

面向分层联邦学习的传输优化研究

邹赛兰, 李卓, 陈昕

引用本文

邹赛兰, 李卓, 陈昕. 面向分层联邦学习的传输优化研究[J]. 计算机科学, 2022, 49(12): 5-16.

ZOU Sai-lan, LI Zhuo, CHEN Xin. [Study on Transmission Optimization for Hierarchical Federated Learning](#) [J]. Computer Science, 2022, 49(12): 5-16.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于迁移学习和模型压缩的玉米病害识别方法](#)

Identification Method of Maize Disease Based on Transfer Learning and Model Compression
计算机科学, 2022, 49(11A): 211200009-6. <https://doi.org/10.11896/jsjcx.211200009>

[基于知识蒸馏模型ELECTRA-base-BiLSTM的文本分类](#)

Text Classification Based on Knowledge Distillation Model ELECTRA-base-BiLSTM
计算机科学, 2022, 49(11A): 211200181-6. <https://doi.org/10.11896/jsjcx.211200181>

[基于多阶段多生成对抗网络的互学习知识蒸馏方法](#)

Mutual Learning Knowledge Distillation Based on Multi-stage Multi-generative Adversarial Network
计算机科学, 2022, 49(10): 169-175. <https://doi.org/10.11896/jsjcx.210800250>

[基于YOLOv4的目标检测知识蒸馏算法研究](#)

Study on Knowledge Distillation of Target Detection Algorithm Based on YOLOv4
计算机科学, 2022, 49(6A): 337-344. <https://doi.org/10.11896/jsjcx.210600204>

[基于无标签知识蒸馏的人脸识别模型的压缩算法](#)

Compression Algorithm of Face Recognition Model Based on Unlabeled Knowledge Distillation
计算机科学, 2022, 49(6): 245-253. <https://doi.org/10.11896/jsjcx.210400023>

面向分层联邦学习的传输优化研究

邹赛兰^{1,2} 李卓^{1,2} 陈昕²

1 网络文化与数字传播北京市重点实验室(北京信息科技大学) 北京 100101

2 北京信息科技大学计算机学院 北京 100101

(zousailan@163.com)

摘要 与传统机器学习相比,联邦学习有效解决了用户数据隐私和安全保护等问题,但是海量节点与云服务器间进行大量模型交换,会产生较高的通信成本,因此基于云-边-端的分层联邦学习受到了越来越多的重视。在分层联邦学习中,移动节点之间可采用 D2D、机会通信等方式进行模型协作训练,边缘服务器执行局部模型聚合,云服务器执行全局模型聚合。为了提升模型的收敛速率,研究人员对面向分层联邦学习的网络传输优化技术展开了研究。文中介绍了分层联邦学习的概念及算法原理,总结了引起网络通信开销的关键挑战,归纳分析了选择合适节点、增强本地计算、减少本地模型更新上传数、压缩模型更新、分散训练和面向参数聚合传输这 6 种网络传输优化方法。最后,总结并探讨了未来的研究方向。

关键词: 分层联邦学习;传输优化;通信开销;节点选择;模型压缩

中图法分类号 TP393

Study on Transmission Optimization for Hierarchical Federated Learning

ZOU Sai-lan^{1,2}, LI Zhuo^{1,2} and CHEN Xin²

1 Beijing Key Laboratory of Internet Culture and Digital Dissemination Research(Beijing Information Science & Technology University), Beijing 100101, China

2 School of Computer Science, Beijing Information Science & Technology University, Beijing 100101, China

Abstract Compared with traditional machine learning, federated learning effectively solves the problems of user data privacy and security protection, but a large number of model exchanges between massive nodes and cloud servers will produce high communication costs. Therefore, cloud-edge-side layered federated learning has received more and more attention. In hierarchical federated learning, D2D and opportunity communication can be used for model cooperation training among mobile nodes. Edge server performs local model aggregation, while cloud server performs global model aggregation. In order to improve the convergence rate of the model, the network transmission optimization technique for hierarchical federated learning is studied. This paper introduces the concept and algorithm principle of hierarchical federated learning, summarizes the key challenges that cause network communication overhead, summarizes and analyzes six network transmission optimization methods, such as selecting appropriate nodes, enhancing local computing, reducing the upload number of local model updates, compressing model updates decentralized training and parameter aggregation oriented transmission. Finally, the future research direction is summarized and discussed.

Keywords Hierarchical federated learning, Transmission optimization, Communication overhead, Node selection, Model compression

1 引言

随着机器学习的兴起和智能移动设备的普及,近年来在无线边缘网络中进行机器学习逐渐引起了研究人员的广泛关注。传统的机器学习中用户需要将本地数据上传至计算能力强大的云服务器以进行集中模型训练,包含一个主云和若干收集数据的参与节点,如图 1 所示。在这种场景下,大量的数据上传过程会产生大量的能耗和通信时延,同时对于隐私敏

感的参与节点也存在着隐私泄露的风险。针对此问题,研究人员针对分布式机器学习展开研究,谷歌于 2017 年首次提出了联邦学习^[1](Federated Learning, FL)技术,之后联邦学习在学术界一直受到极大的关注^[2]。

联邦学习在参与节点的私有数据集上进行机器学习模型训练,并将本地模型参数而非本地数据上传至云服务器以进行全局模型聚合,不必依赖云服务器来集中培训机器学习模型,形成了一个星型拓扑的网络通信结构,有效地解决了敏感

到稿日期:2022-03-21 返修日期:2022-07-06

基金项目:国家自然科学基金(61872044);北京市青年拔尖人才项目

This work was supported by the National Natural Science Foundation of China(61872044)and Beijing Municipal Program for Top Talent.

通信作者:李卓(lizhuo@bistu.edu.cn)

用户的数据隐私保护和数据共享需求之间的矛盾,具体工作流程如图2所示。联邦学习中存在着两种重要的实体,即产生和收集数据的参与节点、进行全局模型聚合的云服务器。参与节点使用本地数据来训练本地模型并将其上传给云服务器,云服务器接收上传的本地模型进行全局聚合以得到全局模型,经过多轮次迭代后得到使整个模型训练损失函数最小的模型 w 。联邦学习模型训练的一次迭代过程如下(t 表示进行的是第 t 次迭代过程)。

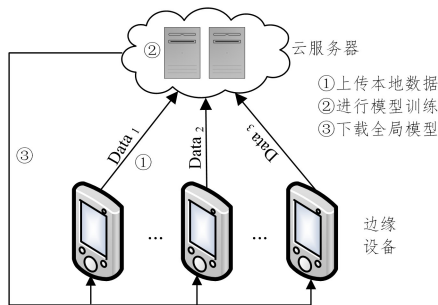


图1 传统的机器学习架构

Fig. 1 Architecture of traditional machine learning

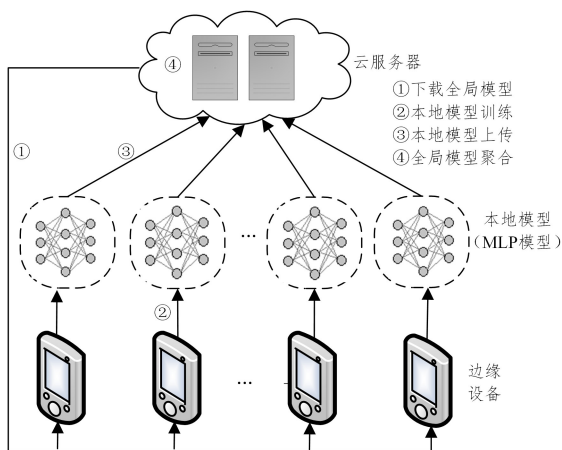


图2 联邦学习架构

Fig. 2 Architecture of federated learning

(1)参与节点从云服务器下载全局模型 $w(t-1)$ 。

(2)参与节点 m 使用本地数据进行本地模型训练,以得到本地模型更新 $w_m(t)$ 。

(3)各参与节点将本地模型更新上传至云服务器。

(4)云服务器将接收的本地模型进行加权平均聚合,以得到全局模型 $w(t)$ 。

从以上迭代过程可以看出,联邦学习中存在着两个关键性挑战:1)部分云服务器可能会根据参与节点上传的模型参数推测出其本地数据,从而泄露用户的数据隐私;2)云服务器与参与节点之间需要进行大量的通信轮次以交换模型参数。特别是在大规模网络上训练和管理数据密集、延迟敏感的机器学习任务时,传统的联邦学习并不适用。为了解决以上问题,近两年研究人员提出了分层联邦学习^[3-5]。

