SwarmAvg - A Novel Approach to Fully Distributed Machine Learning

Josh Pattman

May 2023

University Of Southampton

Problems

Privacy

What is the problem:

- Data stored in multiple locations
- Cannot share the data between locations for privacy reasons
- *Medical records*

Why is this a problem:

- Cannot train models on all available data
- May cause models to have lower performance

Performance

What is the problem:

- Machine learning needs lots of processing power
- Available hardware may be slow
  - May have assess to many slow computers
  - *Company with many unused work computers during the night*

Why is this a problem:

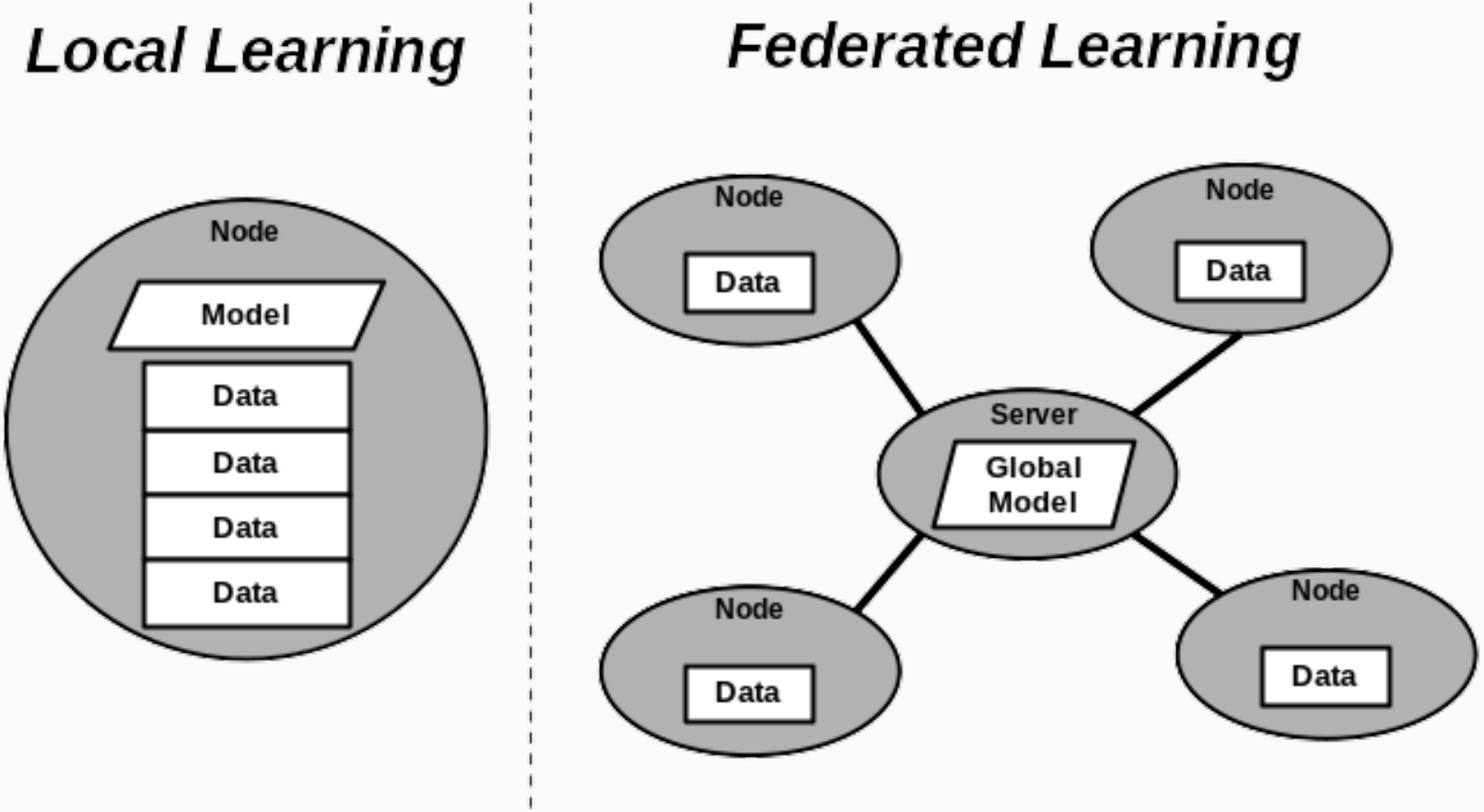
- Slow to train models on available hardware
- Unused processing power

Federated Learning

Federated Learning - The Current Solution

- A single model is stored on the server
- Server controls many nodes - computers that can perform training
- Each node has its own dataset
    - This is not shared with other nodes or the server
- **Goal:** Perform machine learning by only sharing the model, not the data

Federated Learning



Federated Learning - Variations

- Many variations of federated learning
  - One of the originals is *Federated Averaging (FedAvg)*
  - Many other algorithms are based off this

FedAvg - How Does It Work?

