计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(1):1−10 [doi: 10.15888/j.cnki.csa.009346]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于宽度网络架构的单模型主导联邦学习①

文家宝, 陈泯融

(华南师范大学 计算机学院, 广州 510631)
通信作者: 陈泯融, E-mail: chenminrong@scnu.edu.cn

摘　要: 联邦学习是一种分布式机器学习方法, 它将数据保留在本地, 仅将计算结果上传到客户端, 从而提高了模型传递与聚合的效率和安全性. 然而, 联邦学习面临的一个重要挑战是, 上传的模型大小日益增加, 大量参数多次迭代, 给通信能力不足的小型设备带来了困难. 因此在本文中, 客户端和服务器被设置为仅一次的互相通信机会. 联邦学习中的另一个挑战是, 客户端之间的数据规模并不相同. 在不平衡数据场景下, 服务器的模型聚合将变得低效. 为了解决这些问题, 本文提出了一个仅需一轮通信的轻量级联邦学习框架, 在联邦宽度学习中设计了一种聚合策略算法, 即 FBL-LD. 算法在单轮通信中收集可靠的模型并选出主导模型, 通过验证集合理地调整其他模型的参与权重来泛化联邦模型. FBL-LD 利用有限的通信资源保持了高效的聚合. 实验结果表明, FBL-LD 相比同类联邦宽度学习算法具有更小的开销和更高的精度, 并且对不平衡数据问题具有鲁棒性.

关键词: 联邦学习; 宽度网络; 单轮通信; 隐私保护; 机器学习

引用格式: 文家宝,陈泯融.基于宽度网络架构的单模型主导联邦学习.计算机系统应用,2024,33(1):1–10. http://www.c-s-a.org.cn/1003-3254/9346.html

Single Model Dominant Federation Learning Based on Broad Network Architecture

WEN Jia-Bao, CHEN Min-Rong

(School of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract: Federated learning is a distributed machine learning approach that enables model delivery and aggregation without compromising the privacy and security of local data. However, federated learning faces a major challenge: the large size of the models and the parameters that need to be communicated multiple times between the client and the server, bringing difficulties for small devices with insufficient communication capability. Therefore, this study set up the client and server to communicate with each other only once. Another challenge in federated learning is the data imbalance among different clients. The model aggregation for servers becomes inefficient in data imbalance. To overcome these challenges, the study proposes a lightweight federated learning framework that requires only one-shot communication between the client and the server. The framework also introduces an aggregation policy algorithm, FBL-LD. The algorithm selects the most reliable and dominant model from the client models in a one-shot communication and adjusts the weights of other models based on a validation set to achieve a generalized federated model. FBL-LD reduces the communication overhead and improves aggregation efficiency. Experimental results show that FBL-LD outperforms existing federated learning algorithms in terms of accuracy and robustness to data imbalance.

Key words: federated learning; broad network; one-shot communication; privacy protection; machine learning

① 基金项目: 国家自然科学基金 (61872153, 61972288)
收稿时间: 2023-06-28; 修改时间: 2023-07-27; 采用时间: 2023-08-08; csa 在线出版时间: 2023-11-24
CNKI 网络首发时间: 2023-11-28

大数据时代背景下, 数据量快速增长, 同时数据规模急剧扩大, 然而这并没有带来数据训练的质量和效率的提升, 反而出现了数据孤岛问题. 具体地说, 由于商业模式、隐私法律、传输限制等因素的影响, 并非所有数据都能被有效地利用和共享, 很多数据被孤立在不同用户设备上. 这就导致了机器学习中最重要最本质的资源, 即数据, 无法被充分地聚合和利用. 同时, 数据的收集者和利用数据进行机器学习训练的开发者通常不属于同一体系, 这也增加了数据训练的难度和成本. 这些问题不仅限制了机器学习模型的质量和效率, 也阻碍了人工智能技术在各行各业的广泛应用. 为了解决“数据孤岛”问题, 研究人员提出了一种分散数据训练的方案, 联邦学习 (federated learning, FL). 联邦学习通过各种深度神经网络, 将各个数据碎片训练的成果联合起来, 在处理分散数据方面表现出了优异的性能[1].

联邦学习结合深度神经网络广泛应用于人脸识别[2]、无线通信[3]、车辆物联网[4]等领域. 尽管神经网络在许多领域表现出优异的性能, 但由于深度神经网络的多层网络结构, 训练深度神经网络模型是一个非常耗时且重复的过程, 此外, 联邦学习支持在大量边缘设备上进行训练, 但这些设备的算力往往比较有限[5], 这给联邦学习带来了新的挑战和需求. 一方面, 随着智能设备用户的增多, 他们逐渐成为数据贡献的重要一环. 为了根据用户偏好制定商业策略, 娱乐购物应用对用户数据的需求快速增长. 另一方面, 万物互联成为趋势, 无论是工业生产设备还是家用电器设备, 不同设备之间需要处理对方收集的数据进行智能化工作. 因此, 小型设备的联邦训练场景日益增加, 联邦学习需要更加简洁快速的网络结构.

近年来也有以快速高效为特点的神经网络, 如宽度学习系统[6], 可以通过伪逆计算和增量计算, 快速处理大量数据并生成网络参数. 宽度学习系统并非在深度上扩展网络, 而是横向扩展, 并且可以使用增量学习方法不断增加网络节点, 以实现更好的性能. 本文将宽度学习系统引入联邦学习, 以减少大量参数的通信开销. 目前在联邦宽度学习上的研究存在以下 3 个问题.

(1) 陈旧更新问题: 为了节省时间, 采用异步策略的联邦学习可能会导致陈旧问题. 即客户端和服务器的模型版本不同步, 从而使用了过时的模型来更新参数.

(2) 通信消耗问题: 一方面, 在联邦学习中, 通信资源是有限的, 尤其是部分联邦设备不具备较强的通信能力, 因此需要上传参数模型尽可能小, 通信次数尽可能少. 另一方面, 大量参数多次迭代的通信很容易发生隐私泄露问题. 因此需要利用有限的通信资源来提高全局模型的优化效率, 将客户端服务器的通信次数尽可能降低.