与传统的两层联邦学习不同,分层联邦学习将边缘计算应用到联邦学习中,引入边缘服务器作为参与节点和云服务器之间通信的中介。边缘服务器接收参与节点的本地模型参数并进行边缘模型聚合,再将边缘聚合模型上传至云服务器

以进行全局模型聚合,减少了上行链路数据的传输量。但是,节点的移动性、通信链路的不稳定性也为分层联邦学习模型训练带来了挑战。同时,自联邦学习被提出以来,研究人员对其中存在的隐私安全、信任与激励机制、网络传输效率等问题展开了一系列研究。

为了解决用户隐私安全问题,学术界通过大量研究来增强隐私安全。文献[6-8]分别对移动边缘网络中的用户数据及隐私保护问题进行了讨论。文献[9-11]表明,恶意服务器可以通过参与节点上传的梯度信息间接推出本地数据集信息,从而泄露用户的隐私数据。Li等^[12]研究了基于差分隐私、同态加密及安全多方计算的用户隐私保护方法。类似地,Shi等^[13]提出了一种基于局部差分隐私(LDP)理论的隐私保护方案,以实现客户端和边缘服务器层级的严格差分隐私保护。文献[14]提出了一种差分隐私异步联邦学习(DPAFL)算法,在本地模型更新过程中引入高斯噪声来保护隐私。而Lu等^[15]将该DPAFL算法应用到车联网系统,以解决敏感数据泄露的问题。文献[16]将联邦学习和隐私保护技术应用于医学领域,并展开相关研究。悉尼科技大学、澳大利亚迪肯大学和中国的西安电子科技大学、北京邮电大学、电子科技大学等知名高校,均对存在的隐私安全问题展开了相关研究。以上研究均聚焦于联邦学习的安全与隐私保护这一研究领域,将这些隐私保护技术应用到联邦学习中能够提供强大的安全性,但同时也存在着一些问题。例如,结合联邦学习和同态加密、安全多方计算技术,能够增强隐私保护,但同时也可能造成较大的模型训练通信传输开销^[17-19]。

为了解决信任与激励机制问题,研究人员通过大量研究来选择信誉值高的优质用户及激励更多用户参与模型训练。针对网络中执行环境不可信及用户恶意攻击等信任问题,文献[20]利用多层和多链结构存储大量参与节点的本地数据,以减小网络环境不可信和恶意攻击带来的影响,但是没有考虑用户激励机制设计。文献[21]提出了一种异步联邦学习方案,并采用混合区块链架构实现共享数据和模型训练的可靠性。针对模型训练中激励机制缺乏的问题,文献[22]将参与者和模型所有者之间的交互过程建模为Stackelberg博弈。Zhan等^[23]将深度强化学习应用到激励机制设计中,根据边缘节点的参与程度和历史贡献来决定报酬。Deng等^[24]提出了基于拍卖机制激励质量感知的学习质量最大化问题。文献[25]则针对cross-silo联邦学习设计激励机制,在满足个体理性和预算平衡的约束下,保证社会福利最大化。而文献[26]采用区块链技术设计模型训练场景,以激励更多用户参与训练。内容分发激励机制设计在INFOCOM,WCNC,TWC,JSAC等国际知名学术期刊会议上是热点研究问题。南洋理工大学、麻省理工学院和国内的清华大学、上海交通大学、西安电子科技大学等著名院校都对该领域展开了深入的研究和探索。

边缘网络中参与模型训练的用户通常有成千上百个,大量节点与服务器之间需要不断地进行模型更新传输,使得网络传输效率成为了模型训练收敛的关键瓶颈^[27]。在整个模型训练过程中,大量参与节点的本地模型更新及模型更新上传会造成网络传输开销过大,而且上述提到的隐私安全研究领域应用的同态加密和安全多方计算等隐私技术都会引起模

型训练的网络传输开销较大。此外,当训练模型较为复杂时,激励机制研究中采用区块链存储的模型规模过大,也会导致模型参数的加密和传输过程耗费大量时间。因此,研究网络传输优化技术以降低模型训练中的网络传输开销,在联邦学习研究中是十分必要的。

本文深入分析面向分层联邦学习的网络传输优化技术,对相关研究的进展和现状进行了讨论,并按照发表时间对主要的传输优化研究进展进行了整理(见图3),最后提出未来

的研究方向。分层联邦学习在人工智能领域的顶级会议 IJ-CAI, AAAI, NeurIPS 等上均有相关论文讨论,并且组织有专门的 workshop 进行研讨;面向分层联邦学习的网络优化在网络领域的顶级会议 IEEE Infocom 和顶级期刊 IEEE JSAC, IEEE TPDS 等上均有相关论文讨论。美国斯坦福大学、普渡大学、德克萨斯大学和中国的清华大学、中南大学、北京邮电大学、香港科技大学、香港理工大学等知名高校,均对面向分层联邦学习的网络优化开展了相关研究。

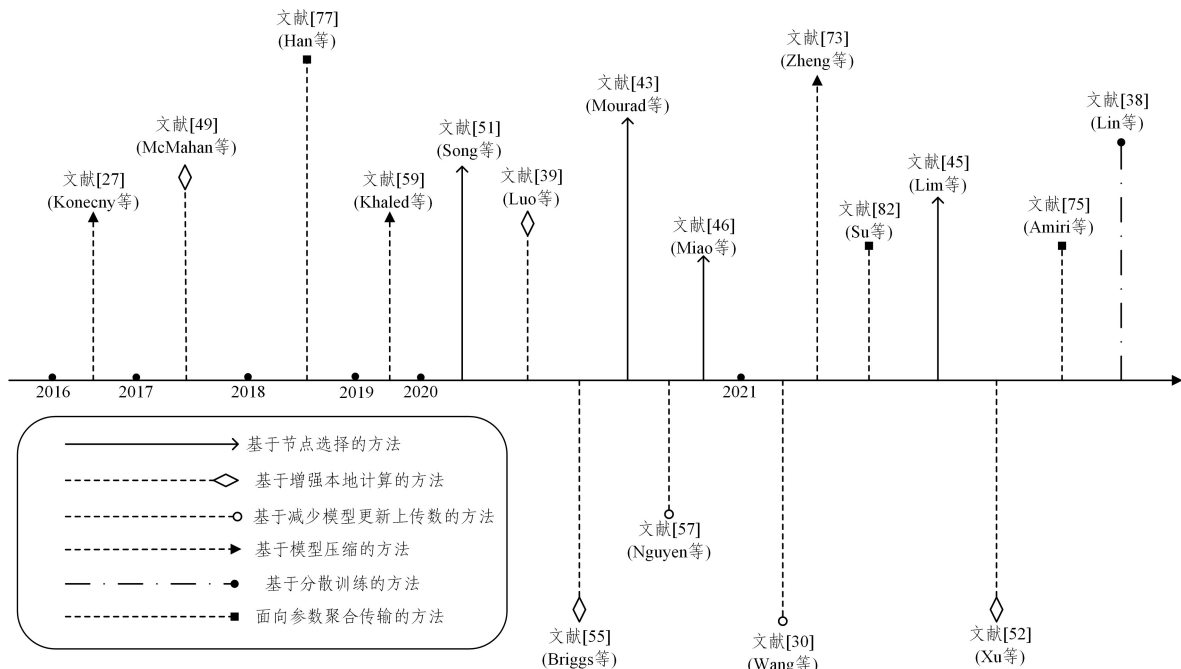


图3 传输优化方法

Fig. 3 Transmission optimization methods

本文第2节概述了分层联邦学习相关知识;第3节分析了分层联邦学习中引起网络传输瓶颈的关键挑战;第4节针对现有主要研究成果进行了讨论分析;第5节对网络传输优化研究的未来发展方向进行了展望;最后总结全文。

2 分层联邦学习的基本概念

2.1 分层联邦学习

在传统的两层联邦学习中,参与节点使用本地数据进行本地模型训练,并将其上传给云服务器以进行全局模型聚合,整个模型训练的目标函数通常是求得使损失函数达到最小值的全局模型 w 。使损失函数最小的目标函数通常定义为:

$$\min_w F(w) \quad (1)$$

$$F(w) = \frac{\sum_{m=1}^M |D_m| F_m(w)}{|D|} \quad (2)$$

$$F_m(w) = \frac{\sum_{j=1}^{|D_m|} f_j(w)}{|D_m|} \quad (3)$$

其中, $D = \{x_j, y_j\}_{j=1}^{|D|}$ 为总训练数据集, $|D|$ 为所有本地参与节点数据量总和, x_j 为第 j 次输入样本, y_j 为第 j 次输入样本对应的输出标签, M 为参与模型训练的节点总数。用 $F_m(w)$ 表示参与节点 m 的本地损失函数, $f_j(w)$ 表示具有参数 w 的

