Oort

Efficient Federated Learning via Guided Participant Selection

Fan Lai, Xiangfeng Zhu,
Harsha V. Madhyastha, Mosharaf Chowdhury
Emerging Trend of Machine Learning

Edge devices generate massive data
Emerging Trend of Machine Learning

Edge devices generate massive data

Increasing resource on edge device

Model inference latency

- 2018:
  - 85 ms
  - 48 ms
  - 19 ms
- 2021:
  - 8.4 ms
Emerging Trend of Machine Learning

Edge devices generate massive **data**

Increasing **resource** on edge device

**Model inference latency**

- 2018: 85 ms
- 2021: 8.4 ms
- 2021: 48 ms
- 2021: 19 ms

**ML needs fresh and large real-life datasets**
Emerging Federated Learning on the Edge

- On-device machine learning helps
- Reduce data migration/privacy risk
- Learn on fresh real-world data
- …

Mistify: Automating DNN Model Porting for On-Device Inference at the Edge
Towards Federated Learning at Scale: System Design
MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications
Applied Federated Learning: Improving Google Keyboard Query Suggestions
Many others …
Emerging Federated Learning on the Edge

- On-device machine learning helps
  - Reduce data migration/privacy risk
  - Learn on fresh real-world data
  - ...

- Federated **training and testing**
  - Run model across millions of edge clients
Execution of Federated Learning (FL)

Primary Objective

Better *time to accuracy*:
- Less time for target acc. under the same setting
Execution of Federated Learning (FL)

Primary Objective

Better time to accuracy:
- Less time for target acc. under the same setting
Execution of Federated Learning (FL)

O(100) Rounds:
- Client selection
- In-situ Execution
- Result aggregation

Round i

Client Pool

Submit Job
- Model
- Config

Coordinator
- Execution Driver
- Client Manager

① Client Selection

Model

Config
Execution of Federated Learning (FL)

O(100) Rounds:
- **Client selection**
- **In-situ Execution**
- **Result aggregation**
Execution of Federated Learning (FL)

O(100) Rounds:
- Client selection
- In-situ Execution
- Result aggregation

Round i

Submit Job

Client Pool

Model Config

Execution Driver

Coordinator

Client Manager

② Execution
Execution of Federated Learning (FL)

Coordinator

Execution Driver → Client Manager

Submit Job
Model, Config

Client Pool

Round \( i \)

O(100) Rounds:

- Client selection
- In-situ Execution
- Result aggregation
Execution of Federated Learning (FL)

O(100) Rounds:
- Client selection
- In-situ Execution
- Result aggregation

Round i
Execution of Federated Learning (FL)

Primary Objective

Better *time to accuracy*:
- Less time for target acc. under the same setting

O(100) Rounds:

- Client selection
- In-situ Execution
- Result aggregation
Challenges in Federated Learning

<table>
<thead>
<tr>
<th>System</th>
<th>FL</th>
<th>In-cluster ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous</td>
<td>Homogeneous</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneous system speed
# Challenges in Federated Learning

<table>
<thead>
<tr>
<th></th>
<th>FL</th>
<th>In-cluster ML</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System</strong></td>
<td>Heterogeneous</td>
<td>Homogeneous</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>Heterogeneous</td>
<td>Homogeneous via shuffling</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>$O(1M)$</td>
<td>$O(10)$</td>
</tr>
<tr>
<td><strong>Dynamics</strong></td>
<td>Client can drop out/rejoin</td>
<td>Few</td>
</tr>
</tbody>
</table>

### Heterogeneous data distribution

- **Client A**
- **Client B**

...
Challenges in Federated Learning

<table>
<thead>
<tr>
<th></th>
<th>FL</th>
<th>In-cluster ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>Heterogeneous</td>
<td>Homogeneous</td>
</tr>
<tr>
<td>Data</td>
<td>Heterogeneous</td>
<td>Homogeneous via shuffling</td>
</tr>
<tr>
<td>Scale</td>
<td>$O(1M)$</td>
<td>$O(10)$</td>
</tr>
<tr>
<td>Dynamics</td>
<td>Client can drop out/rejoin</td>
<td>Few</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- Existing work optimize for better
  - System efficiency
    - Reduce round duration
  - Statistical efficiency
    - Reduce # of rounds needed
  - ...

