Privacy-preserving deep learning

Reza Shokri, Vitaly Shmatikov
The University of Texas at Austin; Cornell Tech.

Shilong Zhao [†]

December 5, 2022

[†] Institute of Computing Technology
University of Chinese Academy of Sciences

Outline

Introduction

Introduction

联邦学习开山之作：Privacy-Preserving Deep Learning, CCS'15.

研究背景:

1. 基于神经网络的深度学习算法在语音、图像、文本识别、机器翻译等领域取得了突破;
2. 然而，大规模数据集的训练带来了巨大的隐私问题。

隐私问题:

1. 收集这些数据的公司可以永久保存数据;
2. 图像和语音记录通常包含意外捕获的敏感信息，如人脸、车牌等;
3. 公司保存的用户数据会受到相关情报机构的无授权监视。

贡献:

1. 在训练模型的同时保护数据的隐私性;
2. 定量地测量模型的准确性和隐私性之间权衡;
3. 高效性，最大限度降低通信成本和计算成本。

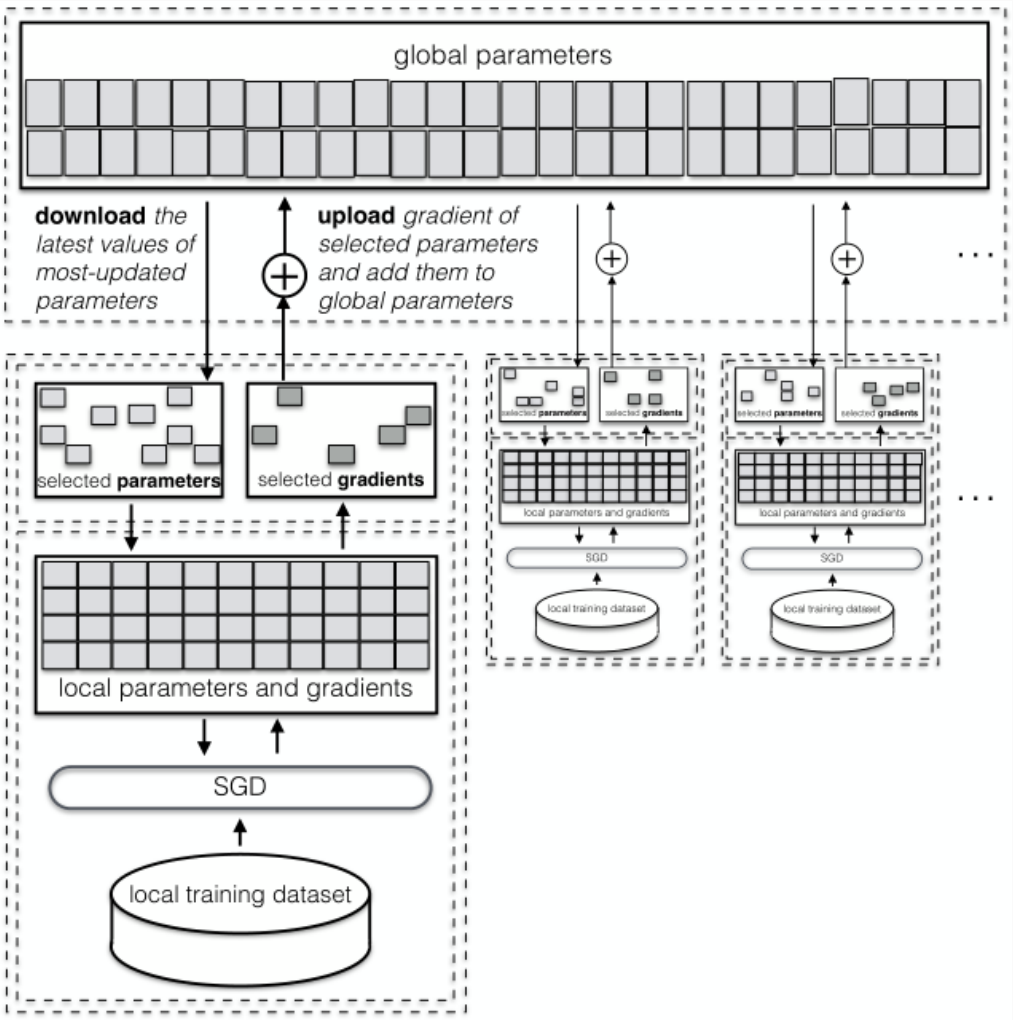
Proposed Method

Distributed Selective SGD

选择性参数更新：
在随机梯度下降过程中，某些参数对神经网络目标函数的贡献更大。

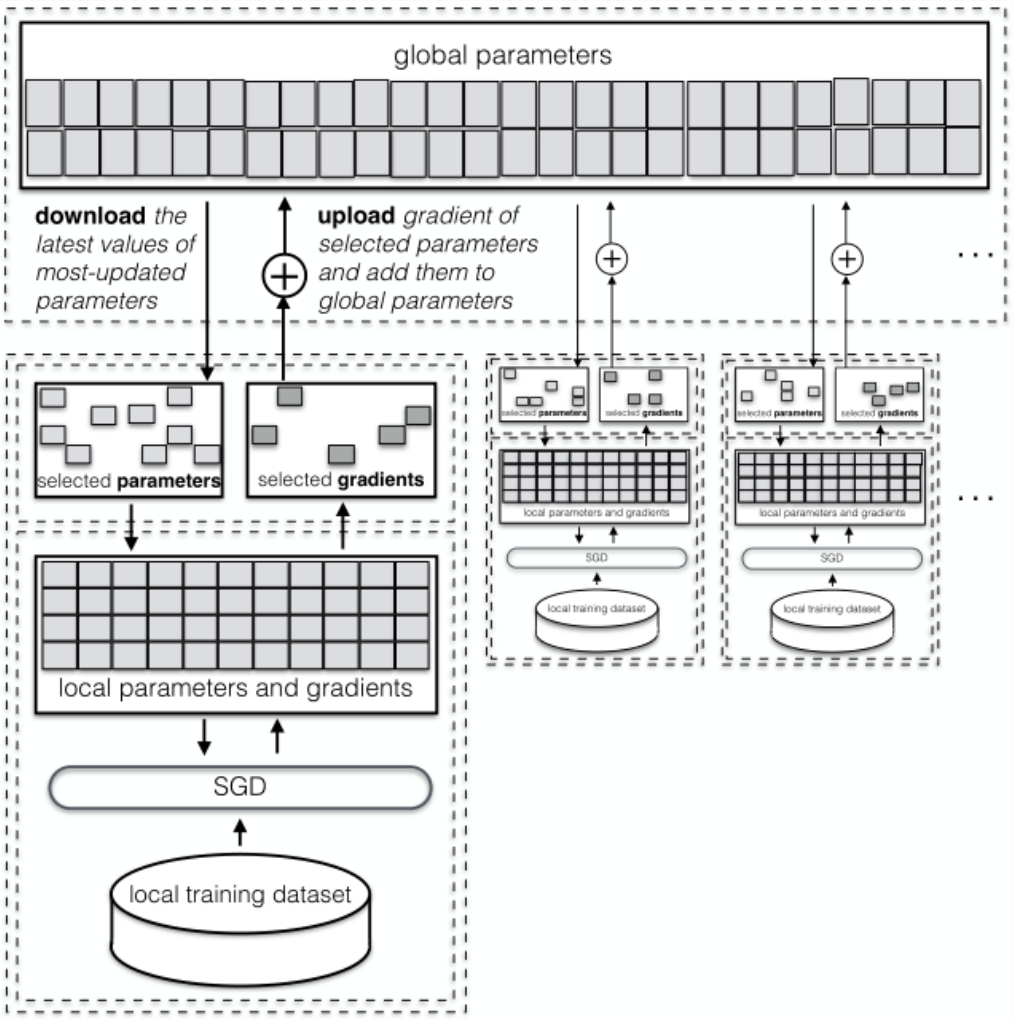
分布式协作机器学习：
参与者可以直接交换梯度，或者通过一个受信任的中央服务器交换梯度。
每个参与者从服务器下载参数的子集，并使用它们更新自己的本地模型。

System Overview



- 系统中存在一个参数服务器，它负责维护所有各方可用的最新参数值。
- 系统中的参与方数量记为 N，每个参与方都有一个本地的私有数据集。

System Overview



每个参与方在本地运行 SGD 算法，优化参数，然后执行参数交换协议。

参与方将选定的参数上传至服务器后，服务器进行运算后更新全局参数，以供参与方下次下载。

DSSGD FOR PARTICIPANT

Choose initial parameters $\mathbf{w}^{(i)}$ and learning rate α.

