论集群学习的反垄断性

*费利克斯·萨特勒*1*，克劳斯·罗伯特·穆勒*2*,*3*,*4*，托马斯·威根*1*,*2*，沃伊切赫·萨梅克*1 1 德国柏林夫琅和费-海因里希-赫兹研究所23

图柏林，德国柏林，马克斯普朗克信息学研究所，萨尔布吕肯，德国4

韩国首尔韩国大学

# 摘要

联邦学习（FL）是目前最广泛采用的隐私约束下（deep）机器学习模型协同训练的框架。尽管它很受欢迎，但有人观察到，如果本地客户的数据分布出现分歧，联合学习会产生次优的结果。最近提出的集群联合学习框架解决了这个问题，它基于参数更新之间的成对余弦相似性将客户群分成不同的组。在这项工作中，我们研究了CFL在拜占庭环境中的应用，其中一部分客户表现出不可预测的行为，或者试图以一种定向或无定向的方式干扰联合培训工作。我们用deep神经网络对常见的联邦学习数据集进行了实验，结果表明CFL（无需修改）能够可靠地检测出拜占庭式客户机并将其从训练中移除。

*索引术语*-鲁棒学习，分布式学习，联合学习，多任务学习，聚类

# 1.    简介

联合学习是一种分布式的培训框架，它允许多个客户机（通常是移动或物联网设备）以高效的通信方式，就其组合数据联合培训单个深度学习模型，不要求任何参与者将其私人培训数据透露给集中实体或彼此[1][2][3][4][5]。联合学习依赖于这样一个假设：X→Y能够同时拟合所有客户端的数据生成分布。这一假设可以正式表述如下：*fθɕi*

假设1。（“联合学习”）：存在一个参数配置∈Θ*θ*∗*，这将（局部）最小化*

|  |
| --- |
| 这项工作得到了德国教育和研究部柏林大数据中心（01IS14013A）和柏林中心的支持  机器学习（01IS18037I）。确认DFG的部分资助  （EXC 2046/1，项目编号：390685689）。这项工作还得到了韩国政府资助的信息和通信技术规划与评估（IITP）赠款（编号：2017-0-00451）的支持。 |

*所有客户数据同时生成分布的风险*

*里*（θ\*）≤Ri（θ）∀θ∈Bε（θ\*），i=1，…，m（一）

*据此里*（θ） =（fθ（x），y）dψi（x，y）R*我风险函数是否与分配有关ψi和l是一个合适的损失函数。*

很容易看出，这种假设并不总是令人满意的。具体地说，如果（a）客户有不一致的条件分布（y | x）6=ψj（y | x）或（b）模型不够表达，无法同时拟合所有分布，则违反了这一点。这两种情况的简单反例见[6]。在下面，我们将调用两个客户及其分配和（关于和）如果他们满足假设1，如果他们不满足不一致。*ɕifθɕiψj全等式f我*

最近提出的集群联邦学习框架（CFL）[6]概括了联邦学习假设1，并且能够解决客户持有来自遵循聚类结构的不一致分布的数据的联合学习问题：

假设2。（“集群联合学习”）：存在客户群体的划分，这样每个客户机子集*c*∈C*满足传统的联合学习假设。*

聚类联邦学习（图1）利用FL损失曲面的几何性质来识别聚类结构C。与现有的联邦多任务学习方法相比，CFL不需要对FL通信协议进行任何修改，适用于一般的非凸目标（特别是深度目标神经网络）和强大的数学保证的聚类质量。我们将在第2节中快速回顾CFL的理论基础。

拜占庭式的背景：年已经证明，常规美联储-

在存在错误和恶意客户端的情况下，erated Learning无法收敛[7][8]。一个单一的坏客户机可能会损害整个联合培训模型的性能，并否定所有其他客户机的培训努力。为了缓解这个问题，不同的健壮的联合学习策略

978-1-5090-6631-5/20/$31.00©2020 IEEE ICASSP 2020

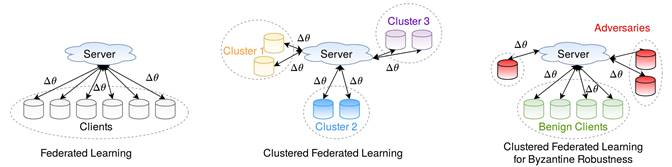


图1：集群联合学习（center）是传统联邦学习框架（left）的扩展，它通过自动将客户机分成可联合训练的数据分布的集群来提高健壮性和灵活性。在这项工作中，我们研究了CFL在拜占庭环境（右）中的应用，它可以被视为集群环境的一个特例，其中只有最大的集群被认为是良性的，而所有其他集群都被认为是敌对的。

