STOR566: Introduction to Deep Learning

Lecture 19: Federated Learning

Yao Li
UNC Chapel Hill

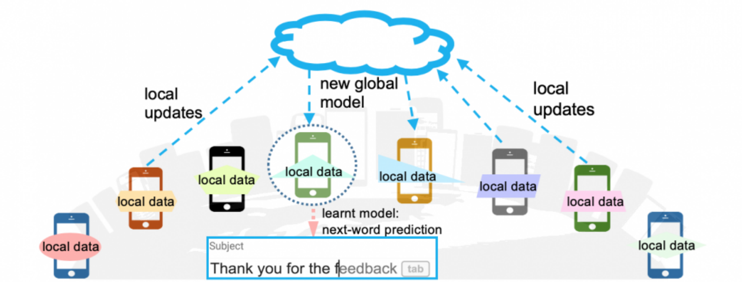
Nov 3, 2022

Some materials are from *Machine Learning and Vision Lab, UNIST*

Federated Learning (FL)

Federated Learning: Overview

- Overview:



- Decentralized data
- Data privacy preserving

Federated Learning

- Examples:
  - Gboard on Android
  - Media playback preferences in Safari
  - Voice assistant in Siri
  - Health care related problems

Example: Gboard on Android

- Gboard on Android:

Example: Voice assistant in Siri

- Voice assistant in Siri:

MIT Technology Review — Sign in — Subscribe

Artificial intelligence / Machine learning

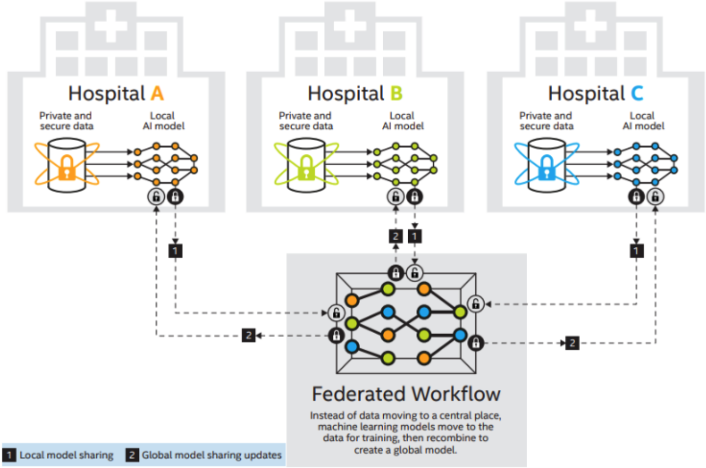
How Apple personalizes Siri without hoovering up your data

The tech giant is using privacy-preserving machine learning to improve its voice assistant while keeping your data on your phone.

by **Karen Hao** — December 11, 2019

Example: Health care

- Privacy-Preserving AI to Identify Brain Tumors:



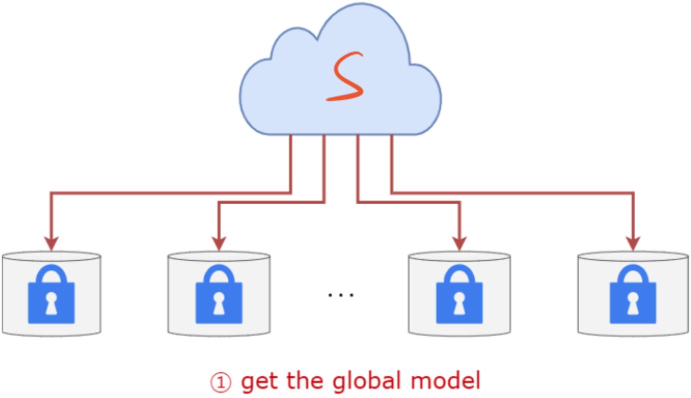
Definition

Federated Learning

Federated learning(FL) is a machine learning setting where multiple clients collaborate in solving a ML problem, under the coordination of a central server. **Each client's raw data is stored locally and not exchanged or transferred**; instead, updates intended for immediate aggregation are used to achieve the learning objective.

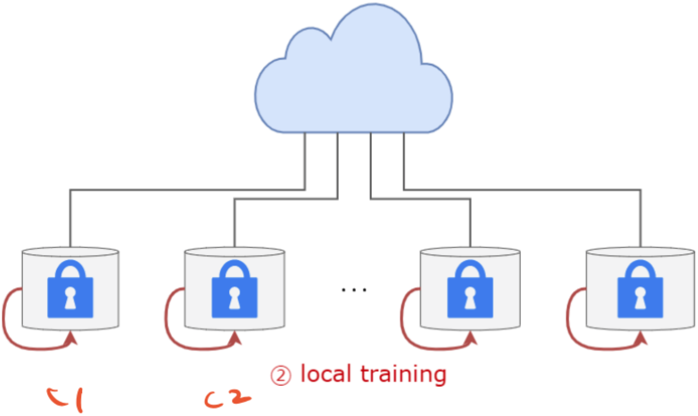
Workflow

- Get the global model:



