



Resource-Adaptive Federated Learning with All-In-One Neural Composition

Yiqun Mei, Pengfei Guo, Mo Zhou, Vishal M. Patel
NeurIPS 2022

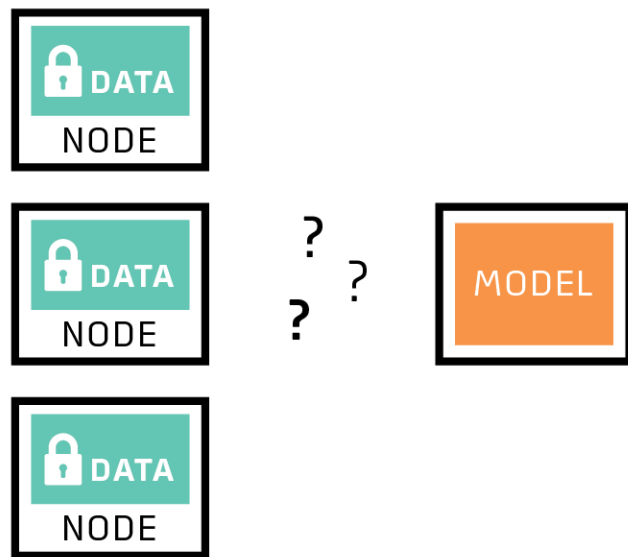
资源自适应的联邦学习 with 一体化神经网络组合

朱姚林

2024-11-12

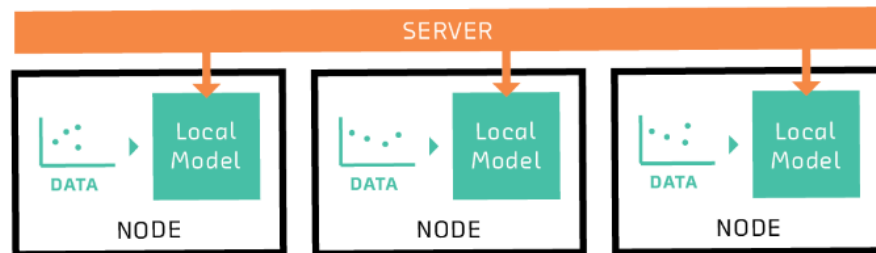
什么是联邦学习

- 数据诞生在边缘：智能手机、工业传感器、.....
- 去中心化数据：
 - 服务器不能收集客户端上的数据以保护隐私
 - 另外数据传输、存储也十分困难
- 联邦学习：
 - 设备在本地训练模型
 - 经过训练的模型被传输到中央服务器进行聚合

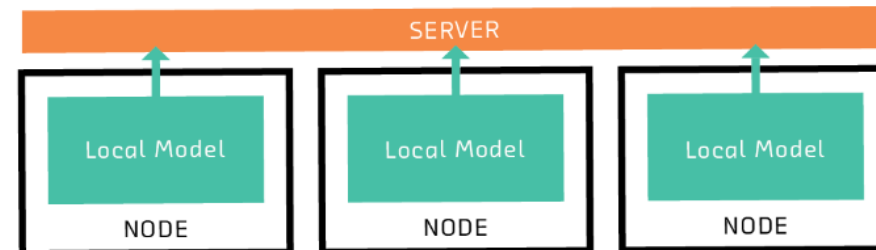


怎么做联邦学习

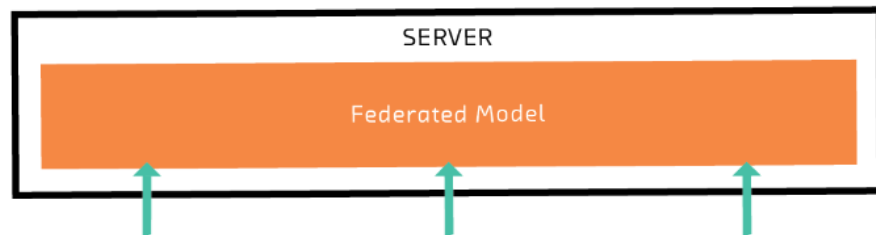
1. 节点从服务器获取模型，并开始训练
2. 节点将训练后的模型发送给服务器
3. 服务器聚合所有的模型
4. 回到1



