软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(2):715−731 [doi: 10.13328/j.cnki.jos.007153] [CSTR: 32375.14.jos.007153]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于注意力机制的联邦无线流量预测模型*

柴宝宝[1], 董安明[2], 王桂娟[2], 韩玉冰[2], 李 浩[1], 禹继国[3]

[1](山东科技大学 计算机科学与工程学院, 山东 青岛 266590)

[2](齐鲁工业大学 (山东省科学院), 山东 济南 250353)

[3](中国石油大学 (华东) 计算机科学与技术学院, 山东 青岛 266580)

通信作者: 禹继国, E-mail: jiguoyu@sina.com

摘 要: 移动数据每天都在不断增长, 如何精准预测无线流量对高效、合理的配置通信和网络资源至关重要. 现有的流量预测方法多采用集中式训练架构, 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. 联邦学习可以在数据本地存储的前提下训练一个全局模型, 保护用户隐私, 有效减轻数据频繁传输负担. 但是在无线流量预测中, 单个基站数据量有限, 且不同基站流量数据模式异构, 流量模式难以捕捉, 导致训练得到的全局模型泛化能力较差. 此外, 传统联邦学习方法在进行模型聚合时采用简单平均, 忽略了客体贡献差异, 进一步导致全局模型性能下降. 针对上述问题, 提出一种基于注意力的“类内平均, 类间注意力”联邦无线流量预测模型, 该模型根据基站的流量数据进行聚类, 更好地捕捉具有相似流量模式基站的流量变化特性; 同时, 设计一个预热模型, 利用少量基站数据缓解数据异构, 提高全局模型的泛化能力; 在模型聚合阶段引入注意力机制, 量化不同客体对全局模型的贡献, 并在模型迭代过程中融入预热模型, 大幅提升模型的预测精度. 在两个真实数据集 (Milano 和 Trento) 上进行大量实验, 结果表明该方法优于所有基线方法. 并且与目前最先进的方法相比, 在两个数据集上的平均绝对误差性能增益最高分别达到 10.1% 和 9.6%.

关键词: 无线流量预测; 联邦学习; 聚类; 注意力机制

中图法分类号: TP393

中文引用格式: 柴宝宝, 董安明, 王桂娟, 韩玉冰, 李浩, 禹继国. 基于注意力机制的联邦无线流量预测模型. 软件学报, 2025, 36(2): 715–731. http://www.jos.org.cn/1000-9825/7153.htm

英文引用格式: Chai BB, Dong AM, Wang GJ, Han YB, Li H, Yu JG. Federated Wireless Traffic Prediction Model Based on Attention Mechanism. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 715–731 (in Chinese). http://www.jos.org.cn/1000-9825/7153.htm

Federated Wireless Traffic Prediction Model Based on Attention Mechanism

CHAI Bao-Bao[1], DONG An-Ming[2], WANG Gui-Juan[2], HAN Yu-Bing[2], LI Hao[1], YU Ji-Guo[3]

[1](College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

[2](Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China)

[3](College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: As mobile data is growing everyday, how to predicate the wireless traffic accurately is crucial for the efficient and sensible allocation of communication and network resources. However, most existing prediction methods use a centralized training architecture, which involves large-scale traffic data transmission, leading to security issues such as user privacy leakage. Federated learning can train a global model with local data storage, which protects users' privacy and effectively reduces the burden of frequent data transmission. However, in wireless traffic prediction, the amount of data from the single base station is limited, and the traffic patterns vary among

* 基金项目: 国家自然科学基金 (62272256, 62202250, 61832012); 山东省自然科学基金基础研究重大基础研究项目 (ZR2022ZD03); 山东省自然科学基金 (ZR2021QF079, ZR2023MF040); 齐鲁工业大学 (山东省科学院) 科教融合创新试点项目 (2022XD001); 齐鲁工业大学 (山东省科学院) 培优基金 (2023PY059); 济南“新高校 20 条”资助项目 (20228093)
收稿时间: 2023-08-09; 修改时间: 2023-10-12; 采用时间: 2023-12-21; jos 在线出版时间: 2024-07-17
CNKI 网络首发时间: 2024-07-21

different base stations, making it difficult to capture the traffic patterns and resulting in poor generalization of the global model. In addition, traditional federated learning methods employ averaging in model aggregation, ignoring the differences in guest contributions, which further leads to the degradation of the global model performance. To address the above issues, this study proposes an attention-based "intra-cluster average, inter-cluster attention" federated wireless traffic prediction model. The model first clusters base stations based on their traffic data to better capture the traffic variation characteristics of base stations with similar traffic patterns. At the same time, a warm-up model is designed to alleviate data heterogeneity by a small amount of base station data to improve the generalization ability of the global model. The study introduces the attention mechanism in the aggregation stage to quantify the contributions of different objects to the global model and incorporates the warm-up model in the model iteration process to improve the prediction accuracy of the model. Extensive experiments are conducted on two real-world datasets (Milano and Trento), and the results show that the DualICA outperforms all baseline methods. The mean absolute error performance gain over the state-of-the-art method is up to 10.1% and 9.6% on the two datasets, respectively.

Key words: wireless traffic prediction; federated learning; clustering; attention mechanism

移动设备的普遍使用直接导致了移动数据流量的急剧增长. 蜂窝数据网络是承载移动互联网数据的主要组成部分, 这为用户友好的移动设备 (如手机、上网本和平板设备) 以及大量的移动应用程序提供了动力. 自 2019 年第 5 代移动通信 (5G) 网络商业化以来, 关于第 6 代移动通信 (6G) 的潜在功能和使能技术的初步研究已经引起了学术界和工业界的广泛关注[1,2]. 无线流量预测可以预估未来的流量数据量, 为通信网络管理和优化提供决策依据[3,4]. 根据预测的流量数据, 通信运营商能够提前采取积极的措施来缓解突发传输造成的网络拥堵和中断. 换言之, 精确预测未来的流量负载, 有助于根据实际流量需求动态配置通信网络资源, 提高通信网络能源利用效率, 在设计绿色流量感知蜂窝网络中发挥重要作用[5,6]. 此外, 通过无线流量预测可以使 6G 通信网络中普遍存在的异质性服务需求得到很好满足[7].

目前大多数的无线流量预测方法都集中在中心化学习策略上, 普遍涉及将大量的原始数据传输到数据中心来学习一个通用的预测模型. 但是, 训练数据的频繁传输和信号开销很容易耗尽网络容量, 这将对有效数据的传输产生负面影响. 不仅如此, 随着隐私保护日渐受到人们的重视, 数据隐私和安全成为世界性趋势, 从多个数据拥有者 (如医院) 手中收集数据来训练预测模型的方式变得不可行, 各组织所掌握的数据被迫独立存储, 形成了数据孤岛. 因此, 对能够应对上述挑战的新的无线流量预测方法的需求变得日渐迫切.

联邦学习 (federated learning, FL)[8−12]在保证用户数据本地保存的同时训练一个全局模型, 它的出现和成功使无线流量预测问题成为可能. 在联邦学习设置中, 众多客户端, 如移动设备、基站 (base station, BS) 或企业等, 在中央服务器的协调下协作训练一个预测模型. 只有通过本地训练获得的中间梯度或模型参数被发送到中央服务器, 而不需要发送原始数据, 从而保证了用户隐私[13,14]. 此外, 联邦学习促进了前所未有的大规模灵活的数据收集和模型训练, 边缘客户端可以在白天主动收集数据, 然后在夜间联合起来更新全局模型, 以提高第 2 天使用预测模型的效率和准确性.

尽管有很好的应用前景, 但基于联邦学习的精准无线流量预测仍然存在问题. 首先, 用户的移动性导致无线流量之间复杂的时空耦合, 流量变化特点很难被捕捉和建模. 不仅如此, 不同的基站流量模式不同, 流量数据高度异构, 由此训练得到的全局模型泛化能力较差. 此外, 联邦学习在进行模型聚合时采用简单的求平均方法, 忽略了不同客体的贡献差异, 在此基础上训练得到的模型想要达到精准预测是非常困难的. 针对上述问题, 本文提出了一个基于注意力机制的"类内平均, 类间注意力"联邦无线流量预测模型 DualICA (intra-cluster average and inter-cluster attention), 以实现精准的无线流量预测. 主要贡献如下.

• 提出了一种流量模式聚类方法, 根据基站的流量数据对基站进行聚类, 把流量模式相似的基站进行归类, 并在联邦模型聚合时采用类内平均的办法来获取类本地模型, 更好的捕捉了具有不同流量变化基站的流量模式特点.

• 设计了一个"预热"模型, 它由参与训练的基站的少部分任意数据训练得到, 主要用来克服因基站之间流量数据高度异构导致的全局模型泛化能力较差的问题. 此外, 使用预热模型取代系统随机初始化操作, 加快模型收敛速度.

• 在模型聚合阶段引入注意力机制, 量化了不同客体的贡献, 根据客体贡献为其分配相应权重, 解决了简单平均造成的不同客体贡献差异被忽略的问题. 同时, 在全局模型的训练过程中融入预热模型, 大幅提高了全局模型的预测准确率.

• 在两个现实世界数据集上对提出的模型进行了实验验证, 实验结果表明与现有的无线流量预测方法相比, 本文所提出的 DualICA 具有更高的预测准确率.

本文第 1 节介绍无线流量预测和联邦学习的相关方法和研究现状. 第 2 节给出问题定义. 第 3 节介绍本文提出的联邦预测模型. 第 4 节介绍实验设置和实验结果, 并通过实验分析阐述所提模型的有效性. 最后, 第 5 节进行全文总结和展望.

1 相关工作

1.1 无线流量预测

由于无线通信中的许多任务都需要精确的流量建模和预测能力, 蜂窝网络中流量预测问题[4]受到了广泛的关注. 现存的流量预测方法众多, 最初人们主要利用统计学和概率的相关理论对无线流量进行建模和预测, 经典方法是自回归综合移动平均法 (ARIMA)[15], 其变体在文献 [16] 中进行了探讨. 还有其他一些基于固有模型演变而来的预测方法, 如 FARIMA 模型[17]、移动性模型[18]、网络流量模型[19]和 α-稳定模型[20]等, 再通过一些适当的方法来探索流量特性. 除了固有模型外, 还有通过现代信号处理技术 (如主成分分析法[21,22], 卡尔曼排序法[23]或压缩感知法[24]) 来捕捉流量模式的演变. 此外, 熵理论[25]和协方差函数[26]也被探索用来进行无线流量预测.

