ARGONNE PRIVACY-PRESERVING FEDERATED LEARNING WITH FUNCX

TRUNG-HIEU HOANG

Ph.D. Candidate
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
Data Science and Learning Division, Argonne National Laboratory
Multi-disciplinary team from ANL, LLNL, Harvard, UChicago and UIUC
KEY MOTIVATIONS & OBJECTIVES

Data shift in Machine Learning

Policy concerns with Biomedicine Data

Key objectives

To develop and implement APPFL (Argonne Privacy-Preserving Federated Learning) framework that implements Differential Privacy algorithms for training Federated Learning models with the biomedical datasets from multiple organizations
FEDERATED LEARNING OVERVIEW

https://www.palisadex.net/appfl
FEDERATED LEARNING RESOURCES MANAGEMENT

Approved Research Team

Identity Management

Researcher A

Institution X

Researcher B

Institution Y

Researcher C

Federated Learning Project

Data Management

Protected Data

Protected Data

Protected Data

Computing Resources Management

Computing Resources

Protected Data

Shared Computing Resources

APPFL
KEY COMPONENTS

IAM Layer

Execution Layer

Fed-learning algorithm

Communication Layer

Globus Authentication Service

APPFL Server

APPFL

func

Client

func

Endpoint

Model's parameters transfer

Control signals

Client's Workers

https://www.globus.org/
https://aws.amazon.com/s3/
SETTING UP AN EXPERIMENT

Users need to define the following files:

- **User-defined Model & Data-loader**
  - model.py
  - dataloader.py
  - Using PyTorch, as regular Python scripts

- **Client Configuration**
  - clients.yml
  - Parameters, funcX-endpoint IDs and data loaders at all clients

- **Server Configuration**
  - server.yml
  - Parameters, fed-learning algorithm at server
### EXAMPLES – FUNCX TASK EXECUTION LOG

<table>
<thead>
<tr>
<th>timestamp</th>
<th>task_name</th>
<th>client_name</th>
<th>status</th>
<th>execution_time</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/16/22 9:00 AM</td>
<td>client_validate_data</td>
<td>uiuc-cig-01-gpu-02</td>
<td>success</td>
<td>13.63</td>
</tr>
<tr>
<td>8/16/22 9:00 AM</td>
<td>client_validate_data</td>
<td>uchicago-gpu</td>
<td>success</td>
<td>39.04</td>
</tr>
<tr>
<td>8/16/22 9:01 AM</td>
<td>client_training</td>
<td>uiuc-cig-01-gpu-02</td>
<td>success</td>
<td>34.95</td>
</tr>
<tr>
<td>8/16/22 9:01 AM</td>
<td>client_training</td>
<td>uchicago-gpu</td>
<td>success</td>
<td>65.08</td>
</tr>
<tr>
<td>8/16/22 9:02 AM</td>
<td>client_training</td>
<td>uiuc-cig-01-gpu-02</td>
<td>success</td>
<td>31.73</td>
</tr>
<tr>
<td>8/16/22 9:02 AM</td>
<td>client_training</td>
<td>uchicago-gpu</td>
<td>success</td>
<td>35.21</td>
</tr>
<tr>
<td>8/16/22 9:03 AM</td>
<td>client_training</td>
<td>uiuc-cig-01-gpu-02</td>
<td>success</td>
<td>31.51</td>
</tr>
<tr>
<td>8/16/22 9:03 AM</td>
<td>client_training</td>
<td>uchicago-gpu</td>
<td>success</td>
<td>35.42</td>
</tr>
<tr>
<td>8/16/22 9:03 AM</td>
<td>client_training</td>
<td>uiuc-cig-01-gpu-02</td>
<td>success</td>
<td>28.75</td>
</tr>
<tr>
<td>8/16/22 9:03 AM</td>
<td>client_training</td>
<td>uchicago-gpu</td>
<td>success</td>
<td>35.63</td>
</tr>
</tbody>
</table>

- **task_name**: client_training
- **endpoint**: uiuc-cig-01-gpu-02
- **start_at**: '2022-08-16 09:01:22.433471'
- **end_at**: '2022-08-16 09:01:54.578533'
- **events**:
  - **start_endpoint_execution**: '2022-08-16 09:01:22.552091'
  - **stop_endpoint_execution**: '2022-08-16 09:01:54.578533'
- **timing**:
  - **load_dataset**: 7.529
  - **download_global_state**: 0.626
  - **load_global_state_to_device**: 3.94
  - **training_client_update**: 19.404
  - **epoch_1**:
    - **val_before_update_val_set**: 0.726
    - **train_one_epoch**: 1.849
  - **epoch_2**:
    - **train_one_epoch**: 1.852
TAKEAWAYS

- **APPFL**: An open-source framework for privacy-preserving federated-learning tasks
- **funcX** helps us to efficiently perform remote task execution at clients **without thinking too much about the underlying computing infrastructure**
- **funcX authentication** via Globus provides a reliable way for managing data/computing resources access for teams of cross-institution researchers
- Our project facilitates collaborations on developing machine learning algorithms in many biomedicine researches. **We are actively looking for collaborations in the future!**

THANKS FOR YOUR ATTENTION

hthieu@illinois.edu