在文献中提出。然而，这些现有的策略需要对联邦通信协议进行修改，并且通常在计算上非常昂贵。

在本文中，我们将探讨CFL在这些拜占庭式设置中的应用，在这些环境中，客户群体的一个子集表现出明显的有害方式。很容易看出，这些设置可以包含在假设2下，方法是声明一个特定的客户机集群∈C为“良性”客户机，而与此群集不一致的所有其他客户机称为“敌对”客户机：C={cbenign}∪Cadv。*C设计*

# 2.    集群联合学习

对于遵循假设2的一般联合学习问题，文献[6]已经证明，在相对温和的条件下，聚类联邦学习能够通过检测余弦相似度来推断客户及其数据的聚类结构C

（二）

在不同客户端的渐变更新之间。结果可归纳为以下定理：

定理1（分离定理+推论[6]）。让

*D*1*，…，Dm为本地培训资料m个不同的客户端，每个数据集从k不同的数据生成分布ϕ*1*，..，ψk，这样Di公司*∼∼I（I）（x，y）（θ）（θ）*. 让经验风险里每一个客户的真实风险里*（一）*在联邦学习目标的每一个平稳解上θ*∗*s、 t。*

（三）

*然后，在每个固定溶液中θ*∗*双联邦学习*

*.* （四）

*总是正确的*[[1]](" \l "_ftn1" \o ") *如果它真的成立的话*



*具有*



                                                      问

*Hi、 j*= −γ*我γj*+1-γ*我*2        1-γ*j*2*.* （六）

通过简化条件（5），可以得到基于最大近似误差的稍微较弱但明显更简单的语句：*γ最大值*

推论1。只要它能坚持

（七）

*方程中的聚类机制*（4） 将始终生成正确的群集。

备注1。不等式（7）的实现是足够的，但不是获得正确聚类所必需的。在实践中，正确的聚类可以在概率≈1的情况下实现*对于更大范围的γmax和k详见[6]。*

算法：聚类联邦学习算法以自上而下的方式分离客户群体：从初始的一组客户={1，…，m}和参数初始化开始，客户使用随机梯度下降迭代（SGD）计算权重更新∆θi。与传统的联合学习一样，这些权重更新随后被传递到一个集中的服务器，在那里它们通过平均值聚合到一个全局模型更新中。在CFL中*cθ*0*n*



算法1：集群联合学习（用于拜占庭稳健性）

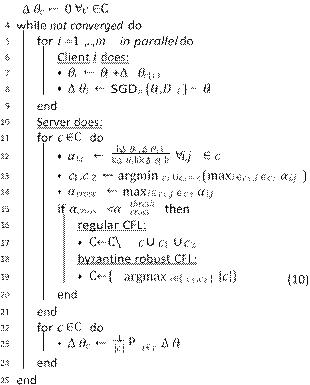


1   输入：初始参数、差异阈值*θ*0

，局部迭代次数/周期*n*

2   输出：良性客户，良性模式*C设计θc温和的*

3   init：设置初始集群C={1，…，m}}，设置初始模型←θ0∀i=1，…，m，设置初始更新*θi*



26 返回，∀c∈c*θc*



然而，在模型聚合之前，服务器计算成对余弦相似度矩阵，并在此基础上计算候选聚类*α*

*.* （八）

如果两个簇之间的最大相似性

（九）

低于一定的阈值，则主簇被二分为两个候选簇。然后分别为每个客户机集群执行模型聚合。在随后的通信回合中，重复这一过程，现有的集群可能会使用相同的机制进一步细化。整个过程在算法1中给出。

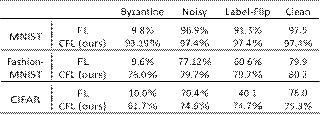
拜占庭设置的CFL：原则上，对于拜占庭设置，不需要修改上述CFL算法。然而，由于我们假设在拜占庭环境下，大多数客户属于一个良性集群，而所有其他客户都被视为敌对客户，我们当然可以通过将所有不属于最大集群的客户从培训中排除来节省计算工作量（10）。

# 3.    相关工作

传统的联合学习[1][2]已经在一致的非iid场景[9][10]中进行了广泛的研究，但无法应对一般客户和敌对客户中数据分布不一致的挑战，如第1节所述。对抗性的联合学习设置可以大致分为两组：在拜占庭式的设置中，假设客户群体的一个子集以任意（随机）的方式行为。这种设置已经被广泛研究，并且已经提出了各种鲁棒的聚集规则，这些规则依赖于梯度相似性[7][8]、几何中值聚集[11]、冗余通信[12]或自适应模型质量估计[13]。虽然其中一些建议的方法在拜占庭环境下提供了收敛保证，但是它们在计算或通信方面也很昂贵，并且常常需要修改联邦通信协议。一个更困难的问题是联邦学习中毒。在这种情况下，（可能有多个）客户尝试在联合培训的模型中引入隐藏的后门功能[14][15][16][17]。由于这种情况下的对手可以根据从服务器接收到的模型更新来调整攻击，因此很难检测到这些攻击，而且到目前为止还没有提出有效的防御策略。