上述方法大多是线性统计方法, 然而, 线性模型在许多实际应用中并不适用. 随着机器学习和人工智能技术的持续发展, 它们在解决基于神经网络的无线流量预测问题方面与上述方案相比有着较为明显的优势[27,28]. 起初, 人们利用如线性回归[29]和支持向量回归 (SVR)[30,31]这些比较浅层的方法进行流量预测. 之后, 深度学习[32]开始快速发展和并得到广泛应用, 如何利用强大的深度学习技术对蜂窝网络进行准确的流量预测[33,34]成为研究热点. 在文献 [35] 中, 作者在自动编码器和长短时记忆网络 (long short-term memory, LSTM) 的基础上设计了一个混合深度学习框架, 以同时捕捉不同小区之间的空间和时间依赖性. 为了对多个小区进行预测, 研究人员还通过使用 LSTM 引入了一个多任务学习框架, 即不使用所有邻近的流量信息, 而是选择与目标 BS 有最高相关系数的最相关的邻居来提供时空信息[36]. Zhang 等人[37]和 Wu 等人[38]在使用深度神经网络架构捕捉时空关系的同时引入了迁移学习来进行跨地域的知识迁移和流量预测. 但深度学习方法缺乏对全局时空相关性的考虑, 为了解决这个问题, Lin 等人[39]提出了一个新的多变量时空预测模型, 使用扩展的图注意力网络来探索蜂窝流量之间的相关性, 并利用注意力机制来提高捕捉空间依赖性的效率. 与之相似, Yao 等人[40]提出了一个新的多视图时空图网络 (multi-view spatial-temporal graph network, MVSTGN), 它将注意力和卷积机制结合到流量模式分析中, 实现了对时空特征的全面挖掘.

上述所有的工作主要集中在集中式的无线流量预测问题上, 难以避免数据传输负担和用户隐私泄露等安全问题. 本文提出的预测模型与上述方案不同, 我们试图通过联邦学习来解决无线流量预测问题.

1.2 联邦学习

联邦学习提供了一个分布式训练架构, 可以通过聚合本地客户端的模型来获得全局模型[41]. 为了获得全局模型, Google 首次提出了一种称为 FedAvg[42]的聚合方法. 研究表明, 当客户端数据服从独立同分布 (independent homogeneous distribution, IID) 时, FedAvg 与集中式学习相比取得了类似的性能. 然而, 当客户端数据为非独立同分布 (non-independent homogeneous distribution, Non-IID) 时, FedAvg 的性能会大大降低. 为了解决这个问题, Zhao 等人[43]提出了一种数据共享策略, 即创建一个小的数据子集, 在所有的客户端设备中全局共享, 这种策略可以解决联邦学习所面临的统计异构性挑战. 在文献 [44] 中, 作者提出了 FedProx, 它可以被视为 FedAvg 的概括和重新参数化, 以解决联合网络中的异质性问题. 此外, 考虑不同客户端对全局模型的贡献存在差异, Ji 等人[45]引入了一个全面的联合聚合方案, 称为 FedAtt. 该方案提高了全局模型的泛化能力, 并成功地用于解决自然语言建模问

题. 同样针对数据异构问题, Shu 等人[46]根据节点模型相似性将节点进行分组, 提高了模型对不同特征的提取能力. 相似的, Yang 等人[47]提出了一个联邦元学习框架, 能够自适应的对节点进行分组, 并通过元学习获得个性化模型. 但是上述方法并不适用于时序数据的预测, 因此无法应用到无线流量预测中. 在文献 [48] 中, Arisdakessian 等人通过选择高信任度的节点联盟, 并定期与其联盟成员共享小部分数据的方式, 降低了整个联盟中数据的异构性. Sun 等人[49]提出了一个新的权重调整机制 AFedSV, 将根据节点模型性能计算得到的沙普利值作为该节点本地参数在模型聚合时的比重, 提升高贡献参与方对全局模型的影响, 缓解了数据异构带来的模型性能下降问题. 但是, AFedSV 在每个训练轮次都需要花费大量时间进行沙普利值的计算, 降低了模型效率. 在文献 [50] 中, Mai 等人提出了一种新的模型学习和融合方法 SCCD (server-client collaborative distillation), 在服务器和节点端都进行模型蒸馏, 提炼训练模型, 减少训练方差, 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.

在联邦无线流量预测[47]研究上, 也存在相似的工作. Zhang 等人[51]提出了一个新的基于双重注意力的联邦学习无线流量预测框架 FedDA, 作者使用迭代聚类的方法同时捕捉基站流量变化的时间和空间依赖关系. 同时, FedDA 设计了一个准全局模型并作为先验知识在不同客户端之间共享, 解决了联邦学习所面临的数据异构问题. 此外, 为了构建全局模型, 作者进一步提出了一个双重注意力方案,根据不同类模型的贡献为其分配权重, 而不是简单地平均本地模型的权重. 但是, 作者提出的方案中全局模型对先验知识的依赖程度较高, 这在实际情况中是不允许的. 此外, 方案中经过迭代聚类后所得全局模型对基站的选取较为敏感. Zheng 等人[52]提出了针对无限流量预测的安全威胁模型, 并给出了对应的安全攻击应对策略. Zhang 等人[53]提出了一种高效的联邦元学习方法 MAML, 同时设计了一个基于距离的加权模型来进行联邦模型聚合, 用来捕捉不同区域之间时空依赖关系.

上述工作针对无限流量预测问题分别给出了不同的解决办法, 但是它们并没有考虑到数据以及流量模式异构问题对模型性能的影响, 因此所得到的模型预测准确度不高. 而数据和流量模式异构问题在无线流量预测场景中是普遍存在的, 解决这两个问题对于实现更加精准的流量预测来说是非常必要的.

2 问题定义

2.1 无线流量预测

在蜂窝网络中, 由 BS 负责接收并转发用户产生的无线流量, 假设共有 M 个 BS, 每个 BS 都有自己的本地无线流量数据, 表示为 $S^m=\{S_1^m,S_2^m,\ldots,S_N^m\}$, 其中 S_n^m 表示第 m 个基站在第 n 个时间间隔内的流量数据. 无线流量预测任务主要是根据已有的流量数据预测未来的流量. 假设 S_n^m 是需要预测的目标流量, 那么无线流量预测问题可以描述为:

$$S_n^m=f(S_{n-1}^m,S_{n-2}^m,\ldots,S_1^m;w) \tag{1}$$

其中, f 表示选用的预测模型, w 表示预测模型对应的参数, 预测模型可以任意选择.

在无线流量预测问题中, 通常使用已知的历史数据作为输入来降低预测复杂度, 并据此输入信息得到预测值作为模型的输出. 因此, 在上面 S^m 的基础上, 采用滑动窗口的方法来获取输入输出对 $\{x_i^m,y_i^m\}_{i=1}^n$. x_i^m 表示第 m 个基站在第 i 个时间间隔的输入数据, x_i^m 表示为:

$$x_i^m=\{S_{N-1}^m,\ldots,S_{N-p}^m,S_{N-\lambda 1}^m,\ldots,S_{N-\lambda q}^m\} \tag{2}$$

与得到的预测数据 y_i^m 对应. 其中, N 表示当前所在的时间间隔编号, p 和 q 表示滑动窗口的大小, 分别用于捕捉无线流量数据的近似性依赖和周期性依赖, λ 用来调节周期性长度. 本文中, 我们只研究下一个时刻的流量预测问题, 故而公式 (1) 所表示的预测问题可以重新定义如下:

$$\hat{y}_i^m=f(x_i^m;w) \tag{3}$$

为了使得所有基站的预测误差都最小, 我们可以通过求解下面的公式来获取参数 w:

$$\arg\min_w\left\{\frac{1}{Mn}\sum_{m=1}^{M}\sum_{i=1}^{n}\Gamma(f(x_i^m;w),y_i^m)\right\} \tag{4}$$

其中, Γ 表示损失函数, 通常使用 $\left|f(x_i^m;w)-y_i^m\right|$ 或者 $\left|f(x_i^m;w)-y_i^m\right|^2$.

2.2 联邦学习基础模型

假设流量数据存储在不同地理位置的基站中, 在联邦学习设置下, 我们的目标是解决上面公式 (4) 中的最优参数计算问题. 首先, 系统通过初始化操作得到全局模型初始参数 w, 之后基站与中心服务器节点之间进行模型参数的交互和迭代更新. 其中, 中心服务器节点可以是任何一个受信任的拥有较大算力的节点. 第 t 次迭代的交互过程如下.

1) 中心服务器节点将参数 w^t 下发给所有基站.

2) 基站收到全局参数 w^t 后, 在此参数基础上根据本地数据进行训练和参数更新. 本地参数更新规则如下 $w_m^{t+1} \leftarrow w_m^t - \eta \nabla w^t \Gamma(f(x^m; w^t), y^m)$, 其中 η 表示学习率, ∇w^t 表示梯度损失.

3) 基站将训练得到的本地模型参数 w_m^{t+1} 发回给中心服务器节点.

4) 中心服务器节点在收到所有参与训练的基站所返回的本地模型参数后, 执行聚合操作. 经典的聚合方式是联邦平均 FedAvg[42], 聚合规则为 $w^{t+1} \leftarrow \frac{1}{M}\sum_{m=1}^{M} w_m^{t+1}$.

3 联邦无线流量预测模型

本节详细介绍所提出的联邦无线流量预测模型 DualICA (intra-cluster average and inter-cluster attention), 图 1 是对所提预测模型 DualICA 的整体框架展示. 该预测模型主要分 3 部分, 第 1 部分: 根据基站本地数据对基站进行聚类得到多个类. 第 2 部分: 由参与训练的基站共享部分数据得到“预热”模型以缓解数据异构. 同时, 利用注意力机制量化不同基站贡献, 以提高模型泛化能力. 第 3 部分: 将“预热”模型和注意力机制相融合, 并经过重复训练, 最终得到流量预测模型.

