框架。首先, 各个边缘计算服务器采集相应区域内的流量数据, 并在本地流量预测模型训练; 其次, 各个边缘计算服务器将本地流量预测模型的模型参数上传给中心云服务器, 由中心云服务器对模型参数进行融合, 并将融合后的模型参数下发给各个边缘计算服务器; 最后, 各个边缘计算服务器在融合后的模型参数基础上继续进行本地模型训练, 直至训练结束, 形成通用流量预测模型。现有流量预测方案普遍采用集中式训练框架, 需要将城域级流量数据汇聚到一台中心服务器上进行模型训练, 通信开销大, 模型复杂度高。本文所提的模型是一种分布式训练框架, 各边缘计算服务器只需采集区域级的流量数据进行模型训练, 因此, 通信开销较小, 模型复杂度较低。此外, 联邦训练预测框架通过融合各区域的模型参数, 间接地扩大了各个边缘计算服务器上的流量训练集, 从而提高了区域级流量预测的准确性。

2) 提出了基于合作博弈的个性化联邦流量预测模型。各个边缘计算服务器在执行完上述联邦训练过程形成通用流量预测模型之后, 各自独立训练出反应本地区特征的个性化的流量预测模型。具体地, 各个边缘计算服务器将本地区的各个区域性特征作为合作博弈的参与者, 通过合作博弈的超可加性准则, 比较不同特征组合对流量预测产生的收益, 根据收益进行本地区特征的筛选。对比现有多源流量预测模型, 通常将流量数据和所有的多源特征数据直接融合在一起, 以提升模型预测准确度。但是, 融合的多源特征过多会加深模型训练的复杂度, 导致预测效果变差。而本文所提的基于合作博弈的区域个性化特征筛选方法, 能够为每个区域引入合适的多源特征数据, 从而达到提升模型预测效果的目的。

1 相关研究

针对蜂窝流量预测问题, 国内外已涌现大量的相关研究。传统算法普遍采用统计概率模型或者时间序列预测模型进行流量预测^[1-2]。预测算法都是针对每个小区进行单独建模, 无法直接适用于所有小区。在实际应用中, 对千万数量级别小区流量进行并行化建模非常困难。

随着深度学习算法在各领域取得突破进展的同时, 能够捕获空间相关性的卷积神经网络(convolutional neural network, CNN)以及能够捕获时间相关性的长短时记忆(long short-term memory, LSTM)网络被逐步应用到通信流量预测领域。深度学习模型由于具有拟合复杂非线性特点, 通常将多个小区的流量数据整体作为训练集进行训练并产生统一模型用于多个小区的预测。然而, 随着流量预测范围扩大到城市全域尺度, 导致预测模型的复杂度提高, 实际预测效果不理想。

为了解决以上问题, 本文在模型中引入联邦学习的思想。联邦学习是一种具有隐私保护的分布式机器学习训练框架, 多个客户端在一个中心服务器的协同之下共同训练一个模型^[3-4]。分布式算法模型可以大范围地进行流量预测, 数据集增多可提高模型预测精度, 同时也不会出现复杂度高、实时性低的问题。

但在联邦学习的研究过程中, 还面临着一个问题, 常见数据集大多都是独立同分布(independent identically distribution, IID), 但在现实场景中, 不同边缘计算服务器面对的数据质和量都不尽相同, 很难满足IID的假设, 而在模型训练时 non-IID数据往往比IID数据表现差。为了解决这一问题, 文献[5]针对 non-IID数据根据每个客户机的贡献确定其参与训练的权重, 模型性能有很大提升。其中对贡献的定义分为两类: 一是根据每个客户机上不同类别数据的数量; 另一类是根据本地模型在验证集上的准确性。文献[6]根据客户机的数据分布的相似性确定客户机集群, 改进了模型的个性化与泛化性能。

通信流量数据大多为 non-IID数据, 也同样面临上述问题。为了解决该问题, 本文在挑选出流量分布相似的区域中引入联邦平均算法, 提高模型的泛化性能, 引入合作博弈对各区域的本地特征进行筛选, 提高模型的个性化。

2 蜂窝流量特征分析

结合“意大利电信大数据挑战赛”中所提供的开源数据集,分析蜂窝流量的在时空范围的相关性和分布特性。该数据集将整个米兰市以 235 m×235 m 的地理粒度划分为 100×100 个栅格(栅格标号从左下角开始到右上角结束,从左至右、从下至上,依次为 1~10000,且定义左下角栅格(编号为 1)的中心坐标是 (0, 0) km, 右上角栅格(编号为 10000 的中心坐标是 (23.265, 23.265) km), 并且统计了栅格内用户的语音、短信以及上网流量。统一以栅格内每小时短消息业务流量为例进行数据分析。

2.1 蜂窝流量的相关性

蜂窝流量在时间和空间两个维度都具有相关性。蜂窝流量是人群活动的反映。人群活动相似性决定了蜂窝流量具有相似性。人群的自然属性(每天的活动规律:早起晚睡)决定了蜂窝流量在时间