# 4.    实验

我们采用文献[13]中的实验装置来研究集群联合学习对拜占庭式和敌对式客户端的鲁棒性。我们在著名的MNIST、Fashion MNIST和CIFAR-10数据集上进行了实验，在这些数据集上我们使用SGD训练卷积深度神经网络，批大小为100。我们考虑一个有100个参与的客户的联合学习环境，我们在其中随机平均地分割培训数据。在每个实验中，我们声明30%的客户群是错误/恶意的。我们考虑三种不同的场景：1）在拜占庭场景中，恶意客户端从具有各向同性协方差矩阵和标准差1的中心高斯分布中提取其权重更新∆θ（而不是使用随机梯度下降计算）。此场景模拟针对协作训练过程的无定向攻击。2）在标签翻转场景中，恶意客户端的训练数据的所有标签都设置为零。这个场景模拟了一个定向攻击，目标是不均衡地将联合训练的模型偏向一个特定的类。3）在噪声场景下，故障客户的训练数据表1：传统的联合学习和CFL在四个被调查场景中的准确度。



通过向所有像素和通道添加独立的均匀噪声ˆ=x+U（−10,10）进行修改。这个场景模拟训练数据的非结构化噪声失真。4.）在干净的场景中，没有对手。在所有客户端都是良性的情况下，一个好的鲁棒训练算法不应该损害收敛性。在每个场景中，我们根据算法1对整个客户群执行集群联合学习，并将不同集群之间必要的余弦差异阈值设置为*十*

.

我们的目标是研究算法1中定义的CFL是否能够在上述场景中检测出故障和恶意客户端，并将它们从主集群中移除。图2显示了在我们的三个数据集和三个不同场景中，70个良性和30个恶意客户端的前200个通信回合的开发。每次的价值*α交叉α交叉*根据（4）将客户分成两个不同的组。这些事件在图中的标记如下：每当敌对客户从良性客户的主要集群中分离时，在图中标记为绿色，并且在这一轮中分离的对手客户的数量。每当良性客户机从主集群中删除时，它在图中标记为红色。

如我们所见，在所有数据集和不同的对抗场景中，所有恶意客户端都需要不到40轮的通信时间才能从培训中移除。在嘈杂和标签翻转场景中，敌对客户机共享相同的数据生成分布，这导致它们被CFL一次性分离出来。在拜占庭场景中，对手客户机通过高斯随机更新进行通信，每个对手形成自己的集群，因此需要几轮集群来过滤所有对手。在干净的场景中，跨簇相似度永远不会低于相似度阈值。因此，在这种情况下，客户群永远不会分离，CFL不会损害收敛性。*α交叉*

在CIFAR上的noise和label flip问题上，经过大约170个通信轮，跨集群相似性降低到阈值以下，这导致良性客户端与主集群分离（标记为红色）。虽然这对模型精度没有直接的负面影响，但它应该

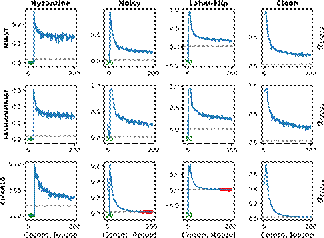


图2：在集群的前200个通信轮上集群候选不同性的发展*α交叉*

联合学习。当*α交叉*主星团下方的倾角被分成两个弧形-

式4。正确的聚类显示为greed，不正确的聚类显示为红色。如果集群正确，则图中会显示正确分离的客户机数量。

在实践中当然要避免。一种可能的补救办法是随着培训的进行降低的价值。

然而，在每一次实验中，所有的恶意客户端都首先被完全分离出来。FL和CFL在200轮沟通后最终达到的精度如表1所示。如我们所见，在对抗性场景中，CFL比常规FL获得了更高的准确度，与干净基线紧密匹配，在传统FL偏离的拜占庭环境中，准确度差异最大。