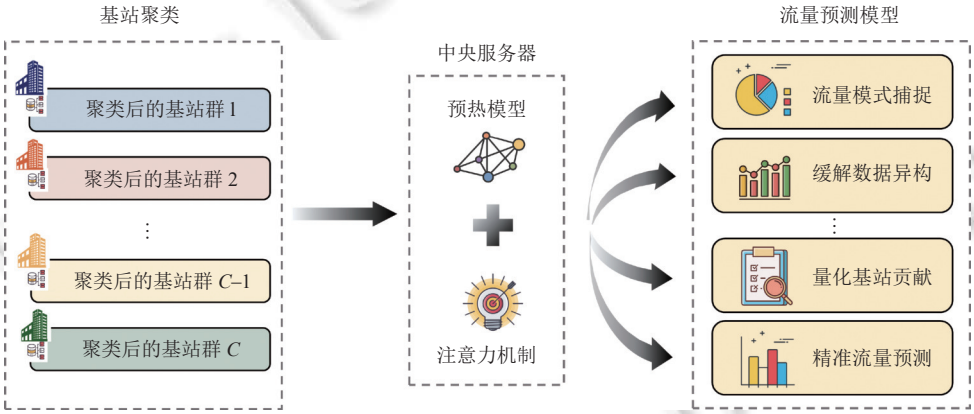
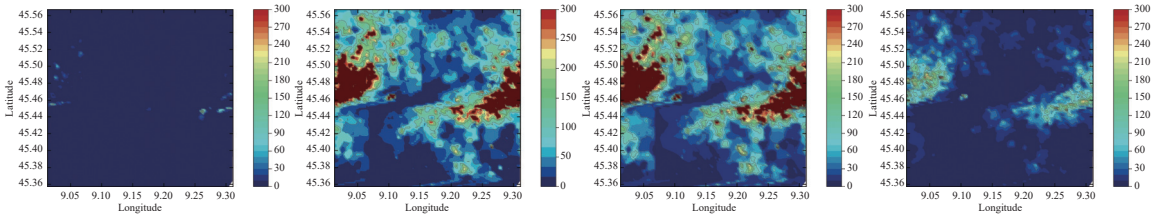


图 1 联邦注意力预测模型框架

在接下来的章节中, 我们针对框架图中的各部分内容给出了详细的解释. 首先, 对流量数据分布进行了分析, 并介绍了可以捕捉不同 BS 流量模式的模式聚类策略. 接着, 说明了所提出的预热模型的由来和组成结构. 最后详细阐述了所提出的基于注意力的类内、类间联邦模型聚合方案.

3.1 流量模式聚类

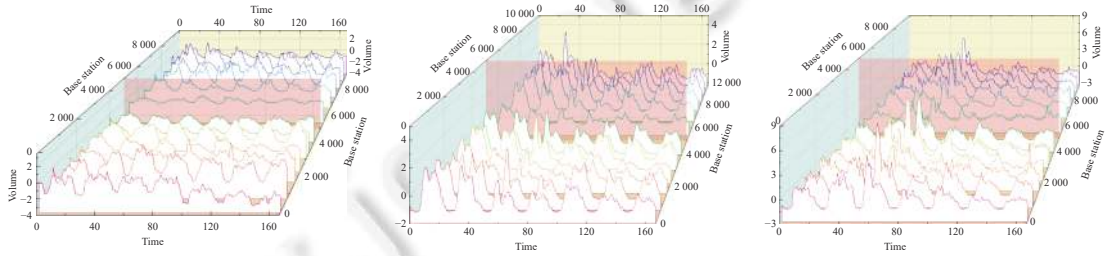
一个地区的基站数量众多, 不同基站在不同时刻的流量模式不尽相同. 如图 2, 以数据集中收集到的米兰地区同一天中不同时刻的 10 000 个基站收集到的通话流量变化为例. 从图 2 中可以看出, 同一基站在不同时刻的流量数据存在差异, 流量的增减主要跟随人们的生活习惯变化, 与时间高度相关. 例如在 4:00 和 22:00 时, 大多数人正处于休息时间, 所以流量数据较少 (图 2 中蓝色区域). 而在 10:00 和 16:00 处于上班时间, 通话数据则比较活跃 (图 2 中红色区域). 不同基站之间流量分布差异更加明显, 同时刻分布图中可以看出基站的流量模式是高度异构的. 即使在 10:00 和 16:00 的上班时间, 仍然存在基站通话数据极少的部分. 换句话说, 基站间流量数据变化特点是不同的, 流量模式也必然存在差异.



(a) 4:00 (b) 10:00 (c) 16:00 (d) 22:00

图 2 不同时刻无线数据流量分布

为了更加仔细地观察不同基站的流量数据变化特点, 我们从 Milan 数据集中随机选取 13 个基站 (基站编号分别为 47, 728, 1647, 2226, 3026, 3288, 4032, 5409, 6537, 7440, 8241, 8935, 9535), 统计并对比了这 13 个基站一周内收集到的流量数据, 它们的变化情况如图 3 所示 (按照基站编号从小到大顺序排列). 为避免采集到的数据过于稀疏, 基站数据每 1 h 采样一次, 共计 7 天的流量数据, 其中包括图 3(a) 上网流量; 图 3(b) 通话流量和图 3(c) 短信流量.



(a) Internet (b) Call (c) SMS

图 3 不同基站无线流量数据变化

在图 3(a) 中不难看出, 不同基站的流量变化不同, 主要表现在流量峰值的大小以及出现的时间不同. 例如基站 47 和基站 1647, 它们所表现出的流量变化特点是不同的. 然而, 在图 3(a) 中, 基站 728 与基站 8241 的流量变化模式则非常相似. 不仅如此, 图 3(b) 中的基站 47, 基站 728, 基站 3026, 基站 7440 虽然编号相差很大, 但是它们的流量变化特点基本一致. 图 3(c) 中基站 47 和基站 728, 基站 3026 等也都存在相似的流量变化特点. 综上分析可知, 不同基站收集到的流量数据受时间、位置等因素影响存在着异构特性, 但它们之间也同样存在一定的相似性. 充分利用不同基站间流量数据变化的相似及异构特性能够使得模型预测更加准确.

为了更高效地捕捉 BS 之间的流量模式变化特性, 训练出一个适合所有 BS 的预测模型, 我们利用聚类的思想进行模型的获取, 如图 4 所示. 根据基站的不同流量模式对 BS 进行聚类操作, 将流量模式相似的基站归为同一类, 充分挖掘相似基站流量模式变化特点. 同时, 将不同流量模式归类处理减少了系统训练量, 降低了全局模型的训练复杂度. 具体来说, 所有 BS 都有它自己收集到的无线流量数据, 如公式 (5). 其中, S 表示所有基站流量数据, S^i 表示第 i 个基站收集到的流量数据. 根据这些数据利用 K-means 算法对基站进行聚类.

$$S=\{S^1,S^2,\ldots,S^m\} \tag{5}$$

$$\text{K-means}\{S^1,S^2,\ldots,S^m\}=\{C_i\}_{i=1}^{C} \tag{6}$$

为了保证算法结果的普适性, 我们使用随机初始化的方式得到 C 个聚类中心, 之后根据基站数据进行聚类, 聚类操作完成后得到新的 C 个聚类中心, 也就是 C 个不同的类. 每个基站都得到一个类标签标识所属类, 聚类的数目可以根据实际情况调整. 经过上述步骤, 流量模式相似的基站被归为同一类, 在每一轮训练开始后, 所有基站首先使用自身流量数据训练得到本地模型参数 $w_{c,i}^{t+1}$, 其中 c 代表基站所属类别, i 代表基站编号, t+1 表示当前训练轮次. 然后根据基站所属类别将这些本地模型进行分类模型聚合. 由于同一类中基站的流量模式是相似的, 故而在类内模型聚合时我们选择使用 FedAvg 方法得到类本地模型.

$$\{w_{L,c}^{t+1}\}_{c=1}^{C}=\frac{1}{n}\sum_{i=1}^{n}w_{c,i}^{t+1} \tag{7}$$

每个类别都会生成一个类本地模型, 这里所说的类本地模型仍然是通过联邦学习的训练方式得到. 接着, 为了捕捉不同类基站间流量模式变化特性, 利用注意力机制对所有类本地模型进行模型聚合, 注意力机制会根据不同类本地模型对全局模型的贡献为其分配权重, 充分考虑了不同类本地模型的贡献差异. 由于注意力机制考虑了类本地模型贡献差异, 因此得到的全局预测模型预测更加准确, 详细的模型聚合在第 3.2 节中介绍.

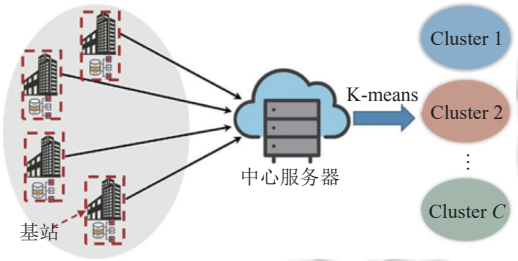


图 4 流量模式聚类

3.2 预热模型

从对图 2 和图 3 的分析中我们发现, 同一基站在不同时刻统计到的流量数据是有差异的, 流量的增减主要跟随人们的生活习惯变化, 与时间高度相关, 同时刻分布图中可以看出基站的流量模式是高度异构的. 同时, 即使不同基站流量变化也可能存在相似性. 因此, 如何在利用相似性的同时降低数据异构的影响成为主要问题. 尤其在联邦学习设置下, 流量数据由不同位置分布的基站自身保存, 中央服务器对这些流量数据没有控制权, 无法避免基站间数据异构对模型的影响, 因此经过训练得到的模型泛化能力普遍较差. 为了解决数据异构导致的模型性能较差的问题, 我们设计了一个预热模型, 如图 5 所示. 在预热模型中, 每个 BS 随机抽取其收集到的流量数据的 α (百分比) 部分发送给中央服务器, 然后中央服务器将获取到的各基站的流量数据集中训练得到预热模型. 如公式 (8)、公式 (9), 其中 w_p^t 表示第 t 次迭代得到的预热模型参数, Φ 代表训练模型, γ 代表学习率. 经过多轮训练和更新, 最终得到预热模型参数 W_p.

$$w_p^t=\Phi(\alpha S) \tag{8}$$

$$w_p^{t+1}=w_p^t+\gamma\Phi(\alpha S,w_p^t) \tag{9}$$

图 5 预热模型

虽然预热模型只包含了各个基站少部分数据, 但是由于这些数据来自不同基站, 包含了各个基站数据变化特性, 因此在一定程度上缓解了数据异构.