维度具有相关性,人群的社会属性(工作日往返于家和办公地点,每周末往返于家和公共娱乐场所)导致了流量都具有空间相关性。

随机选择 3 个不同地区的栅格,处于米兰市中心的栅格 (ID=5051)、处于市区的栅格 (ID=3439) 以及郊区栅格 (ID=444)。首先,通过自相关系数衡量蜂窝流量的时间相关性,如图 1a 所示。3 个栅格的负载时间序列的业务流量,与 1 阶滞后和 2 阶滞后流量的相关性均较高。这表明相邻时刻业务流量有极大相关性。其次,3 个栅格的负载时间序列均存在以天为单位的周期性,并且时间间隔越短,周期性规律越显著。通过皮尔逊相关系数 (pearson correlation coefficient, PCC) 衡量蜂窝流量的空间相关性,如图 1b 所示。以 3 个栅格 (ID=5051、ID=3439、ID=444) 为目标栅格,以空间范围 (2.115×2.115) km 来计算其与临近栅格流量的 PCC 相关系数。可以看出,蜂窝流量具有一定的空间相关性,且相关性依赖于与目标小区的空间距离。

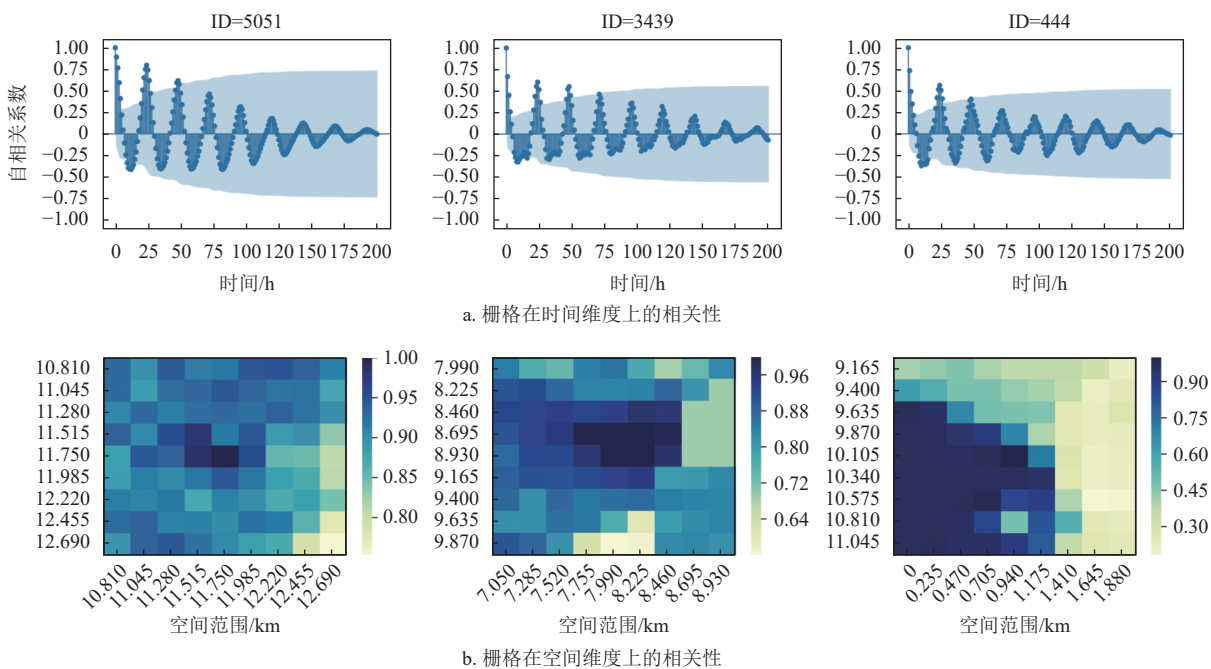


图 1 短消息业务流量在时空维度上的相关性

2.2 蜂窝流量的分布性

大量个体的社会化活动聚集起来形成群体的社会化活动导致在地理空间上形成功能区域(商业区、大学、办公区、公交枢纽)。功能区域反向影响了人群的活动特性,导致不同功能区域的蜂窝流量又呈现出差异性。

本文统计了中心栅格 (ID=5051)、市区栅格

(ID=3439) 以及郊区栅格 (ID=444) 每小时短消息业务的分布,如图 2 所示。可以看出,中心栅格业务量最大,市区栅格业务量次之且拖尾较长,而郊区栅格业务量最低。此外,不同栅格的短消息业务分布也不尽相同。

2.3 蜂窝流量特征对流量预测模型的影响

一方面,蜂窝流量在全市范围内都存在时间、

空间两个维度, 且都具有相关性。因此, 各区域可采用相同的流量预测模型以捕获共同的时空相关性。另一方面, 全市不同区域的蜂窝流量具有相关性 (非独立) 且具有不同的分布, 即 non-IID。因此, 流量预测模型又需要体现出地区的差异性。