(3) 数据不平衡问题: 现实场景下, 每个客户端的数据来源受限于多种因素影响, 例如, 受环境制约而收集数据量悬殊的不同用户之间进行联合训练时, 存在客户端之间的数据规模差距过大的现象. 如果简单地聚合数据量少的用户模型, 将会降低全局模型效果, 如果剔除数据量少的用户, 这将导致全局模型缺失部分特征, 无法在这些用户上发挥作用.

针对陈旧更新, 通信消耗, 数据不平衡问题, 本文首先将客户端与服务器之间限制为一轮通信, 将联邦学习中的通信轮次降至最低, 同时模型参数的缩减也让通信开销减少. 此外, 由客户端控制模型的发送时机, 每个客户端随时可以上传自己的最新模型并获取联邦模型, 在一轮通信的模型更新中, 客户端所获取的一定是当前最新的模型, 避免了陈旧问题. 此外, 在单轮通信中, 从客户端选出主导者, 利用其他客户端的模型参数泛化联邦模型, 通过主导权重步来动态调整各个用户的聚合占比, 使得小数据客户端也能参与聚合并分配到合理的权重, 进一步提高联邦模型的效果. 综上所述, 本文提出了一个新的算法 FBL-LD, 贡献为以下 3 点.

(1) 在单轮通信的限制下, 本文设计的 FBL-LD 通信方法简洁高效, 由客户端掌握主动权, 每个客户端随时可以上传自己的最新模型并获取联邦模型, 若服务器收集的模型充分, 则直接返回给客户端最新的联邦模型, 否则等待服务器充分收集模型, 在这过程中, 客户端可以使用新的模型覆盖上一次的模型重新进行通信, 保证每次收到的联邦模型都使用了客户端最新的模型, 解决了客户端陈旧模型更新的问题.

(2) FBL-LD 将更多工作交给服务器端, 迭代不再发生在客户端与服务器之间, 而是在服务器上完成. 本文设计了一种服务器上的主导者聚合策略, 这是 FBL-LD 的核心思想, 选出一个具有代表性和优势性的主导者, 利用其他模型的参数进行泛化, 动态调整这些模型的权重比直到达到收敛条件.

(3) FBL-LD 能有效应对数据不平衡问题, 不同客户端的数据规模不同时, 算法会动态调整各个客户端

模型的权重, 既不舍弃小数据客户端的部分特征, 也不会让小数据客户端模型对联邦模型的影响过大, 从而提高联邦模型的精度.

1 相关工作

由于数据隐私保护越来越受到关注和重视, 各个数据方不愿意将自己所拥有的数据直接交付给第三方进行联合训练, 联邦学习应运而生. 联邦学习方法在构建模型时, 不需要移动或共享原始数据, 用于上传和聚合的只有网络中的权重参数. 但是模型却得益于每个数据方的贡献, 从而可以在每个用户上发挥作用. 用户在本地用自己的数据训练模型, 并交换优化彼此的模型参数梯度. 例如在训练期间选择性共享模型参数, 随机梯度下降过程中, 在参数共享和局部参数更新之间进行权衡, 让参与者从其他参与者的模型中获益[7]. 这是一种典型的横向联邦场景, 所有的参与方持有不同样本但相同特征的数据. 在模拟场景中, 一个持有大量同类数据的数据方, 将数据分发给各个用户, 每个用户分配到不同的数据, 然后数据方不再参与训练过程, 各个用户在不共享原始数据的情况下联合训练模型, 即横向联邦[8]. 除了横向联邦以外, 联邦学习还有两种类型, 纵向联邦学习和联邦迁移学习[9]. 本文主要聚焦于横向联邦学习, 此类联邦学习中各个客户端数据的区别在于样本, 通常以公司, 集团, 地域, 国家等方式划分, 因此客户端之间存在较大的数据规模差异[10].

联邦学习发展至今, 已细分出多个侧重研究方向, 如数据[11−13], 安全[14,15], 公平[16−18], 多任务[19,20], 压缩[21,22]等. FedAvg 作为经典的基线算法, 使用大批量同步 SGD, 在客户端上进行更多的计算来提高收敛速度. 为了应对数据异构和系统异构, FedProx[23]将 FedAvg 抽象出来, 对原有算法进行了近端修正项的微调, 这一项的改变保留了 FedAvg 的算法特性, 并且可以缓解模型陷入局部最优的问题. 同时根据客户端的情况动态调整训练终止条件, 防止系统异构导致的过度训练影响整体模型的收敛. 文献[24]充分考虑了客户端的个性需求, 通过结合元学习方法, 为每一个客户端设计个性化全局模型, 解决小数据客户端无法初始化模型的问题, 并考虑了更多类型的设备. 文献[25]采取客户端服务器异步策略, 针对本地模型在联邦模型更新后出现的陈旧问题, 将 3 种衰减策略函数应用到模型更新中, 调整陈旧模型的权重, 减少陈旧模型对收敛的影响. 文献[26]考虑了局部和全局损失的差异性, 对于局部的本地训练, 矫正其向全局最优的方向收敛, 利用动态正则化阻止客户端的局部收敛, 从而提升联邦模型的质量. 与上述研究不同, 本文的研究重新设计了一种客户端与服务器的交互方法, 客户端在一轮通信内获取联邦模型, 因此, 客户端不再会因为服务器的超前训练而忽略自己的模型. 这提高了训练过程中的隐私性, 减少了通信开销.

