A2FL: Availability-Aware Selection for Machine Learning on Clients with Federated Big Data

IEEE ICC 2023
Rome, Italy - May 2023

Ahmed M. Abdelmoniem, Yomna M. Abdelmoniem, Ahmed Elzanaty
Data lives at the Edge

• Billions of phones & IoT devices constantly generate data
• Data enables better products and smarter models
• On-device processing (e.g., inference for mobile keyboards)
  • advanced specialized hardware (e.g., GPU and NPUs on mobile/IoT devices)
• Benefits
  • Improved latency
  • Works offline
  • Better battery life
  • Privacy advantages

What about analytics & learning?
Centralized vs Federated Learning

• Centralized Training:
  • Central (Data) server
  • Expensive data movement
  • Communication-intensive
  • Privacy concerns

• Federated Learning:
  • Central (Aggregation) server
  • Model exchange
  • Communication-efficient
  • Privacy-preserving (Differential privacy + secure aggregation)
Practical Use-cases of Federated Learning (FL)

What are good applications for FL?

• On-device data is more relevant than server-side data (or lack of it)
• On-device data is privacy sensitive or large to communicate
• Labels can be inferred naturally from user interaction

Gboard: next-word prediction

Using FL, better next-word prediction accuracy: +24%


Apple: Voice recognition

Medical Imaging

Ng D, Lan X, Yao MM, Chan WP, Feng M. Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. Quant Imaging Med Surg. 2021
Federated Learning Life-cycle

Selection | Reporting deadline | Aggregation

FL server

Straggler (wasted resource)

Learners

Round $i$

Round $i+1$

Heterogeneity impacts quality and time!

- Heterogenous data distributions $\rightarrow$ non-IID setting (quality)
- Diverse hardware and network capabilities $\rightarrow$ stragglers (time)
- Clients are not always available $\rightarrow$ inclusivity is hard (quality)
Non-Practical Selection Methods

• Most existing methods aim to improve the time-to-quality

**Reduce Time**

FedCS (ICC’18), Oort (OSDI’21)

• Biases the client selection to reduce the training time by exploiting the fast learners

**Improve Quality**

AdaPow (AISTATS’22)

• Biases the client selection towards ones with high loss to boost model quality

Disregards clients’ availability

(low inclusivity)

FedCS → T. Nishio, R. Yonetani, Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge, ICC, 2018


AdaPow → Yae Jee Cho, Jianyu Wang, Gauri Joshi, Towards Understanding Biased Client Selection in Federated Learning, AISTATS, 2022
Availability does NOT matters in IID case

- **Availability does not impact the model quality**
  - Oort in IID data distribution $\rightarrow$ client’s data are uniformly distributed
  - Even biased selection (fast learners) $\rightarrow$ still can capture the global data distribution
Motivation – Availability matters in non-IID case

• Availability can impact model quality
  • In non-IID data distribution → every client’s data samples are important
  • Lack of inclusive selection → hard to capture the global data distribution
Takeaway & Proposed Solution

• Existing methods disregard client availability in the selection
• Biased selection can result in low resource diversity

\[ \text{A2FL: Availability-Aware Federated Learning} \]

- **Selection**: prioritize selection of least available learners → Increases diversity
  - Availability Prediction Module: on-device prediction models (no privacy violation)
  - Hybrid Selection Method: the selection leverages both availability prioritization and random sampling

In the paper → more detailed description and discussion of the algorithm
System Design

Flowchart showing the components of the A2FL System.

- **FL Developer** submits a task to the **FL Server**.
- The **Global model** is updated by the **Server module**.
- **Cloud Services** include storage and model updates.
- **Server module** and **Server AP module** interact with **Selection Decisions**.
- **Availability Probability (p)** is used in the decision process.
- **On Device FAP module** and **Client module** are connected to the system.

Legend:
- **Submit FL Task**: Arrow from FL Developer to FL Server.
- **Availability Probabilities**: Arrows from Server module to Server AP module.
- **Selection Decisions**: Arrow from Server module to Server AP module.
- **Availability Probability (p)**: Arrow from Server module to Server AP module.
An Illustrative Example

A2FL selects among online clients using availability info and breaks ties by selecting at random
An Illustrative Example

A2FL may select some stragglers but improves diversity, Oort only selects the fast learners
An Illustrative Example

A2FL selects clients with the least availability, Random selects regardless of future availability.
An Illustrative Example

(a) Learner availability
An Illustrative Example

A2FL is able to achieve higher client diversity which improves the statistical efficiency of the model.

A2FL have high rate of unique clients, while other methods would not cover all the clients.
Experimental Evaluation

- FL Benchmarks using Google’s Speech Recognition task [1]
- Various data distributions: IID, Label-limited (non-IID)

Evaluation of A2FL

- A2FL \(\rightarrow\) best model quality with least amount of resources and time
  - It improves over all the other methods in both IID and non-IID cases

![Graphs showing comparison between Oort, A2FL, and Random methods in IID and Non-IID cases](image)
Evaluation of A2FL

• **Availability prioritization** leads to better diversity
  • This is evident by the high rate of unique clients for A2FL (close to random)
  • This is with even lower number of updates (i.e., higher stragglers).
Takeaways

• Heterogeneity is a major challenge for FL:
  • Model quality degradations are not acceptable, esp. in non-IID settings
  • Behavior heterogeneity impacts the quality even more.

• To tackle heterogeneity \(\rightarrow\) adapt to availability dynamics of the clients
  • A2FL leverages support of on-device availability prediction module and prioritizes the clients with least availability \(\rightarrow\) gains in model quality.

• Future Work & Technical Challenges
  • How to deal with mis-information from malicious/non-faithful learners?
  • How to fine-tune knobs to control the trade-off between efficiency & diversity?
Thanks

Q & A

For further questions, please reach out to ahmed.sayed@qmul.ac.uk

If interested in solving real-world problems!
Join us at https://sayed-sys-lab.github.io