此外, 在传统机器学习中, 模型训练开始前都以随机初始化的形式获得初始参数, 然后经过不断地训练迭代对此参数进行优化, 最终得到预测模型. 然而, 随机初始化的方法随机性较高, 不利于模型快速收敛. 因此, 我们将上

述预热模型参数作为初始化参数, 替代了随机初始化操作, 加速模型收敛. 详细来说, 在得到预热模型后, 中央服务器将预热模型参数 W_p 作为系统初始参数下发给所有 BS, BS 在参数 W_p 基础上训练自身数据. 在一轮训练完成后, 基站对模型参数进行更新, 并将更新后的模型参数返回给中央服务器, 由中央服务器对所有基站上传的模型参数进行聚合. 换言之, 这里的预热模型类似于先验知识, 通过利用先验知识的办法来加快模型的收敛速度. 同时, 为了充分发挥预热模型作用, 我们将预热模型融入全局模型中, 使得预热模型在模型聚合时也发挥着一定作用, 进一步提高了全局模型的预测准确率.

3.3 类内、类间联邦模型聚合

模型聚合是联邦学习中最基础也是最重要的部分, 它主要涉及将来自各个用户的本地模型进行聚合以构建最终的全局模型. 本文提出了一种新的联邦聚合策略, 在模型聚合阶段引入了注意力机制, 同时融入了预热模型, 利用注意力机制对类本地模型和预热模型的贡献进行量化, 根据量化之后的值为其分配权重以获取全局模型.

具体而言, 我们的联邦聚合策略共分为 3 步, 如图 6 所示. 第 1 步对基站进行聚类操作, 根据基站流量模式对基站进行聚类, 以捕捉相似基站的流量变化特点. 第 2 步进行预热模型的训练, 利用预热模型缓解数据异构对全局模型的影响. 同时, 使用预热模型取代系统随机初始化操作, 加快模型收敛. 联邦模型聚合发生在第 3 步, 主要包括类内模型聚合和类间模型聚合. 在进行类内模型聚合时, 由于同一个类内基站的流量模式相似, 我们采用经典的 FedAvg 算法, 使用求平均值的方式获取类本地模型. 在进行类间模型聚合时我们利用注意力机制, 量化不同类本地模型贡献差异, 为不同的类本地模型分配不同的权重. 此外, 在全局模型中还加入了预热模型, 以缓解数据异构问题. 最后, 经过多次训练迭代得到包含所有类流量变化特性的全局模型.

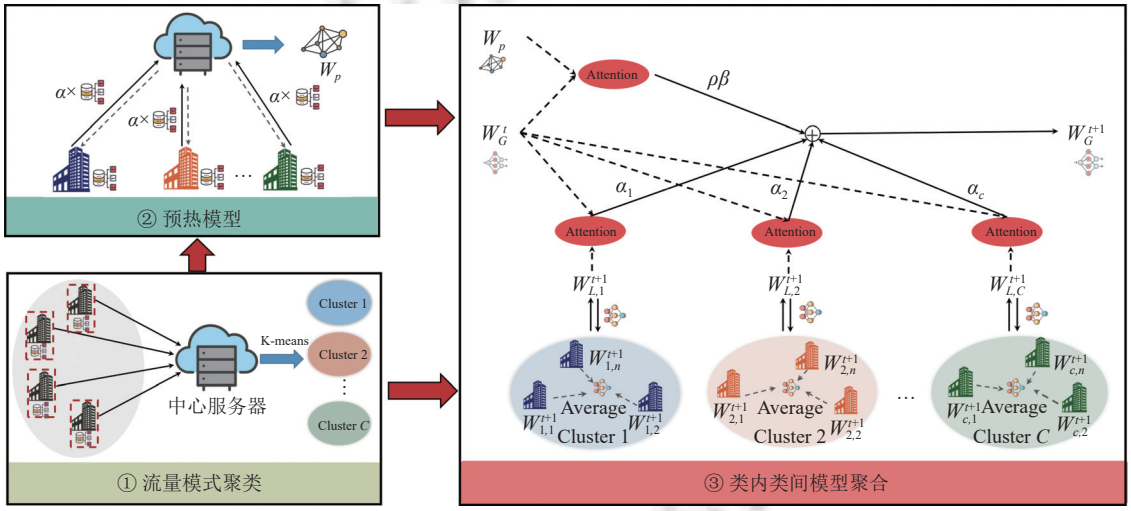


图 6 联邦注意力预测模型

需要注意的是, 在联邦模型训练中, 并不是所有节点都同时参与模型训练. 而是在选定一定数量的节点后, 每一轮随机选取基站进行训练. 考虑到联邦训练特点, 为了避免在聚类过程中出现本身相似度不高, 但是由于选取的基站数量限制而被聚为同一类, 导致类本地模型无法反映类内基站流量变化特性的问题, 首先对所有基站进行聚类操作, 再从聚类后的基站中随机选取一定数量的基站参与联邦模型的训练. 其中类标签由第 1 步根据基站流量数据聚类获得, 聚类完成之后再从带有类标签的基站种随机选取基站参与后续训练过程, 并由这些选到的基站使用自身流量数据执行第 2 步操作产生预热模型.

如图 6 所示, 经过聚类操作后, 得到 C 个不同的类. 其中, $w_{c,n}^{t+1}$ 表示第 c 类中第 n 个基站的本地模型参数, $w_{L,c}^{t+1}$ 表示类别 c 生成的类本地模型参数. 共有 C 个不同的类, 分别对应 C 个类本地模型参数 $\{w_{L,c}^{t+1}\}_{c=1}^{C}$. w_G^{t+1} 表示第 t+1 轮训练聚合类本地模型后得到的全局模型参数. 在中央服务器端进行联邦聚合优化的目的是找到一个最佳的全局

模型, 该模型对所有 BS 的流量数据模式都有很强的泛化能力. 为了达到这个目的, 全局模型应该在捕捉 BS 的类内和类间流量模式变化之间找到一个最佳平衡点. 因此, 在提出的方案中, 将优化问题视为找到一个接近本地模型和预热模型的参数空间的全局模型, 同时还要考虑到它们在模型聚合过程中对全局模型的贡献差异. 因此, 优化目标是通过使用自适应分数作为权重, 使不同模型之间的总加权距离最小. 联合优化问题被正式定义为:

$$\arg\min_{w_G^t}\left\{\sum_{c=1}^{C}\frac{1}{2}\alpha_c\Gamma(w_G^t,w_{L,c}^{t+1})^2+\frac{1}{2}\rho\beta\Gamma(w_G^t,w_p)^2\right\} \tag{10}$$

其中, $\alpha_c=\{\alpha_c^i\}_{i=1}^J$ 和 $\beta=\{\beta^i\}_{i=1}^J$ 代表注意力权重向量, 分别表示第 c 个类本地模型和预热模型每一层参数贡献对应的权重, J 表示参数总层数; ρ 是一个与优化任务相关的正则化参数, 可以根据实验要求手动设置. 为了获得权重 α_c, 我们使用注意力机制, 并将其应用于模型各层的参数中. 对于第 c 个类本地模型, 第 j 层的参数被表示为 $w_{L,c}^{t+1,j}$. 基于层间参数, 第 t+1 次训练的类本地模型 $w_{L,c}^{t+1,j}$ 和上一轮迭代所得全局模型 $w_G^{t,j}$ 各层参数间的距离可以通过它们之差的 Frobenius 准则来计算, 即表示为:

$$d_{L,c}^{t+1,j}=\Gamma(w_{L,c}^{t+1,j},w_G^{t,j})=\left\|w_{L,c}^{t+1,j}-w_G^{t,j}\right\|_2^2 \tag{11}$$

随后, 使用 Softmax 函数对计算所得的非标准化的距离值 $d_{L,c}^{t+1,j}$ 进行映射, 将它们映射为 C 个类本地模型的概率分布. 通过这种方式, 可以确定不同类本模型的贡献. 标准的 Softmax 函数 $\sigma(\cdot)$ 表示为:

$$\alpha_c^{t+1,j}=\sigma(d_{L,c}^{t+1,j})=\frac{\mathrm{e}^{d_{L,c}^{t+1,j}}}{\sum_{c=1}^{C}\mathrm{e}^{d_{L,c}^{t+1,j}}} \tag{12}$$

类似地, 还可以得到 β 的值. 在得到 α_c 和 β 后, 全局模型的参数可以通过梯度下降算法进行更新. 首先计算公式 (5) 关于 w_G^t 的导数, 得到相应的梯度:

$$\nabla=\sum_{c=1}^{C}\alpha_c(w_G^t-w_{L,c}^{t+1})+\rho\beta(w_G^t-w_p) \tag{13}$$

利用得出的梯度, 输出模型参数可以通过以下方式更新:

$$w_G^{t+1}=w_G^t-\gamma\left(\sum_{c=1}^{C}\alpha_c(w_G^t-w_{L,c}^{t+1})+\rho\beta(w_G^t-w_p)\right) \tag{14}$$

其中, γ 是一个预先确定的步长, 控制在全局模型 w_G^t 在每次迭代中应向相反的梯度方向移动长度大小. 利用公式 (9) 对全局模型进行迭代更新, 得到最终的全局模型. 我们对预测模型 DualICA 的整个运行过程进行了简化, 如算法 1 所示. 同时, 为了提升算法的易读性, 我们在算法中进行了注释. 上述所有涉及的参数在第 4 节参数设置中我们都给出了详细的参数说明.

算法 1. DualICA.

输入: Traffic dataset $\{S^1, S^2, \dots, S^r\}$, cluster number C, proportion α and learning rate γ in pre-warm model, regularization parameter ρ, fraction of BSs δ, total number of iteration T;

输出: Global model W_G.

1. $\{C_i\}_{i=1}^{C}\leftarrow$ K-means$\{S^1,S^2,\dots,S^r\}$ /* Clustering BS to C clusters */
2. $w_{L,c_{c=1}}^{t+1^C}=\frac{1}{n}\sum_{i=1}^{n}w_{c,i}^{t+1}$
3. /* Intra-cluster average */
4. $S=\{S^1,S^2,\dots,S^m\}\leftarrow$ Random select m BSs to participant training process ($m<r$)
5. $w_p^{t+1}=w_p^t+\gamma\phi\left(\alpha S,w_p^t\right)$
6. /* Constructing pre-warm model */
7. **for** i=1 **to** T **do**

8. $K \leftarrow \max(m \cdot \delta, 1)$
9. /* Randomly select K BSs */
10. **for** j=1 **to** K BSs **do**
11. $w_G^{t+1} = w_G^t - \gamma\left(\sum_{c=1}^{C} \alpha_c \left(w_G^t - w_{L,c}^{t+1}\right) + \rho\beta\left(w_G^t - w_p\right)\right)$ /* Inter-cluster attention */
12. Obtain global traffic prediction model W_G

4 实验及分析

为了验证本文提出的联邦预测模型 DualICA 的性能, 我们在两个现实世界数据集上进行了一系列实验. 本节主要对数据集, 基线方法和参数设置进行说明, 并对实验结果进行对比和详细分析.