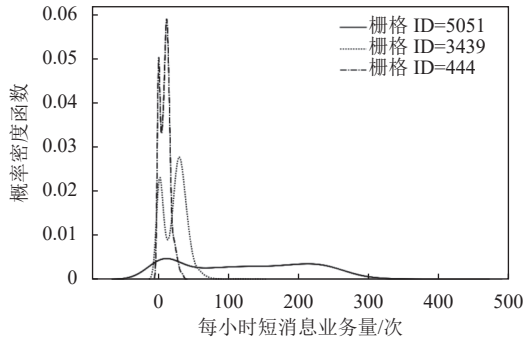


图 2 短消息业务流量的分布

3 基于个性化联邦学习的城市全域尺度流量预测模型

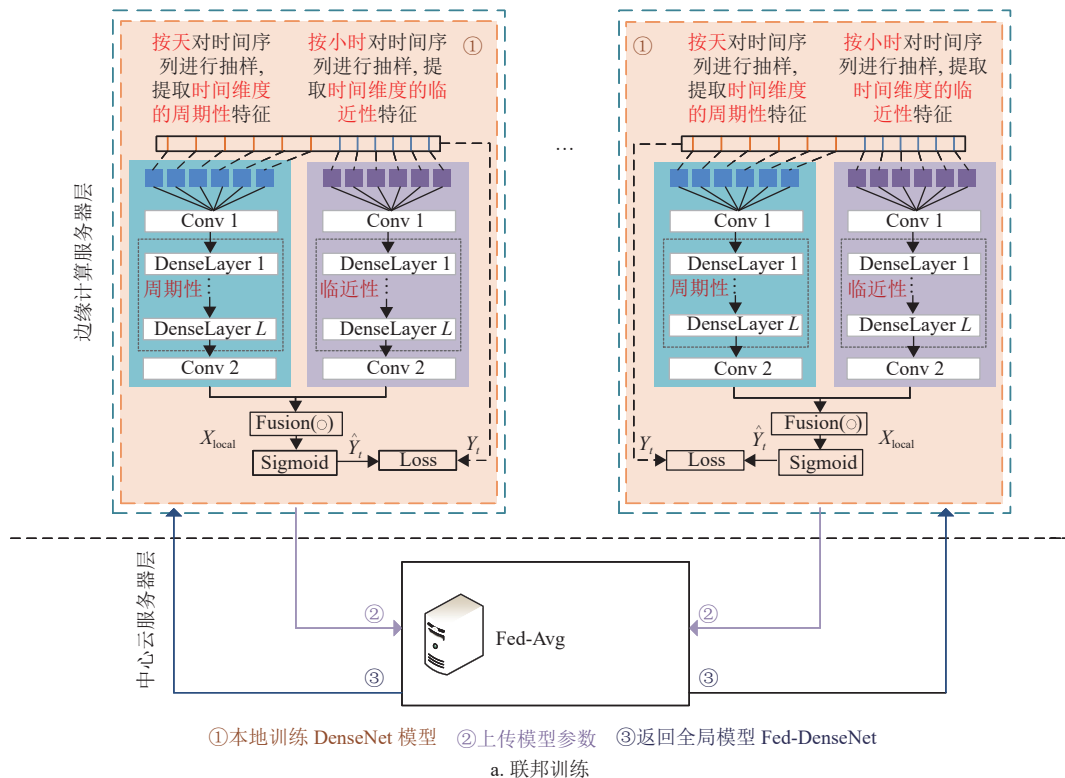
本文提出了基于个性化联邦学习的城市全域通信流量预测模型 (personalized-Fed-DenseNet, p-Fed-DenseNet)。首先, 各个移动边缘计算服务器采集本地的流量数据, 在本地训练 DenseNet 模型, 然后在中心服务器的协同下, 基于联邦平均算法将多个移动边缘计算服务器本地训练的 DenseNet 模型

融合一个通用的流量预测模型 Fed-DenseNet, 如图 3a 所示; 其次, 在 Fed-DenseNet 的全局模型上, 结合本地区的区域性数据 (天气气候、基站分布密度、地理功能区域、社交节日活动等), 将本地区的各个区域性数据特征作为合作博弈的参与者, 通过合作博弈的超可加性准则, 筛选区域数据, 如图 3b 所示。最终每个区域结合筛选出的区域数据训练出 p-Fed-DenseNet 模型, 如图 3c 所示。

本章详细介绍了部署在边缘计算服务器端的时空流量预测 DenseNet 模型、云服务器端联邦平均算法形成 Fed-DenseNet 以及边缘端的个性化联邦模型 p-Fed-DenseNet 训练过程, 分析了模型的复杂度和预测的时效性。

3.1 时空流量预测模型

一方面, 在时间维度上, 相邻时刻的业务流量具有极大相关性, 且短间隔周期内的业务流量也表现出规律性。因此, 本文将这两者分别定义为时空流量预测模型中的临近层及以天为单位的周期层, 以捕获蜂窝流量在时间维度上的依赖性。设定第 t 个时隙的流量是要被预测的目标, 定义参数 c 和 p 分别为临近依赖性与周期依赖性的大小, 按照 c 与 p 选取归一化的历史流量数据作为临近层和周期层的输入, 分别表示为 $[X_{t-c}, X_{t-(c-1)}, \dots, X_{t-1}]$ 与 $[X_{t-pm}, X_{t-(p-1)m}, \dots, X_{t-m}]$, $m = 24$ 。



