

差异私人联合学习：

客户端级别的透视图



Robin C.Geyer，Tassilo Klein，Moin Nabi1*,*211

SAP SE，苏黎世ETH12

geyerr@ethz.ch, 塔斯洛.克莱因@sap.com网站，m。nabi@sap.com∗

# 摘要

联合学习是隐私保护的最新进展。在这种情况下，一个受信任的馆长聚合了由多个客户端以分散方式优化的参数。然后，生成的模型被分发回所有客户机，最终聚合成一个联合代表模型，而不必显式地共享数据。然而，该协议容易受到差分攻击，这种攻击可能来自于联邦优化过程中参与的任何一方。在这种攻击中，通过对分布式模型的分析，揭示了客户在培训过程中的贡献和有关其数据集的信息。针对这个问题，我们提出了一个客户端差分隐私保护联邦优化算法。其目的是在培训期间隐藏客户的贡献，平衡隐私损失和模型性能之间的权衡。实证研究表明，在有足够多的参与客户的情况下，我们提出的方法可以在模型性能上以很小的代价维持客户级别的差异隐私。

# 1        介绍

近年来，机器学习中的安全问题越来越受到人们的关注。这在很大程度上归因于大数据与深度学习相结合的成功，以及为数据挖掘创建和处理越来越大的数据集的冲动。然而，随着越来越多的机器学习服务成为我们日常生活的一部分，利用我们的数据，必须采取特殊措施来保护隐私。不幸的是，仅仅匿名通常是不够的[7,2]，标准的机器学习方法很大程度上忽略了隐私方面。

在联合学习[5]中，模型是由多个客户以分散的方式学习的。学习是转移到客户和只有学习的参数集中在一个可信的馆长。然后，这个策展人将聚合模型分发回客户机。

客户不透露他们的数据是privacyprotection的一个进步，然而，当一个模型以传统的方式学习时，它的参数会显示有关在培训期间使用的数据的信息。为了解决这一问题，文献[1]提出了学习算法的差分隐私（dp）[3]的概念。其目的是确保学习的模型不会揭示某个数据点是否在训练期间被使用。

我们提出了一种将dp预存机制与联合学习相结合的算法。然而，与[1]相反，我们的目标不是仅保护wr.t.单个数据点。相反，我们希望确保所学模型不会揭示客户是否参与了分散式培训。这意味着客户机的整个数据集都受到保护，免受来自其他客户机的差异攻击。



∗备选电子邮件地址：cyrusgeyer@gmail.com

第31届神经信息处理系统会议（NIPS 2017），加利福尼亚州长滩，美国。

我们的主要贡献：首先，我们展示了在联邦学习中，当模型性能保持较高时，客户的参与是可以隐藏的。我们证明，我们提出的算法可以在模型性能损失较小的情况下实现客户端级别的差异隐私。同时发表的一项独立研究[6]提出了客户级dp的类似程序。然而，实验设置有所不同，[6]还包括元素级隐私措施。其次，我们建议在分散训练中动态调整dp保持机制。实证研究表明，通过这种方式可以提高模型的性能。这与具有差别隐私权的集中培训的最新进展形成了鲜明对比，如果这种适应是没有好处的。我们可以将这种差异与以下事实联系起来：与集中学习相比，联合学习中的梯度在整个训练过程中对噪声和批量大小表现出不同的敏感性。

# 2        背景

2.1联合学习

在联合学习[5]中，馆长和客户之间的交流可能会受到限制（例如移动电话）和/或容易被拦截。联邦优化的挑战是学习一个客户机和管理员之间的信息开销最小的模型。此外，客户的数据可能是非IID的、不平衡的和大规模分布的。最近[5]提出的“联合平均”算法解决了这些挑战。在馆长和客户之间的多轮沟通中，培训一个中心模型。在每一轮沟通中，馆长会将当前的中心模型分发给一小部分客户。然后，客户机执行局部优化。在一个小批量通信过程中，可能会采取多个梯度下降的步骤来减少客户之间的沟通。接下来，优化后的模型被发送回馆长，馆长将其聚合（例如平均）以分配一个新的中心模型。根据新的中央模式的性能，培训要么停止，要么开始新一轮的交流。在联合学习中，客户机从不共享数据，只共享模型参数。