- FedAvg has repeated training steps. Each step:
  1. Server sends model to a set of nodes
  2. Nodes perform training on the model
  3. Nodes send their models back to server
  4. New model is the average (mean) of all nodes models

FedAvg - Issues

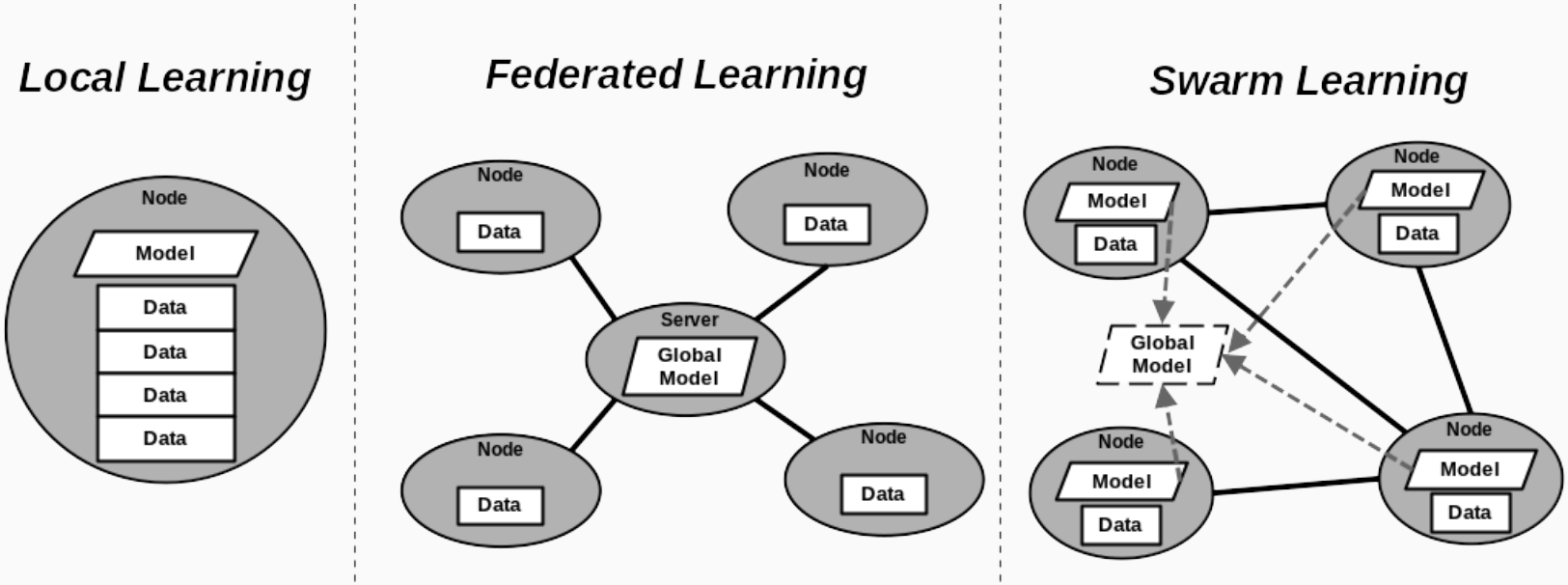
- Vulnerable to central server going down
- Requires that every node has direct access to the server
- Scalability issues with bottleneck

Swarm Learning

Swarm Learning

- No central server/node
- Each node has a distinct model, called the *local model*
  - Must keep all local models similar to each other
- Each node has its own dataset
  - This dataset cannot be shared with any other nodes
- The goal is to train all local models together using all available data

Swarm Learning



SwarmAvg

- SwarmAvg - *Swarm Learning using Averaging*
- Inspired by FedAvg
    - Model averaging is the core mechanic
- No blockchain
- Local models of nodes are usually slightly different - no consensus on a global model

SwarmAvg - How Does It Work?

- Repeated Training Steps. Each node each step:
  1. Perform training on the local model
  2. Send trained model to all neighbours
     - This will get saved on the neighbour
  3. New local model is the combination of all neighbours most recent local models

SwarmAvg - Specifics

- Different combination methods
  - Combine by mean
  - Combine with learning rate (weighted mean)
- Only combine neighbours who have done more training than this node
- Wait for certain number of neighbours to catch up with this node

Swarm Learning vs Issues of Federated Learning

- ~~Vulnerable to central server going down~~
  - No central server - to stop training you would have to take out every node
- ~~Requires that every node has direct access to the server~~
  - Swarm learning can function on sparse networks of nodes
- ~~Scalability issues with bottleneck~~
  - As the number of nodes in the network scales, the number of connections per node can stay consistent

Experiments

What Did I Test?

Task was classification of MNIST fashion items, and results are accuracy of classification

- Tested SwarmAvg against FedAvg:
  - In densely connected networks with varying amounts of available data
  - In densely connected networks with varying amounts of available data and limited classes
- Tested SwarmAvg:
  - In sparsely connected networks with varying amounts of available data
  - In sparsely connected networks with varying amounts of available data and limited classes

SwarmAvg Findings

- In densely connected networks, SwarmAvg is slightly slower than FedAvg, and does not reach quite as high a final accuracy.
- With a high enough volume of data, SwarmAvg is only slightly affected by sparse network conditions, when no class restrictions are imposed.
- The effect of network sparsity on SwarmAvg is increased when class restrictions are introduced.

Thanks for listening! Any questions?