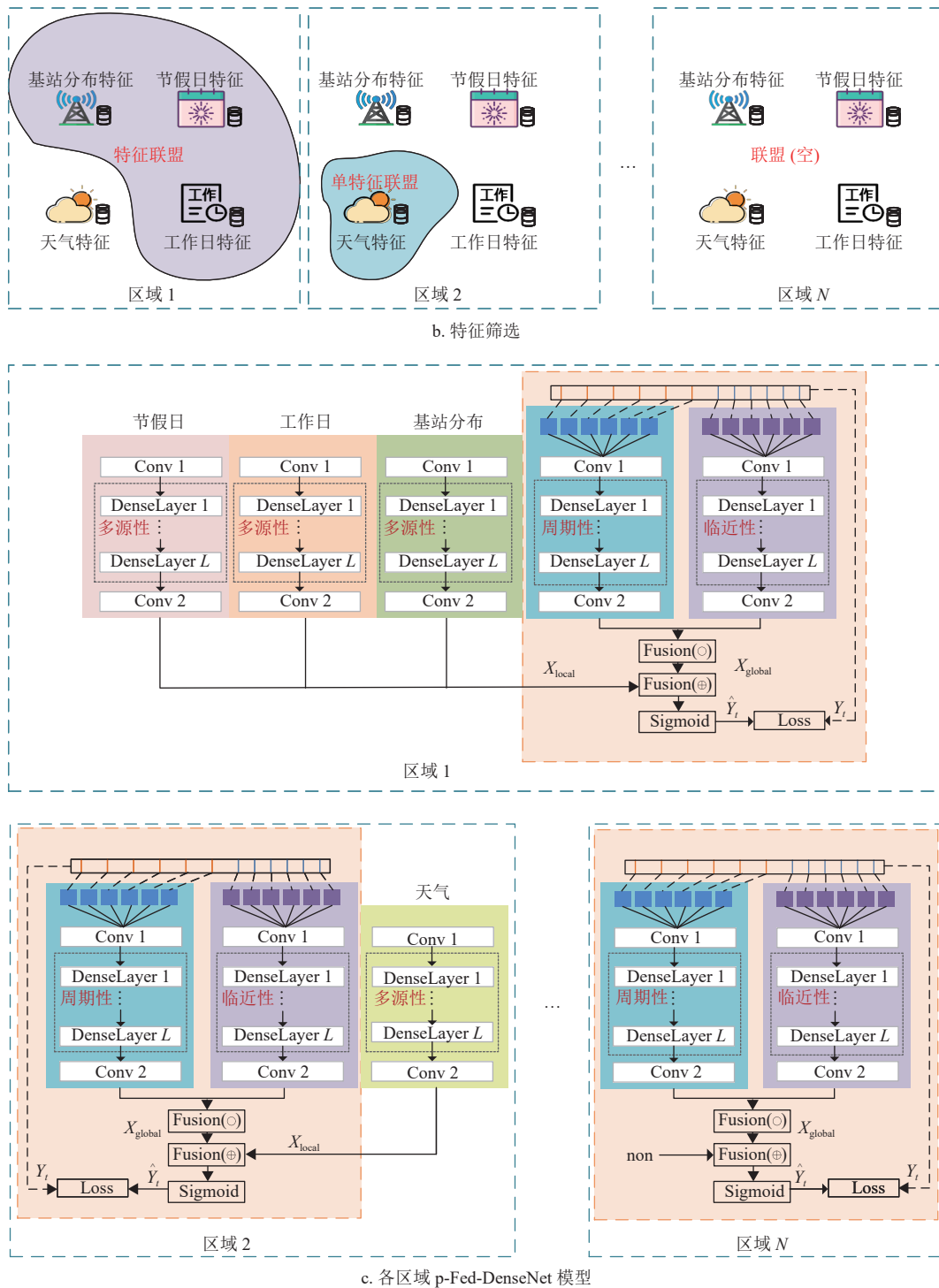


图 3 基于个性化联邦学习的全域流量预测模型 p-Fed-DenseNet 形成过程

另一方面，在空间维度上，地理上相邻的地区业务流量也具有相关性。因此，采用密集卷积操作以充分捕获区域之间的空间依赖性。具体地，采用密集连接神经网络训练参数，将输入送入相互独立的神经网络单元 DenseUnit，进行时空维度特征的提取。DenseUnit 依次包含 3×3 卷积层、DenseBlock 架构和 1×1 卷积层，其中，第一

层 3×3 卷积层能够遍历并融合所覆盖区域的流量数据，提取出该区域的抽象特征；DenseBlock 架构以 L 层全连接层构成，对特征进行非线性函数的复合计算； 1×1 卷积层对 DenseBlock 的输出再次提取特征与降低维度，得到每个 DenseUnit 的输出结果。

最后，将临近层和以天为单位的周期层所提取

到的时间相关性特征结果分别定义为 F_c 和 F_p , 以不同的权重 α_c 和 α_p ($\alpha_c, \alpha_p \in [0,1]$) 对两个时间层的结果进行基于参数矩阵的融合, 得到融合结果为:

$$Y_g = \alpha_c \circ F_c + \alpha_p \circ F_p \quad (1)$$

式中, \circ 表示哈达玛乘积, 权重矩阵与特征输出矩阵相同位置分别相乘。

通过 Sigmoid 激活函数, 将 Y_g 映射到 $[0,1]$, 从而保证与真实流量标签范围一致, 得到时空流量预测模型在 t 时刻的输出, 定义为:

$$\widehat{Y}_t = \sigma(Y_g) \quad (2)$$

3.2 个性化联邦学习算法

通过 2.2 节分析可知, 不同区域的蜂窝流量具有相关性 (非独立) 且具有不同的分布, 即 non-IID。城市全域范围内蜂窝流量 non-IID 问题会导致本地训练的流量预测模型之间出现较大差异性, 如果直接采用联邦平均融合算法, 融合后的全局模型性能并不佳。针对移动边缘计算服务器上区域流量预测模型差异性导致融合模型性能变差的问题, 提出了基于合作博弈的个性化联邦学习算法。