2.2差异隐私学习

在集中学习模型的情况下，对数据级的dp保护进行了大量的研究。这可以通过在学习过程中加入dp保持随机机制（例如高斯机制）来实现。

我们在随机机制中对dp的定义与[1]相同：

一个随机机制：D→R，具有域和范围，满足（ǫ，δ）-微分隐私，如果对任意两个相邻的输入∈D和输出的任何子集⊆R，则认为[M（D）∈S]≤eǫPr[M（D′）∈S]+δ。在这个定义中，解释了普通差别隐私被破坏的可能性。*米DRd、 d*′*SPδǫ*

高斯机制（GM）近似实值函数：D→R，具有微分私有机制。具体地说，GM将高斯噪声校正到函数数据集灵敏度上。该灵敏度定义为绝对距离kf（d）−f（d′）k2的最大值，其中和是两个相邻的输入。GM定义为（d）=f（d）+N（0，σ2Sf2）。*f平方英尺d*′*d米*

在下面，我们假设和是固定的，并评估对GM关于（d）的单一近似的询问。我们可以根据≤-dp被破坏的概率≤*σǫfǫδ*

（文献[4]中定理3.22）。值得注意的是，如果*δ*

因此，为了保护隐私，会计师会跟踪。一旦达到一定的门槛，总经理将不再回答任何新的询问。*δδ*

最近，[1]提出了一种差分私有随机梯度下降算法（dp-SGD）。dpSGD的工作原理类似于小批量的梯度下降，但是梯度平均步骤是用GM来近似的，另外，通过随机抽样来分配小批量。为了得到修复，隐私会计师会跟踪并在达到阈值后停止培训。直观地说，这意味着一旦学习模型揭示某个数据点是否是训练集的一部分的概率超过某个阈值，训练就停止。*ǫδ*

2.3联邦优化中的客户端差异隐私

我们建议在联合学习中加入随机机制。然而，与[1]相反，我们并不旨在保护单个数据点在学习模型中的贡献。相反，我们的目标是保护整个客户机的数据集。也就是说，我们希望确保学习的模型不会显示客户是否在保持高模型性能的同时参与了分散式培训。

# 3        方法

在联邦优化[5]的框架中，中央馆长在每轮沟通后平均客户模型（即权重矩阵）。在我们提出的算法中，我们将用随机机制改变和近似这个平均值。这样做是为了将单个客户的贡献隐藏在聚合中，从而隐藏在整个分散学习过程中。

我们用来近似平均值的随机机制包括两个步骤：

1.    随机子抽样：取客户总数。在每个通信轮中，对大小≤K的随机子集进行抽样。然后，馆长将中心模型分发给这些客户。中心模型由客户对其数据进行优化。客户在*KZt公司公吨重量Zt公司*现在拥有独特的本地模型。优化后的局部模型与中心模型之间的差异称为客户机*k*&apos;s update∆wk=wk−wt。在每一轮沟通结束时，更新信息将发送回中央馆长。

2.    扭曲：高斯机制用于扭曲所有更新的总和。这需要了解集合相对于求和运算的灵敏度。我们可以通过使用缩放版本而不是真正的更新来增强一定的敏感性：△w´k=

. 缩放可确保第二个范数被限制为∀k，k△w´kk2<S。因此，缩放更新相对于求和运算的灵敏度上限为。GM现在将噪声（按比例缩放到灵敏度）添加到所有缩放更新的总和中。用通用汽车公司的产出除以，可以得到一个近似于所有客户更新的真实平均值，同时防止有关个人的关键信息泄露。*SS公吨*

通过将此近似值添加到当前中心模型中，可以分配一个新的中心模型。*重量*+1 *重量*

剪辑的更新总数*S*

{

近似更新和的高斯机制

在将1/mt分解为高斯机制时，我们注意到平均值的失真是由噪声方差控制的。但是，这种失真不应超过一定的限度。否则，过多的子采样平均值信息会被附加噪声破坏，并且不会有任何学习进度。GM和随机子抽样都是随机机制。（实际上，[1]正是在dp-SGD中使用了这种平均近似。但是，它用于梯度平均，在每次迭代中隐藏一个数据点的梯度）。因此，还定义了当随机机制提供平均近似值时产生的隐私损失。*S*2*σ*2*/米σ米*