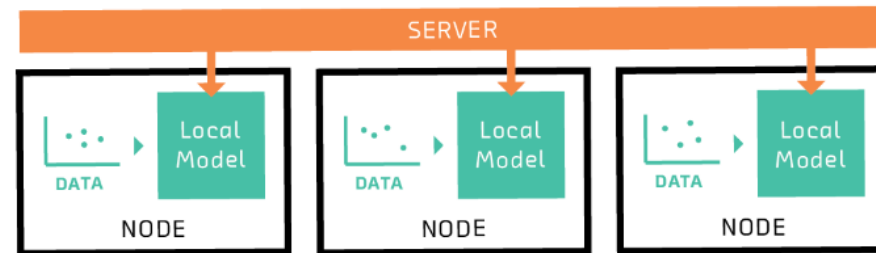
1. Nodes receive model from server and start training.



2. Nodes send partially trained models to server.

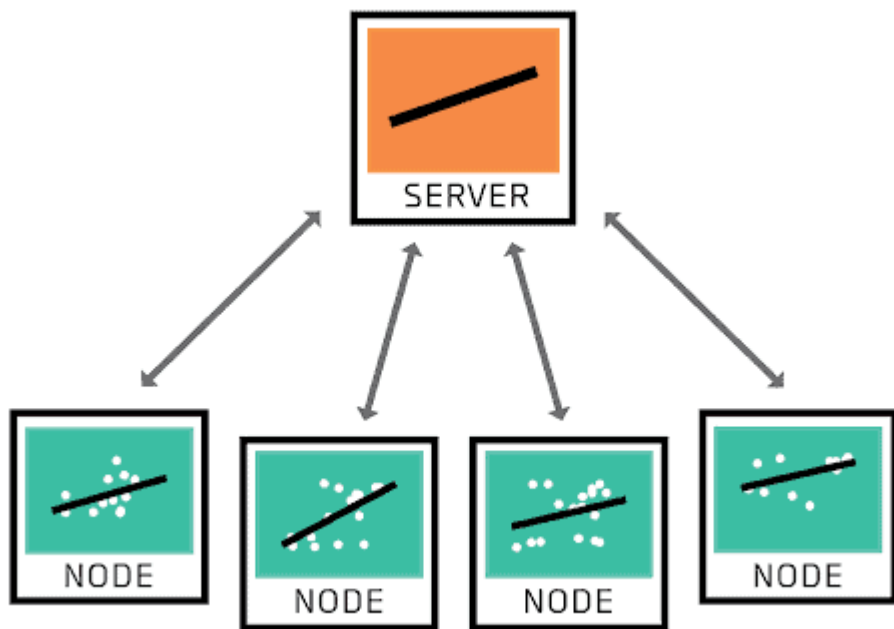


3. The server combines those models to make a federated model.

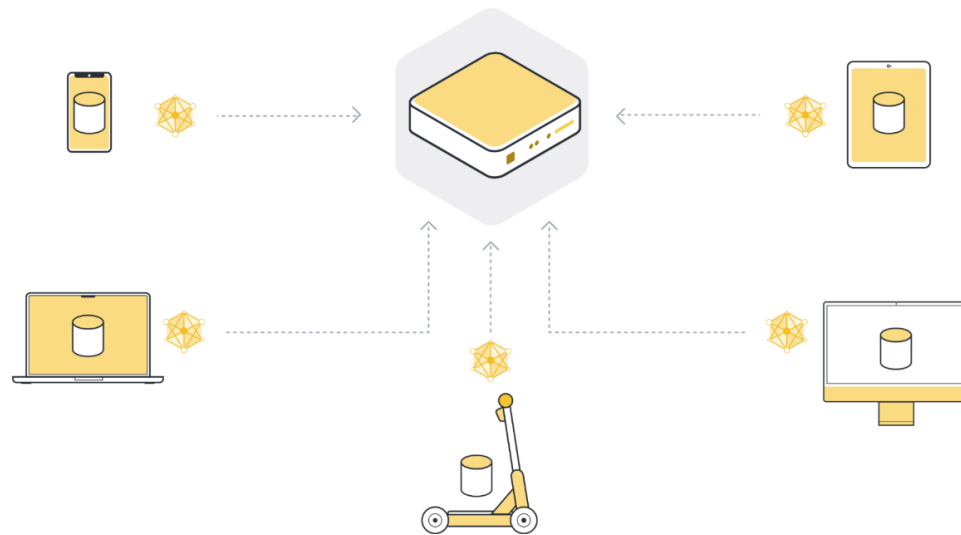