深度学习的许多网络中包含了大量的层和连接参数, 消耗大量的时间和资源, 为了给深度学习提供更多视角, 提出了更为简洁的宽度学习系统 (broad learning system, BLS)[6], 将输入数据通过两次激活函数的变换, 并为一层, 不断横向扩展, 直到网络参数学习到数据特征与标签的关系. 具体地说, BLS 源于随机向量函数链神经网络 (random vector functional link neural network, RVFLNN)[27], 在其基础上新增加了一层提取原始数据的特征, 将原网络中数据与增强层的连接升级为映射层与增强层的连接, 可以直接对接其他网络的特征提取器, 在扩展了网络适用性能力的同时也增加了网络的性能. BLS 具有灵活性高, 扩展性强, 计算速度快等优点, 能够有效解决过拟合的问题[28]. 例如, Lei 等人提出了 ConvBLS[29], 这是一种用于图像分类的高效递增卷积宽度学习系统, 它采用了球形 K-means 算法和两阶段多尺度特征融合的卷积技术, 能够提取出更有效的多尺度特征. Yuan 等人则将 BLS 应用于迁移学习领域, 并提出了 DABLS-LLE[30], 这是一种基于局部线性嵌入的领域适应性宽度学习系统, 它可以利用目标域的少量标记数据和源域的全部标记数据学习分类模型.

一个 BLS 的主要结构如图 1[6], 输入数据 X 乘以一组随机的权重并加上随机的偏差, 通过激活函数的非线性变换得到许多组新节点, 这些节点构成映射层, 如式 (1)[6]:

$$Z_i = [\Phi(X_k W_{ei} + \beta_{ei})] \tag{1}$$

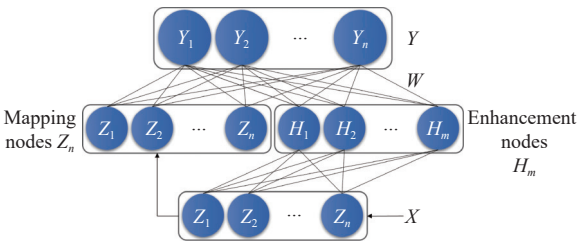


图 1 BLS 基本网络模型

对映射层使用相同的方法, 如式 (2)[6], 当然, 在这过程中可以设置不同的超参数, 这样就得到了增强层.

$$H_j = \left[\xi\left(Z_n W_{hj} + \beta_{hj}\right)\right] \tag{2}$$

将映射层和增强层水平拼接, 获得输入矩阵. 输入矩阵与特定的权重 W_m 相乘即可变为输出层 Y, 权重 W_m, 又称参数矩阵, 可以通过输入矩阵的伪逆与 Y 计算得到. 如式 (3)[6]. 参数矩阵决定了 BLS 的网络性能, 也就是宽度网络的模型参数, 可以通过增量学习不断微调.

$$W = [Z_n, H_m]^{+} Y \tag{3}$$

BLS 网络与联邦宽度学习的结合是近年来一个新兴的研究领域, 已经有许多研究在这方面取得了显著的进展. 例如, Le 等人提出了 FCL-BL[31], 这是一种基于宽度网络结构的联邦持续学习方法. 为了满足连续学习的需求, 它采用了加权处理策略来应对灾难性遗忘问题, 并使用了批处理异步方法来减少客户端和服务器端之间的交互次数, 同时利用了增量学习技术来加快新数据的处理速度和节省计算资源. Ren 等人提出了 FBL-ET[32], 这是一种基于事件触发机制的联邦宽度学习框架, 它为每个客户端设定了动态触发条件来筛选上传模型的质量, 并设计了一种激励机制来提升客户端的参与度和活跃度, 从而吸引更多高质量的客户端加入全局聚合过程. Ge 等人提出了一种基于垂直联邦学习思想的宽度学习系统[33], 它将两个客户端的数据特征集成到第三方进行模型训练, 从而利用更丰富的信息来提高模型性能. 但是这种方法需要将标签数据共享给第三方, 存在标签信息泄露的问题. 同样, 宽度学习思想已应用于一些联邦学习的具体实例, 如通用的轻量级联邦学习框架 FedDBL[34], 将深度学习特征提取器与宽度学习联邦聚合方法结合起来, 用于组织病理学; BiBLS[35]使用双向连接方式改进宽度网络结构, 并将联邦学习模式应用到车联网领域.

2 FBL-LD 方法

本节中详细介绍了 FBL-LD 模型. 首先介绍 FBL-LD 的通信方法, 然后介绍 FBL-LD 在服务器上所使用的聚合策略, 讨论主导权重步这一参数的作用. 最后通过两个算法描述 FBL-LD 的总体框架, 给出全部具体的步骤.

2.1 通信方法

联邦学习中, 通信资源是宝贵的. 为了减少通信开销和提高隐私性, 各种研究尝试缩减通信模型的大小, 加快收敛速度, 减少通信轮次. 这样既能节省通信资源, 又能提高模型在传输中的安全性. 本文将通信次数限制为单轮, 使得算法能适用于低通信力的设备. 此外, 部分用户出于对隐私保护的考虑, 不愿意频繁地在客户端和服务器之间交换模型数据, 单轮通信同样满足此类用户的需要. 在通信方式中, 最常见的模式是服务器主导训练, 服务器随机地选择客户端并接受这些客户端的模型, 聚合后分发联邦模型再次训练迭代, 但是客户端并不都是可靠的, 存在掉线客户端或中断传输的客户端, 这就造成了客户端模型的陈旧和服务器的等待延迟. 一些研究为了克服这样的问题, 将测试集数据分发给客户端, 在客户端上进行模型的改进, 但是测试集的分发又会涉及通信开销和隐私问题[32]. 为了解决上述问题, 本文提出的通信方法如下: 客户端被设置为通信的发起者, 任何时候都可以向服务器发送最新的本地模型, 而服务器在接受足够的模型后进行聚合并下发给客户端.