4.1 数据集介绍

本文使用的数据集来自意大利电信公司发起的大数据挑战[54], 主要包含意大利的米兰和特伦蒂诺两个地区的详细流量记录[55,56]. 在该数据集中, 上述两个地区被划分为大小为 $H \times W$ (长×宽) 的网格, 网格中每个方块被称为一个单元. 其中, 米兰地区被划分为 10 000 个单元, 而特伦蒂诺被划分为 6 575 个单元. 在每个单元中, 用户的电信活动由 BS 提供服务和记录, 因此我们用 BS 来表示 1 个单元. 数据集中包括 3 种类型的无线流量, 分别对应于短信、语音通话和互联网服务. 数据集记录了 2013 年 11 月 1 日–2014 年 1 月 1 日两个月内的流量数据, 每 10 min 记录一次流量信息. 在后续几个章节的实验中, 流量数据被重新采样为每小时记录 1 次, 避免单个时间间隔内采集到的数据量过少影响模型训练效果.

4.2 基准方法和评价指标

我们将提出的无线流量预测模型与下面 7 种现存的主流预测方法进行比较.

• SVR[31]: SVR 是经典的机器学习分类算法之一, 已经成功用于流量预测.

• LSTM[36]: LSTM 对时间序列数据集有很强的建模能力, 通常比线性模型和浅层机器学习模型有更好的预测性能.

• FedAvg[42]: FedAvg 是随着联邦学习这一概念被提出时产生的联邦平均模型, 它在进行全局模型聚合时采用了取本地模型平均值的方法来获取全局模型.

• FedAtt[45]: FedAtt 与 FedAvg 相似, 唯一不同点在于进行模型聚合时, FedAtt 引入了注意力机制, 它会根据不同本地模型的贡献差异来为其分配权重, 进而更新全局模型, 它考虑到了不同本地模型的变化差异, 有效提升了模型预测性能.

• FedDA[51]: FedDA 在 FedAtt 的基础上做出了进一步改进, 它加入了一个准全局模型来缓解数据异构. 此外, FedDA 设计了一个迭代聚类模型同时捕捉时间和空间依赖特性, 更加贴合无线流量预测场景.

• AD+MKrum[52]: AD+MKrum 针对无线流量预测场景下存在恶意的客户端, 导致训练阶段模型性能容易受到威胁的问题, 提出了针对无线流量预测的实用的威胁模型, 并设计了相应的扰动策略来应对模型攻击.

• MAML[53]: MAML 是一种高效的联邦元学习方法, 它使用从不同节点收集的数据来学习全局模型. 同时, 设计了一个基于距离的加权模型来进行联邦模型聚合, 用来捕捉不同区域之间时空依赖关系.

本文采用平均绝对误差 (mean absolute error, *MAE*) 和平均均方误差 (mean square error, *MSE*) 作为评价指标, 计算公式如下:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|,$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

4.3 参数设置

所有实验均在 Intel (R) Core (TM) i5-4200H CPU @ 2.80 GHz 2.79 GHz, 内存 8 GB 的笔记本上进行, 实验涉及

代码使用 PyTorch 编写. 在不失一般性的情况下, 每次训练开始时, 我们从聚类后的基站集合中随机选择 100 个基站参与训练, 并对这些基站收集到的 3 种无线流量数据进行实验. 其中, 将数据集中前 7 周的流量作为训练集来训练预测模型, 最后 1 周的流量作为测试集用来测试模型性能. 在使用滑动窗口方案构建训练样本时, 紧密性依赖的长度 p 和周期性依赖的长度 q 都设定为 3, 周期性长度 λ 设为 7. 我们使用 LSTM 作为网络模型, 其中包含两个 LSTM 层, 每层各有 64 个隐藏神经元. 最后, 经过一个线性层将特征映射为一个输出作为预测值. 为了公平起见, 除浅层学习算法外, 所有基线方法都采用相同的网络结构. 除非另有说明, 否则我们在本地客户端和中央服务器之间默认进行 100 轮通信, 并返回最终模型. 正则化项 ρ 是通过网格搜索的方法来确定的, 数值范围为−0.3 至 0.3, 步长为 0.1, 用来调整预热模型对全局模型的影响. 聚类大小 C 设定为 16. 与 FL [42]中的标准设置类似, 本地迭代轮数和本地采样批次大小的值分别被设置为 1 和 20. 在每一轮通信中, 随机从已选中的 100 个基站中选择 10% 的基站参与了模型训练. 我们使用随机梯度下降 (stochastic gradient descent, SGD) 作为优化器来更新我们的模型, 学习率为 0.01. 为了模拟数据分布式存储, α 取 10. 其中, α 的取值代表每个基站共享的数据占基站总数据量的比例, 即共享数据占基站总数据量的 10%. 其中, α 的取值可以进行灵活调整, 它的取值越大, “预热”模型所包含的各基站流量变化模式越多, 全局模型的性能则越好. 反之, 当 α 的取值过小时, 所包含的信息有限, 无法起到缓解数据异构问题的作用. 因此, 在经过不断尝试后, 我们将 α 设置为 10, 这样既可以有效发挥 α 的作用, 同时又能保证基站所共享的数据量不会泄露基站整体流量变化模式. 换言之, α 取 10 时, 在现实情况下是允许的, 用户只需要分享任意 10% 不涉及隐私的数据即可. 在这种意义上, 这些不涉及隐私的数据即使被其他恶意用户获取也无法得知任何有效信息.

4.4 实验结果及分析

对不同的基线方案在相同实验条件下进行了实验, 实验结果见表 1. 需要注意的是, 在提出的方案中加入了预热模型, 为了使得方案符合现实情况, 将模型中先验知识的比例设置为 10, 即每个基站都分出自身任意 10% 的数据来训练预热模型, 这在现实情况中也是可以接受的. 此外, 可以通过调整 ρ 的值来改变预热模型在全局模型中所占的比例. 下面除非另有说明, 否则所有结果都是在 α=10 的情况下得到的. 从表 1 中可以看出, 我们的 DualICA 在两个数据集上的各种无线流量上的预测性能都优于所有的基线方法.

表 1　不同方法在两个数据集上预测性能比较

Methods	Milano						Trento					
	MSE			*MAE*			*MSE*			*MAE*		
	SMS	Call	Internet	SMS	Call	Internet	SMS	Call	Internet	SMS	Call	Internet
SVR	0.4144	0.0919	0.1036	0.3528	0.1852	0.2220	5.2285	1.7919	5.9080	1.0390	0.5656	1.0470
LSTM	0.5608	0.1379	0.1697	0.4287	0.2458	0.2936	3.6947	1.1378	4.6976	0.9426	0.5013	1.1193
FedAvg	0.3744	0.0776	0.1096	0.3386	0.1838	0.2319	2.2287	1.6048	4.7988	0.7416	0.5319	1.0668
FedAtt	0.3667	0.0774	0.1096	0.3375	0.1837	0.2321	2.1558	1.5967	4.7645	0.7444	0.5306	1.0629
FedDA	0.3481	0.0753	0.1062	0.3321	0.1810	0.2275	2.0719	1.1699	3.9266	0.7320	0.4543	0.9504
AD+MKrum	0.4342	0.0816	0.1158	0.3583	0.1867	0.2359	4.2458	1.7817	5.1978	0.9830	0.5600	1.1011
MAML	0.3472	0.0743	0.1051	0.3301	0.1783	0.2243	2.0515	1.1419	3.9056	0.7212	0.4432	0.9413
DualICA-W	0.3565	0.0806	0.1219	0.3363	0.1841	0.2358	3.7893	1.2510	4.1231	0.8899	0.4658	0.9612
DualICA	**0.3164**	**0.0692**	**0.0912**	**0.3198**	**0.1710**	**0.2062**	**1.9183**	**0.9011**	**3.5812**	**0.7072**	**0.4201**	**0.8543**
↑ (%)	9.1	8.1	14.1	3.7	5.5	9.4	7.4	23	8.8	3.4	7.5	10.1

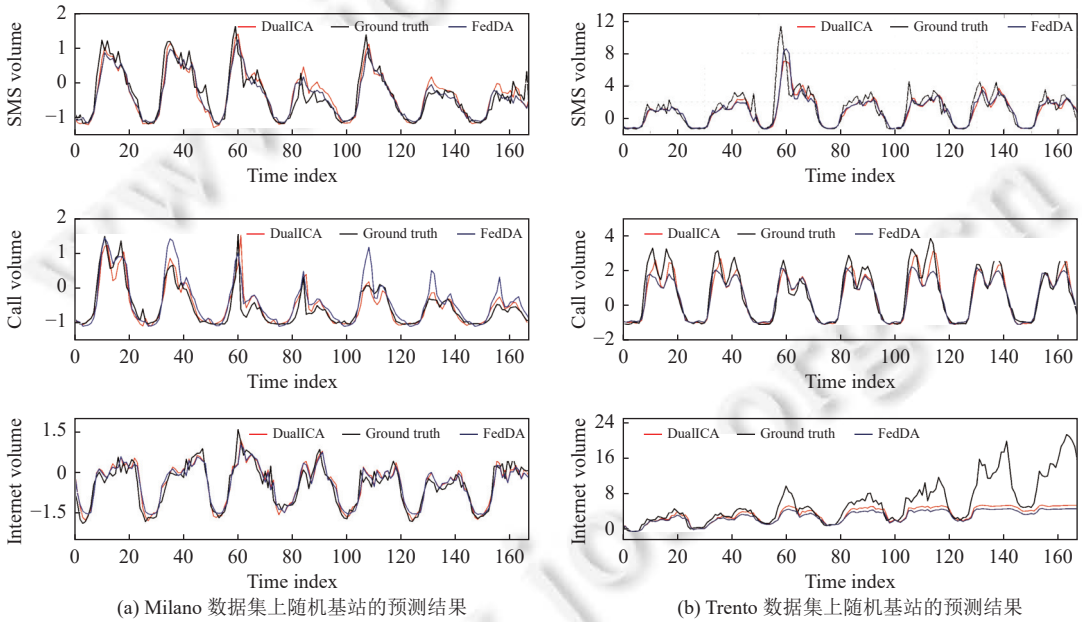
注: ↑表示DualICA与FedDA对比性能增益

具体来说, 对于 Milano 数据集的短信、电话和互联网服务 3 种无线流量, 我们主要与基线中表现最好的方法 FedDA 做对比, DualICA 在上面 3 种无线流量上取得的 *MSE* 增益分别为 4.3%、8% 和 14.2%. 同样的, 在 Trento 数据集上, DualICA 在电话流量上性能提升最为明显, 达到了 23%, 在短信和互联网服务上的提升与 Milano 数据集上相似, 分别为 7.4% 和 8.8%. 就 *MAE* 指标而言, 虽然改进效果不如 *MSE* 明显, 但是在 2 个数据集的 3 种无线流量上仍然有着不小的提升. 在 Milano 数据集上, 3 种无线流量数据在 *MAE* 上的增益分别为 2.7%、5.5%、9.6%. 在 Trento 数据集上, 3 种无线流量数据的增益则分别来到 3.4%、7.5%、10.1%. 总体而言, 虽然先验知识的比例仅

有 10%, 但是在性能上的提升是较大的, 出现这种结果的原因除了预热模型外, 注意力机制在模型聚合时发挥的作用也是不可忽略的. 与其他几种基线方法相比, DualICA 取得成功的原因归结如下.

