成绩_____

苏州大学

课　　程　　文献阅读和科技写作

学院(部)　　计算机科学与技术学院

类　　型　　综合作业

学　　号　　1927405160

姓　　名　　张昊

日　　期　　2022 年 5 月 25 日

综合作业

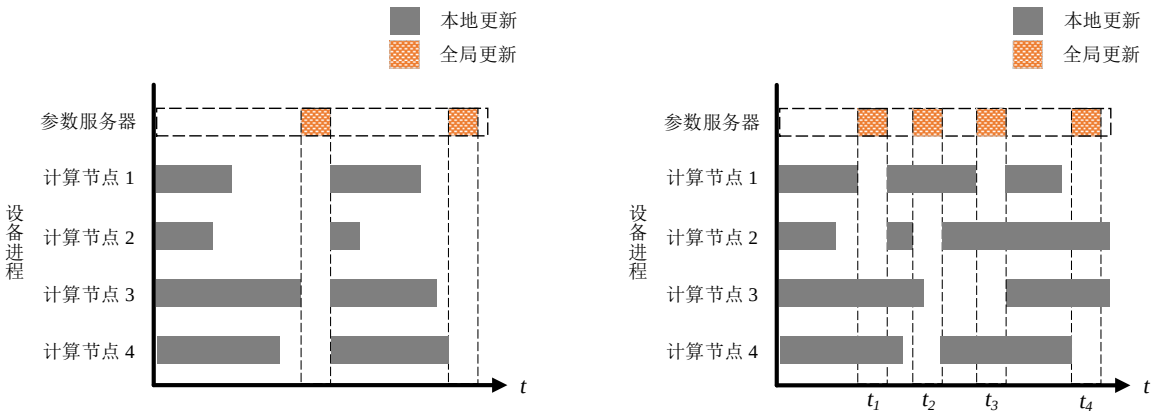
1927405160 张昊

[1)](苏州大学 计算机科学与技术学院，江苏 苏州 215006)

1. 图、表及公式的制作

1.1. 图的制作

联邦学习是一类特殊的分布式机器学习．一个联邦学习系统通常由一个或多个参数服务器和大量计算节点组成，计算节点向参数服务器发送本地更新，并从参数服务器接收更新的全局模型[1].



(*a*) 同步联邦学习　　(*b*) 异步联邦学习

图 1 固定的时间段情况下两种联邦学习方案的本地更新和全局更新示意图.

图 1 修改自文献[1]，假设边缘计算系统中有一个参数服务器和十个计算节点，选择四个计算节点参与模型训练任务．图 1 使用两个子图在固定长度的时间段内分别展示了同步联邦学习方案和异步联邦学习方案（Communication-Efficient Asynchronous Federated Learning，CE-AFL）的本地（灰色）和全局（橙色）更新．纵轴表示不同设备的进程，横轴表示模型训练时间．对于子图(*a*)中的同步方案，只有在参数服务器收到来自四个计算节点的所有本地更新后，才会执行模型聚合以计算得到更新的全局模型．当计算节点收到全局模型后将继续使用本地数据进行训练．对于子图(*b*)中的异步方案，参数服务器接收到了来自任意两个计算节点的本地更新（CE-AFL 的参数$\alpha = 0.5$）就执行模型聚合；如果服务器在聚合期间收到了本地更新

的模型，那么其将在下一次全局更新中聚合这些本地更新．可见在给定固定的时间段中，异步方案中有四个全局更新，而同步方案中只有两个全局更新．因此使用此图可以直观地展示异步方案在相同的时间预算约束下收敛速度更快的优势．

1.2. 表的制作

文献[2]提出了一种实用的联邦学习算法 FederatedAveraging（FedAvg），适用于在通信成本的约束下利用不平衡且非独立同分布（Independent and Identically Distributed，IID）数据训练模型．该算法有三个关键的超参数：每轮执行计算的客户端比例C；每轮每个客户端使用本地数据集的训练次数E；用于客户端更新的本地批次大小（local mini-batch size）B．特别地，$B=\infty$表示将整个本地数据集被当作单个批次（mini-batch）处理．$C=0$表示每轮仅使用一个客户端．