为了跟踪这种隐私损失，我们利用Abadi等人提出的矩会计。[1] 一。与标准合成定理（3.14 in[4]）相比，这种计算方法对所产生的隐私损失提供了更严格的限制。每次馆长分配一个新的模型，会计评估给定的，和。一旦达到某个阈值，即客户贡献被披露的可能性过高，应停止培训。的阈值的选择取决于客户端的总数*δǫσ米δδK*. 为了确保许多人的隐私不会以泄露少数人的全部信息为代价，我们必须确保，更多细节请参考[4]第2.3章。

选择：当削减贡献时，有一个权衡。一方面，应选择较小的噪声，使噪声方差保持较小。另一方面，人们希望尽可能多地保持最初的贡献。按照文献[1]提出的方法，在每个通信轮中，我们计算所有未截取贡献的中值范数，并以此作为剪裁界=中值{△wk}k∈Z。我们没有使用随机机制来计算中值，严格来说，这是对隐私的侵犯。然而，通过中位数的信息泄漏是很小的（未来的工作将包含这样的隐私措施）。*SSSt*

选择和：对于固定，比率=σ2/m控制失真和隐私丢失。因此，越高和越低，隐私损失就越高。隐私会计师告诉我们，对于固定值=σ2/m，也就是说，对于相同程度的失真，隐私损失较小，两者都很小。因此，失真率的上限和子抽样客户数量的下限将导致选择。然而，很难估计其下限。也就是说，因为联合设置中的数据是非IID的，来自客户端的贡献可能非常不同。因此，我们将客户端之间的差异定义为客户端更新之间的相似性度量。*σ米Srσ米rσ米r米σ米风险投资*

定义。设△wi，j（i，j）△w∈R*定义-表单更新中的th参数问*×p公司*，在某个沟通环节t、 为了清楚起见，我们现在将不再使用通信轮的具体索引。*

*参数方差*（i，j）*全部K客户定义为，*

*K*

                                                                                     1 X 2个*k*

*变量*[△w]=（△w−µ），K*i、 ji、 ji、 j*

*k*=0

*哪里µi、 j*=P=1△w*K*1 *Kk公司i、 jk。*

*然后我们定义Vc作为更新矩阵中所有参数方差的和，*

*.*

*此外，更新比例美国的定义是，*

*.*

# 4        实验

为了测试我们提出的算法，我们模拟了一个联邦设置。为了便于比较，我们选择了与文献[5]相似的实验装置。我们将排序后的MNIST集分成碎片。因此，每个客户端都有两个碎片。这样一来，大多数客户只会有两位数的样本。因此，单个客户机无法根据其数据训练一个模型，使其对所有10位数字都达到高分类精度。

我们正在研究∈{100100010000}场景下联邦设置中的差异隐私。在每一个设置中，客户机可以得到600个数据点。对于∈{100010000}，重复数据点。*KK*

对于所有三个场景∈{100100010000}，我们对以下参数执行了交叉验证网格搜索：*K*

•每个客户的批次数*B*

•在每个客户端上运行的时间段*E*

•每轮参与的客户数量*米*

•GM参数*σ*

根据[1]，我们将值定为8。在培训期间，我们会跟踪会计师隐私的损失。一旦100名、1000名和10000名客户分别达到−3、e−5、e−6，培训即停止。此外，我们还分析了培训过程中客户之间的差异。上述实验代码可从以下网址获得：https://github.com/cyrusgeyer/DiffPrivate\u-FedLearning。*ǫδe*