与传统的基础分类算法 SVR 和考虑长短时间序列的 LSTM 算法相比, DualICA 使用滑动窗口模式更好的捕捉了时间依赖性. 同时, DualICA 加入了模式聚类来捕捉不同类无线流量变化特点, 因此在性能上更加优越. 与联邦学习算法 FedAvg 和 FedAtt 相比, DualICA 加入了预热模型, 有效克服了因基站之间流量数据模式高度异构而导致的全局模型泛化能力差的问题. DualICA 与 FedDA 相似, 但是在细节上仍然有着较大区别. DualICA 先对整个数据集中的基站进行聚类操作, 之后从带有类标签的基站中进行随机选择, 缓解了 FedDA 方案中因选取的基站数量较少, 从而出现聚类后再进行随机抽取时存在的基站选择敏感的问题. 此外, 使用模式聚类来捕捉不同基站间的流量模式变化, 从而具有了更加准确的预测. 方案 AD+MKrum 在考虑流量预测的同时还解决了训练过程中存在恶意节点时的模型安全问题, 但是由于扰动策略的加入也影响了模型的预测准确率. MAML 利用基于距离的加权模型进行模型聚合, 相对 FedDA 来说有所提升, 但是提升效果有限. 同时, 为了印证所设计的预热模型对最终模型的影响, 在 DualICA-W 中进行了实验, 结果表明去掉预热模型后, 模型预测性能明显下降.

此外, 从表 1 中结果还可以看出, 与中心化算法相比, 基于联邦学习的算法可以实现更好的预测. 其中, FedAtt 引入了注意力机制, 其预测性能要优于经典的 FedAvg 算法. 但是它们缺乏先验知识, 直接导致了全局模型预测精度的损失. 为了更加直观的评估不同算法的预测性能, 还对不同算法的预测值和真实值进行了比较. 从表 1 中可以看到, FedDA、MAML 预测性能相似. 但是, FedDA 与本文方法更为相似, 同样采用注意力机制, 而 MAML 所涉及的元学习方法与本文无关. 因此, 综合模型预测性能和方案设计方法, 为观察不同算法长期的预测表现, 我们选择 FedDA 进行预测值和真实值对比, 图 7 展示了 DualICA 与基线方法 FedDA 的对比结果.



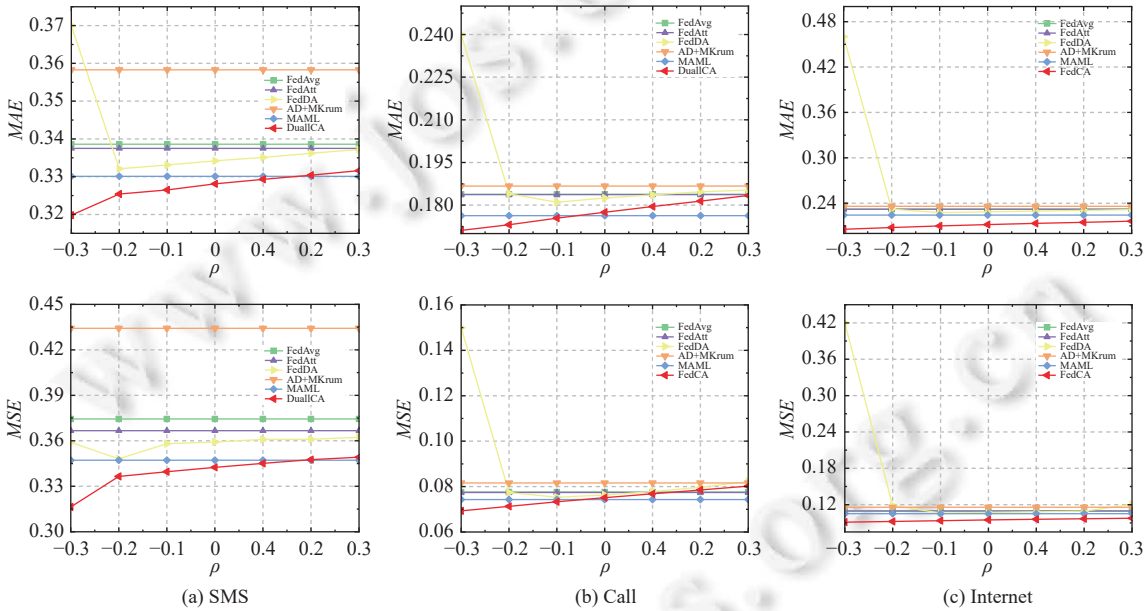
(a) Milano 数据集上随机基站的预测结果　　(b) Trento 数据集上随机基站的预测结果

图 7　预测值与真实值对比

图 7(a) 和图 7(b) 分别展示了 DualICA 及对比方案 FedDA 在 2 个数据集上 3 种无线流量数据下预测值与真实值的关系. 以数据集中最后一周的数据用来测试模型性能, 可以看出在两个数据集中无线流量都与时间序列有关, 呈现规律变化. 详细来说, 在 Milano 数据集的 3 种无线流量数据的预测上, DualICA 都表现出了更加优异的性能, 尤其在电话和上网服务流量的预测较 FedDA 更加精确. 相似的, 对于 Trento 数据集, DualICA 在短信流量上的预测性能与 FedDA 相差并不明显. 然而, 在电话和上网服务两种无线流量数据上的预测性能的差距则较为凸显. 在两个数据集上的预测性能对比还有一个有趣的共同点, 方案 DualICA 与方案 FedDA 的差别都出现在流量峰值

和转折处, 流量变化较为平缓处的差距几乎不可见. ρ 决定了预热模型在全局模型中所占比重, 对模型的预测性能影响巨大. 通过改变 ρ 的值来观察预热模型对模型性能影响, 实验结果如图 8 和图 9 所示. 图 8 和图 9 分别展示了模型在 Milano 和 Trento 数据集上随 ρ 变化时的预测性能变化, 并且在其上下两部分的子图中展示了 *MSE* 和 *MAE* 指标结果. ρ 变化范围从−0.3 到 0.3, 步长 0.1.

除了本文提出的模型, 还对比了其他基线方法在 ρ 改变时的模型性能变化. 其中, FedAtt、FedAvg、AD+MKrum 以及 MAML 方法虽然也融合了联邦学习, 但是并没有加入先验知识, 所以它们的结果在固定随机种子的情况下是固定不变的, 并不会随 ρ 的改变而发生改变. 在图 8 Milano 数据集上, DualICA 达到了最小的误差. 同时, 随着 ρ 增大, DualICA 得到的 *MAE* 和 *MSE* 都有了微小的增加, 处于相对稳定的状态. 这表明 DualICA 性能并不完全依赖于预热模型. FedDA 在最开始出现了剧烈下降, 并在后面也出现随着 ρ 增大而增大的情况. 对比其他基线方法 FedAtt 和 FedAvg, 由于缺乏先验知识, 其模型误差虽然稳定不变, 但是都高于 FedDA. 在 Trento 数据集上, FedDA 波动明显, 除了短信流量, 在电话和上网服务流量上的模型误差受 ρ 影响较大, 表明了 FedDA 对先验知识的依赖程度较高. 而 DualICA 则相反, 基本处于稳定状态, 受 ρ 的影响较小. 综上分析可知, 在模型中加入先验知识对模型性能确实是有正面影响的. 然而, DualICA 和 FedDA 两种方法在都加入了先验知识情况下, 从实验结果和分析中可以看出 FedDA 模型性能受先验知识的影响较大, 而 DualICA 则在取得更小误差的情况下整体趋于稳定状态, 明显优于其他基线方法.