Repeat until an approximate minimum is obtained:

1. Download $\theta_d \times |\mathbf{w}^{(i)}|$ parameters from server and replace the corresponding local parameters.
2. Run SGD on the local dataset and update the local parameters $\mathbf{w}^{(i)}$ according to (1).
3. Compute gradient vector $\Delta\mathbf{w}^{(i)}$ which is the vector of changes in all local parameters due to SGD.
4. Upload $\Delta\mathbf{w}_S^{(i)}$ to the parameter server, where S is the set of indices of at most $\theta_u \times |\mathbf{w}^{(i)}|$ gradients that are selected according to one of the following criteria:
   - *largest values*: Sort gradients in $\Delta\mathbf{w}^{(i)}$ and upload θ_u fraction of them, starting from the biggest.
   - *random with threshold*: Randomly subsample the gradients whose value is above threshold τ.

   The selection criterion is fixed for the entire training.

DSSGD ON THE SERVER

Choose initial global parameters $\mathbf{w}^{(global)}$.

Set vector **stat** to all zero.

EVENT: A participant uploads gradients $\Delta \mathbf{w}_S$.

- For all $j \in S$:
  - Set $\mathbf{w}^{(global)} := \mathbf{w}^{(global)} + \Delta \mathbf{w}_j$
  - Set $stat_j := stat_j + 1$

EVENT: A participant downloads θ parameters.

- Sort **stat**, and let I_θ be the set of indices for **stat** elements with largest values.
- Send $\mathbf{w}^{(global)}_{I_\theta}$ to the participant.

Parameter exchange protocol

参数交换协议：

1. 轮询: round robin
   参与者按固定顺序执行协议，每个人从服务器下载最新参数的一部分，运行本地训练，并上传选定的梯度，下一个参与者按照顺序执行相同操作。
2. 随机: random order
   参与者以随机顺序下载、学习和上传，但对服务器的访问是原子性的，即参与者在读取之前锁定服务器，写入之后释放锁。

异步性: asynchronous
参与者不协调。当一个参与者在训练一组参数时，其他人可能会在训练结束前在服务器上更新这些参数。

Pravicy

Preventing direct leakage

在训练模型时防止直接泄漏。与传统深度学习不同的是，在我们的系统中，参与者不会向任何人透露他们的训练数据集，从而确保了数据的高度隐私性。本地数据集是动态的，大小是保密的，在每一轮 SSGD 中可以使用不同的数据样本。参与者也可以随时删除他们的训练数据。

在使用模型时。所有参与者都学习模型，因此可以在本地和私下使用它，无需与其他参与者进行任何通信，也无需向任何人透露输入数据或模型的输出。因此，与传统深度学习相比，该模型在使用过程中没有泄漏。

Preventing indirect leakage

1. 参与方可以决定共享哪部分参数，因此可以避免共享涉及隐私的参数；
2. 使用差分隐私来确保参数更新不会泄露关于训练数据集中任何单个点的太多信息。

定义

一般将满足差分隐私的函数称为一个机制（Mechanism）。如果对于所有临近数据集（Neighboring Dataset）x 和 x' 和所有可能的输出 S，机制 F 均满足

$$\frac{\mathrm{Pr}[F(x) = S]}{\mathrm{Pr}[F(x') = S]} \leq e^{\epsilon} \tag{4.1}$$

