Non-IID Quantum Federated Learning with One-shot Communication Complexity

Code on GitHub

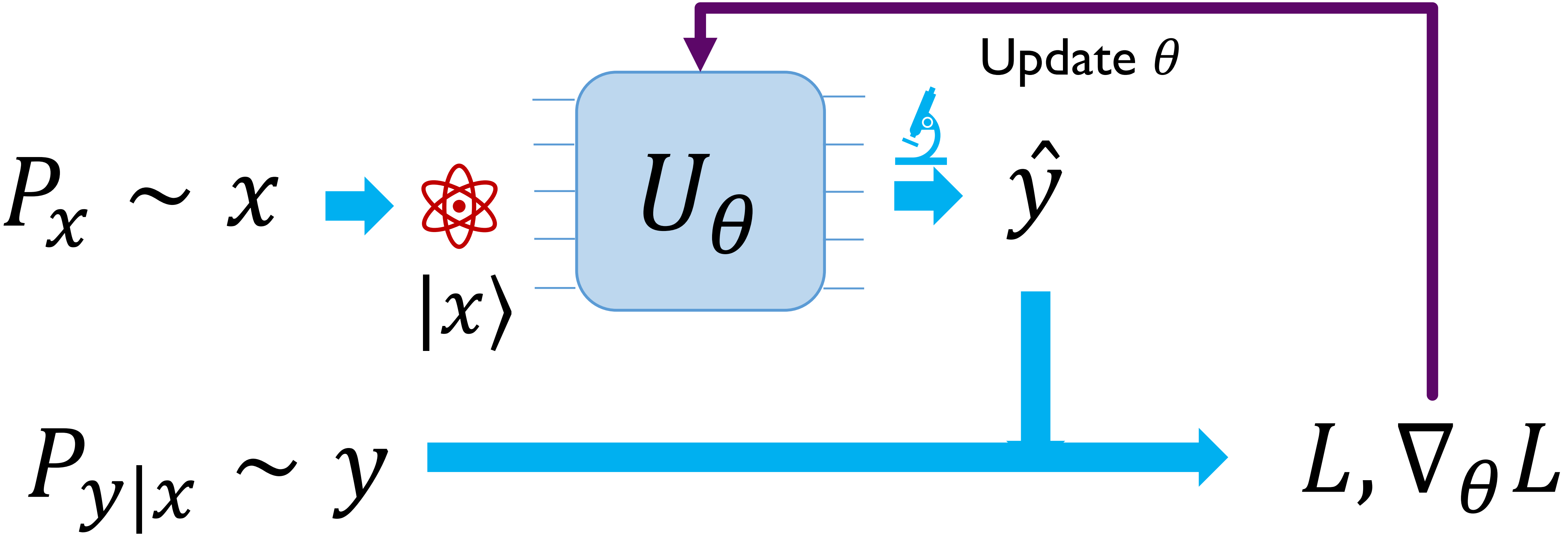
Haimeng Zhao
haimengzhao@icloud.com
Tsinghua University

arXiv:2209.00768
Quantum Machine Intelligence 5,3.

2023-11-20, QTML 2023

Motivation

- Quantum machine learning (supervised, variational)



Motivation

- Data in reality
  - Collected by different clients
    - Non-IID:

    a different distribution for each client

x_i^A

Client A

x_i^B

Client B

x_i^C

Client C

Motivation

- Data in reality
  - Collected by different clients
  - Privacy
    - x_i^A: Record of patient i at hospital A
    - No sharing across clients!
    - Cyber-physical securities, IoT

$$x_i^A$$

Hospital A

$$x_i^B$$

Hospital B

$$x_i^C$$

Hospital C

Motivation

- Data in reality
  - Collected by different clients
  - Privacy
  - Hard to transmit
    - Huge volume
    - Fragile data (quantum states)

x_i^A

Lab A

x_i^B

Lab B

x_i^C

Lab C

How to learn from decentralized private data?

Quantum Federated Averaging (qFedAvg)

- Each client C keeps a record of its own parameters θ_C
  - (1). Local updates for T steps

    Each client updates its θ_C using gradient descent on its own data x^C
  - (2). Global averaging

    $\theta \leftarrow \sum_C p_C \theta_C$, prior p_C = #data in C / total #data
  - (3). Broadcast

    $\theta_C \leftarrow \theta$ for all C
- Application to QML is straightforward!

  (Li et al. 2021, Xia et al. 2021, Chen et al. 2021, etc.)