1) 全局模型统一训练过程

① 初始化。在中心云服务器, 随机生成一个初始化全局模型参数 ω , 并将该全局模型推送至各个参与联邦训练的区域。

② 中心融合过程。区域 r 接收到 ω 后, 根据本地流量数据训练全局模型并更新参数, 得到更新后的本地模型参数 ω_t^r 。所有区域同步训练, 将模型参数上传至中心云服务器后, 中心云服务器根据联邦平均算法对全局模型进行一次融合更新:

$$\omega_t = \frac{1}{R} \sum_{r=1}^R \omega_t^r \quad \forall t \quad (3)$$

式中, ω_t 表示更新后的全局模型。

③ 本地迭代过程。将本地模型的一次参数更新称为一次迭代, 设定联邦训练迭代次数, 达到次数后停止参数更新。从训练样本中选取 b 表示一个批量, 本地模型根据梯度下降法进行参数更新, 以使得损失函数最小化, 即:

$$\omega_t^r \leftarrow \omega_t^r - \frac{\eta}{|b|} \nabla f(\omega_t^r) \quad (4)$$

式中, $f(\omega_t^r)$ 表示模型的损失函数; η 表示学习率, $\eta \in [0,1]$ 。损失函数可以利用 L_2 范数表示:

$$f(\omega_t^r) = \|Y_t - \widehat{Y}_t\|_2^2 \quad (5)$$

式中, Y_t 表示流量真实值。

2) 本地模型个性化训练过程

为了捕捉特定区域的通信流量需求, 每个移动边缘计算服务器都基于全局模型和本地个性化数据 (基站分布、工作日、节假日、天气等) 进行训练, 根据训练的结果选择合适的本地特征, 建立一个最优的本地化流量预测模型。本地模型的训练过程如下。

将临近层和周期层融合后的输出矩阵与本地化特征输出矩阵做一个拼接操作, 得出模型 t 时刻的预测输出:

$$\widehat{Y}_t = \sigma(X_{\text{global}} \oplus X_{\text{local}}) \quad (6)$$

式中, \oplus 表示拼接操作; $\sigma(\cdot)$ 表示 Sigmoid 激活函数。输入数据已经过 Min-Max 归一化处理为 $[0,1]$, 但在通过模型的复杂运算后, 输出预测数据的范围产生了变化, 而真实流量标签的范围为 $[0,1]$, 为方便利用预测数据与真实数据进行误差计算, 需要选取 Sigmoid 激活函数将预测结果对应映射到 $[0,1]$ 。

根据预测结果与真实结果, 可利用 L_2 范数来表示损失函数, 对整个模型进行训练过程中的参数更新:

$$L(\theta) = \|Y_t - \widehat{Y}_t\|_2^2 \quad (7)$$

式中, θ 指模型所训练的使得损失函数最小的参数集合; Y_t 表示真实值。

3) 个性化特征选择过程

为了对每个边缘服务器建立最优的本地化流量预测模型, 在全局模型基础上, 结合本地进行个性化特征的选择, 进一步训练本地化模型。原因有如下两方面: 一方面, 不同的区域之间存在各种差异, 如基站数量、兴趣点 (points of interest, POI) 数量、天气等, 不同的区域特征对各个区域流量的影响不同。另一方面, 同时引入多个特征会增加模型处理的复杂度, 不能提高预测精度。因此, 为每个边缘服务器选择合适的本地特征就显得尤为重要。

合作博弈^[7-8]研究多个个体如何达成合作, 以及达成合作后的收益分配问题。本文所提的个性化特征选择过程, 同样可以看成是各个特征的合作问题。将各个本地特征 (基站数量、POI 数量、天气等) 视为合作博弈的参与者, 这些本地特征相互合作共同完成流量预测的目标, 并以预测误差作为评价这些本地特征共同合作的收益。

本文定义本地个性化特征的合作博弈如下:

$G = [F, V]$ 是 N 个本地特征合作博弈，其中， $F = \{f_1, f_2, \dots, f_N\}$ 表示 N 个本地特征的全集， $S = \{f_1, f_2, \dots, f_n\}$ ($n \leq N$) 是隶属于 F 的子集 $S \subseteq F$ ，表示一个联盟。 $V(S)$ 是联盟 S 的特征函数，表示 n 个本地特征合作所获得的收益。在流量预测问题中，收益为预测的均方根误差 (root mean squared error, RMSE)。

本地个性化特征合作博弈 G 要解决的首要问题是合作的参与者 (本地特征) 如何结成稳定的合作联盟。本文引入合作博弈中超可加性以解决合作博弈 G 的稳定性问题。超可加性表示两个互不相交的联盟合作之后获得的收益要大于这两个联盟各自收益之和，即对于任意联盟 S 和 T ， $S, T \subseteq F$ 且 $S \cap T = \emptyset$ ，有 $V(S \cup T) \geq V(S) + V(T)$ 。由于本文是以预测误差表示多个特征的合作收益，即 $\text{RMSE}(S \cup T) \leq \text{RMSE}(S) + \text{RMSE}(T)$ 。因此，在本地个性化特征筛选中，多个本地特征只有满足超可加性，其结成的合作联盟才是稳定的。