# 5.    结论

集群联合学习是最近提出的对常规联邦学习框架的扩展，它通过自动将客户机分成可联合训练的数据分布的集群来提高健壮性和灵活性。在这项工作中，我们研究了CFL在拜占庭场景中的应用，拜占庭场景可以看作是一个特殊的客户数据分布，其中一个特定的分布被声明为良性分布，而所有其他分布都被认为是对抗性的。在三个不同的数据集和三种不同类型的对抗性场景下的实验中，我们发现CFL（无需任何修改）能够在相对较少的通信回合内过滤出敌方客户。这表明，即使在客户没有形成明显的集群结构的情况下，CFL也可以比常规的联合学习提供显著的优势。

# 6.    参考文献

[1] H Brendan McMahan，Eider Moore，Daniel Ramage，Seth Hampson等人，“从分散数据中高效学习深层网络”，arXiv预印本arXiv:1602.056292016年。

[2] Jakub Konecnˇy、H Brendan McMahan、Daniel Ram-`age和Peter Richtárik，“联邦优化：设备智能的分布式机器学习”，arXiv预印本arXiv:1610.025272016年。

[3] Keith Bonawitz、Vladimir Ivanov、Ben Kreuter、Antonio Marcedone、H Brendan McMahan、Sarvar Patel、Daniel Ramage、Aaron Segal和Karn Seth，“隐私保护机器学习的实用安全聚合”，IACR加密电子打印档案，2017年第281页。

[4] Keith Bonawitz，Hubert Eichner，Wolfgang Grieskamp，

Dzmitry Huba，Alex Ingerman，Vladimir Ivanov，Chloe Kiddon，Jakub Konecny，Stefano Mazzochi，H Brendan McMahan等人，“在规模上走向联合学习：系统设计”，arXiv预印本arXiv:1902.010462019年。

[5] Tian Li，Anit Kumar Sahu，Ameet Talwalkar和Virginia Smith，“联合学习：挑战、方法和未来方向”，arXiv预印本arXiv:1908.078732019年。

[6] Felix Sattler、Klaus Robert Müller和Wojciech Samek，“集群联合学习：隐私约束下的模型无关分布式多任务优化”，arXiv预印本arXiv:1910.01991年2019年。

[7] Peva Blanchard，Rachid Guerraoui，Julien Stainer等，“对手的机器学习：拜占庭容忍梯度下降”，神经信息处理系统进展，2017年，第119-129页。

[8] El Mahdi El Mhamdi、Rachid Guerraoui和Sébastien Rouault，“拜占庭分布式学习的隐藏漏洞”，arXiv预印本arXiv:1802.079272018年。

[9] Felix Sattler、Simon Wiedemann、Klaus Robert Müller和Wojciech Samek，“非iid数据的鲁棒和通信效率联邦学习”，IEEE神经网络和学习系统事务（出版），2019年。

[10] 赵悦，李孟，赖良珍，纳文素达，达蒙·西文和维卡斯·钱德拉，“非iid数据的联合学习”，arXiv预印本arXiv:1806.005822018年。

[11] 陈玉东，苏丽丽，徐嘉明，“对抗环境下的分布式统计机器学习：拜占庭梯度下降”，计算机系统测量与分析学术论文集，第1卷，第2期，第44页，2017年。

[12] 陈凌娇，王洪义，查理查尔斯，和迪米特里斯帕帕帕利奥普洛斯，“德拉科：通过冗余梯度的拜占庭弹性分布式训练”，arXiv预印本arXiv:1803.098772018年。

[13] Luis Muñoz González，Kenneth T Co和Emil C Lupu，“通过自适应模型平均进行拜占庭鲁棒联邦机器学习”，arXiv预印本arXiv:1909.051252019年。

[14] Luis Muñoz González、Battista Biggio、Ambra Demontis、Andrea Paudice、Vasin Wongrasamee、Emil C Lupu和Fabio Roli，“用反向梯度优化向深度学习算法下毒”，发表于第10届ACM人工智能与安全研讨会论文集。ACM，2017年，第27-38页。

[15] 冯克莱门特，郑永元，和伊万贝沙特尼克，“减轻联邦学习中毒中的西比尔斯”，arXiv预印本arXiv:1808.048662018年。

[16] Eugene Bagdasaryan，Andreas Veit，Yiqing Hua，Deborah Estrin，和Vitaly Shmatikov，“如何后门联合学习”，arXiv预印本arXiv:1807.004592018年。

[17] Arjun Nitin Bhagoji、Supriyo Chakraborty、Prateek Mittal和Seraphin Calo，“通过对抗性视角分析联合学习”，arXiv预印本arXiv:1811.124702018年。

[[1]](" \l "_ftnref1" \o ")我们称之为双划分，iff∈c1，j∈c2⇒∃c∈cst.∈c，即iff只将客户机从不同的集群中分离出来。（c.f.[6]）。*c*1*c*2 *对的我i、 j*