McMahan et al. 2017

Limitations of (q)FedAvg

- Non-IID quagmire:
  - Performance deteriorates significantly when data are non-IID across clients
- Gradient inversion attack:
  - Gradient updates can be easily reversed engineered => private data leakage
- Large communication overhead:
  - Communication cost $\propto$ #iterations $\times$ #parameters $\times$ #clients

Main Results

- Non-IID quagmire: ?, this work: × qFedAvg √ qFedInf (NISQ)
  - Performance deteriorates significantly when data are non-IID across clients
- Gradient inversion attack: √, Li et al. 2021: quantum blind computing
  √, this work: a NISQ alternative
  - Gradient updates can be easily reversed engineered => private data leakage
- Large communication overhead: ?, this work: √ qFedInf
  Communication cost $\propto$ #iterations × #parameters × #clients
  ~~#iterations ×~~ #parameters × #clients: one-shot

Non-IID quagmire of qFedAvg

(NISQ)

Prop. 1 (informal). Assuming no entanglement among clients. Under certain assumptions on loss function and the smoothness of its gradient, at iteration m, the deviation of parameters learned by qFedAvg from centralized SGD is

$$\underbrace{\Delta_m \leq \sum_C p_C\, a_C^T\, \Delta_{m-1}}_{\text{past deviation}} + \underbrace{\eta \sum_C p_C\, g_C\, \mathrm{EMD}_C}_{\text{non-IID}},$$

where EMD_C is the earth mover distance between the data distribution in Client C and the total distribution, $a_C > 1$ and $g_C > 0$ depends on the loss function, η is the learning rate.

Quantum Federated Inference (qFedInf)

Idea (classical version) explained in one formula:

$$\boxed{f_y(x)} = \mathbb{P}[y|x] = \sum_C \mathbb{P}[y|x \text{ from Client } C]\mathbb{P}[\text{from Client } C|x]$$

global model

Bayes

$$= \sum_C \boxed{f_y^C(x)} \frac{\mathbb{P}[x|\text{from Client } C]\mathbb{P}[\text{from Client } C]}{\mathbb{P}[x]}$$

model of Client C

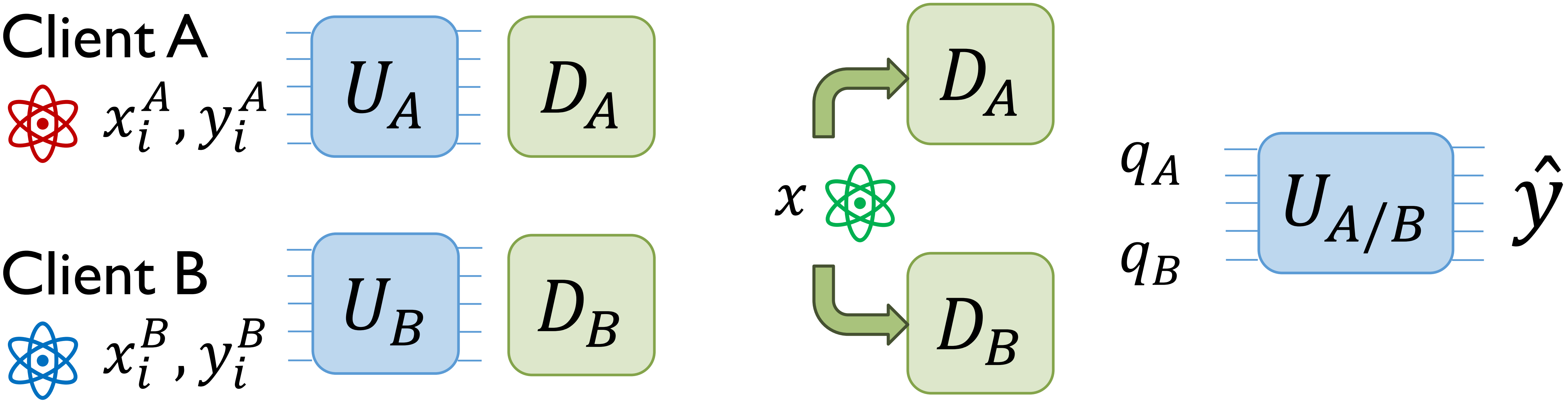
$$= \sum_C \boxed{f_y^C(x)} \frac{\boxed{D_C(x)}p_C}{\sum_{C'} \boxed{D_{C'}(x)}p_{C'}}.$$

density estimator of client C

Quantum Federated Inference (qFedInf)