对 n 个本地特征所结成的合作联盟的满足超可加性的判断过程如算法 1 所示。

算法 1 本地特征结成联盟的稳定性判断过程

输入：对于某区域，其本地特征的集合 $I = \{f_1, f_2, \dots, f_n\}$ ($n \geq 2$)

输出：本地特征形成的所有稳定联盟的集合 S

$A_N = \{V_k | k=1, 2, \dots, C_n^N = \{f_m | m=1, 2, \dots, N\}\}$ ($N \in [2, n]$, $m \in [1, n]$) // A_N 为所有 N 特征联盟的集合

$L = \{\}, S = \{\}, N = 2$ // L 存放不稳定联盟， S 存放稳定联盟

while $N \leq n$ do // 开始判定

for $V_k \in A_N$ do

if $N > 2$ then

if $\exists L_q \in L, L_q \subseteq V_k$ then

add V_k to L

break from line 4

end if

end if

if 对于 $\forall T_j$ 都满足 $\text{RMSE}(V_k) \leq \sqrt{\sum_{j=1}^J \text{MSE}(T_j) / N}$

(其中 $T_j \subseteq V_k$ ，且 $T_1 \cup T_2 \cup \dots \cup T_J = V_k$ ， T_1, T_2, \dots, T_J 两两相交为 \emptyset) then

联盟 V_k 稳定，add V_k to S

else

联盟 V_k 不稳定，add V_k to L

end if

end for

$N \leftarrow N + 1$

$A_N \leftarrow A_{N+1} // (N+1)$ 特征联盟集合

end while

return S // 返回包含所有稳定联盟的集合

在所有稳定联盟的集合 S 中选择 RMSE 最小的联盟，该联盟中的本地特征即是每个区域的最优特征。

3.3 模型复杂度与时效性分析

本文所提模型训练过程包括 3 个阶段。

1) 各个边缘计算服务器收集区域内流量数据进行本地模型训练。在该阶段，各个服务器只需采集本区域的流量数据，相较集中式算法，通信开销少，时效性高。

2) 各个边缘计算服务器在中心云服务器的协同下进行联邦训练。联邦训练框架采用的基模型为 DenseNet。基于联邦训练，各个边缘计算服务器形成通用流量预测模型 Fed-DenseNet。分布式 Fed-DenseNet 模型仅需要负责区域级流量特征捕获，集中式 DenseNet 模型需要负责城域级流量特征捕获。流量特征越复杂多变，就要求 DenseNet 模型的网络深度越深。因此，分布式 Fed-DenseNet 模型比集中式 DenseNet 模型复杂度低。

3) 各个边缘计算服务器在通用流量预测模型的基础上，基于合作博弈的超可加性，对本区域多源特征进行筛选，形成个性化流量预测模型。基于合作博弈中的超可加性对区域级的多源特征进行筛选，算法复杂度为 $O(N^2)$ ， N 为区域级的多源特征总数。由于区域级的多源特征数量有限，所以，基于合作博弈搜索稳定的特征联盟过程的时间复杂度在可接受的范围之内。

综上分析，相比集中式模型，本文提出的模型复杂度更低，通信开销更低。

4 模型性能分析

从 4 个方面展开实验对比。1) 在个性化联邦学习中选择不同本地特征进行训练的结果对比，验证了本地特征对流量预测的影响以及进行特征选择的必要性。2) 将本文所提 p-Fed-DenseNet 与传统 Fed-DenseNet、DenseNet 进行对比，验证了个性化联邦学习性能优于传统联邦学习。3) 将 p-Fed-DenseNet 与现有流量预测模型进行对比，验证了联邦所代表的分布式学习优于传统集中式学习。4) 选取一个区域对比不同算法绝对预测误差的分布

函数图, 具体展现了本文所提的 p-Fed-DenseNet 模型的预测性能。

4.1 数据集描述

在原数据集将米兰市划分成 100×100 个栅格的基础上, 再进一步将米兰市按郊区、市区和市中心划分为 25 个区域, 每个区域由 20×20 个栅格组成, 区域编号信息如图 4 所示。

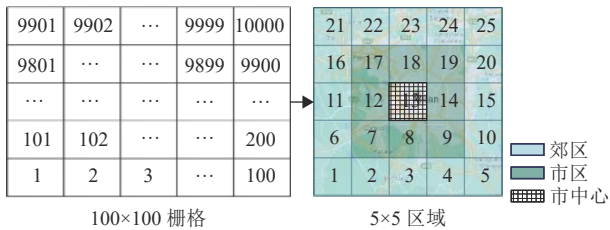


图 4 米兰市区域信息

基于上述划分, 整个米兰市被划分为 3 个行政区: 市中心 (区域 13)、市区 (区域 7、8、9、12、14、17、18、19) 以及郊区 (剩余的 16 个区域)。各个行政区中区域流量比较接近, 使用联邦平均算法

训练全局模型。

4.2 参数设置和实验环境

在训练模型过程中, 先从训练集中取出 10% 作为验证集, 剩下的 90% 用以初次训练模型。在训练过程中逐步调整参数达到最优模型, 最终参数设置为: 以小时和天为单位的历史数据序列长度取 3, 训练过程中学习率取 0.001, Batch size 设为 32, Epoch size 设为 300。

实验环境的硬件配置为: Intel(R)Core(TM) i5-8250U CPU @ 1.60 (8 CPUs)~1.8 GHz, 128 G 内存, Intel(R) UHD Graphics 620, GPU 数量为 2。软件配置为: Python3.7.1, Pytorch1.8.1, CUDA11.3, cuDNN8.1.1。