则称机制 F 满足差分隐私。

$\mathrm{F}()$ 计算参数的梯度，并选择其中哪些与其他参与者共享。

Preventing indirect leakage

这可能造成有两个种泄漏：

1. 如何选择用于共享的梯度；
2. 共享梯度的实际值。

为了减轻这两种类型的泄漏，文章使用了稀疏向量技术：

1. 随机选择值高于阈值的一小部分梯度;
2. 给准备共享的梯度值增加噪声。

- Let ϵ be the total privacy budget for one epoch of participant i running DSSGD, and let Δf be the sensitivity of each gradient
- Let $c = \theta_u |\Delta \mathbf{w}|$ be the maximum number of gradients that can be uploaded in one epoch
- Let $\epsilon_1 = \frac{8}{9}\epsilon, \epsilon_2 = \frac{2}{9}\epsilon$
- Let $\sigma(x) = \frac{2c\Delta f}{x}$

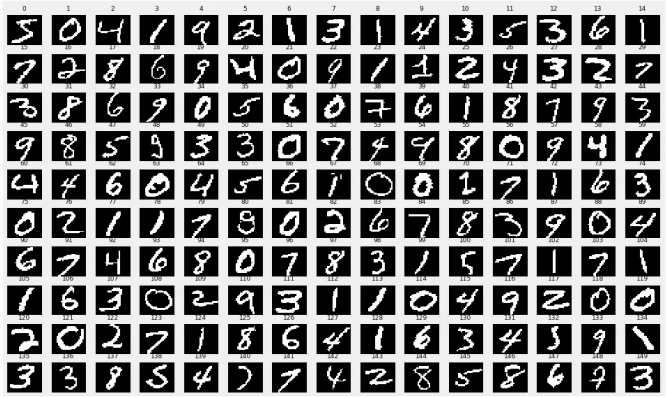
1. Generate fresh random noise $r_\tau \sim \text{Lap}(\sigma(\epsilon_1))$
2. Randomly select a gradient $\Delta w_j^{(i)}$
3. Generate fresh random noise $r_w \sim \text{Lap}(2\sigma(\epsilon_1))$
4. If $\text{abs}(\text{bound}(\Delta w_j^{(i)}, \gamma)) + r_w \geq \tau + r_\tau$, then
   (a) Generate fresh random noise $r'_w \sim \text{Lap}(\sigma(\epsilon_2))$
   (b) Upload $\text{bound}(\Delta w_j^{(i)} + r'_w, \gamma)$ to the parameter server
   (c) Charge $\frac{\epsilon}{c}$ to the privacy budget
   (d) If number of uploaded gradients is equal to c, then Halt Else Goto Step 1
5. Else Goto Step 2

Experiment

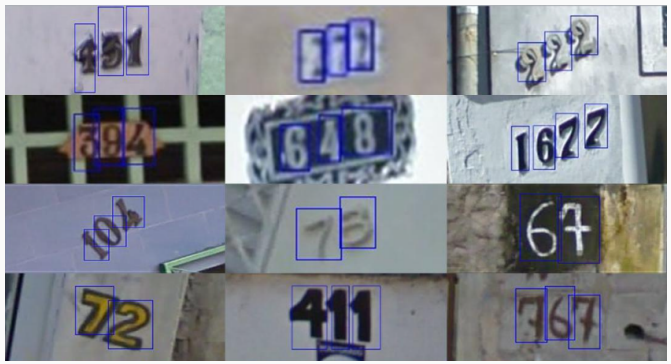
Experiment

数据集: MINST & SVHN

MINST (手写数字数据集)



SVHN (门牌号码数据集)



	MNIST	SVHN
train	60,000	100,000
test	10,000	10,000

Table 2: Size of training and test datasets

	MNIST	SVHN
MLP	140,106	402,250
CNN	105,506	313,546

Table 3: Number of neural-network parameters

Conclusion

Conclusion

文章设计、实现并评估了一个实用的协作式深度学习系统（联邦学习框架），该系统在实用性和隐私性之间提供了一个有吸引力的折衷方案。

为了最大限度地减少隐私泄漏，文章展示了如何使用稀疏向量技术将差分隐私应用于参数更新，从而减少由于参数选择（即选择共享哪些参数）和共享参数值而造成的隐私损失。

Q&A

Q & A