算法1客户端差异私有联合优化。是参与客户机的数量；是本地小批量大小，是本地时间段的数量，*KBEη*是学习率，是GM的方差集。确定参与客户的数量

在每一轮交流中。定义我们的目标dp。是-dp被破坏的概率的阈值。是超过的通信轮数。B是一个集合，用于保存客户机的数据，这些数据被分成B大小的批*ǫ问δǫTδ问*



1： 过程服务器执行

2： 初始化：会计（ǫ，K）⊲初始化权重和主会计师3：每轮=1,2，。。。做*w*0*,t*

4： ←会计（mt，σt）⊲会计返回当前第5轮的私人损失：如果那么返回如果隐私预算被花费，则返回当前模型*δδ>Q重量*

6： ←随机集客户端随机分配一组K个以上的客户端7个：对于每个客户端∈Zt并行do*Zt公司公吨⊲公吨k*

第八章：ClientUpdate（）客户端k的更新和更新的范数9：=客户端更新规范的中值{ζk}k∈Zmedian*k、 重量⊲St⊲*

*k*

十：更新中心模型

11： 函数CLIENTUPDATE（）*k、 重量*

12： ←重量*w*

13： 对于每个本地历元=1,2，…E do*我*

14： 批处理∈B do*b*

15： ←w−η∇L（w；b）⊲小批量梯度下降*w*

16： △wt+1=w−wt⊲客户端本地模型更新

17： =k△wt+1k2⊲更新范数*ζ*

18： return△wt+1，ζ⊲返回剪裁更新和更新范数



|  |
| --- |
|  |
|  |  |

表1：差异私有联合学习（Dp）的实验结果以及与非差异私有联合学习（non-Dp）的比较。隐私预算=8的情况下有100、1000和10000个客户。差异隐私被破坏的最高可接受概率。一旦到达，训练停止。准确度表示为“ACC”，通信次数表示为“CR”，通信成本表示为“CC”。*ǫδ*′*ǫδ*′

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 客户 | *δ*′ | ACC公司 | CR公司 | 科科斯群岛 |
| 非dp | 100 | - | 0.97 | 380 | 38000 |
| Dp | 100  1000 | e-3和e-5 | 0.78  0.92 | 11  54 | 550  11880 |
|  | 10000 | e-6型 | 0.96 | 412 | 209500 |

# 5        结果

在交叉验证网格搜索中，我们寻找那些在保持在各自边界以下时达到最高精度的模型。另外，当多个模型达到相同的精度时，需要较少通信轮数的模型是首选的。*δ*

表1保存了∈{100100010000}的最佳模型*K*. 我们列出了准确度（ACC）、所需的沟通轮数（CR）以及由此产生的沟通成本（CC）。沟通成本定义为客户在培训过程中发送模型的次数，即。。此外，作为一个基准，表1还包含了表现最好的非差异私有模型的ACC、CR和CC*K*=100。在图1中，在整个培训过程中，所有四个表现最好的模型的准确性都被描绘出来了。

在图2中，对于=100的非差异私有联邦优化的精确度，再次描述了在训练过程中客户端之间的差异和更新比例。*K*

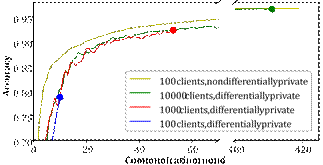


图1：在分散化培训过程中，根据客户持有的非IID MNIST数据进行数字分类的准确性。对于差异私有联合优化，精度曲线末端的点表示达到了-阈值，因此停止了训练。*δ*

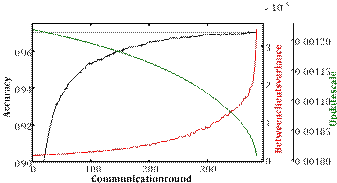


图2:100个客户机的场景，非差异私有：准确性，客户机之间的差异和联邦优化过程中的更新规模。

# 6        讨论

正如直觉所预期的，参与客户的数量对所实现的模型性能有重大影响。对于100个和1000个客户机，模型精度不收敛，并且显著低于非差异私有性能。然而，对于∈{1001000}的78%和92%的准确率仍然远远高于客户仅在自己的数据上进行培训时能够达到的任何目标。在这个数量级和差别隐私是最重要的领域，这种模型仍然会使任何参与的客户受益。例如医院。几百人可以共同学习一个模型，而有关某个特定医院的信息却被隐藏起来。此外，联合学习的模型可以用作进一步客户端培训的初始化。*KK*