4.3 个性化联邦学习中本地特征选择

挑选米兰市区中流量分布相近的 7、8、9、12、17 这 5 个区域, 对这 5 个区域进行全局模型训练, 在形成的全局模型基础上, 在不同的本地特征组合下进行训练, 依据合作博弈的超可加性进行特征筛选。训练结果如表 1 所示。

表 1 不同本地化特征组合下个性化联邦学习模型预测误差 (RMSE)

本地特征类型	市区区域7	市区区域8	市区区域9	市区区域12	市区区域17
无	9.69	9.84	7.11	30.84	15.03
工作日/节假日	9.88	10.23	7.01	29.17	14.35
天气	9.48	9.85	7.02	29.01	14.35
基站密度	9.45	9.87	6.96	28.34	14.21
天气+基站密度	9.44	10.06	7.05	28.7	14.22
工作日/节假日+天气	9.90	10.24	7.08	29.35	14.39
工作日/节假日+基站密度	9.99	10.21	7.16	28.87	14.10
工作日/节假日+天气+基站密度	9.93	10.13	7.15	29.01	14.22

观察表 1, 大多数区域在全局预测模型的基础上融合本地特征 (节假日、天气 (气温和降雨) 和基站密度) 后, 流量预测的精度都得到提高。但是, 不同的本地特征对不同区域流量的影响程度各不相同, 如区域 7 的流量仅受天气和基站密度的影响, 区域 8 的流量不受任何本地特征的影响。

通过 3.2 和 3.3 节超可加性的判定方法, 就可确定: 区域 7 选择融合天气和基站密度特征, 区域 8 不融合任何特征, 区域 9 和区域 12 融合基站密度特征, 对于区域 17, “天气+基站密度” 和 “工作日/节假日+基站密度” 均满足超可加性, 但 “工作日/节假日+基站密度” 的误差更小, 故选择融合工作日/节假日和基站密度特征。由此可

建立最优的个性化流量预测模型, 达到最佳的流量预测精度。

4.4 集中式学习、联邦学习与个性化联邦学习对比

为验证本文个性化联邦学习的有效性, 分别对集中式学习、普通联邦学习与个性化联邦学习在相同的区域上 (米兰市区中流量分布相近的 7、8、9、12、17 这 5 个区域) 进行预测性能对比。其中, 集中式学习为 5 个区域的流量汇聚起来, 采用一个全局的 DenseNet 模型进行整体流量预测; 普通联邦学习为 5 个区域各自训练自己的 DenseNet, 并通过联邦平均学习框架进行模型参数融合, 形成 Fed-DenseNet; 个性化联邦学习为在 Fed-DenseNet 的全局模型上, 进一步结合本地特征,

训练出 p-Fed-DenseNet。

如图 5 所示,可以看出 p-Fed-DenseNet 预测效果最好, Fed-DenseNet 其次, DenseNet 最差。相比集中式 DenseNet 学习, 个性化联邦学习后, 5 个区域的 RMSE 都呈下降趋势, 区域 7 下降了 9.1%, 区域 8 下降了 3.8%, 区域 9 下降了 4%, 市区 12 下降了 9.5%, 市区 17 下降了 9.1%。原因是, 多区域联邦学习的训练样本数据量增加, 能够更准确地捕获各区域共有流量的时空特性, 在联邦学习的基础上进行个性化学习, 能够更准确地为每个区域引入与流量相关的外部特征。

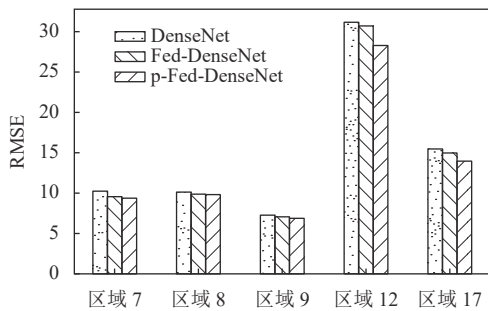


图 5 预测效果对比

另外, 市区区域 12 相比其他区域, 误差较高, 由于 RMSE 描述的是一种离散程度, 因此可能是因为市区 12 的流量还受到除基站密度、气象条件以外的其他未知因素影响, 导致个别预测值与真实值有较大出入, RMSE 也会更高, 而其他区域预测结果则比较稳定, RMSE 较低。

4.5 联邦学习与集中式学习算法对比

本文对比了个性化联邦学习算法与其他常用集中式学习算法的性能。将 LSTM^[9]、ConvLSTM^[10]、DenseNet^[11] 模型在相同数据集下的预测效果与本文提出的 Fed-DenseNet 和 p-Fed-DenseNet 的预测误差进行对比, 对所有个性化模型的均方误差取平均值, 再利用均方根误差与均方误差的关系, 得到 p-Fed-DenseNet 的预测误差, 如图 6 所示。

在短消息业务流量数据集上, Fed-DenseNet 的性能相较于 ConvLSTM 模型提升了 17.1%, 比 LSTM 模型提升 10.7%, 比 DenseNet 模型提升了 1.7%。相对于已有模型, 本文提出的 p-Fed-DenseNet 性能最好。首先, DenseNet 是时空流量预测模型, 有利于捕捉各区域中流量的共性时空特征, 预测性能优于现有集中式模型。其次, Fed-DenseNet 是分布式模型, 通过提升训练样本数据, 性能优于集中式 DenseNet 算法。最后, 本文所提 p-Fed-DenseNet