模型对数据集 D_m 中的实例 (x_j, y_j, w) 产生的损失函数。

在联邦学习系统中,参与节点通常采用随机梯度下降(SGD)算法来进行本地模型更新, η 表示学习率,则第 t 轮迭代中本地模型的更新如下:

$$w_m(t) = w_m(t-1) - \eta \nabla F_m(w_m(t-1)) \quad (4)$$

第 t 轮迭代中云服务器的全局模型的聚合如下:

$$w_t = \frac{\sum_{m=1}^M |D_m| w_m(t)}{|D|} \quad (5)$$

根据上述联邦学习模型训练过程,参与节点和云服务器之间需要交换大量的模型更新,这需要大量的网络资源并导致上行链路的通信成本增加。分层联邦学习将边缘计算应用到联邦学习中,利用边缘服务器的计算、传输能力进行部分模型聚合,减少上行链路传输的数据量,从而有效解决上述问题。

本节考虑典型的云-边-端三层分层联邦学习场景,其工作流程如图4所示。分层联邦学习中存在着3个主要的实体:参与节点、边缘服务器、云服务器。参与节点产生或收集数据,每个节点都有自己的私有数据集;边缘服务器接收参与节点上传的本地模型参数并进行边缘模型聚合;云服务器存储原始模型并与所有参与节点共享全局模型,接收边缘服务器上传的边缘模型以进行全局模型聚合。

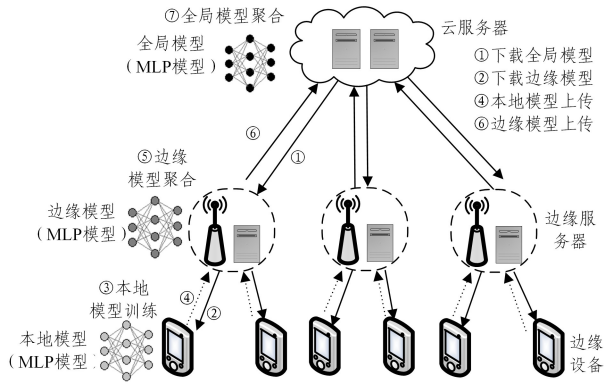


图4 分层联邦学习架构

Fig. 4 Architecture of hierarchical federated learning

分层联邦学习的一次全局迭代过程如下。

(1)初始化:云服务器将初始全局模型广播给边缘服务器及参与训练的移动节点。

(2)本地模型计算:参与节点使用本地数据集进行本地模型训练,并将本地模型更新上传至边缘服务器。

(3)边缘模型聚合:边缘服务器接收上传的本地模型更新并进行边缘模型聚合,然后将边缘聚合模型上传至云服务器。

(4)全局模型聚合:云服务器接收上传的边缘聚合模型并进行全局模型聚合;然后将更新后的全局模型共享给边缘服务器及边缘节点以进行下一轮次的模型训练。

2.2 分层联邦学习算法的原理

边缘设备使用本地数据集 D_m^k 进行本地模型计算,其损失函数及本地模型参数更新可表示为:

$$F_m^k(w_m^k(t)) = \frac{\sum_{j=1}^{|D_m^k|} f_j(w_m^k(t))}{|D_m^k|} \quad (6)$$

$$w_m^k(t) = w_m^k(t-1) - \eta \nabla F_m^k(w_m^k(t-1)), \eta > 0 \quad (7)$$

其中, $\mathcal{K} = \{k: k=1, 2, \dots, K\}$ 为边缘服务器集合, $\mathcal{M} = \{m: m=1, 2, \dots, M\}$ 为每个边缘服务器 k 下的边缘设备集。边缘设备将本地模型更新上传至边缘服务器 k ,边缘服务器将收到的本地模型进行平均以得到边缘聚合模型,其中 $\sum_{m=1}^M |D_m^k| = |D_k|$ 为边缘服务器 k 上的数据量,可表示为:

$$w_k(t) = \frac{\sum_{m=1}^M |D_m^k| w_m^k(t)}{\sum_{m=1}^M |D_m^k|} = \frac{\sum_{m=1}^M |D_m^k| w_m^k(t)}{|D_k|} \quad (8)$$

边缘服务器将边缘聚合模型上传至云服务器,云服务器将收到的边缘模型进行平均以得到全局模型,其中 $\sum_{k=1}^K |D_k| = \sum_{k=1}^K \sum_{m=1}^M |D_m^k| = |D|$ 为模型训练的总数据量,可表示为:

$$w(t) = \frac{\sum_{k=1}^K \sum_{m=1}^M |D_m^k| w_k(t)}{|D|} = \frac{\sum_{k=1}^K |D_k| w_k(t)}{|D|} \quad (9)$$

从上文中的分层联邦学习的工作流程和算法原理可以看出,模型参数以上传、下载的方式在参与节点、边缘服务器、云服务器之间进行多轮次传输,通信时延、通信轮次数、单一通信轮传输的数据量较高,会造成较大的通信开销。分层联邦学习越来越多地部署在通信带宽有限的移动设备上,当参与模型训练的边缘节点较多和通信链路不稳定时,会降低通信

效率,从而给分层联邦学习模型训练带来挑战。

3 分层联邦学习网络传输挑战

(1)海量的参与节点。当分层联邦学习系统中拥有大量的参与节点时,机器学习模型可以在更多的本地数据集上进行训练,从而提升学习性能和模型准确度,这是大量参与节点进行模型训练的优势^[28]。然而,大量的参与节点在与边缘服务器共享本地模型更新时,会增加参与节点与边缘服务器之间的通信轮次,导致整个模型训练的通信开销较大。在互联网环境中,大量参与节点的本地模型更新和上传可能导致模型训练的总计算成本增加和服务器通信开销过大,从而无法满足应用的正常需求^[29-31]。

(2)网络带宽受限。现在的分层联邦学习实际应用场景大多都部署在不确定的无线环境(动态信道和干扰)和有限的无线资源(带宽和功率)上^[32],且边缘网络分布着计算和通信资源受限的边缘节点。在参与节点与边缘服务器、边缘服务器与云服务器的通信过程中,参与节点和边缘服务器并不总是具有所需的网络带宽。因网络带宽分配不同而引起的上行传输速度和下行传输速度之间的差异会导致移动节点向边缘服务器、边缘服务器向云服务器上传模型的延迟,从而引起分层联邦学习模型训练的通信瓶颈。为了提高通信效率,进一步保持通信带宽仍然是十分重要的^[29]。

(3)数据异构性。模型训练中引起网络传输瓶颈的另一个挑战是数据异构性^[33]。分层联邦学习中,参与节点使用本地数据进行本地模型计算和更新。当不同参与节点收集的本地数据是非独立同分布时,每个参与模型训练的移动节点的数据分布不能代表其他移动节点的数据总体分布,而且节点间收集和分布的数据大小通常会有很大差异^[34]。本地数据较大的移动节点可能需要较长时间来进行本地模型更新,而本地数据较小的移动节点则需要很短的时间来完成本地模型更新。完成本地模型更新花费较短时间的那些节点需要等待本地模型更新完成较慢的参与节点,即待最慢的边缘节点完成本地模型更新之后,才会同步将本地模型更新上传至边缘服务器以进行边缘模型聚合。因此,节点的本地数据大小不同可能会导致节点间本地模型更新的延迟,从而造成模型训练过程难以收敛和网络传输轮次过多等问题^[1,27]。

(4)计算能力异构。在分层联邦学习的实际应用场景中,本地模型计算依赖于参与节点,而不是强大的GPU和CPU^[28]。同时,部署在网络边缘的参与节点可能在计算、存储和链路带宽^[35]方面有限制。不同边缘节点所具有的计算能力可能存在着很大的不同,具有较强计算能力的边缘节点在进行本地模型更新时所耗费的时间更短,而计算能力较弱的节点可能需要耗费较长时间才能完成本地模型更新。因此,执行单个本地更新所需的时间因节点而异^[36],完成本地计算较快的节点需要等待本地计算缓慢的节点完成本地训练后,才能与边缘服务器进行模型传输,从而增加模型训练的通信成本。

(5)节点协作的自私性。在有限的网络带宽资源和节点计算能力约束下,资源受限节点与资源丰富节点间可进行协作通信来节省功率和带宽^[37],即资源受限的参与节点可将本地

数据或模型参数传递给资源丰富的参与节点以协作进行本地训练和模型上传。然而,协作训练会带来通信开销,存在自私节点不愿参与协作训练并贡献其通信和计算资源^[38],从而增加模型训练延迟,影响整体网络的性能。