4. The federated model is sent to the nodes. Repeat as necessary.

联邦学习的异质性



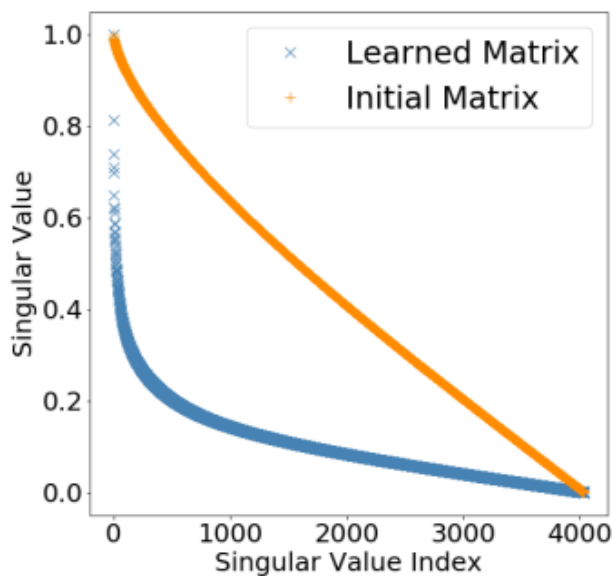
数据异质性：经设备训练后的模型偏差很大



系统异质性：设备的性能各不相同

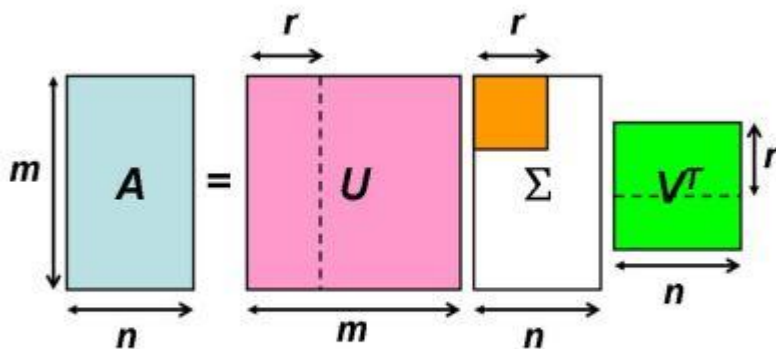
低秩分解

- 基于假设：矩阵往往不是满秩的
- 后置分解：不需要重新训练



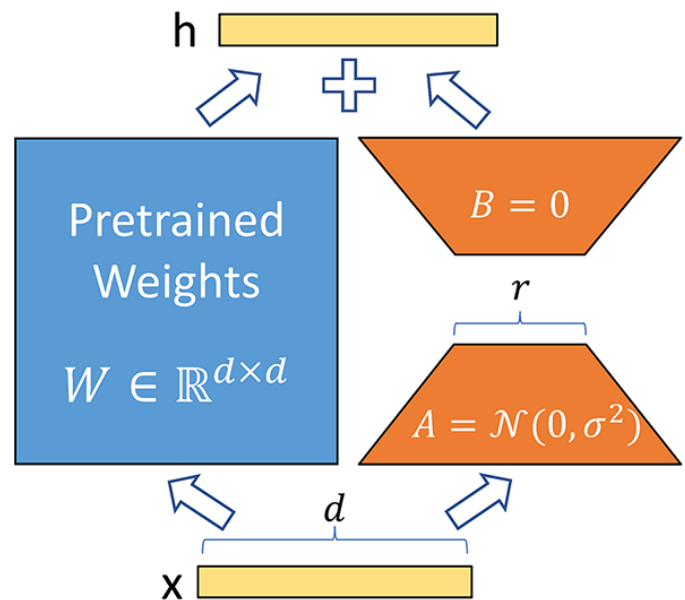
(a) Singular Values

通过低秩近似减少计算量
而不会明显损失精度



SVD分解，特征值分解，
梯度下降分解.....

