

# 论文分享20211217

《Federated Optimization for Heterogeneous Networks》

## 背景

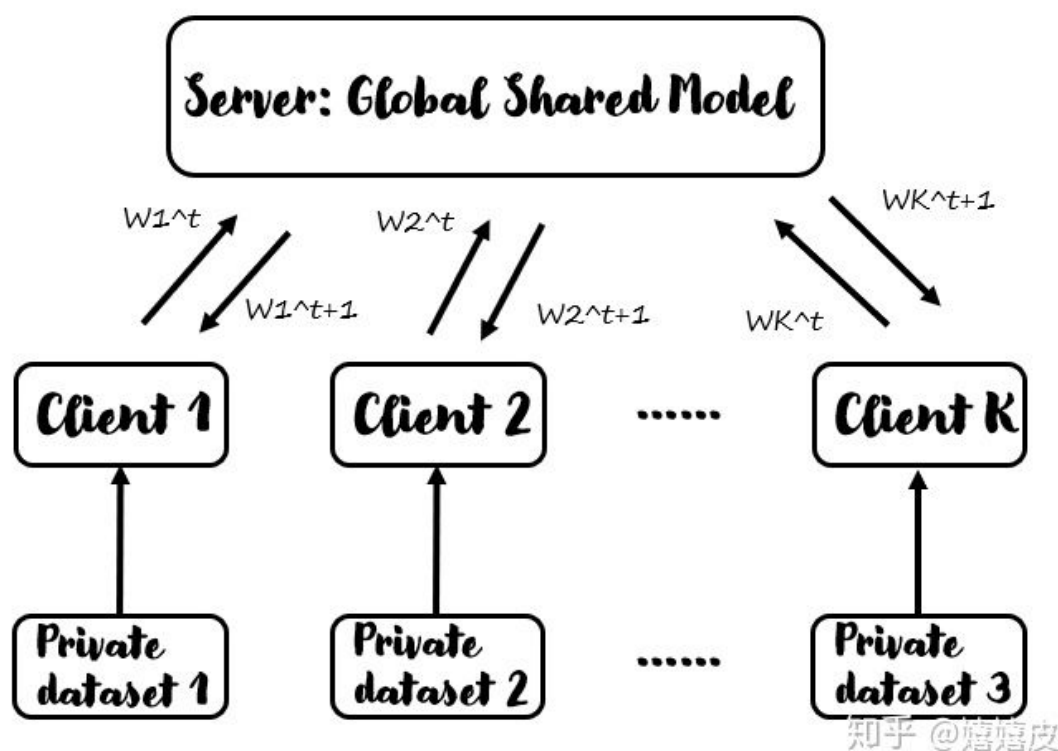
### 1. 联邦学习

一种**分布式**机器学习技术，能有效帮助多个机构在满足用户隐私保护、数据安全和政府法规的要求下，进行数据使用和机器学习的共同建模。联邦学习最早在 2016 年由谷歌提出，原本用于解决安卓手机终端用户在本地更新模型的问题；

### 2. 联邦学习的挑战

**数据异构**（主要是用户间数据Non-IID）：Non-IID非独立同分布，每个设备上的数据情况不一样  
**系统异构**（设备间通信和计算能力的差异）

### 3. FedAvg



FedAvg 的一般步骤：在每个 Communication Round 内，参与更新的  $n$  个设备在本地 SGD 迭代  $E$  epochs，然后在将模型上传到 Server 端进行聚合。一方面，本地迭代次数  $E$  的增大能减少通信成本；另一方面，不同 local objectives 在本地迭代次数过多后容易偏离全局最优解，影响收敛。

并且 FedAvg 这种固定  $E$  的操作没考虑到不同硬件间的差异，如果在固定时间内未完成  $E$  epochs 的迭代就会被系统 drop 掉。

### 4. Introduction

在实际的联邦设置中，当跨统计异质设备（即每个设备以不同的方式收集数据）学习时，它的表现并没有得到很好的理解。此文引入了一个处理统计异质性的框架FedProx，它将FedAvg作为一个特例。我们通过设备差异性假设为FedProx提供了收敛保证，相对于FedAvg，广义FedProx框架在异质网络中学习更具有稳健性和稳定性。



# Related Work


文章指出直接 drop 掉这些用户或者单纯把他们未迭代完成的模型进行聚合都会严重影响收敛的表现。因为丢掉的这些设备可能导致模型产生 bias，并且减少了设备数量也会对结果精度造成影响。



文章提出了 proximal term 来保证对这些未完成计算的 partial information 进行聚合

传统联邦学习的优化目标是最小化经验损失 Empirical risk:

作者在此处引入 proximal term( $u=0$ 时就是FedAvg):

将原来的 [公式] 变为 [公式] 的目的: 使得本地更新不要太过远离初始 global model, 在容忍系统异构性的前提下减少 Non-IID 的影响。

同时定义了 [公式] -inexact solution, 通过对 local function 的非精确求解, 动态调整本地迭代次数, 保证对异构系统的容忍度。

如果 [公式] 满足下式则称为 [公式] -inexact solution.