(6)参与节点与边缘服务器的边缘关联问题。在分层联邦学习系统中,密集分布的参与节点通常能够与多个边缘服务器通信,同时边缘服务器希望与尽可能多的参与节点通信进行边缘模型聚合,以提高学习精度。但当更多的参与节点都选择与同一个边缘服务器通信时,每个参与节点分得的通信资源会变少,从而导致更长的通信延迟^[39]。因此,参与节点的计算和通信资源分配以及它们的边缘关联问题也会引起整个模型训练通信延迟的增加^[40]。

针对以上分层联邦学习中存在的传输挑战,通常可以通过减少通信总轮次/降低通信频率、减少单一通信轮中需要传输的数据量等方式进行优化,以提高通信效率。

4 分层联邦学习传输优化技术

本节将着重介绍分层联邦学习中有关传输优化的已有研究和进展,将其按照方法原理进行分类梳理,并列举出关键优化技术(见图5),包括选择合适节点、增强本地计算、减少模型更新上传数、压缩模型更新、分散训练、面向参数聚合传输等。

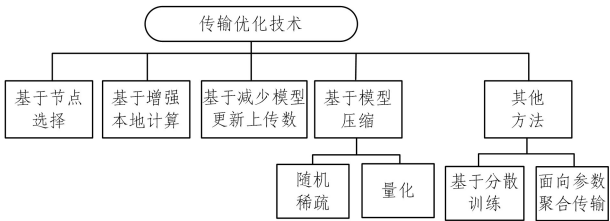


图5 典型的传输优化技术

Fig. 5 Typical transmission optimization techniques

4.1 基于节点选择的优化方法

在许多机器学习任务中,边缘节点的数量通常在数千到数十亿之间,且不同边缘节点的计算和通信能力具有异质性和有限性,可能会导致一些节点在模型训练时遇到障碍,导致大量通信轮次无效。因此,参与模型训练的节点选择至关重要。分层联邦学习中的边缘节点选择是通过限制参与模型训练的节点数量及选择更加可靠、更加强大的节点参与模型训练来降低模型通信成本,从而减少网络通信开销。

为了保证在一定时间内模型收敛且达到一定的准确性,Song等^[41]考虑计算和通信之间的权衡,研究了在有限时间

范围内基于边缘计算的分布式机器学习中保证收敛和达到一定模型精度所需的最优边缘节点参与数量。文献[42]提出了一种轻量级节点选择算法,在考虑异构网络下节点的本地计算和通信资源的同时,迭代选择边缘节点参与模型训练,来有效地执行学习任务。而文献[43]提出了一种客户端选择算法,在考虑客户端的计算和通信资源有限性的同时,也实现了最大化每轮次模型训练的客户端数量。通过考虑客户端的CPU、内存、能量及时间等资源,来预测它们是否能够执行并完成模型训练任务。经过实验对比,该方法在减少通信轮次、最大化每轮模型训练的客户端数量及达到预期模型精度方面具有很大的有效性。

文献[39]提出了分层联邦边缘学习框架(Hierarchical Federated Edge Learning, HFEL),并在该框架下制定了一个全局学习成本最小化的计算和通信资源分配及边缘关联的优化问题;并将其分解为了两个子问题:为每个边缘服务器分配一组预定设备的资源分配,及服务器上节点的边缘关联;利用单个边缘服务器下一组的资源分配子问题的最优策略,通过迭代的全局成本降低调整过程,直至收敛到一个稳定的系统点。该框架实现了更高的全局模型精度并降低了模型训练的开销和损失。

上述有关研究都集中在处于网络边缘的节点选择上。文献[44]提出了一种新颖的半异步分层联邦学习 SHEL 框架。该框架在考虑边缘节点关联和资源分配问题的同时,还选择具有最大数据重要性的部分边缘服务器来参与模型训练,以最大化通信效率。

在分层联邦学习中边缘服务器没有动机与模型所有者共享资源,并且现有的边缘关联和资源分配都侧重于静态的方法。文献[45]提出了分层的博弈框架来研究边缘关联与资源分配的动态问题。在两层的博弈中,低层采用演化博弈方法,高层采用 stackelberg 差分博弈,边缘服务器给定带宽分配策略,而模型所有者决定最优的报酬方案。文献[46-47]提出了一种群体感知网络分层激励机制框架,使用契约理论和联邦博弈论方法,根据模型所有者的边际贡献来奖励它们,以获得存在信息不对称的不同模型所有者类型的高质量数据。而 Lim等^[48]提出了基于演化博弈与拍卖的激励机制和资源分配框架,采用应用演化博弈论来模拟集群选择过程,采用基于深度学习的拍卖机制,在满足个体理性和激励相容约束的前提下保证卖家收益最大化。基于节点选择的主要研究对比如表1所列。

表1 基于节点选择的传输优化

Table 1 Transmission optimization based on node selection

主要文献	应用领域				针对挑战						优化目标		
	通用/ 理论研究	物联网	智能交通 系统	车联网	海量参与 节点	网络带宽 受限	数据 异构性	计算能力 异构	节点协作 的自私性	边缘关联 问题	提高 效率	降低 能耗	提高模型 精度
[41]	✓				✓	✓		✓			✓		✓
[42]		✓			✓	✓	✓	✓			✓		
[43]		✓			✓	✓	✓	✓			✓		✓
[39]		✓			✓	✓		✓			✓	✓	✓
[44]			✓		✓	✓				✓	✓		✓
[45]	✓				✓	✓			✓	✓	✓		
[46]		✓			✓				✓	✓	✓		
[47]				✓	✓				✓		✓		
[48]	✓				✓	✓			✓		✓		

分层联邦学习中与参与模型训练的节点进行上行链路和下行链路通信十分重要,必须与拥有可靠网络带宽和资源的节点进行通信。上面提到的相关节点选择技术有助于实现更高质量的全局训练模型,并有效地降低实现可靠的全局模型所需的总体通信成本。在选择合适的参与节点进行模型训练之后,可以通过增强节点的本地计算量,来降低参与节点与边缘服务器之间的通信频率,从而减小通信开销。

4.2 基于增强本地计算的优化方法

本地模型更新的总体目标是使用本地数据集在移动设备上进行机器学习模型训练,但是在单一的通信轮进行本地模型计算并将本地模型更新共享到边缘服务器上,会造成通信效率不高。针对这一问题,一种传输优化的方法就是增强本地计算以降低通信频率,从而减少模型训练所需的通信轮次。

文献[49-51]提出尽可能多地增加节点的本地计算,以减少模型训练所需的通信轮次。McMahan 等^[49]提出,增加节点在每轮迭代中的本地计算次数,即每个节点在提交模型更新前进行多次本地迭代。Yao 等^[50]采用了迁移学习中常用的双流模型来改进每个参与设备上的计算操作,以加快全局模型的收敛速度。文献[51]提出了分层联邦学习算法,边缘服务器汇聚从参与节点提交的局部模型,仅当边缘服务器汇聚若干次之后,才向云服务器提交一次全局模型更新,从而降低整体的通信开销。而文献[39]结合了上述两方面的方法,提出了一种新的分层联邦边缘学习(HFEL)框架。在该框架中,参与节点经过多次迭代的本地模型计算来达到本地模型精度,再将其上传给边缘服务器;边缘服务器经过多次迭代的边缘模型聚合来达到边缘模型精度,再将其上传给云服务器

以进行一次全局模型聚合。与该文献研究的整体思想类似,文献[52]也规定参与节点将本地模型更新上传至边缘服务器之前要进行多轮次的本地迭代,以减少与服务器之间的通信次数,且边缘聚合模型在上传至云服务器之前被进行多次边缘迭代,从而降低通信开销,减少模型训练时间和边缘设备的能耗。

然而,当模型训练任务较复杂,大量边缘节点参与模型训练时,其与边缘服务器、云服务器之间模型参数的频繁交换,可能会增加通信开销。文献[53]提出了一种基于周期平均和量化的算法,其周期平均特性使移动设备在上传本地模型更新之前进行多次本地迭代计算,而在服务器上只进行周期性平均,这样可以减少通信轮次,从而降低训练过程中的通信成本。

上述很多学者提出的通过增强本地计算来减少通信轮次的方法,在节点本地数据分布上考虑更多的是数据独立同分布(Independently Identically Distribution, IID)。但是文献[54]指出,当节点的本地数据为非独立同分布(Non-IID)时,可能会在设备本地模型更新中引入“漂移”,从而导致不稳定和缓慢的收敛,并针对此问题提出一种新的随机算法,通过使用控制变量(方差减少)来纠正本地更新中的“客户漂移”。考虑将模型训练扩展到大量的数据样本和集群上,文献[55]提出了一种为客户端子集训练专门模型的分层聚类联邦学习。首先根据客户端更新后的本地模型来判断客户端之间的相似度,然后采用分层聚类算法迭代合并最相似的客户端以进行聚类,从而使得在 Non-IID 设置下也能减少模型训练过程中达到模型收敛的通信轮数,提高模型测试的准确性。基于增强本地计算的主要研究对比如表 2 所列。