Training Phase

Inference Phase

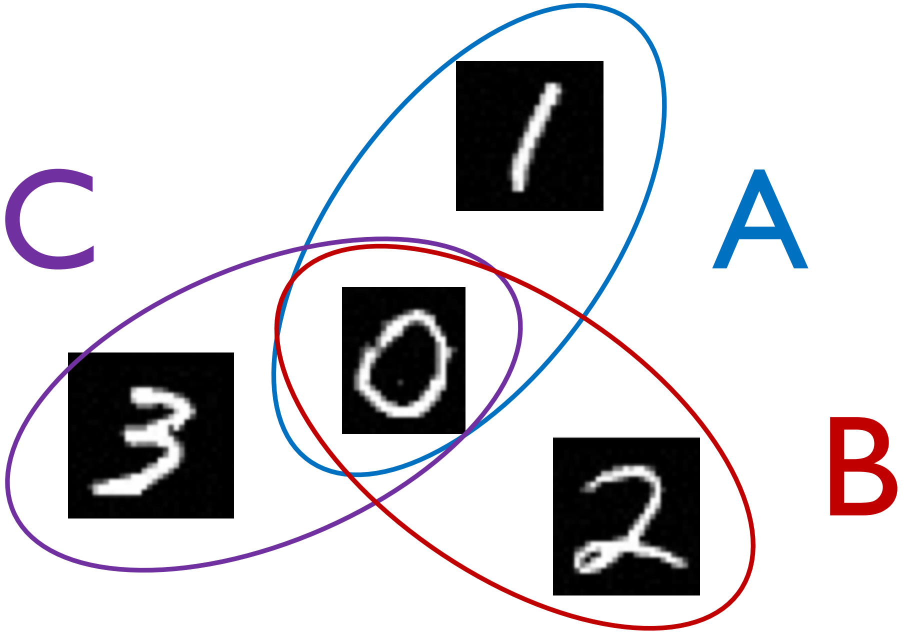


Standard local training + density estimator
Transmit *U, D* : one-shot, anonymized