算法 1 描述了客户端所做的工作, 其输入是客户端 k 拥有的本地数据 (X_k, Y_k), 其输出是经过训练的本地模型参数 W_k. 每个客户端 k 都利用自己的数据构建本地模型, 并不需要向服务器端传输数据. 具体的过程如下: 首先, 客户端 k 根据式 (1) 使用激活函数对数据进行变换, 得到 n 个映射层; 然后, 根据式 (2) 再次使用激活函数对映射层进行变换, 得到 m 个增强层; 接着, 根据式 (3) 利用伪逆方法计算出 BLS 网络的本地模型参数 W_k; 最后, 客户端 k 将 W_k 发送给服务器端. 所有客户端都重复上述步骤, 直到服务器端收集到足够数量的本地模型.

算法 1. 客户端本地训练

1) **for** $k = 0; k < N$ **do**
2) **for** $i = 0; i < n$ **do**
3) Random W_{ei}, β_{ei};
4) Calculate $Z_i = [\Phi(X_k W_{ei} + \beta_{ei})]$;
5) **end for**
6) Set the mapping nodes $Z_{nk} = [Z_1, \cdots, Z_n]$;
7) **for** $j = 0; j < m$ **do**
8) Random W_{hj}, β_{hj};
9) Calculate $H_j = [\xi(Z_n W_{hj} + \beta_{hj})]$;
10) **end for**
11) Set the enhancement nodes $H_{mk} = [H_1, \cdots, H_m]$;
12) Set $A_k = [Z_{nk}, H_{mk}]$;
13) Calculate A_k^{+};
14) $W_k = A_k^{+} Y_k$;
15) Send W_k to server;
16) **end for**

对于服务器端来说, 在完成联邦聚合后, 它会立即将全局模型参数 W_{fed} 分发给各个客户端. 如果某个客户端在服务器停止模型接收之前已经更新了自己的本地模型, 则该客户端可以用新的本地模型参与下一轮联邦聚合, 并且放弃旧的本地模型, 将新的本地模型参数 W_{k2} 发送给服务器端. 这样就保证了服务器下发的模型所使用的都是最新的本地模型, 避免了陈旧模型问题. 关于服务器的工作将在第 2.2 节讨论.

2.2 聚合策略

在一般的联邦学习中, 服务器在收集到足够的客户端模型后, 会将模型聚合成新模型下发, 在分配本地模型的参与度时, 会根据样本数, 本地精度, 时间滞后性等因素进行加权平均, 这样的分配在使用梯度下降的深度网络中往往有效. 在单轮的宽度模型聚合中, 本文提出了主导者策略. 首先, 服务器在接受了一定数量的客户端模型后, 将一个模型作为主模型, 分配固定的最大权重, 其余模型则平分剩余的权重. 在聚合得到新的模型参数后, 在验证集上得到新模型的精度并记录. 然后更换主模型重复此过程, 将精度最大的主模型确立为最新的主模型. 最后, 在主模型确立后利用其他客户端的模型参数泛化联邦模型, 通过主导权重步这一参数来动态调整其他用户的泛化程度, 重复此过程直到收敛或最大验证数达到最大轮次, 便得到了最终的联邦模型, 下发给用户即完成了此次联邦训练.

2.3 主导权重步

联服务器在首轮聚合后确定了基准模型, 之后将其他模型参数均匀地分配给主模型, 以改进其对于所有数据的适用性. 在这过程中, 主导者的权重会随着主导权重步依次下降直到收敛.

算法 2 描述了服务器所做的工作, 其输入是客户端 k 训练得到的本地模型 W_k, 其输出是经过联邦聚合的全局模型参数 W_{fed}. 首先, 服务器端需要等待一定比例的客户端完成训练并上传本地模型. 然后服务器端将收集到的 N 个本地模型分别作为主模型, 并与其他本地模型进行 N 轮联邦聚合. 在每轮聚合中, 主模型和其他本地模型分别赋予最大权重 M 和次要权重 Y. 同时, 服务器端根据每轮聚合后的验证精度来更新主模型所属的客户端编号, 选择验证精度最高的客户端作为主导者. 在确定了主导者之后, 服务器端将逐渐降低主模型的最大权重 M, 按照主导权重步长 $Step$ 进行衰减, 并继续进行联邦聚合和验证精度检测, 直至验证精度 θ 达到预设阈值 α 或轮次达到最大迭代轮数 $epoch$ 为止. 值得注意的是, 在服务器开始聚合前的任何时候, 客户端都可以更新自己的模型, 让服务器接受最新的客户端模型, 但是当服务器收集足够数量的本地模型后, 服务器将不再接受客户端的模型, 此时服务器会快速计算出联邦模型并完成此次训练, 确保训练不会被客户端的响应无限延长, 并避免本地模型的陈旧.

算法 2. FBL-LD 算法

1) Initialize $epoch = 0$, $\theta = 0$;
2) **for** $m = 0$; $m < N$ **do**
3) **for** $n = 0$; $n < N$ **do**
4) $Y_n = (1 - M)/(N - 1)$;
5) **end for**
6) $Y_m = M$;
7) Calculate $W_{\text{fed}} = \sum_{k=1}^{N} Y_k W_k$;
8) Update $k_{\max}$;
9) **end for**
10) **while** $\theta \leqslant \alpha$ and $epoch \leqslant \gamma$ **do**
11) $M = M - Step$;
12) **for** $n = 0$; $n < N$ **do**
13) $Y_n = (1 - M)/(N - 1)$;
14) **end for**
15) $Y_{k\max} = M$;
16) Calculate $W_{\text{fed}} = \sum_{k=1}^{N} Y_k W_k$;
17) Update W_{fed}
18) $\theta \leftarrow$ Calculate prediction accuracy in validation set;
19) $epoch = epoch + 1$;
20) **end while**

2.4 算法步骤

FBL-LD 单模型主导联邦学习算法如算法 3 所示.

算法 3. FBL-LD 单模型主导联邦学习算法

1) 客户端用本地数据构建宽度学习系统, 通过映射层和增强层的合并伪逆计算得到网络参数, 然后发送给服务器.
2) 客户端在服务器下发联邦模型前, 可以重新上传本地模型, 覆盖之前自己上传过的客户端模型.
3) 服务器接受本地模型直到满足聚合数量, 停止对客户端请求的响应.
4) 服务器通过验证精度选出主导者模型, 并通过主导权重步对主导模型的权重进行调整直到模型收敛.
5) 服务器下发模型给客户端用于后续测试.