表 1 使用 FedAvg 算法训练的模型在测试集达到目标准确率所需的通信次数．

C	模型*	IID	$B=\infty$**	$B=10$	C	模型	IID	$B=\infty$	$B=10$
0.0	2NN	是	1455	316	0.0	CNN	是	387	50
0.1	2NN	是	1474 (1.0×)	87 (2.4×)	0.1	CNN	是	339 (1.1×)	18 (2.8×)
0.2	2NN	是	1658 (0.9×)	77 (4.1×)	0.2	CNN	是	337 (1.1×)	18 (2.8×)
0.5	2NN	是	— (—)***	75 (4.2×)	0.5	CNN	是	164 (2.4×)	18 (2.8×)
1.0	2NN	是	— (—)	70 (4.5×)	1.0	CNN	是	246 (1.6×)	16 (3.1×)
0.0	2NN	否	4278	3275	0.0	CNN	否	1181	956
0.1	2NN	否	1796 (2.4×)	664 (4.9×)	0.1	CNN	否	1100 (1.1×)	206 (4.6×)
0.2	2NN	否	1528 (2.8×)	619 (5.3×)	0.2	CNN	否	978 (1.2×)	200 (4.8×)
0.5	2NN	否	— (—)	443 (7.4×)	0.5	CNN	否	1067 (1.1×)	261 (3.7×)
1.0	2NN	否	— (—)	380 (8.6×)	1.0	CNN	否	— (—)	97 (9.9×)

* MINIST 图像分类模型 2NN（2 个隐藏层，每层 200 个单元的多层感知器）：参数$E=1$，目标准确率 97 %；MINIST 图像分类模型 CNN：参数$E=5$，目标准确率 99 %.

** 将$C=0$作为基线方法，括号中为相对于基线方法加速的倍率．

*** 标“—”的数据表示在五次运行中均未在规定时间内达到目标精度．

实验部分首先基于 MINIST 数据集分析 FedAvg 算法的超参数C、E、B对实验结果的影响．表 1 节选自文献[2]提高并行性实验的数据（实验设定可参考原文献）．实验中固定超参数E的取值，分析超参数C和B的变化对测试集达到准确率所需通信次数的影响，并将$C=0$作为基线方法，计算了相对于基线方法加速的倍率．

1.3. 公式的制作

下面的公式为文献[1]中证明定理 5 的第一步：

$$\begin{aligned}\mathbb{E}[F(\widehat{w}^{d})-F(w^{*})] &\leq F(\widehat{w}^{d-1})-F(w^{*})-\eta\mathbb{E}[\langle\nabla F(\widehat{w}^{d-1}),\nabla f(\widehat{w}^{d-1};q_{d})\rangle]\\ &\quad+\frac{L\eta^{2}}{2}\mathbb{E}[||\nabla f(\widehat{w}^{d-1};q_{d})||^{2}]\\ &\leq F(\widehat{w}^{d-1})-F(w^{*})-\frac{\eta}{2}||\nabla F(\widehat{w}^{d-1})||^{2}\\ &\quad+\frac{\eta}{2}\mathbb{E}[||\nabla F(\widehat{w}^{d-1})-\nabla f(\widehat{w}^{d-1};q_{d})||^{2}]\\ &\leq F(\widehat{w}^{d-1})-F(w^{*})-\frac{\eta}{2}||\nabla F(\widehat{w}^{d-1})||^{2}+\frac{\eta\mathcal{P}}{2}\\ &\leq F(\widehat{w}^{d-1})-F(w^{*})-\eta\mu[F(\widehat{w}^{d-1})-F(w^{*})]+\frac{\eta\mathcal{P}}{2}\\ &\leq(1-\eta\mu)[F(\widehat{w}^{d-1})-F(w^{*})]+\frac{\eta\mathcal{P}}{2}.\end{aligned}\tag{1}$$

其中，全局损失函数F是 L-光滑的和μ-强凸的（$L>0,\mu\geq 0$且均为常数）；D为每个计算节点在向参数服务器报告更新的模型之前执行的本地更新次数；对于每个本地更新$d\in\{1,\cdots,D\}$，$\widehat{w}^{d}$表示d次本地更新后的模型参数，w^{0}表示初始模型参数；w^{*}是损失函数$F(w)$的最优值；η是步长，也称为学习率；$\mathcal{P}$为有界梯度方差假设中的正上界.