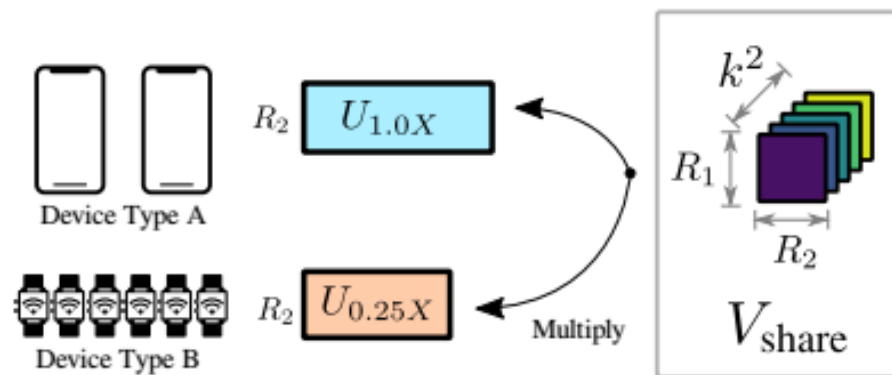
- 前置分解：从网络结构上进行拆解



将Linear拆成两个Linear，
将Conv拆成两个Conv

FLANC的前置分解

- All-In-One Neural Composition: 能否将参数拆成两个部分, 一个部分由所有客户端共享知识 (共享基), 另一个用于自适应客户端的资源能力 (系数) ?
- 将参数 W_p 拆成 V_{share} 和 U_p
 - W_p 是拥有特定处理容量为 p 的客户端能够训练的最大网络。
 - 共享神经基 V_{share} 是一个容量无关的张量, 将在所有的客户端上进行训练并实现知识的共享。
 - 系数 U_p 则对应所有拥有特定处理容量 p 的客户端, 用于适应现实中散布的各种边缘设备, 它只在同组设备间实现知识共享。



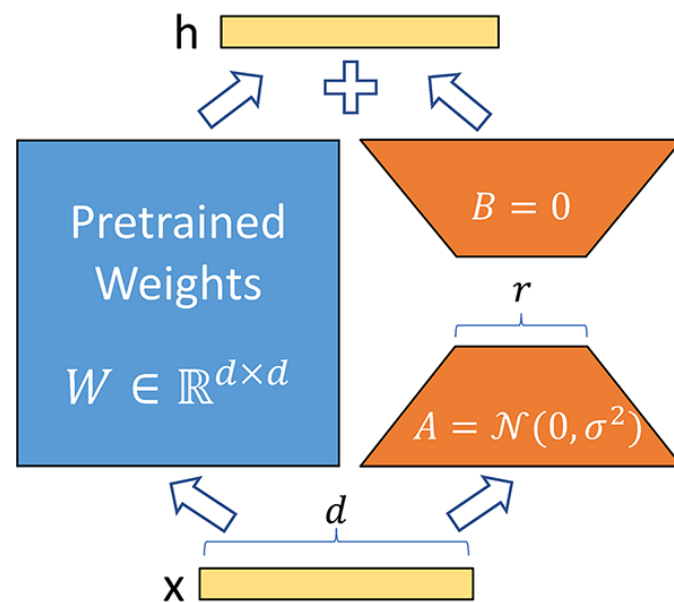
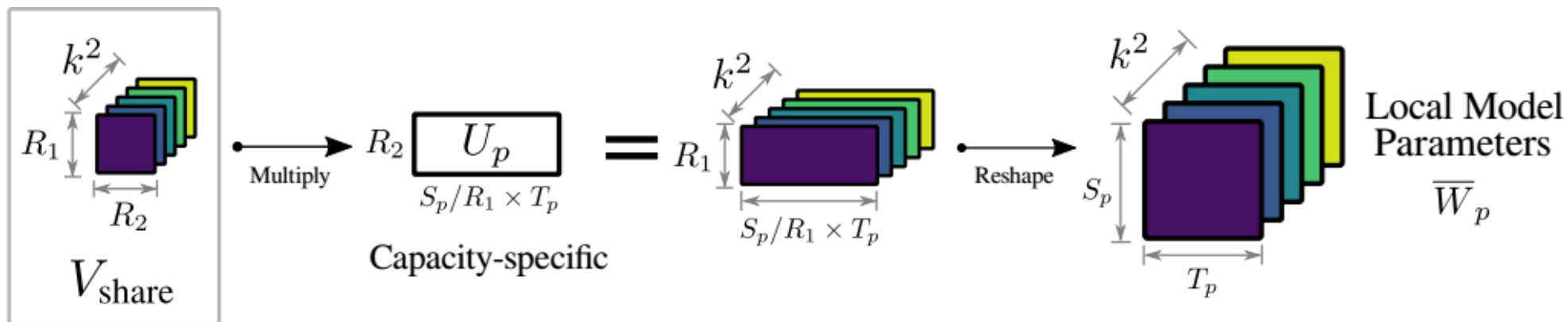
FLANC的前置分解