图 2 展示了 FBL-LD 模型的工作流程, FBL-LD 被分为客户端和服务器端两部分. 客户端可以随时使用本地的数据独立训练各自的模型上传, 并向服务器请求返回当前的联邦模型. 服务器接受到客户端的模型后, 若模型数量达到预期, 则立刻完成聚合并下发, 若模型数量不足, 则等待更多的本地模型, 在这过程中, 本地模型随时可以更新自己的模型防止陈旧模型.

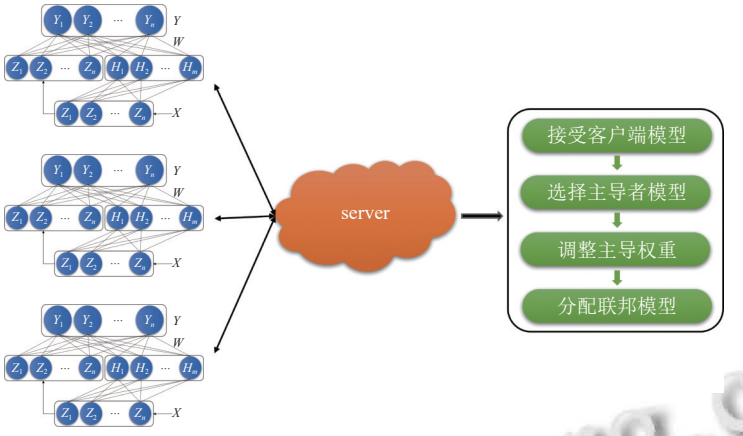


图 2 FBL-LD 框架

2.5 算法特点

FBL-LD 算法具有以下几个特点.

(1) 客户端可以随时上传本地训练得到的最新模型, 而无需从服务器端不断迭代下载和更新模型. 由于本地模型是由轻量级的 BLS 网络构建的, 层数浅参数少, 因此上传的模型小, 对通信资源的消耗低.

(2) 服务器端可以实时更新本地模型, 并利用它们进行联邦聚合. 这样可以避免使用过时的本地模型, 提高联邦模型的准确性.

(3) 算法对客户端设备的要求很低, 不需要频繁通信或处理大量数据. 算法的主要计算工作都在服务器端完成, 而服务器端有足够的计算能力和存储空间.

综上所述, FBL-LD 算法具有结构轻量、即时性和低限制等特点.

3 实验分析

本节中介绍了实验所使用的数据集以及数据集的划分方式. 然后通过几组实验对 FBL-LD 的参数进行调整, 并评估其性能表现. 最后将 FBL-LD 与基准算法做出比较, 以显示其在时间和通信上的优越性.

3.1 实验设置

MNIST 是一个经典的手写数字识别数据集, 有 60000 个样本的训练集和 100 000 个样本的测试集, 对应 10 个标签. Fashion-MNIST 是一个与 MNIST 设置相仿的数据集, 拥有更高的挑战性. Fashion-MNIST 将对数字的识别升级到对服饰的识别, 同样有 10 个标签, 相同的数据集大小. 本文采用了 2 种数据划分方式: (1) 平衡划分, 将数据集随机打乱并均匀地分配给所有客户. 所有客户持有的样本数量是相同的. (2) 非平衡划分: 将数据集随机打乱, 采取不均等比例的循环二分法划分, 直到数据分区大于要求的客户端数, 由于分配比例是不均等的, 客户端的样本数目具备明显的差距, 采用这样的数据划分以显示 FBL-LD 对小样本设备和差异性设备的适用性, 实验将每次二分比例设置为 1/3 和 2/3, 两个客户端的样本差距最大会达到 64 倍.

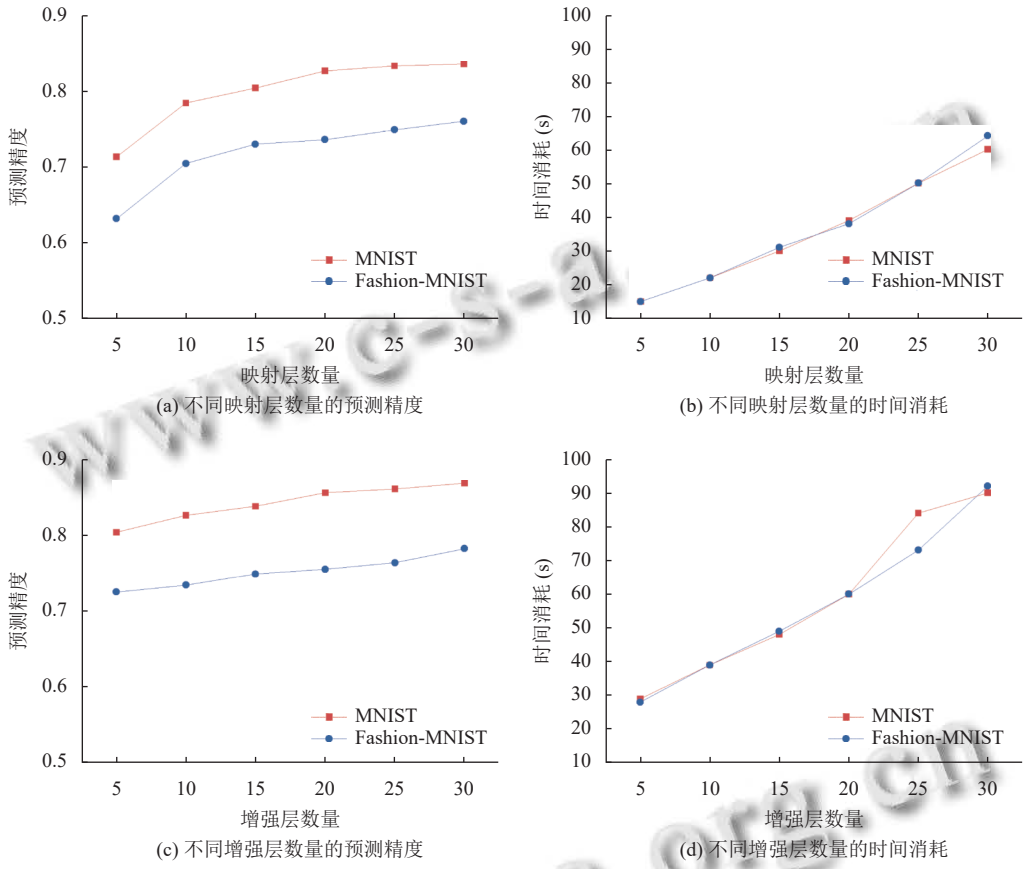
在实验中, FBL-LD 每组节点数设置为 100, 组数通过实验确定为 20, 映射层和增强层均使用 ReLU 激活. 设置主导权重步为 0.01, 然后设置最大验证数 γ 为 10. 对于比较算法, 选择联邦学习的基本算法 FedAvg[2], 及结合联邦学习和宽度学习的算法 FCL-BL[31], FBL-ET[32].