当=10000时，差分私有模型几乎达到了无差别私有模型的精度。这表明，对于涉及多个参与方的场景，差异隐私几乎不会影响模型性能。这些场景包括移动电话和其他消费设备。*K*

在交叉验证网格搜索中，我们还发现在训练过程中提高模型性能。当观察一个单一的早期通信回合时，同时降低并且以保持恒定的方式，几乎不会影响该轮中的精度增益。但是，当这两个参数都降低时，隐私损失会减少。这意味着在以后的训练中，在隐私预算耗尽之前，可以进行更多的交流。在随后的通信回合中，为了获得准确度，不可避免地需要大量的数据，为了改进模型，必须考虑更高的隐私成本。*公吨公吨σtσt*2*/公吨公吨*

这一观察结果可以与信息理论在学习算法方面的最新进展联系起来。如图2所示，Shwartz-Ziv和Tishby[8]指出，我们可以区分两个不同的训练阶段：标签拟合和数据拟合阶段。在标签拟合阶段，客户机的更新是相似的，因此是低的，如图2所示。但是在这个初始阶段是高的，因为对随机初始化的权重进行大的更新。在数据拟合阶段上升。个别更新△wk看起来不太相似，因为每个客户机都对其数据集进行了优化。然而，当全局模型的局部最优值被逼近时，精度会收敛，贡献在一定程度上相互抵消。图2显示了和的这些依赖关系。我们可以得出这样的结论：）在早期的沟通环节中，小部分客户可能仍会贡献一个平均更新△wt代表真实的数据分布），在后期阶段，需要一个平衡（因此更大）的客户来达到更新的特定代表性。）高使早期更新不易受噪音影响。*风险投资坎特伯雷大学风险投资坎特伯雷大学风险投资坎特伯雷大学我二三坎特伯雷大学*

# 7        结论

我们能够通过第一次实证研究表明，在客户层面上的差异隐私是可行的，并且当足够多的参与者参与时，可以达到高模型精度。此外，我们还表明，仔细调查数据和更新分布可以优化隐私预算。在未来的工作中，我们计划从信噪比、数据代表性和客户方差之间的关系等方面推导出最优界，并进一步研究与信息论的联系。

# 工具书类

[1] M.Abadi，A.Chu，I.Goodfello，H.Brendan McMahan，I.Mironov，K.Talwar和L.Zhang。深度学习与差异隐私。ArXiv电子印刷品，2016年。

[2] L.Backstrom、C.Dwork和J.Kleinberg。你为什么是3579X？：匿名社交网络、隐藏模式和结构隐写术。第16届万维网国际会议记录，WWW&apos;07，181-190页，美国纽约，2007年。ACM公司。

[3] C.德沃克。差别隐私。第33届国际自动机会议记录，语言和编程-第二卷，ICALP&apos;06，第1-12页，柏林，海德堡，2006年。斯普林格·维拉格。

[4] C.德沃克和A.罗斯。差异隐私的算法基础。找到了。趋势理论。计算机。《科学》，9（3&#8211；4）：211–4072014年8月。

[5] B.麦克马汉，E.摩尔，D.拉马奇，S.汉普森和B.A.y.阿尔卡斯。从分散数据中有效地学习深层网络。编辑A.Singh和J.Zhu，《第20届国际人工智能与统计会议论文集》，机器学习研究论文集第54卷，第1273-1282页，美国佛罗里达州劳德代尔堡，2017年4月20-22日。PMLR。

[6] H.B.McMahan，D.Ramage，K.Talwar和L.Zhang。在不损失准确性的情况下学习不同的私人语言模型。更正，abs/1710.069632017年。

[7] A.Narayanan和V.Shmatikov。大型稀疏数据集的鲁棒去匿名化。2008年IEEE安全与隐私研讨会（sp 2008），第111-125页，2008年5月。

[8] R.S.Ziv和N.Tishby。通过信息打开深层神经网络的黑匣子。更正，abs/1703.00810，2017年。