公式（1）利用平滑性和强凸性的假设推导了 CE-AFL 机制中本地更新的收敛结果．上述描述与推导过程涉及到的一些定理可以参考文献[1]，这里受限于篇幅不作具体说明.

2. 摘要撰写

近年来我们见证了众多编程语言的发展和壮大，这些语言的共同目标是更方便、高效地实现人与机器之间的交流和沟通．尽管在众多领域已经取得了许多进展，但仍然存在一些未解决的问题．例如 C 和 Java 等强大的语言往往复杂，学习成本高，开发效率低；而如 ABC 之类的可读性更强的语言却平台迁移能力弱，难以进行扩展．针对这些问题，本文介绍了解释型脚本语言 Python，作为 ABC 语言的继承．Python 是一种解释的、交互式、开源、跨平台的编程语言，其代码的可读性高，语法简洁，用更少的代码可以完成同样的工作．Python 也是一门强大的、可扩展的语言，包含动态类型等高级特性，支持面向对象编程等多种编程范式．与现有语言相比，Python 简单直观，同样强大，适用于短期开发的日常任务.

3. 阅读心得

联邦学习本质上是一种分布式机器学习技术，广泛应用于边缘计算场景中，其目标为在保证数据隐私安全及合法合规的基础上，利用本地数据训练机器学习模型．在联邦学习中，用户对设备有绝对的控制权，设备计算能力不同且负载不均衡，通信代价往往很高，其数据并非 IID，这些与传统分布式机器学习的不同带来了很多有研究价值的问题．

联邦学习最重要的研究方向是对联邦学习算法进行改进和优化，现有工作[1][2][3][4]可分为同步和异步联邦学习算法优化．这类问题的关键是在于改进联邦学习算法，减少通信次数或优化其他关键指标，使用设备端非 IID 的数据在带宽等资源约束下加快模型训练的收敛速度，提升模型精度．此外，联邦学习在隐私保护和鲁棒性方面也存在某些问题．例如现有工作[5][6]可以反向推断设备端隐私信息，针对数据或模型的攻击[7][8]可以很容易地被改造来攻击联邦学习等．现有解决方案往往会影响模型精度，或依赖于 IID 的数据[9][10]，未来值得进一步研究．

参 考 文 献

[1] Liu J, Xu H, Xu Y, *et al*. Communication-efficient asynchronous federated learning in resource-constrained edge computing[J]. Computer Networks, 2021, 199: 108429.

[2] 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.

[3] Mills J, Hu J, Min G. Communication-efficient federated learning for wireless edge intelligence in IoT. IEEE Internet of Things Journal, 2019, 7(7): 5986-5994.

[4] Wang S, Tuor T, Salonidis T, *et al*. Adaptive federated learning in resource constrained edge computing systems. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1205-1221.

[5] 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.

[6] 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.

[7] Shafahi A, Huang W R, Najibi M, *et al*. Poison frogs! targeted clean-label poisoning attacks on neural networks[J]. Advances in neural information processing systems, 2018, 31.

[8] Bhagoji A N, Chakraborty S, Mittal P, *et al*. Analyzing federated learning through an adversarial lens[C]//International Conference on Machine Learning. PMLR, 2019: 634-643.

[9] Yin D, Chen Y, Kannan R, *et al*. Byzantine-robust distributed learning: Towards optimal statistical rates[C]//International Conference on Machine Learning. PMLR, 2018: 5650-5659.

[10] Yin D, Chen Y, Kannan R, *et al*. Defending against saddle point attack in Byzantine-robust distributed learning[C]//International Conference on Machine Learning. PMLR, 2019: 7074-7084.