(a) SMS (b) Call (c) Internet

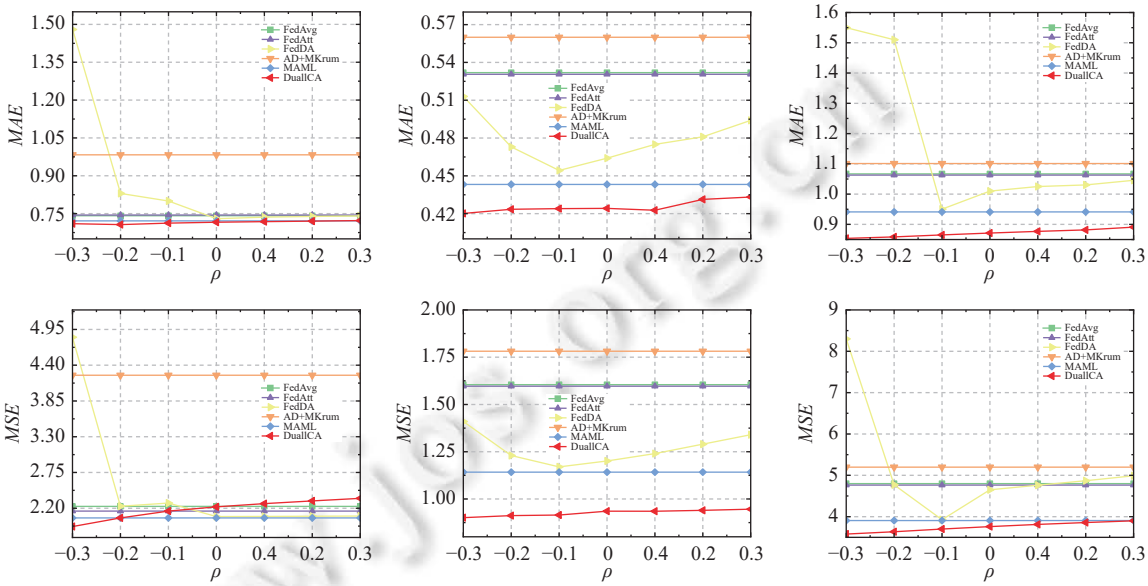
图 8 在 Milano 数据集上 ρ 对预测性能的影响

为了捕捉不同基站的无线流量模式, 缓解基站选择敏感的问题, 本文提出了模式聚类方法, 先对整个数据集中所有基站进行聚类操作之后再进行后续的训练过程. 其中, 聚类大小 C 决定了最终参与全局模型的聚合的类本地模型数量. 因此, C 的值也会影响最终模型的预测性能. 为了探讨聚类大小如何影响 DualICA 的预测性能, 在不同聚类大小情况下进行了实验, 结果见图 10. 图 10(a) 和图 10(b) 分别展示了在 Milano 和 Trento 数据集上的实验结果. 我们主要考虑 3 种情况, 即 C=1、C=16 和 C=32. 其中, C=1 与不采用聚类是等价的. 可以看到, C 的选择对 DualICA 的预测性能产生了不同的影响. 在大多数情况下, 引入聚类策略确实可以降低预测误差. 从图 10(a) 中观察到, 当聚类大小为 16 或 32 时, DualICA 在短信和呼叫流量方面取得了一定的性能改进. 对于互联网流量, 虽然 C=16 时性能略有下降, 但 C=32 时则有所提高. 在 Trento 数据集中, 引入模式聚类策略在短信和上网服务流量上的改善程度较大, 尤其是在聚类大小取 16 时, 效果有了明显提升. 但是在聚类数量增加到 32 时, 反而对模型性能

起到了反作用. 总的来说, 图 10 的结果证明了在 DualICA 中引入模式聚类的优越性. 这是因为聚类的大小 C 控制了参与全局模型聚合的类本地模型数量, 决定了模型所捕捉到的流量变化类别数量, 从而影响全局模型性能.

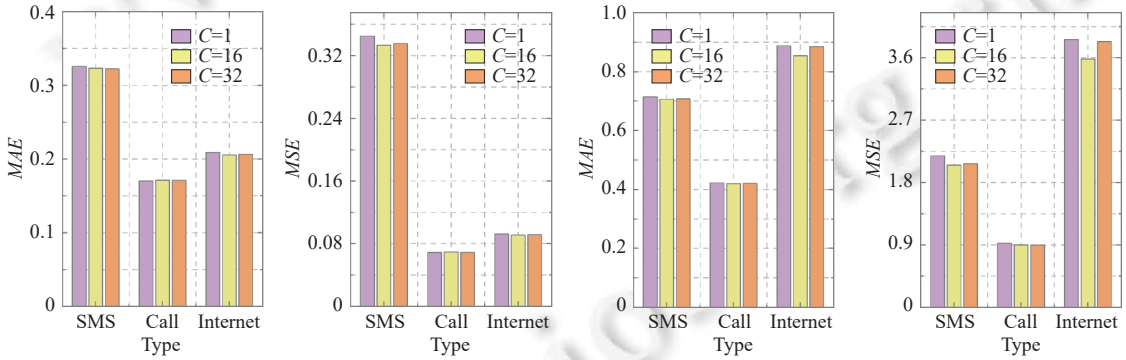
相较于其他预测方案, 所有的数据都混合在一起生成全局模型, 隐藏在数据中不同的流量模式难以被捕获, 制约了模型性能. 同时, 我们也发现聚类数量并不是越大越好, 需要根据实际情况灵活调整.

总体而言, DualICA 模型可以更加有效的捕捉无线流量变化特点, 在不依赖于先验知识的前提下, 实现更加精准预测.



(a) SMS (b) Call (c) Internet

图 9 在 Trento 数据集上 ρ 对预测性能的影响



(a) Milano 数据集上不同聚类大小的预测结果 (b) Trento 数据集上不同聚类大小的预测结果

图 10 聚类大小对预测性能的影响

5 总 结

本文研究了无线流量预测问题, 提出了一种基于注意力机制的“类内平均, 类间注意力”联邦无线流量预测模型 DualICA. DualICA 主要包括 3 部分: (1) 提出了模式聚类方法, 把流量模式相似程度较大的基站归为同类, 并使用经典平均的方法获取类本地模型, 更好的捕捉了不同流量模式的变化特点. (2) 设计了一种预热模型, 该模型由基站的少部分数据训练得到, 解决了因基站之间流量数据异构导致的全局模型泛化能力差的问题, 提高了全局模

型泛化能力. (3) 在模型聚合阶段引入了注意力机制, 量化不同客体贡献, 解决了联邦学习中模型聚合时不同客体贡献差异被忽略的问题, 有效提升了模型性能. 最后, 在两个现实世界数据集上进行了大量实验, 实验结果表明, 对比最先进的预测方法, DualICA 在两个数据集上都取得了较大幅度的预测性能提升.

但是, DualICA 仍存在不足. 虽然本文使用流量模式聚类的方法更好的捕捉基站的流量模式变化, 但是并没有利用基站的位置信息. 在现实情况中基站的空间依赖关系通常与流量变化有着很大关联, 对于流量预测必然有着极大的帮助, 所以如何利用基站间的空间依赖关系来开展无线流量预测问题会是我们下一步要研究的内容.

References:

[1] David K, Berndt H. 6G vision and requirements: Is there any need for beyond 5G? IEEE Vehicular Technology Magazine, 2018, 13(3): 72–80. [doi: 10.1109/MVT.2018.2848498]

[2] Zong BQ, Fan C, Wang XY, Duan XY, Wang BJ, Wang JW. 6G technologies: Key drivers, core requirements, system architectures, and enabling technologies. IEEE Vehicular Technology Magazine, 2019, 14(3): 18–27. [doi: 10.1109/MVT.2019.2921398]

[3] Wang Z, Wong VWS. Bayesian meta-learning for adaptive traffic prediction in wireless networks. IEEE Trans. on Mobile Computing, 2023, 23(6): 6620–6633. [doi: 10.1109/TMC.2023.3325301]

[4] Jiang WW. Cellular traffic prediction with machine learning: A survey. Expert Systems with Applications, 2022, 201: 117163. [doi: 10.1016/j.eswa.2022.117163]

[5] Niu ZS, Wu YQ, Gong J, Yang ZX. Cell zooming for cost-efficient green cellular networks. IEEE Communications Magazine, 2010, 48(11): 74–79. [doi: 10.1109/MCOM.2010.5621970]

[6] Niu ZS. TANGO: Traffic-aware network planning and green operation. IEEE Wireless Communications, 2011, 18(5): 25–29. [doi: 10.1109/MWC.2011.6056689]

[7] Kato N, Mao BM, Tang FX, Kawamoto Y, Liu JJ. Ten challenges in advancing machine learning technologies toward 6G. IEEE Wireless Communications, 2020, 27(3): 96–103. [doi: 10.1109/MWC.001.1900476]

[8] Yang Q, Liu Y, Chen TJ, Tong YX. Federated machine learning: Concept and applications. ACM Trans. on Intelligent Systems and Technology, 2019, 10(2): 12. [doi: 10.1145/3298981]

[9] Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 2020, 37(3): 50–60. [doi: 10.1109/MSP.2020.2975749]

[10] Kairouz P, McMahan HB, Avent B, *et al*. Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 2021, 14(1–2): 1–210. [doi: 10.1561/2200000083]

[11] Yang Q. Federated learning: The last on kilometer of artificial intelligence. CAAI Trans. on Intelligent Systems, 2020, 15(1): 183–186 (in Chinese with English abstract).

[12] Yang Q. AI and data privacy protection: The way to federated learning. Journal of Information Security Research, 2019, 5(11): 961–965 (in Chinese with English abstract). [doi: 10.3969/j.issn.2096-1057.2019.11.003]

[13] Nasr M, Shokri R, Houmansadr A. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In: Proc. of the 2019 IEEE Symp. on Security and Privacy (SP). San Francisco: IEEE, 2019. 739−753. [doi: 10.1109/SP.2019.00065]

[14] Chen H, Zhu T, Zhang T, Zhou WL, Yu PS. Privacy and fairness in federated learning: On the perspective of tradeoff. ACM Computing Surveys, 2023, 56(2): 1–37. [doi: 10.1145/3606017]

[15] Shu YT, Yu MF, Yang O, Liu JK, Feng HF. Wireless traffic modeling and prediction using seasonal ARIMA models. IEICE Trans. on Communications, 2005, 88(10): 3992–3999.