表 2 基于增强本地计算的传输优化
Table 2 Transmission optimization based on enhanced local computing

主要文献	应用领域		针对挑战		优化目标		特点:以计算代价换取通信开销
	通用/理论研究	物联网	网络带宽受限	数据异构性	提高效率	提高模型精度	
[49]	✓		✓		✓		✓
[50]		✓	✓		✓		✓
[52]		✓	✓		✓	✓	✓
[53]	✓		✓		✓	✓	✓
[54-55]	✓		✓	✓	✓	✓	✓

本地模型更新是分层联邦模型训练过程中至关重要的环节。每个参与节点都需要在生成的本地数据集上进行本地模型计算和更新,以便生成更好的全局模型。就目前已有的与本地更新技术相关的研究而言,通过提高本地更新效率,可以有效地提高分层联邦学习系统中的通信效率。但是这些已有研究都是通过牺牲本地计算代价换取通信开销,即增加节点的本地计算量或提高并行性来降低本地模型更新频率,从而减少模型训练所需的通信轮次。除了通过增强本地计算来提高通信效率的方法之外,还可以通过减少上传至边缘服务器的本地模型更新数来降低通信开销。

4.3 基于减少模型更新上传数的优化方法

在模型训练过程中,边缘节点下载全球模型之后,使用本地数据集进行本地模型训练并将本地模型更新上传给边缘服务器以进行边缘模型聚合。因此,选择更少但更加有效的模型更新可以减少因模型更新上传而引起的通信开销。

Tao 等^[56]针对模型参数的重要性差异,提出在每一轮的模型更新操作中仅选择部分重要梯度上传。类似地,Wang 等^[30]提出了一种框架,节点在每一轮的模型更新操作中仅上传相关的本地模型更新,且该相关性由比较本地模型更新与上一轮全局模型更新的差异来进行评价。通过避免向服务器上传无关的模型更新,可以保证整个模型训练在达到收敛的同时大大减小通信开销。

文献[57]提出,在每一轮模型训练中对参与节点进行智能采样,以达到预期的收敛速度。在考虑设备计算和通信异构性的同时,根据设备对模型更新贡献程度的估计来调整聚合,从而提高整个模型训练过程中的模型精度并加快收敛速度。文献[58]提出了一种节点最佳采样和 D2D 卸载优化相结合的方案。对模型训练做出最大贡献的设备进行采样,未被选择的设备将数据卸载给被采样设备。通过使用真实数据集,将提出的优化方案与随机采样、启发式采样进行对比,该

方法在资源利用、模型训练收敛时延及模型精度方面具有明显的优势。基于减少模型更新上传数的主要研究对比如表 3 所列。

表 3 基于减少模型更新上传数的传输优化

主要文献	应用领域		针对挑战		优化目标	
	通用/ 理论研究	物联网	网络带宽受限	计算能力异构	提高效率	提高模型精度
[56]		✓	✓		✓	✓
[57]	✓		✓	✓	✓	✓
[30,58]	✓		✓		✓	✓

上文提到的相关研究都是通过对本地模型更新进行操作来减少通信轮次,以计算代价换取通信代价,而每轮次通信回合中所传输的数据量没有变化。但是,在许多机器学习任务中,模型参数大小可能达到数十或数百兆字节,会产生较大的通信开销。因此可以采用基于模型压缩的优化方法,通过减少单一通信轮中所需传输的总比特数来提高通信效率。

4.4 基于模型压缩的优化方法

模型压缩即稀疏化,是一种通信技术,其中具体压缩方案可以是随机稀疏、量化等方法的一种或者组合方法。各种压缩方案可以在分层联邦学习的不同阶段执行,且模型压缩的目标包括:减少服务器广播给边缘节点的全局模型大小(下行链路);减少边缘节点上传到边缘服务器、边缘服务器上传到云服务器的模型更新大小(上行链路)。

为了减小下行链路的通信开销,文献[59]对服务器迭代广播至客户端的全局模型进行压缩,并分析了压缩操作对模型收敛的影响,通过在模型压缩过程中引入差错反馈机制^[34,60-61],有效地提升了模型的收敛速率。文献[62]将量化技术应用在云服务器广播给全部设备的全局模型上,即在全局模型广播和与设备共享之前将其量化,在提高通信效率方面具有很好的有效性。下面将分别从随机稀疏、量化两个角度对上行链路模型压缩方案的已有研究和进展进行阐述。

表 4 基于模型压缩的传输优化

Table 4 Transmission optimization based on model compression

主要文献	应用领域			压缩方式		压缩阶段		针对挑战			优化目标	
	通用/理论研究	物联网	工业物联网	随机稀疏	量化	上行链路压缩	下行链路压缩	网络带宽受限	数据异构性	边缘关联问题	提高效率	提高模型精度
[59]	✓				✓		✓	✓			✓	✓
[62]	✓				✓		✓	✓	✓		✓	✓
[27,63-64]	✓			✓		✓		✓			✓	
[65]		✓		✓		✓		✓			✓	
[66]	✓			✓		✓		✓			✓	
[67]	✓				✓	✓		✓		✓	✓	✓
[68]			✓		✓	✓		✓			✓	

现有的大部分压缩方法只在边缘节点的本地数据分布为 IID 的情况下有很好的效果,且通过减少每次通信回合中的传输数据量来降低网络通信开销,以降低模型精度来换取通信开销。但当参与模型训练的边缘节点数量非常大时,训练的本地模型参数也随之增多,压缩方法存在一定的限制。

4.5 基于分散训练的优化方法

传统的两层联邦学习中,网络通信拓扑通常是星型拓扑,这会导致边缘节点与云服务器间的通信开销太大。而分层联

(1)随机稀疏

为了减小上行链路的通信开销,文献[27,63-64]采用稀疏化等方法,对客户端的模型更新进行压缩,以减少单一通信轮中上传至服务器的数据量。而文献[65-66]考虑了优化通信效率时存在的隐私安全问题,并提出了相应的解决方案,以达到提高通信效率和保护数据隐私的平衡。文献[65]提出了一种基于 SDM-DSGD 的分布式随机梯度边缘学习方法。在该方法中,提出了一种交换计算节点之间的稀疏差分的广义微分编码(DSGD)更新,使模型以较小的梯度稀疏化传输概率收敛,并引入噪声来考虑数据隐私保护,从而建立了边缘学习中的数据隐私和通信效率的性能保障。文献[66]提出了一种新的机遇稀疏编码 DP 机制和服务器自适应更新的方案,以提高隐私精度和通信效率。在该方案中,将随机稀疏化和梯度扰动结合在每个代理上以增强隐私保护,且进一步引入加速技术来帮助降低隐私成本。

(2)量化

量化技术指将模型更新压缩到合适的大小,即减少每次更新中传输的总比特数,从而使通信更加有效。文献[53]提出每个本地模型更新在上行链路上均采用量化技术进行量化,从而降低整个学习过程的通信开销。文献[67]充分利用云服务器和边缘服务器的优势,提出了一个云-边-端的分层量化联邦学习框架(HQFL)。在该框架中,采用了一种通信高效的模型训练算法(Local-QSGD),结合了本地聚合和云服务器的较少聚合及模型更新上传过程中的权重量化,以提高分层联邦学习中的通信效率。文献[68]提出了一种新的分布式层次张量深度计算模型,将高维空间中的模型参数压缩到一组低维子空间中,通过直接计算低维参数的梯度,减轻了模型训练的通信带宽负担。但是这种通过量化本地计算梯度将梯度量化为低精度值而非直接上传原始梯度值的方法,在降低每个单一通信轮次中通信数据量的同时,也会降低模型精度,增加计算能耗。基于模型压缩的主要研究对比如表 4 所列。

邦学习中的分散拓扑在低带宽或高时延网络上比星型拓扑训练速度更快^[69-71],如图 6 所示。本节将简单介绍一些相关的学术研究。文献[72]表示,在传统的数据中心环境和网络面临低带宽和高延迟的约束时,分散式训练比集中式训练更加快速。文献[73]提出了一种分散的、基于梯度的优化算法,以解决掉队者的延迟和通信开销。在算法的每次迭代过程中,对每个节点的本地梯度计算添加一个期限,然后节点交换本地模型更新的等价版本。文献[38]提出了一种半分散的分层