- ...
## Challenges in Federated Learning

<table>
<thead>
<tr>
<th></th>
<th>FL</th>
<th>In-cluster ML</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System</strong></td>
<td>Heterogeneous</td>
<td>Homogeneous</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>Heterogeneous</td>
<td>Homogeneous via shuffling</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>$O(1M)$</td>
<td>$O(10)$</td>
</tr>
<tr>
<td><strong>Dynamics</strong></td>
<td>Client can drop out/rejoin</td>
<td>Few</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- Existing work optimize for better
- System efficiency
- Reduce round duration
- Reduce # of rounds needed
- ...  

Existing federated learning relies on random participant selection
Existing Client Selection: Suboptimal Efficiency

Image classification task on OpenImage dataset

**Problem #1**

*Overlook heter. client utility*
Existing Client Selection: Suboptimal Efficiency

Image classification task on OpenImage dataset

Problem #1

Overlook heter. client utility

Existing Client Selection: Suboptimal Efficiency

Problem #1

Overlook heter. client utility

Suboptimal training convergence

Image classification task on OpenImage dataset

Accuracy (%) vs. Training Rounds

Performance upper bound

Accuracy drops

In-cluster ML

Prox[1]

YoGi[2]

FL settings

ShuffleNet Model

Enforcing selection criteria is crucial in FL testing

- “Give me 4k representative samples”
- “Give me $x$ samples of class $y$”

(Hypothetical) model testing on all clients $ightarrow$ ground truth
• Enforcing selection criteria is crucial in FL testing
  • “Give me 4k representative samples”
  • “Give me x samples of class y”

Existing Client Selection: Unable for Selection Criteria
Existing Client Selection: Unable for Selection Criteria

- Enforcing selection criteria is crucial in FL testing
  - “Give me 4k representative samples”
  - “Give me x samples of class y”
  - …

Problem #2

Overlook specified selection criteria

Useless testing results
Problem #2

Overlook specified selection criteria

Useless testing results

• Enforcing selection criteria is crucial in FL testing
  • “Give me 4k representative samples”
  • “Give me x samples of class y”
  • …

Existing Client Selection: Unable for Selection Criteria
Oort: Guided Participant Selection for FL

Diagram:
- User submits job to Coordinator
- Coordinator sends execution driver to Client Pool
- Model and Config are exchanged with clients
- Clients perform local computations and return results
Oort: Guided Participant Selection for FL

Design Overview

- **Enable faster FL training**
  - Adaptively explore and exploit high-utility clients

- **Support interpretable FL testing**
  - Enforce developer-specified data selection criteria at scale
**Oort**: Guided Participant Selection for FL

**Design Overview**

- **Enable faster FL training**
  - Adaptively explore and exploit high-utility clients

- **Support interpretable FL testing**
  - Enforce developer-specified data selection criteria at scale
Anatomy of Time to Accuracy in Training

- **System efficiency (round duration)**
  - Determined by client *system speed*

- **Statistical efficiency (round to accuracy)**
  - Determined by client *data*
Anatomy of Time to Accuracy in Training

• **System efficiency (round duration)**
  - Determined by client *system speed*

• **Statistical efficiency (round to accuracy)**
  - Determined by client *data*
Anatomy of Time to Accuracy in Training

- **System efficiency (round duration)**
  - Determined by client *system speed*

- **Statistical efficiency (round to accuracy)**
  - Determined by client *data*

Client utility

- **System utility** (round duration)
- **Statistical utility**
Anatomy of Time to Accuracy in Training

- **System efficiency (round duration)**
  - Determined by client *system speed*

- **Statistical efficiency (round to accuracy)**
  - Determined by client *data*

Client utility

- **System utility** (round duration)
- **Statistical utility:** *how data helps round to accuracy?*
Challenge 1: Identify **Heterogeneous Client Utility**

- **Statistical utility**
  - Capture how the client data can help to improve the model
Challenge 1: Identify Heterogeneous Client Utility

• **Statistical utility**
  - Capture how the client data can help to improve the model