在 Fed-DenseNet 基础上, 又引入了本地化特征, 因此, 预测性能进一步得到提升。

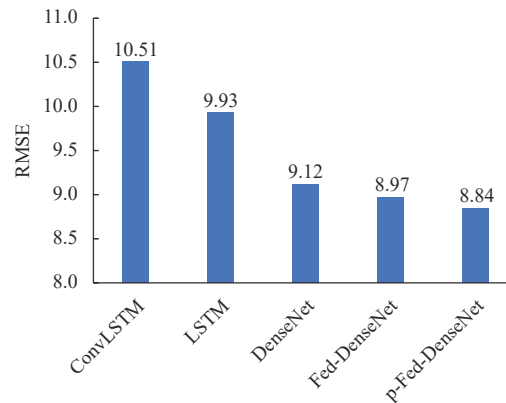


图 6 不同模型短消息业务流量误差对比

4.6 真实值与预测值对比

图 7 展现了随机挑选的一个区域在不同算法的绝对预测误差的累积分布函数图。本文的个性化联邦学习模型在短消息流量数据集上的预测性能最好, 预测误差较小。

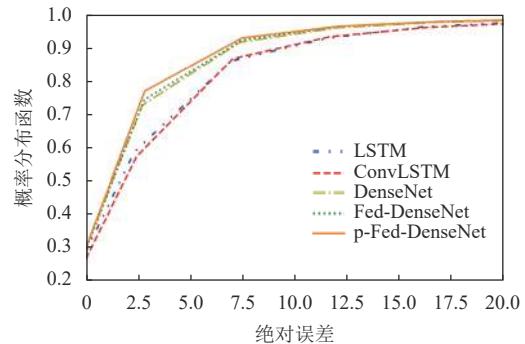


图 7 短消息业务流量真实值和预测值对比

5 结束语

本文提出了基于联邦学习的城市全域流量预测模型 Fed-DenseNet 和基于合作博弈的个性化联邦流量预测模型 p-Fed-DenseNet。前者将各个边缘服务器端时空流量预测的结果上传至中心服务器, 在中心服务器的协同下训练出能够反映各区域流量共性特征的全局模型, 解决了“城市全域尺度”的通信流量预测复杂性、时效性问题。后者在前者的全局模型上, 引入符合条件的本地区域性数据, 在全局模型泛化性的基础上保留各区域模型差异性, 解决了“城市全域尺度”的通信流量预测准确性问题。

参考文献

- [1] 张杰, 白光伟, 沙鑫磊, 等. 基于时空特征的移动网络流量

- 预测模型[J]. *计算机科学*, 2019, 46(12): 108-113.
- ZHANG J, BAI G W, SHA X L, et al. Mobile traffic forecasting model based on spatio-temporal features[J]. *Computer Science*, 2019, 46(12): 108-113.
- [2] WANG K, MA C, QIAN Y, et al. A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction[J]. *Physica A: Statistical Mechanics and Its Applications*, 2021, 583(12): DOI: [10.1016/j.physa.2021.126293](https://doi.org/10.1016/j.physa.2021.126293).
- [3] 张家晨, 左兴权, 黄海, 等. Prophet 混合模型应用于基站网络流量长期预测[J]. *计算机工程与应用*, 2022, 58(14): 80-88.
- ZHANG J C, ZUO X Q, HUANG H, et al. Application of prophet mixture model on long-term prediction of base station cell network traffic[J]. *Computer Engineering and Applications*, 2022, 58(14): 80-88.
- [4] 朱江, 潘成胜, 孔志翔, 等. 基于 Prophet 融合 MGF 提取的网络流量预测[J]. *计算机仿真*, 2021, 38(7): 475-480.
- ZHU J, PAN C S, KONG Z X, et al. Network traffic prediction based on prophet fusion multi-grain feature extraction[J]. *Computer Simulation*, 2021, 38(7): 475-480.
- [5] KAIROUZ E, MCMAHAN H B. Advances and open problems in federated learning[J]. *Foundations and Trends in Machine Learning*, 2021, 14(1): 1-210.
- [6] BAO W, WU C, GULENG S, et al. Edge computing-based joint client selection and networking scheme for federated learning in vehicular IoT[J]. *China Communications*, 2021, 18(6): 39-52.
- [7] KRABS W. A cooperative treatment of an n-person cost-goal-game[J]. *Mathematical Methods of Operations Research*, 2003, 57(2): 309-319.
- [8] SHIROKIKH V A, LEZHNINA E A. Coalition-Formation problem for cooperative inventory routing game[J]. *Automation and Remote Control*, 2019, 80(7): 1358-1367.
- [9] ZHENG Z, CHEN W, WU X, et al. LSTM network: A deep learning approach for short-term traffic forecast[J]. *IET Intelligent Transport Systems*, 2017, 11(2): 68-75.
- [10] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. [S.l.]: MIT Press, 2015: 802-810.
- [11] ZHANG C T, ZHANG H X, YUAN D F, et al. Citywide cellular traffic prediction based on densely connected convolutional neural networks[J]. *IEEE Communications Letters*, 2018, 22(8): 1656-1659.

编辑 税红