联邦学习框架,使用 D2D 通信来交换节点之间的模型参数,增强节点集群之间的协作共识过程来节约网络资源,降低通信开销。文献[20]提出了一个多层分布式计算防御框架,通过数据层、边缘层、云层的协同训练,来降低云服务器的通信负担,以解决海量数据集中通信开销大的问题。但是这种方式并不适用于所有的场景,因为这种物理层次可能不存在,或者不可能预先知道。

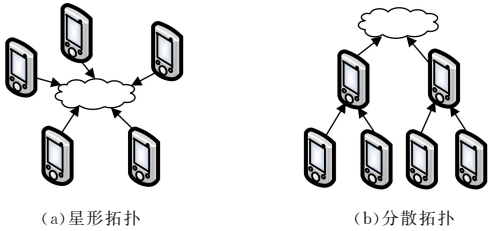


图 6 两种不同的网络拓扑结构

Fig. 6 Two different network topologies

以上相关研究中,考虑处于网络边缘的参与节点是静态的,即在整个模型训练中的地理位置是固定不变的。然而,实际的分层联邦学习任务大多都部署在具有高度移动性的可穿戴设备、智能手机、车载传感器等移动设备上。基于此,文献[74]提出了一种机会联邦学习,移动设备并不是孤立地进行本地模型训练,而是结合了它们在与邻居设备的机会相遇中获得的学习经验。在这种方法中,边缘设备与邻居设备相遇后,将本地模型参数共享给邻居设备,邻居设备使用本地数据集进行模型训练并将模型梯度共享给该设备,该设备使用梯

度聚合算法将本地模型参数与从相遇的邻居设备获得的模型梯度进行聚合,以此来进行本地模型更新,以便得到个性化的健壮模型。

4.6 面向参数聚合传输的优化方法

参与节点通常需要无线接入并建立起与服务器之间的通信,无限信道状态的动态变化、同频信道间的干扰、噪声等因素都会影响参与节点与服务器、参与节点与参与节点之间的通信效率,进而影响模型训练时延和整个系统的可靠性。

Amiri 等^[75]在带宽受限的无限链路上进行收敛性分析,并针对下行链路,分别考虑了数字传输和模拟传输两种方式。文献[76-80]针对模型准确性和通信开销之间的最优权衡展开理论研究,分析了在通信约束下进行分布式统计估算与学习的最优极小极大速率。Yang 等^[81]综合考虑了节点调度机制和小区间干扰等因素的影响,建立无线通信网络中联邦学习收敛速度分析模型,并比较了随机调度、轮询调度、比例公平这 3 种不同调度方式下的模型收敛速度。文献[82]提出了一种新的频谱效率的分层联邦学习框架,该框架在传输过程中使用了 MRT 空中聚合方案,可以自然地利用信道失真和干扰来降低通信复杂度。仿真结果说明,该框架可以在有限的通信资源下传输更多的信息,从而提高收敛速度、测试精度及通信效率。为了模拟汇聚传输模式下资源受限无线通信对 FL 算法的影响,文献[83]提出了一种用户选择和功率控制的联合优化方案,可以减轻无线通信对 FL 算法收敛性和性能的影响。

针对以上分层联邦学习中传输优化技术的研究,按照作用原理进行对比分析,结果如表 5 所列。

表 5 传输优化方法对比

Table 5 Comparison of transmission optimization methods																
优化方法	应用领域				压缩方式		针对挑战						优化目标			主要文献
	通用/ 理论研究	物联网	工业 物联网	车联网	随机 稀疏	量化	海量参与 节点	网络带宽 受限	数据 异构性	计算能力 异构	节点协作 的自私性	边缘关联 问题	提高效率	降低 能耗	提高模型 精度	
节点选择	✓						✓	✓		✓			✓		✓	[41]
		✓					✓	✓	✓	✓			✓		✓	[43]
		✓					✓	✓		✓			✓	✓	✓	[39]
	✓						✓	✓				✓	✓			[45]
		✓					✓				✓		✓			[46]
				✓			✓				✓		✓			[47]
	✓						✓	✓			✓		✓			[48]
增强本地计算	✓							✓					✓			[49]
		✓						✓					✓		✓	[52]
	✓							✓	✓				✓		✓	[54-55]
减少模型更新上传数	✓							✓		✓			✓		✓	[57]
	✓							✓					✓		✓	[58]
模型压缩	✓					✓		✓					✓		✓	[59]
	✓				✓			✓					✓			[27]
		✓			✓			✓					✓			[63-64]
	✓				✓	✓		✓	✓				✓		✓	[65-66]
	✓				✓	✓		✓		✓			✓		✓	[62]
				✓	✓	✓		✓				✓	✓		✓	[67]
					✓		✓					✓			[68]	
分散训练	✓							✓	✓				✓	✓	✓	[38]
面向参数聚合传输		✓				✓		✓	✓				✓			[75]
	✓							✓					✓			[77]
		✓						✓					✓		✓	[82]

注:“✓”表示对应问题在主要文献中涉及

针对面向分层联邦学习的传输优化问题,博弈论、非凸优化理论、深度强化学习理论等理论工具分别被引入该领域的研究。文献[46]使用基于模型所有者边际贡献的博弈论方法

来解决恶意攻击带来的不稳定问题。文献[84]用博弈论激励大规模车辆网络中的知识共享行为。类似地,文献[45, 48, 85]用博弈论解决网络中边缘关联和资源分配的最优解问题。

Lou 等^[37,39-40,52,67]使用非凸优化理论展开相关研究。其中,文献[39]使用非凸优化理论解决每个边缘服务器资源分配的最优解问题;文献[67]将非凸优化理论应用于精度延迟权衡和边缘客户关联问题的系统优化中。而文献[86]基于深度强化学习理论来最小化物联网系统中边缘缓存和内存访问延迟的冗余。与使用的优化原理相同,Wang 等^[87-89]均使用深度强化学习理论对分层联邦学习传输优化领域展开研究。

5 研究展望

分层联邦学习作为一种新兴的分布式机器学习技术,尽管应用场景越来越广泛且有关网络优化方面的已有研究层出不穷,但是该领域仍然存在着许多更深层次的有研究意义的内容。

(1)多层模型聚合下的传输优化研究

典型的云-边-端三层分层联邦学习中,仅引入一层边缘服务器作为传输中介,边缘服务器向云服务器提供的上游维度会较多,可以通过引入多层边缘服务器来进行多层边缘模型聚合,以提高通信效率。因此,在多层模型聚合下的实际应用场景中,边缘网络的拓扑结构如何构建及哪些边缘层参与多层模型聚合,是未来值得研究的问题。

(2)针对节点移动性的在线优化方法研究

实际的分层联邦学习大多都部署在可穿戴设备、智能手机、车载传感器等边缘设备上,这些边缘设备的移动性使其在模型训练各个阶段的地理位置不是固定的,网络拓扑会随着模型训练过程发生改变,从而影响整个模型的训练性能。因此,未来的研究方向可以着重集中在针对边缘节点的高度移动性设计动态的在线优化方法,以提升网络性能。

(3)面向多目标的网络传输优化技术研究

在实际的分层联邦学习场景下进行模型训练,需要综合考虑网络的整体性能,而不能像增强本地计算或压缩模型更新那样一味地牺牲计算代价和模型精度来换取通信开销。因此,未来该领域的研究可以继续深化联合网络的整体性能,在计算开销和模型精度的约束下使通信开销最小化,在系统多目标下展开传输优化工作。

(4)新一代无线通信技术支持下的传输优化研究

随着高速率、低时延的 5G 移动通信技术的日益普及,越来越多分层联邦学习应用于无线网络领域,应用场景部署到 5G 移动边缘设备上,这类设备的通信带宽和电源有限,且无线通信中复杂的噪声和干扰会影响模型的传输过程。因此,面向新一代无线通信技术,开发适用于特定应用中模型聚合的无线通信的分层联邦学习算法具有突出的研究意义。

结束语 研究分层联邦学习中的传输优化问题,对分层联邦学习在实际场景中的应用具有很重要的现实意义。本文首先基于分层联邦学习的概念及算法原理,重点关注模型训练过程中的传输优化研究进展;然后针对分层联邦学习中面临的 6 个传输挑战,分别从选择合适节点、增强本地计算、减少模型更新上传数、压缩模型更新、分散训练、参数聚合传输优化六大方面对传输优化的已有研究展开介绍;最后展望了分层联邦学习的发展前景。

参 考 文 献

[1] MCMAHAN B,MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data [C]// Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.