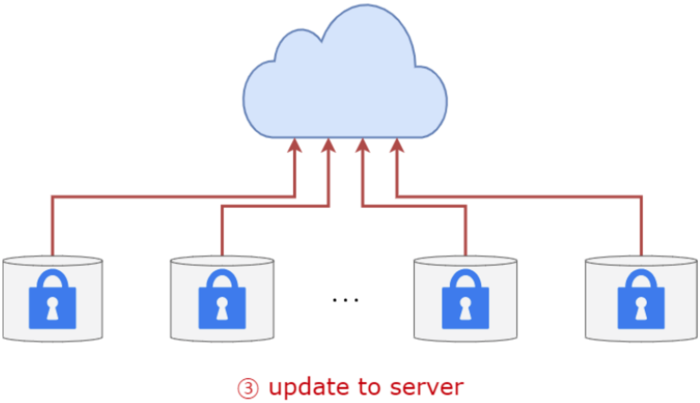
Workflow

- Local training:



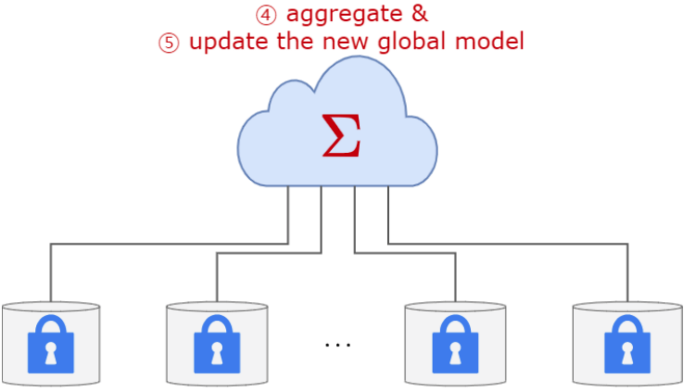
Workflow

- Send updates to server:



Workflow

- Aggregation:



- In general, not all the local users will be selected to participate the aggregation.

Aggregation Algorithms

Algorithms

- McMahan et al. *Communication-efficient learning of deep networks from decentralized data.* PMLR, 2017.
  - FedSGD
  - FedAVG
- Yin et al. *Byzantine-robust distributed learning: Towards optimal statistical rates.* ICML, 2018.
  - Median
  - Trim-mean

FedSGD

- FedSGD: Update the model locally for one epoch then send back to the central server.
- FedSGD: The global model update: $\boldsymbol{w}^{t+1} \leftarrow \boldsymbol{w}^t - \eta \cdot \sum_k^K \frac{n_k}{n} \boldsymbol{g}_k$

  $\boldsymbol{w}^t$: weight of the global model at round t

  η: learning rate

  K: number of local users selected to participate the aggregation

  n_k: number of samples on user k

  n: $\sum_k^K n_k$

  $\boldsymbol{g}_k$: gradient from user k

FedAVG

- FedAVG: Update the model locally for several epochs then send back the new model
- FedAVG:

  Each user first do: $\boldsymbol{w}^{t+1,k} \leftarrow \boldsymbol{w}^{t} - \eta \boldsymbol{g}_k$ (multiple times)

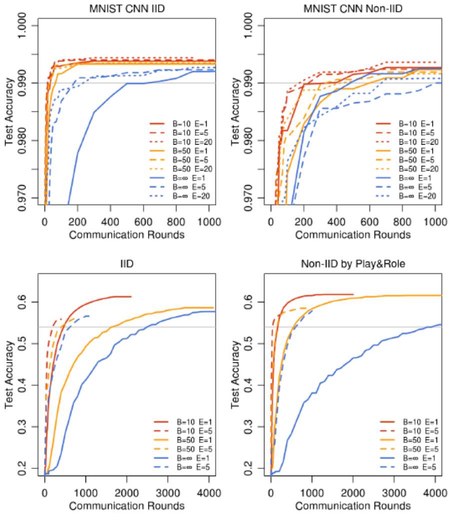
  The global model update: $\boldsymbol{w}^{t+1} \leftarrow \sum_k^K \frac{n_k}{n} \boldsymbol{w}^{t+1,k}$

  $\boldsymbol{w}^{t+1}$: weight of the global model at round $t+1$

  $\boldsymbol{w}^{t+1,k}$: weight of the local model on user k at round $t+1$

Performance

- E: number of local epochs. $E = 1$: FedSGD
- B: batch size of local training



Issues

- Security: no control of the data
- Data heterogeneity: violation of I.I.D. assumption (Non-IID)

Robust Aggregation

What if some local data are mislabelled?

Robust Aggregation Methods:

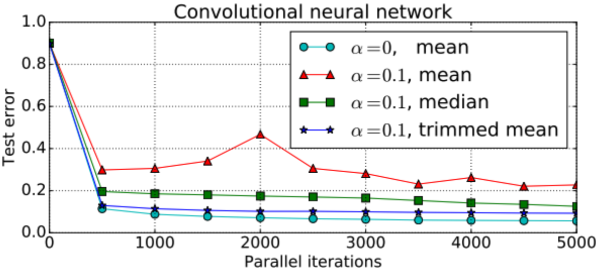
1. **Median:** Yin et al. (2018), coordinate-wise median among the weight vectors of selected users.
2. **Trim-mean:** Yin et al. (2018), coordinate-wise mean with trimmed values.

Problem:

- performance degradation

Robust Aggregation: Performance

- α: proportion of wrong data



- When $\alpha > 0$, robust aggregation methods perform better

Conclusions

- Federated learning
- Aggregation algorithms

Questions?