Randomized inference

Testbed: generating non-IID quantum datasets

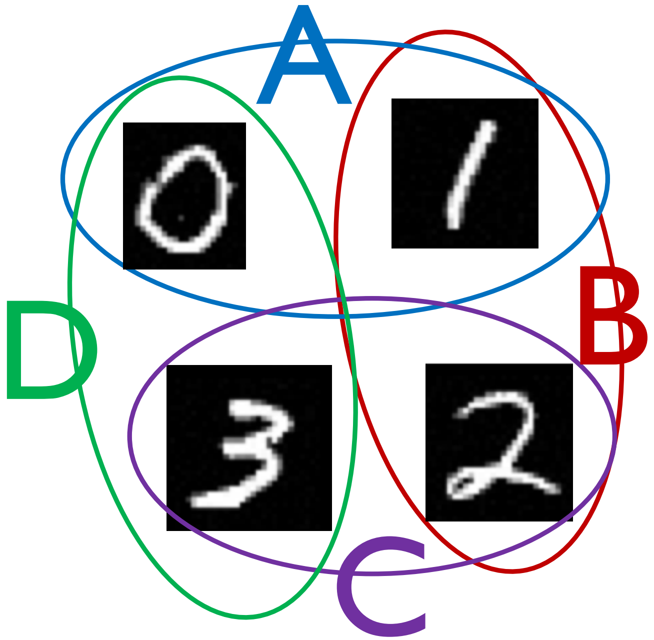
- Star structure



Star

- Cycle-m structure

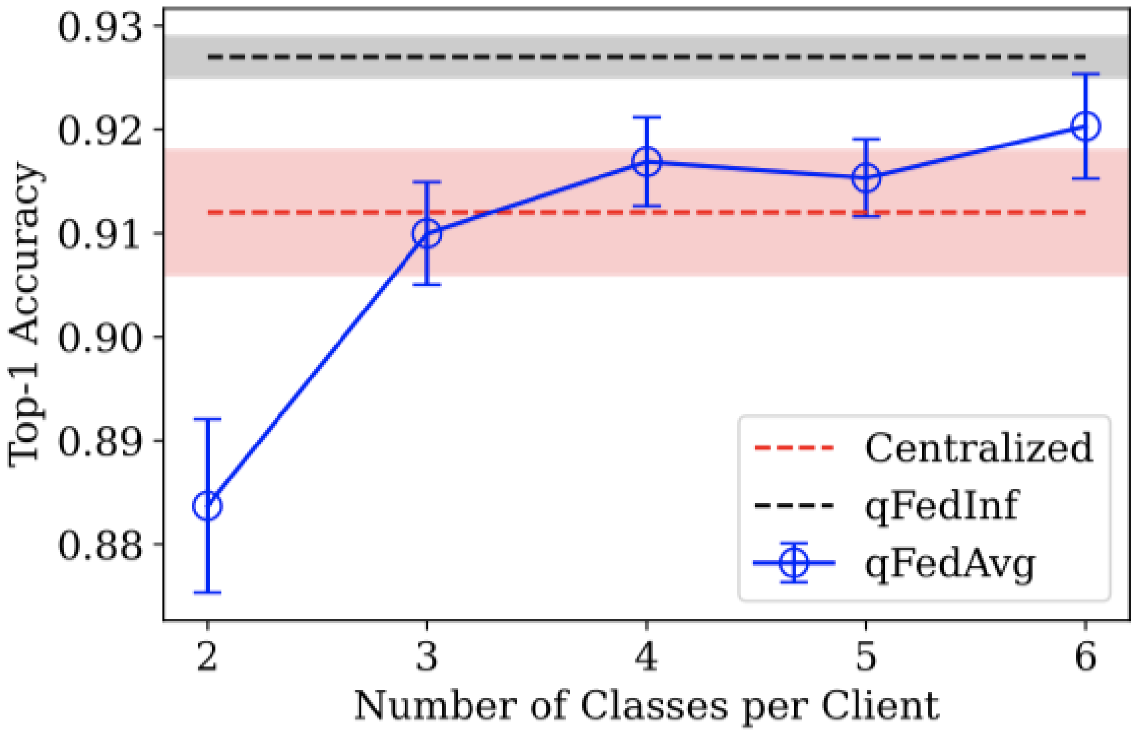
m: level of non-IID

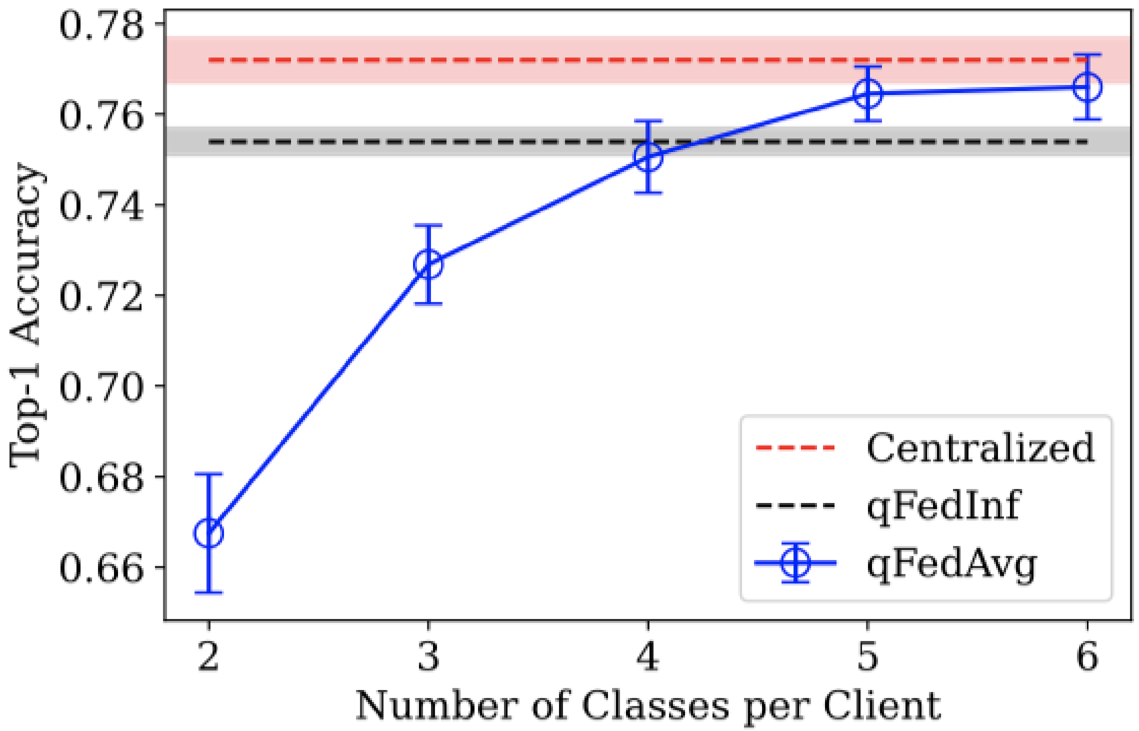


Cycle-2

Performance Comparison

quantum model: variational quantum circuits
density estimator: Gaussian mixture models





Summary

GitHub
Code

- qFedInf vs qFedAvg

  Non-IID quagmire √, gradient inversion attack √, communication √

  (not covered) mixture of experts, ensemble learning, generative learning

- Outlook

  Go beyond NISQ and consider entanglement among clients

  Possible quantum advantage:

    computation/communication/privacy/robustness to non-IID