[2] PARK J, SAMARAKOOM S, BENNIS M, et al. Wireless network intelligence at the edge[J]. Proceedings of the IEEE, 2019, 107(11): 2204-2239.

[3] LIU L, ZHANG J, SONG S H, et al. Client-edge-cloud hierarchical federated learning[C]// IEEE International Conference on Communications(ICC). IEEE, 2020: 1-6.

[4] ABDELLATIF A A, MHAISEN N, MOHAMED A, et al. Communication-efficient hierarchical federated learning for IoT heterogeneous systems with imbalanced data[J]. Future Generation Computer Systems, 2022, 128: 406-419.

[5] SAADAT H, ABOUMADI A, MOHAMED A, et al. Hierarchical federated learning for collaborative IDS in IoT applications [C]// 2021 10th Mediterranean Conference on Embedded Computing(MECO). IEEE, 2021: 1-6.

[6] ZHANG J, LI M, ZENG S, et al. A survey on security and privacy threats to federated learning[C]// 2021 International Conference on Networking and Network Applications(NaNA). IEEE, 2021: 319-326.

[7] YIN X, ZHU Y, HU J. A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions[J]. ACM Computing Surveys (CSUR), 2021, 54(6): 1-36.

[8] LIM W Y B, LUONG N C, HOANG D T, et al. Federated learning in mobile edge networks: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 2031-2063.

[9] SONG M, WANG Z, ZHANG Z, et al. Analyzing user-level privacy attack against federated learning[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(10): 2430-2444.

[10] MELIS L, SONG C, DE CRISTOFARO E, et al. Exploiting unintended feature leakage in collaborative learning [C] // 2019 IEEE Symposium on Security and Privacy (SP). IEEE, 2019: 691-706.

[11] HITAJ B, ATENIESE G, PEREZ-CRUZ F. Deep models under the GAN: information leakage from collaborative deep learning [C] // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017: 603-618.

[12] LI Z, SHARMA V, MOHANTY S P. Preserving data privacy via federated learning: Challenges and solutions[J]. IEEE Consumer Electronics Magazine, 2020, 9(3): 8-16.

[13] SHI L, SHU J, ZHANG W, et al. HFL-DP: Hierarchical Federated Learning with Differential Privacy[C]// 2021 IEEE Global Communications Conference(GLOBECOM). IEEE, 2021: 1-7.

[14] LU Y, HUANG X, DAI Y, et al. Differentially private asynchronous federated learning for mobile edge computing in urban informatics[J]. IEEE Transactions on Industrial Informatics, 2019, 16(3): 2134-2143.

[15] LU Y, HUANG X, DAI Y, et al. Federated learning for data pri-

- vacancy preservation in vehicular cyber-physical systems[J]. *IEEE Network*, 2020, 34(3): 50-56.
- [16] KAISSIS G A, MAKOWSKI M R, RUCKERT D, et al. Secure, privacy-preserving and federated machine learning in medical imaging[J]. *Nature Machine Intelligence*, 2020, 2(6): 305-311.
- [17] GEYER R C, KLEIN T, NABI M. Differentially private federated learning; A client level perspective[J]. *arXiv*: 1712. 07557, 2017.
- [18] BONAWITZ K, IVANOV V, KREUTER B, et al. Practical secure aggregation for privacy-preserving machine learning[C]// *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 2017: 1175-1191.
- [19] MCMAHAN H B, RAMAGE D, TALWAR K, et al. Learning differentially private recurrent language models[J]. *arXiv*: 1710. 06963, 2017.
- [20] SHARMA P K, PARK J H, CHO K. Blockchain and federated learning-based distributed computing defence framework for sustainable society[J/OL]. *Sustainable Cities and Society*, 2020, 59: 102220. <https://DOI.ORG/10.1016/J.SCS.2020.102220>.
- [21] LU Y, HUANG X, ZHANG K, et al. Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(4): 4298-4311.
- [22] SARIKAYA Y, ERCETIN O. Motivating workers in federated learning: A stackelberg game perspective[J]. *IEEE Networking Letters*, 2019, 2(1): 23-27.
- [23] ZHAN Y, LI P, QU Z, et al. A learning-based incentive mechanism for federated learning[J]. *IEEE Internet of Things Journal*, 2020, 7(7): 6360-6368.
- [24] DENG Y, LYU F, REN J, et al. Fair: Quality-aware federated learning with precise user incentive and model aggregation[C]// *IEEE INFOCOM 2021 — IEEE Conference on Computer Communications*. IEEE, 2021: 1-10.
- [25] TANG M, WONG V W S. An incentive mechanism for cross-silo federated learning: A public goods perspective[C]// *IEEE INFOCOM 2021 — IEEE Conference on Computer Communications*. IEEE, 2021: 1-10.
- [26] KIM Y J, HONG C S. Blockchain-based node-aware dynamic weighting methods for improving federated learning performance[C]// *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2019: 1-4.
- [27] KONECNY J, MCMAHAN H B, YU F X, et al. Federated learning; Strategies for improving communication efficiency[J]. *arXiv*: 1610. 05492, 2016.
- [28] SHAHID O, POURIYEH S, PARIZI R M, et al. Communication Efficiency in Federated Learning: Achievements and Challenges [J]. *arXiv*: 2107. 10996, 2021.
- [29] MOTHUKURI V, PARIZI R M, POURIYEH S, et al. A survey on security and privacy of federated learning[J]. *Future Generation Computer Systems*, 2021, 115: 619-640.
- [30] WANG L P, WANG W, LI B. CMFL: Mitigating communication overhead for federated learning[C]// *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019: 954-964.
- [31] CHEN Y, SUN X, JIN Y. Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 31(10): 4229-4238.
- [32] CHEN M, GUNDUZ D, HUANG K, et al. Distributed learning in wireless networks: Recent progress and future challenges[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(12): 3579-3605.
- [33] LI T, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[J]. *Proceedings of Machine Learning and Systems*, 2020, 2: 429-450.
- [34] SATTLER F, WIEDEMANN S, MULLER K R, et al. Robust and communication-efficient federated learning from non-IID data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 31(9): 3400-3413.
- [35] ALI S, SAAD W, RAJATHEVA N, et al. 6G white paper on machine learning in wireless communication networks[J]. *arXiv*: 2004. 13875, 2020.
- [36] HOSSEINALIPOUR S, BRINTON C G, AGGARWAL V, et al. From federated learning to fog learning: Towards large-scale distributed machine learning in heterogeneous wireless networks[J]. *arXiv*: 2006. 03594, 2020.
- [37] HOSSEINALIPOUR S, BRINTON C G, AGGARWAL V, et al. From federated to fog learning: Distributed machine learning over heterogeneous wireless networks[J]. *IEEE Communications Magazine*, 2020, 58(12): 41-47.
- [38] LIN F P C, HOSSEINALIPOUR S, AZAM S S, et al. Semi-decentralized federated learning with cooperative D2D local model aggregations[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(12): 3851-3869.
- [39] LOU S, CHEN X, WU Q, et al. HFEL: Joint edge association and resource allocation for cost-efficient hierarchical federated edge learning[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(10): 6535-6548.
- [40] MHAISEN N, ABDELLATIF A A, MOHAMED A, et al. Optimal user-edge assignment in hierarchical federated learning based on statistical properties and network topology constraints [J]. *IEEE Transactions on Network Science and Engineering*, 2021, 9(1): 55-66.
- [41] SONG J, KOUNTOURIS M. Optimal number of edge devices in distributed learning over wireless channels[C]// *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2020: 1-5.
- [42] CHEN Z, LIAO W, HUA K, et al. Towards asynchronous federated learning for heterogeneous edge-powered internet of things [J]. *Digital Communications and Networks*, 2021, 7(3): 317-326.
- [43] ABDULRAHMAN S, TOUT H, MOURAD A, et al. FedMCCS: Multicriteria client selection model for optimal IOT federated learning[J]. *IEEE Internet of Things Journal*, 2020, 8(6): 4723-4735.
- [44] CHEN Q, YOU Z, JIANG H. Semi-asynchronous Hierarchical Federated Learning for Cooperative Intelligent Transportation