3.2 结果

实验验证了映射层增强层的数量, 客户端数, 客户端接受率等参数对预测准确性的影响. 实验使用了 MNIST 和 Fashion-MNIST 两种数据集, 并分别对其进行了平衡和不平衡的数据划分. 实验首先在平衡数据划分上进行了基准测试, 然后在不平衡数据划分上采用了相同的参数设置, 以评估算法在处理不平衡数据时的鲁棒性. 实验还将本文提出的 FBL-LD 算法与联邦宽度学习领域的其他算法进行了对比, 包括最初的 FedAvg 算法和近 3 年提出的 FCL-BL 和 FBL-ET 算法. 实验以预测精度和时间消耗为主要性能指标, 并通过上传模型的大小来估计通信开销. 实验结果表明, FBL-LD 算法在各项指标上优于其他算法, 体现了其在联邦宽度学习中的优势.

映射层增强层的数量: 在图 3(a) 中, 给客户端设置了不同的映射层, 从 5 层到 30 层, 比较了不同映射层的预测精度. 实验表明, 随着映射层的增加, 预测精度

也会增加. 还可以从图 3(a) 观察到, 映射层从 5 到 10 层的增加对预测精度提升最为显著, 从 10 到 20 层也有稳定的增长趋势, 20 层到 30 层的增长趋势最为缓慢. 从整体上看, 映射层的调整在两个数据集上都显示了类似的结果. 在图 3(c) 中, 在 20 层映射层的基础上给客户端设置了不同的增强层, 同样是从 5 层到 30 层, 比较了不同增强层的预测精度. 实验表明, 随着增强层的增加, 预测精度也会增加. 但是与图 3(a) 不同的是, 增强层对预测精度提升几乎是线性增长的, 这在两个数据集上都有所体现.



(a) 不同映射层数量的预测精度 (b) 不同映射层数量的时间消耗

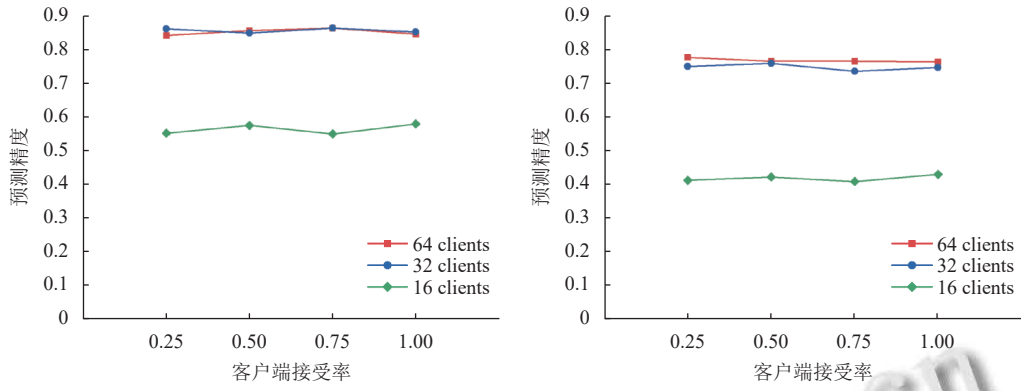
(c) 不同增强层数量的预测精度 (d) 不同增强层数量的时间消耗

图 3 不同网络层数量的预测精度和时间消耗

在图 3(b) 和图 3(d) 中, 考虑了映射层和增强层的增加对时间消耗的影响. 可以观察到, 无论是映射层还是增强层, 时间消耗都随着层数的增加稳步上升, 在 20 层后时间消耗的上升速度会有所提高, 因此综合预测精度和时间消耗, 将映射层增强层的数量都设置为 20.

客户端数和客户端接受率: 实验比较了不同客户端数量和客户端接受率对预测精度的影响, 在图 4 中可以看到随着客户端数的增加, 预测精度也会增加, 而客户端接受率对预测精度没有太大的影响, 不同的接受率下, 服务器的整体精度没有太大差别. 从整体上看, 客户端数从 16 增加到 32 时, 预测精度提升显著, 而增加到 64 时已经没有太大的提升, 但是在图 5 中, 客户端数为 64 的时间消耗要明显大于 32, 因此, 32 个客户端是合适的, 可以保持全局模型的高预测精度, 同时减少聚合时间消耗.

数据划分: 为了比较算法在平衡和不平衡数据上的性能差异, 实验采用了两种不同的数据划分方式, 分别应用于两种数据集. 不平衡数据由于存在数据规模和特征提取的问题, 通常会导致联邦模型的性能下降. 从图 6 可以看出, 对于 MNIST 数据集, 无论是平衡还是不平衡的数据划分, 算法的预测精度都稳定在 0.85 左右. 对于 Fashion-MNIST 数据集, 不平衡数据划分虽然使得算法的预测精度降低, 约为 0.05, 但并没有下降太多. 这说明 FBL-LD 算法具有较强的鲁棒性, 能够在不同的数据划分方式下都达到较好的预测效果.