用 [公式] 作为本地迭代的 proxy, 值越小更新精度越高。

---

## Algorithm 2 FedProx (Proposed Framework)


---

**Input:**  $K, T, \mu, \gamma, w^0, N, p_k, k = 1, \dots, N$   
**for**  $t = 0, \dots, T - 1$  **do**  
    Server selects a subset  $S_t$  of  $K$  devices at random (each device  $k$  is chosen with probability  $p_k$ )  
    Server sends  $w^t$  to all chosen devices  
    Each chosen device  $k \in S_t$  finds a  $w_k^{t+1}$  which is a  $\gamma_k^t$ -inexact minimizer of:  $w_k^{t+1} \approx \arg \min_w h_k(w; w^t) = F_k(w) + \frac{\mu}{2} \|w - w^t\|^2$   
    Each device  $k \in S_t$  sends  $w_k^{t+1}$  back to the server  
    Server aggregates the  $w$ 's as  $w^{t+1} = \frac{1}{K} \sum_{k \in S_t} w_k^{t+1}$   
**end for**

---

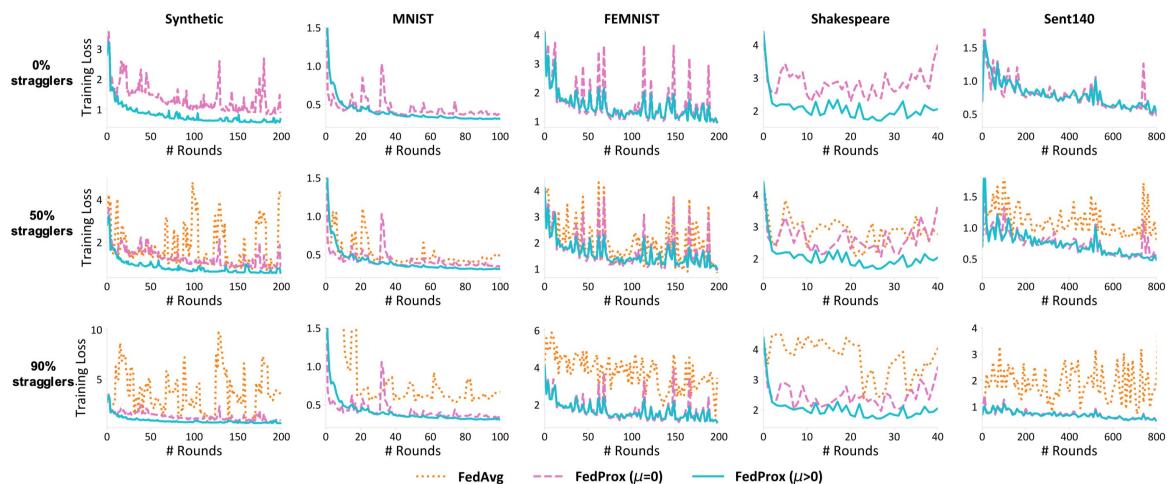
算法流程

替换了原本的 local epoch E,

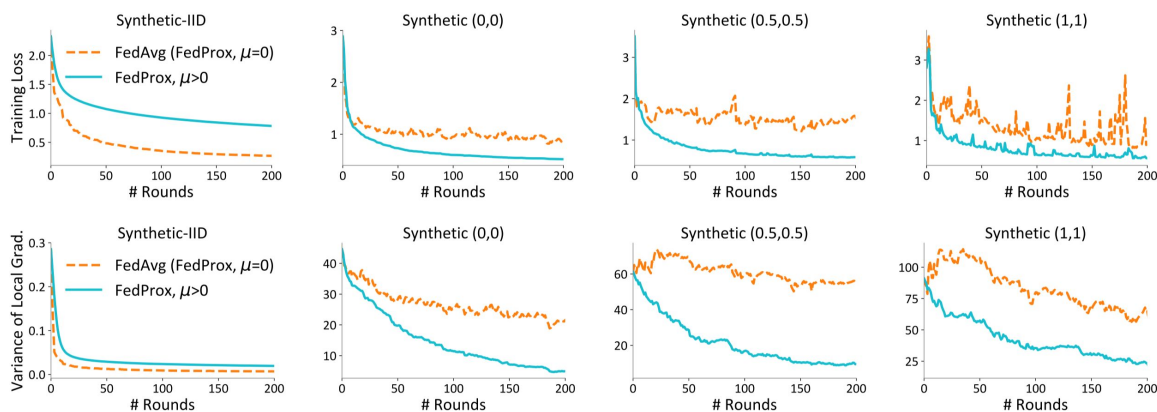
改为求解 [公式] -inexact minimizer.

---

# Experiment



验证 drop 那些 stragglers 对 FedAvg 和 FedProx 的影响



在相同更新 Epochs 下，验证异构数据对算法的影响

## Conclusion

文章将 FedAvg 作为 FedProx 的特殊情况，对原有算法进行了微调，考虑到了数据异构和系统异构的情况，并推导了收敛证明。

通过加入 proximal term 修正项，提高了整体收敛的稳定性。

通过对本地设备动态调整迭代轮数，保证了对系统异构的容忍性。

证明 FedProx 框架可以改善异质联邦网络中的收敛情况

## 思考

优点：巧妙的小方法、做了充分的实验去分析和验证。

缺点：非精确解  $\gamma$  的影响未知