[16] Zhou B, He D, Sun Z. Traffic predictability based on ARIMA/GARCH model. In: Proc. of the 2nd Conf. on Next Generation Internet Design and Engineering. Valencia: IEEE, 2006. 207. [doi: 10.1109/NGI.2006.1678242]

[17] Cappé O, Moulines E, Pesquet JC, Petropulu AP, Yang XS. Long-range dependence and heavy-tail modeling for teletraffic data. IEEE Signal Processing Magazine, 2002, 19(3): 14–27. [doi: 10.1109/79.998079]

[18] Ashtiani F, Salehi JA, Aref MR. Mobility modeling and analytical solution for spatial traffic distribution in wireless multimedia networks. IEEE Journal on Selected Areas in Communications, 2003, 21(10): 1699–1709. [doi: 10.1109/JSAC.2003.815680]

[19] Tutschku K, Tran-Gia P. Spatial traffic estimation and characterization for mobile communication network design. IEEE Journal on Selected Areas in Communications, 1998, 16(5): 804–811. [doi: 10.1109/49.700914]

[20] Li RP, Zhao ZF, Zheng JC, Mei CL, Cai YM, Zhang HG. The learning and prediction of application-level traffic data in cellular networks. IEEE Trans. on Wireless Communications, 2017, 16(6): 3899–3912. [doi: 10.1109/TWC.2017.2689772]

[21] Zhang Y, Roughan M, Willinger W, Qiu LL. Spatio-temporal compressive sensing and Internet traffic matrices. In: Proc. of the 2009

ACM SIGCOMM Conf. on Data Communication. Barcelona: ACM, 2009. 267–278. [doi: 10.1145/1592568.1592600]

[22] Soule A, Lakhina A, Taft N, Papagiannaki K, Salamatian K, Nucci A, Crovella M, Diot C. Traffic matrices: Balancing measurements, inference and modeling. In: Proc. of the 2005 ACM SIGMETRICS Int'l Conf. on Measurement and Modeling of Computer Systems. Banff: ACM, 2005. 362–373. [doi: 10.1145/1064212.1064259]

[23] Sharma S, Majumdar A, Elvira V, Chouzenoux E. Blind Kalman filtering for short-term load forecasting. IEEE Trans. on Power Systems, 2020, 35(6): 4916–4919. [doi: 10.1109/TPWRS.2020.3018623]

[24] Li RP, Zhao ZF, Wei Y, Zhou X, Zhang HG. GM-PAB: A grid-based energy saving scheme with predicted traffic load guidance for cellular networks. In: Proc. of the 2012 IEEE Int'l Conf. on Communications (ICC). Ottawa: IEEE, 2012. 1160–1164. [doi: 10.1109/ICC.2012.6364637]

[25] Li RP, Zhao ZF, Zhou X, Palicot J, Zhang HG. The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice. IEEE Communications Magazine, 2014, 52(6): 234–240. [doi: 10.1109/MCOM.2014.6829969]

[26] Chen XM, Jin YH, Qiang SW, Hu WS, Jiang KD. Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale. In: Proc. of the 2015 IEEE Int'l Conf. on Communications (ICC). London: IEEE, 2015. 3585–3591. [doi: 10.1109/ICC.2015.7248881]

[27] Liu H, Xu B, Lu DJ, Zhang GJ. A path planning approach for crowd evacuation in buildings based on improved artificial bee colony algorithm. Applied Soft Computing, 2018, 68: 360–376. [doi: 10.1016/j.asoc.2018.04.015]

[28] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature, 2015, 521(7553): 452–459. [doi: 10.1038/nature14541]

[29] Su XG, Yan X, Tsai CL. Linear regression. WIREs Computational Statistics, 2012, 4(3): 275–294. [doi: 10.1002/wics.1198]

[30] Sapankevych NI, Sankar R. Time series prediction using support vector machines: A survey. IEEE Computational Intelligence Magazine, 2009, 4(2): 24–38. [doi: 10.1109/MCI.2009.932254]

[31] Feng HF, Shu YT, Wang SY, Ma MD. SVM-based models for predicting WLAN traffic. In: Proc. of the 2006 IEEE Int'l Conf. on Communications. Istanbul: IEEE, 2006. 597–602. [doi: 10.1109/ICC.2006.254860]

[32] Zuo C, Qian JM, Feng SJ, Yin W, Li YX, Fan PF, Han J, Qian KM, Chen Q. Deep learning in optical metrology: A review. Light: Science & Applications, 2022, 11(1): 39. [doi: 10.1038/s41377-022-00714-x]

[33] Kato N, Fadlullah ZM, Mao BM, Tang FX, Akashi O, Inoue T, Mizutani K. The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective. IEEE Wireless Communications, 2017, 24(3): 146–153. [doi: 10.1109/MWC.2016.1600317WC]

[34] Mao Q, Hu F, Hao Q. Deep learning for intelligent wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2595–2621. [doi: 10.1109/COMST.2018.2846401]

[35] Wang J, Tang J, Xu ZY, Wang YZ, Xue GL, Zhang X, Yang DJ. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In: Proc. of the 2017 IEEE Conf. on Computer Communications (INFOCOM 2017). Atlanta: IEEE, 2017. 1–9. [doi: 10.1109/INFOCOM.2017.8057090]

[36] Qiu C, Zhang YY, Feng ZY, Zhang P, Cui SG. Spatio-temporal wireless traffic prediction with recurrent neural network. IEEE Wireless Communications Letters, 2018, 7(4): 554–557. [doi: 10.1109/LWC.2018.2795605]

[37] Zhang CT, Zhang HX, Qiao JP, Yuan DF, Zhang MG. Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1389–1401. [doi: 10.1109/JSAC.2019.2904363]

[38] Wu Q, He KW, Chen X, Yu S, Zhang JS. Deep transfer learning across cities for mobile traffic prediction. IEEE/ACM Trans. on Networking, 2022, 30(3): 1255–1267. [doi: 10.1109/TNET.2021.3136707]

[39] Lin CY, Su HT, Tung SL, Hsu WH. Multivariate and propagation graph attention network for spatial-temporal prediction with outdoor cellular traffic. In: Proc. of the 30th ACM Int'l Conf. on Information & Knowledge Management. ACM, 2021. 3248–3252. [doi: 10.1145/3459637.3482152]

[40] Yao Y, Gu B, Su Z, Guizani M. MVSTGN: A multi-view spatial-temporal graph network for cellular traffic prediction. IEEE Trans. on Mobile Computing, 2023, 22(5): 2837–2849 [doi: 10.1109/TMC.2021.3129796]

[41] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, Ramage D, Segal A, Seth K. Practical secure aggregation for privacy-preserving machine learning. In: Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security. Dallas: ACM, 2017. 1175−1191. [doi: 10.1145/3133956.3133982]

[42] McMahan B, Moore E, Ramage D, Hampson S, Arcas BAY. Communication-efficient learning of deep networks from decentralized data. In: Proc. of the 20th Int'l Conf. on Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2017. 1273–1282.

[43] Zhao Y, Li M, Lai LZ, Suda N, Civin D, Chandra V. Federated learning with Non-IID data. arXiv:1806.00582, 2018.

[44] Li T, Sahu K, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. In: Proc. of the 2020 Machine Learning and Systems. Austin: MLSys, 2020. 429–450.

[45] Ji SX, Pan SR, Long GD, Li X, Jiang J, Huang Z. Learning private neural language modeling with attentive aggregation. In: Proc. of the 2019 Int'l Joint Conf. on Neural Networks (IJCNN). Budapest: IEEE, 2019. 1–8. [doi: 10.1109/IJCNN.2019.8852464]

[46] Shu JG, Yang TT, Liao YX, Chen FR, Xiao Y, Yang K, Jia XH. Clustered federated multitask learning on Non-IID data with enhanced privacy. IEEE Internet of Things Journal, 2023, 10(4): 3453–3467. [doi: 10.1109/JIOT.2022.3228893]

[47] Yang L, Huang Jm, Lin Wy, Cao JN. Personalized federated learning on Non-IID data via group-based meta-learning. ACM Trans. on Knowledge Discovery from Data, 2023, 17(4): 49. [doi: 10.1145/3558005]

[48] Arisdakessian S, Wahab OA, Mourad A, Otrok H. Coalitional federated learning: Improving communication and training on Non-IID data with selfish clients. IEEE Trans. on Services Computing, 2023, 16(4): 2462–2476. [doi: 10.1109/TSC.2023.3246988]

[49] Sun QH, Li X, Zhang JY, Xiong L, Liu WR, Liu JF, Qin Z, Ren K. ShapleyFL: Robust federated learning based on shapley value. In: Proc. of the 29th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. Long Beach: ACM, 2023. 2096–2108. [doi: 10.1145/3580305.3599500]

[50] Mai WM, Yao JC, Chen G, Zhang Y, Cheung YM, Han B. Server-client collaborative distillation for federated reinforcement learning. ACM Trans. on Knowledge Discovery from Data, 2023, 18(1): 9. [doi: 10.1145/3604939]

[51] Zhang CT, Dang SP, Shihada B, Alouini MS. Dual attention-based federated learning for wireless traffic prediction. In: Proc. of the 2021 IEEE Conf. on Computer Communications (INFOCOM 2021). Vancouver: IEEE, 2021. 1–10. [doi: 10.1109/INFOCOM42981.2021.9488883]

[52] Zheng TH, Li BC. Poisoning attacks on deep learning based wireless traffic prediction. In: Proc. of the 2022 IEEE Conf. on Computer Communications (INFOCOM 2022). London: IEEE, 2022. 660–669. [doi: 10.1109/INFOCOM48880.2022.9796791]

[53] Zhang L, Zhang CT, Shihada B. Efficient wireless traffic prediction at the edge: A federated meta-learning approach. IEEE Communications Letters, 2022, 26(7): 1573–1577. [doi: 10.1109/LCOMM.2022.3167813]

[54] Barlacchi G, De Nadai M, Larcher R, Casella A, Chitic C, Torrisi G, Antonelli F, Vespignani A, Pentland A, Lepri B. A multi-source dataset of urban life in the city of Milan and the province of trentino. Scientific Data, 2015, 2: 150055. [doi: 10.1038/sdata.2015.55]

[55] Telecom Italia. Telecommunications—SMS, Call, Internet—TN, 2015. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QLCABU [doi: 10.7910/DVN/QLCABU]

[56] Telecom Italia. Telecommunications—SMS, Call, Internet—MI, 2015. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EGZHFV [doi: 10.7910/DVN/EGZHFV]

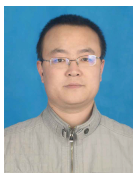
附中文参考文献:

[11] 杨强. 联邦学习: 人工智能的最后一公里. 智能系统学报, 2020, 15(1): 183–186.

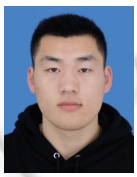
[12] 杨强. AI 与数据隐私保护: 联邦学习的破解之道. 信息安全研究, 2019, 5(11): 961–965. [doi: 10.3969/j.issn.2096-1057.2019.11.003]

柴宝宝(1994－), 男, 博士生, 主要研究领域为联邦学习, 区块链共识算法, 数据共享.

韩玉冰(1987－), 男, 博士, 讲师, 主要研究领域为物联网, 模式识别.



董安明(1982－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为无线通信, 物联网, 机器学习.



李浩(1995－), 男, 博士生, 主要研究领域为张量分解, 推荐系统, 模型压缩.

王桂娟(1990－), 女, 博士, CCF 专业会员, 主要研究领域为数据中心网络, 物联网, 机器学习, 图论.

禹继国(1972－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为无线网络通信, 物联网, 隐私计算, 区块链.