(a) MNIST 数据集 (b) Fashion-MNIST 数据集

图 4 不同客户端接受率的预测精度

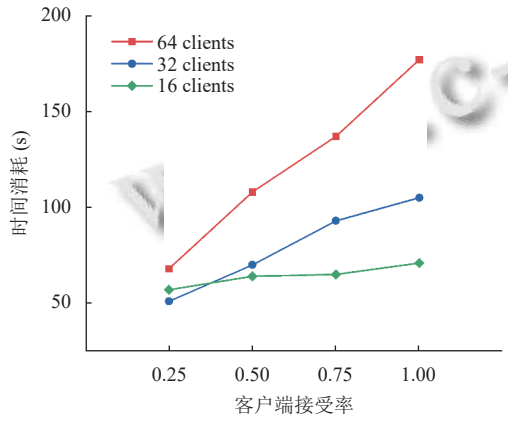


图 5 不同客户端接受率的时间消耗

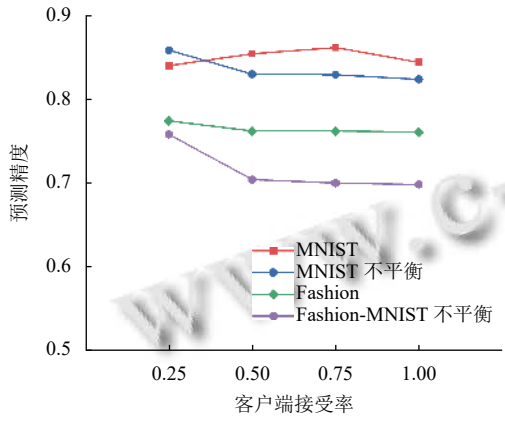


图 6 不同数据划分的 FBL-LD 预测精度

预测精度比较: 使用 FedAvg, FBL-ET, FCL-BL 和本文算法 FBL-LD 进行比较. 其中 FedAvg 使用了 10 个局部历时的 SGD. 图 7 的实验结果显示, 在一轮通信的环境下, FBL-LD 的预测精度要高于其他算法.

通信消耗: 通信消耗的计算方法与文献[31]中的方法类似, 都是根据客户端上传至服务器的模型权重来计算. 上传模型的大小直接影响通信资源的消耗, 是评价算法优劣的重要指标之一. FBL-ET 和 FCL-BL 算法都致力于压缩模型的大小, 以降低通信开销. 而 FBL-LD 算法由于主要在服务器端执行, 因此只需要客户端上传轻量级的宽度网络模型. 在表 1 中显示了模型的大小. 无论是 MNIST 还是 Fashion-MNIST 数据集, FBL-LD 传输的模型都最小, 而且 FBL-LD 只需要一次通信. 所以, FBL-LD 的通信消耗较少.

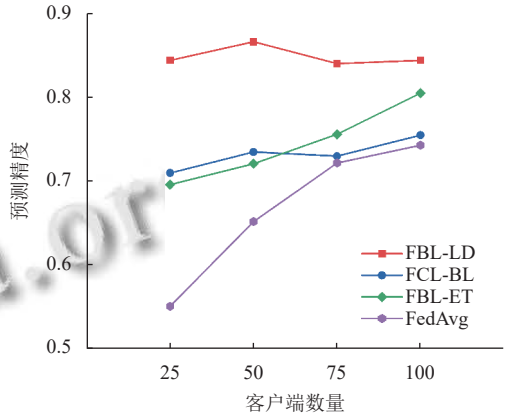


图 7 FBL-LD 与其他算法的性能对比

表 1 光扰检测与光扰消除实验中相关参数设置对照

算法	MNIST	Fashion-MNIST
FBL-LD	**10 000**	**10 000**
FBL-ET	13 000	13 000
FCL-BL	31 000	31 000
FedAvg	21 750	800 000

时间消耗: 将 FBL-ET 与 FCL-BL 和 FedAvg 与本文算法 FBL-LD 进行比较. 根据图 8, 在客户端使用 BL 进行训练可以大大减少时间消耗. FBL-LD 的时间消耗比 FCL-BL 少, 但预测精度更高.

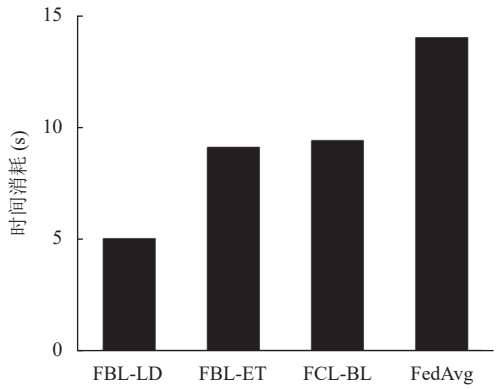


图 8 FBL-LD 与其他算法的时间消耗对比

4 结论

本文提出了一种基于宽度网络架构的单模型主导联邦学习框架 FBL-LD. FBL-LD 具有简洁高效的通信方式, 客户端可以随时加入或退出当前的联邦训练, 并仅通过一轮通信即可获得较好的联邦模型. 此外, 算法将模型参数的迭代从多次通信中解放, 只需在服务器上通过主导聚合策略动态调整直到收敛. 实验结果显示了 FBL-LD 在不同层数量, 客户端数, 客户端接受率的条件下预测准确性的变化, 并且在不同的数据集和不同的数据分布上具有鲁棒性. 与基线算法在相同条件下的对比, 证明了算法在预测性能以及时间、通信消耗方面的优势.

参考文献

1 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. Fort Lauderdale: PMLR, 2017. 1273–1282.