• **Metric:** aggregate training loss of client data
  - Higher loss $\rightarrow$ higher stats utility (proof in paper)
Challenge 1: Identify Heterogeneous Client Utility

• Statistical utility
  • Capture how the client data can help to improve the model
  • Metric: aggregate training loss of client data
    • Higher loss $\rightarrow$ higher stats utility \[\text{(proof in paper)}\]

• Utility of a client $\frac{\text{stats}_\text{util} (i)}{\text{round}_\text{duration} (i)}$
  • i.e., speed of accumulating stats utility in round $i$

Heterogeneity: Statistical utility
Scalability: System utility
Dynamics: Stats. utility
Robustness: AND

18
Challenge 2: Select High-Utility Clients at Scale

- How to identify high-utility clients from millions of clients?
  - *Spatiotemporal* variation: heterogeneous utility across clients over rounds
Challenge 2: Select High-Utility Clients at Scale

- How to identify high-utility clients from millions of clients?
  - Spatiotemporal variation: heterogeneous utility across clients over rounds

- Exploration + Exploitation
  - Explore not-tried clients
Challenge 2: Select High-Utility Clients at Scale

• How to identify high-utility clients from millions of clients?
  • *Spatiotemporal* variation: heterogeneous utility across clients over rounds

• Exploration + Exploitation
  • Explore not-tried clients
  • Exploit known *high-utility* clients
Challenge 3: Select High-Utility Clients Adaptively

- How to account for stale utility since last participation?
  - Utility changes due to dynamics
Challenge 3: Select High-Utility Clients Adaptively

• How to account for **stale** utility since last participation?
  • Utility changes due to dynamics

1. **Aging**: add uncertainty to utility
  • \( \text{current}\_\text{utility} = \text{last}\_\text{observed}\_\text{utility} + \text{observation}\_\text{age} \)
Challenge 3: Select High-Utility Clients Adaptively

• How to account for stale utility since last participation?
  • Utility changes due to dynamics

1. **Aging**: add uncertainty to utility → *Re-discover missed good clients*
  • current_utility = last_observed_utility + \textit{observation\_age}
Challenge 3: Select High-Utility Clients Adaptively

- How to account for *stale* utility since last participation?
  - Utility changes due to dynamics

1. **Aging**: add uncertainty to utility \( \rightarrow \text{Re-discover missed good clients} \)
   - \( \text{current}\_\text{utility} = \text{last}\_\text{observed}\_\text{utility} + \text{observation}\_\text{age} \)

2. **Probabilistic selection** by utility values
   - Prioritize high-utility clients
   - Robust to outliers and uncertainties

Exploited Clients

Probabilistic sampling by utility
• How to respect privacy

• How to be robust to corrupted clients

• How to enforce diverse selection criteria
  • Fairness, data distribution for **FL testing**
Evaluation

Oort as a lib to support TensorFlow Federated / PySyft

Experiment setting

• Testbed w/ 68 GPUs
• Realistic FL Benchmark\[1\]
  • Heter. speed/data
  • Dynamics of devices
  • 1300 participants/round

\[1\] FedScale: Benchmarking Model and System Performance of Federated Learning
Time-to-Accuracy (TTA) Performance

Image classification (OpenImage dataset)

- Prox
- YoGi
- Oort + Prox
- Oort + YoGi

FL Runtime (hours)

ShuffleNet Model
Time-to-Accuracy (TTA) Performance

Image classification (OpenImage dataset)

- Prox
- YoGi
- Oort + Prox
- Oort + YoGi

ShuffleNet Model

Next-word prediction (Reddit Corpus)

- Prox
- YoGi
- Oort + Prox
- Oort + YoGi

Albert Model

Oort improves **TTA by 14X** and **final accuracy by 9%**
Oort achieves close to upper-bound statistical performance
Oort

https://github.com/SymbioticLab/Oort

Participant selection framework for

- **Faster** convergence in FL training
- **Interpretable** data selection in FL testing

Client selection for

\[
\begin{align*}
\text{utility-aware FL training w/ adaptive exploration-exploitation} \\
\text{criteria-aware FL testing to enforce specified data selection}
\end{align*}
\]

Thank you!