- 传统的分解不能用于资源异构场景

- All-In-One Neural Composition:

FLANC 将维度为 $k^2 \times m_p \times n_p$ 的 W_p 分解为维度为 $k^2 \times r_1 \times r_2$ 的 V_{share} 和 $r_2 \times m_p/r_1 \times n_p$ 的 U_p 。

- 如 $100 * 50$ 的线性层，拆分为 $10*5 + 5 * 500$ 的线性层串联



正则化列向量以提高表示能力

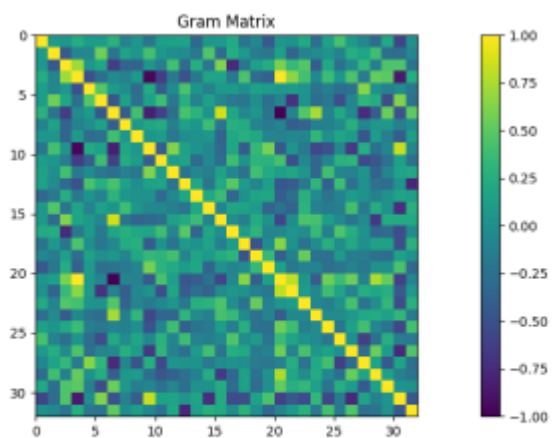
- $W_p = V_{share} U_p$

W_p 的列向量可以看做是 V_{share} 的列向量的线性组合，或 V_{share} 张成的子空间。如果让 V_{share} 列向量线性无关，那么表达能力就更强。

- 引入正则化： $\lambda \|VV^T - I\|_2$ ，使得 V_{share} 的列向量尽可能正交

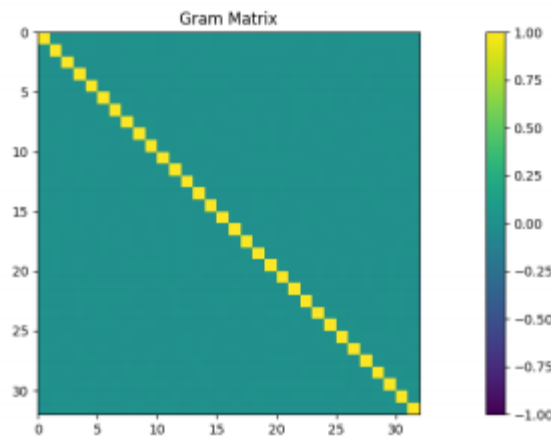
正则化列向量以提高表示能力

- 引入正则化: $\lambda \|VV^T - I\|_2$, 使得 V_{share} 的列向量尽可能正交



(a) Layer 1 (w/o Orth.)

引入正则化



(e) Layer 1 (w/ Orth.)

实验和消融实验

- 实验：在四个数据集上与 HeteroFL, FjORD 进行 TOP-1 ACC 对比
- 消融实验：验证了 r_1 （基向量的维度）和 r_2 （基向量的个数）的影响