2 Meng Q, Zhou F, Ren HN, *et al*. Improving federated learning face recognition via privacy-agnostic clusters. Proceedings of the 10th International Conference on Learning Representations. OpenReview.net, 2022.

3 Yang ZH, Chen MZ, Wong KK, *et al*. Federated learning for 6G: Applications, challenges, and opportunities. Engineering, 2022, 8: 33–41. [doi: 10.1016/j.eng.2021.12.002]

4 Du ZY, Wu C, Yoshinaga T, *et al*. Federated learning for vehicular Internet of Things: Recent advances and open issues. IEEE Open Journal of the Computer Society, 2020, 1: 45–61. [doi: 10.1109/OJCS.2020.2992630]

5 Abreha HG, Hayajneh M, Serhani MA. Federated learning in edge computing: A systematic survey. Sensors, 2022, 22(2): 450. [doi: 10.3390/s22020450]

6 Chen CLP, Liu ZL. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(1): 10–24. [doi: 10.1109/TNNLS.2017.2716952]

7 Shokri R, Shmatikov V. Privacy-preserving deep learning. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver: ACM, 2015. 1310–1321. [doi: 10.1145/2810103.2813687]

8 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]

9 Yang Q, Liu Y, Chen TJ, *et al*. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1–19. [doi: 10.1145/3298981]

10 Konečný J, McMahan HB, Ramage D, *et al*. Federated optimization: Distributed machine learning for on-device intelligence. arXiv:1610.02527, 2016.

11 Hsu TMH, Qi H, Brown M. Federated visual classification with real-world data distribution. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 76–92. [doi: 10.1007/978-3-030-58607-2_5]

12 Augenstein S, McMahan HB, Ramage D, *et al*. Generative models for effective ML on private, decentralized datasets. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020.

13 Hsieh K, Phanishayee A, Mutlu O, *et al*. The Non-IID data quagmire of decentralized machine learning. Proceedings of the 37th International Conference on Machine Learning. JMLR.org, 2020. 408.

14 Mothukuri V, Parizi RM, Pouriyeh S, *et al*. A survey on security and privacy of federated learning. Future Generations Computer Systems, 2021, 115: 619–640. [doi: 10.1016/j.future.2020.10.007]

15 Bonawitz K, Ivanov V, Kreuter B, *et al*. Practical secure aggregation for privacy-preserving machine learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas: ACM, 2017. 1175–1191. [doi: 10.1145/3133956.3133982]

16 Li T, Hu SY, Beirami A, *et al*. Ditto: Fair and robust federated learning through personalization. Proceedings of the 38th International Conference on Machine Learning.

PMLR, 2021.
17 Li T, Sanjabi M, Beirami A, *et al*. Fair resource allocation in federated learning. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020.
18 Mohri M, Sivek G, Suresh AT. Agnostic federated learning. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019.
19 Hu ZO, Shaloudegi K, Zhang GJ, *et al*. Federated learning meets multi-objective optimization. arXiv:2006.11489, 2020.
20 Smith V, Chiang CK, Sanjabi M, *et al*. Federated multi-task learning. arXiv:1705.10467, 2017.
21 Lin YJ, Han S, Mao HZ, *et al*. Deep gradient compression: Reducing the communication bandwidth for distributed training. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018.
22 田金箫. 提升联邦学习通信效率的梯度压缩算法. 计算机系统应用, 2022, 31(10): 199–205. [doi: 10.15888/j.cnki.csa.008748]
23 Li T, Sahu AK, Zaheer M, *et al*. Federated optimization in heterogeneous networks. Proceedings of the 2020 Machine Learning and Systems 2020. Austin: MLSys.org, 2020.
24 Jiang YH, Konečný J, Rush K, *et al*. Improving federated learning personalization via model agnostic meta learning. arXiv:1909.12488, 2019.
25 Xie C, Koyejo S, Gupta I. Asynchronous federated optimization. arXiv:1903.03934, 2019.
26 Acar DAE, Zhao Y, Navarro RM, *et al*. Federated learning based on dynamic regularization. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
27 Pao YH, Park GH, Sobajic DJ. Learning and generalization characteristics of the random vector functional-link net. Neurocomputing, 1994, 6(2): 163–180. [doi: 10.1016/0925-2312(94)90053-1]
28 卢海鹏, 韩莹, 张凯, 等. 基于 VMD-BiLSTM-BLS 模型的短时交通流预测. 计算机系统应用, 2022, 31(5): 238–245. [doi: 10.15888/j.cnki.csa.008469]
29 Lei CY, Chen CLP, Guo JF, *et al*. ConvBLS: An effective and efficient incremental convolutional broad learning system for image classification. arXiv:2304.00219, 2023.
30 Yuan C, Ren CE. Domain adaptation broad learning system based on locally linear embedding. arXiv:2106.14367, 2021.
31 Le JQ, Lei XY, Mu NK, *et al*. Federated continuous learning with broad network architecture. IEEE Transactions on Cybernetics, 2021, 51(8): 3874–3888. [doi: 10.1109/tcyb.2021.3090260]
32 Ren CE, An RQ, Xuan ZH. FBL-ET: A federated broad learning framework based on event trigger. Knowledge-based Systems, 2023, 265: 110366. [doi: 10.1016/j.knosys.2023.110366]
33 Ge JR, Wang XJ, Li FY, *et al*. A broad learning system based on the idea of vertical federated learning. Proceedings of the 4th International Conference on Machine Learning for Cyber Security. Guangzhou: Springer, 2023. 565–574. [doi: 10.1007/978-3-031-20099-1_47]
34 Deng TP, Huang YQ, Shi ZW, *et al*. FedDBL: Communication and data efficient federated deep-broad learning for histopathological tissue classification. arXiv: 2302.12662, 2023.
35 Yuan XM, Chen JH, Zhang N, *et al*. A federated bidirectional connection broad learning scheme for secure data sharing in Internet of vehicles. China Communications, 2021, 18(7): 117–133. [doi: 10.23919/jcc.2021.07.010]

(校对责编: 牛欣悦)