- Systems[J]. arXiv:2110.09073,2021.
- [45] LIM W Y B, NG J S, XIONG Z, et al. Dynamic edge association and resource allocation in self-organizing hierarchical federated learning networks[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(12): 3640-3653.
 - [46] LIM W Y B, XIONG Z, MIAO C, et al. Hierarchical incentive mechanism design for federated machine learning in mobile networks[J]. IEEE Internet of Things Journal, 2020, 7(10): 9575-9588.
 - [47] LIM W Y B, XIONG Z, NIYATO D, et al. Incentive mechanism design for federated learning in the Internet of vehicles[C]//2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall). IEEE, 2020: 1-5.
 - [48] LIM W Y B, NG J S, XIONG Z, et al. Decentralized edge intelligence: A dynamic resource allocation framework for hierarchical federated learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 33(3): 536-550.
 - [49] MCMAHAN B, MOORE E, RAMAGE D, et al. Communication-efficient Learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.
 - [50] YAO X, HUANG C, SUN L, et al. Two-stream federated learning: Reduce the communication costs[C]//2018 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2018: 1-4.
 - [51] LIU L, ZHANG J, SONG S H, et al. Edge-assisted hierarchical federated learning with non-IID data[J]. arXiv: 1905.06641, 2019.
 - [52] WANG Z, XU H, LIU J, et al. Resource-Efficient Federated Learning with Hierarchical Aggregation in Edge Computing[C]//IEEE INFOCOM 2021—IEEE Conference on Computer Communications. IEEE, 2021: 1-10.
 - [53] REISIZADEH A, MOKHTARI A, HASSANI H, et al. Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization[C]//International Conference on Artificial Intelligence and Statistics. PMLR, 2020: 2021-2031.
 - [54] KARIMIREDDY S P, KALE S, MOHRI M, et al. Scaffold: Stochastic controlled averaging for federated learning[C]//International Conference on Machine Learning. PMLR, 2020: 5132-5143.
 - [55] BRIGGS C, FAN Z, ANDRAS P. Federated learning with hierarchical clustering of local updates to improve training on non-IID data[C]//IEEE International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-9.
 - [56] TAO Z Y, LI Q. ESGD: Communication Efficient Distributed Deep Learning on the Edge[C]//Workshop on Hot Topics in Edge Computing. 2018: 1-6.
 - [57] NGUYEN H T, SEHWAG V, HOSSEINALIPOUR S, et al. Fast-convergent federated learning[J]. IEEE Journal on Selected Areas in Communications, 2020, 39(1): 201-218.
 - [58] WANG S, LEE M, HOSSEINALIPOUR S, et al. Device sampling for heterogeneous federated learning: Theory, algorithms, and implementation[C]//IEEE INFOCOM 2021—IEEE Conference on Computer Communications. IEEE, 2021: 1-10.
 - [59] KHALED A, RICHTARIK P. Gradient descent with compressed iterates[J]. arXiv:1909.04716, 2019.
 - [60] TANG H, LIAN X, QIU S, et al. DeepSqueeze: Parallel stochastic gradient descent with double-pass error-compensated compression[J]. arXiv:1907.07346, 2019.
 - [61] KARIMIREDDY S P, REBJOCK Q, STICH S, et al. Error feedback fixes sign SGD and other gradient compression schemes[C]//International Conference on Machine Learning. PMLR, 2019: 3252-3261.
 - [62] AMIRI M M, GUNDUZ D, KULKARNI S R, et al. Federated learning with quantized global model updates[J]. arXiv: 2006.10672, 2020.
 - [63] ALISTARH D, GRUBIC D, LI J, et al. QSGD: Communication-efficient SGD via Gradient Quantization and Encoding[J]. Advances in Neural Information Processing Systems, 2017, 30: 1709-1720.
 - [64] SURESH A T, FELIX X Y, KUMAR S, et al. Distributed mean estimation with limited communication[C]//International Conference on Machine Learning. PMLR, 2017: 3329-3337.
 - [65] ZHANG X, FANG M, LIU J, et al. Private and communication-efficient edge learning: a sparse differential gaussian-masking distributed SGD approach[C]//Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing. 2020: 261-270.
 - [66] HU R, GONG Y, GUO Y. CPFed: communication-efficient and privacy-preserving federated learning[J]. arXiv: 2003.13761, 2020.
 - [67] LIU L, ZHANG J, SONG S, et al. Hierarchical quantized federated learning: Convergence analysis and system design[J]. arXiv: 2103.14272, 2021.
 - [68] ZHENG H, GAO M, CHEN Z, et al. A distributed hierarchical deep computation model for federated learning in edge computing[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 7946-7956.
 - [69] LU X, LIAO Y, LIO P, et al. Privacy-preserving asynchronous federated learning mechanism for edge network computing[J]. IEEE Access, 2020, 8(99): 48970-48981.
 - [70] LIU W, CHEN L, CHEN Y, et al. Accelerating federated learning via momentum gradient descent[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 31(8): 1754-1766.
 - [71] XIAO P, CHENG S, STANKOVIC V, et al. Averaging is probably not the optimum way of aggregating parameters in federated learning[J]. Entropy, 2020, 22(3): 314.
 - [72] LI T, SAHU A K, TALWALKAR A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.
 - [73] REISIZADEH A, TAHERI H, MOKHTARI A, et al. Robust and communication-efficient collaborative learning[J/OL]. Advances in Neural Information Processing Systems, 2019, 32. <https://doi.org/10.48550/arXiv.1907.10595>.
 - [74] LEE S, ZHENG X, HUA J, et al. Opportunistic Federated Learning: An Exploration of Egocentric Collaboration for Pervasive Computing Applications[C]//2021 IEEE International Conference on Pervasive Computing and Communications (Per-

Com). IEEE, 2021; 1-8.

[75] AMIRI M M, GUNDUZ D, KULKARNI S R, et al. Convergence of federated learning over a noisy downlink [J]. IEEE Transactions on Wireless Communications, 2021, 21(3): 1422-1437.

[76] BRAVERMAN M, GARG A, MA T, et al. Communication lower bounds for statistical estimation problems via a distributed data processing inequality [C] // Proceeding of the 48th Annual ACM Symposium on Theory of Computing. 2016; 1011-1020.

[77] HAN Y, OZGUR A, WEISSMAN T. Geometric lower bounds for distributed parameter estimation under communication constraints [C] // Conference On Learning Theory. PMLR, 2018: 3163-3188.

[78] ACHARYA J, CANONNE C L, TYAGI H. Inference under information constraints I: Lower bounds from chi-square contraction [J]. IEEE Transactions on Information Theory, 2020, 66(12): 7835-7855.

[79] BARNES L P, HAN Y, OZGUR A. Lower bounds for learning distributions under communication constraints via fisher information [J]. Journal of Machine Learning Research, 2020, 21(236): 1-30.

[80] KONECNY J, RICHTARIK P. Randomized distributed mean estimation: Accuracy vs. communication [J]. Frontiers in Applied Mathematics and Statistics, 2018, 4: 62-73.

[81] YANG H H, LIU Z, QUEK T Q S, et al. Scheduling policies for federated learning in wireless networks [J]. IEEE Transactions on Communications, 2019, 68(1): 317-333.

[82] SU L, LAU V K N. Distributed Edge Learning with Inter-Type and Intra-Type Over-the-Air Collaboration [C] // ICC 2021 — IEEE International Conference on Communications. IEEE, 2021: 1-6.

[83] FAN X, WANG Y, HUO Y, et al. Joint optimization of communications and federated learning over the air [J]. IEEE Transactions on Wireless Communications, 2022, 21(6): 4434-4449.

[84] CHAI H, LENG S, CHEN Y, et al. A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in internet of vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(7): 3975-3986.

[85] LIM W Y B, NG J S, XIONG Z, et al. Dynamic Resource Allocation for Hierarchical Federated Learning [C] // 2020 16th International Conference on Mobility, Sensing and Networking (MSN). IEEE, 2020: 153-160.

[86] MAJIDI F, KHAYYAMBASHI M R, BAREKATAIN B. Hfdrl: An intelligent dynamic cooperate caching method based on hierarchical federated deep reinforcement learning in edge-enabled iot [J]. IEEE Internet of Things Journal, 2021, 9(2): 1402-1413.

[87] ANH T T, LUONG N C, NIYATO D, et al. Efficient training management for mobile crowd-machine learning: A deep reinforcement learning approach [J]. IEEE Wireless Communications Letters, 2019, 8(5): 1345-1348.

[88] NG J S, LIM W Y B, XIONG Z, et al. A Hierarchical Incentive Design Toward Motivating Participation in Coded Federated Learning [J]. IEEE Journal on Selected Areas in Communications, 2021, 40(1): 359-375.

[89] WANG X, GARG S, LIN H, et al. Towards accurate anomaly detection in industrial internet-of-things using hierarchical federated learning [J]. IEEE Internet of Things Journal, 2021, 9(10): 7110-7119.



ZOU Sai-lan, born in 1997, postgraduate. Her main research interests include edge computing and so on.



LI Zhuo, born in 1983, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include mobile wireless network and distributed computing.

(责任编辑:喻葵)