Understanding Server-Assisted Federated Learning in the Presence of Incomplete Client Participation

Haibo Yang[1], Peiwen Qiu[2], Prashant Khanduri[3], Minghong Fang[4], Jia Liu[2]

[1]GCCIS, Rochester Institute of Technology [2]Dept. of ECE, The Ohio State University
[3]Dept. of CS, Wayne State University [4]Dept. of CSE, University of Louisville

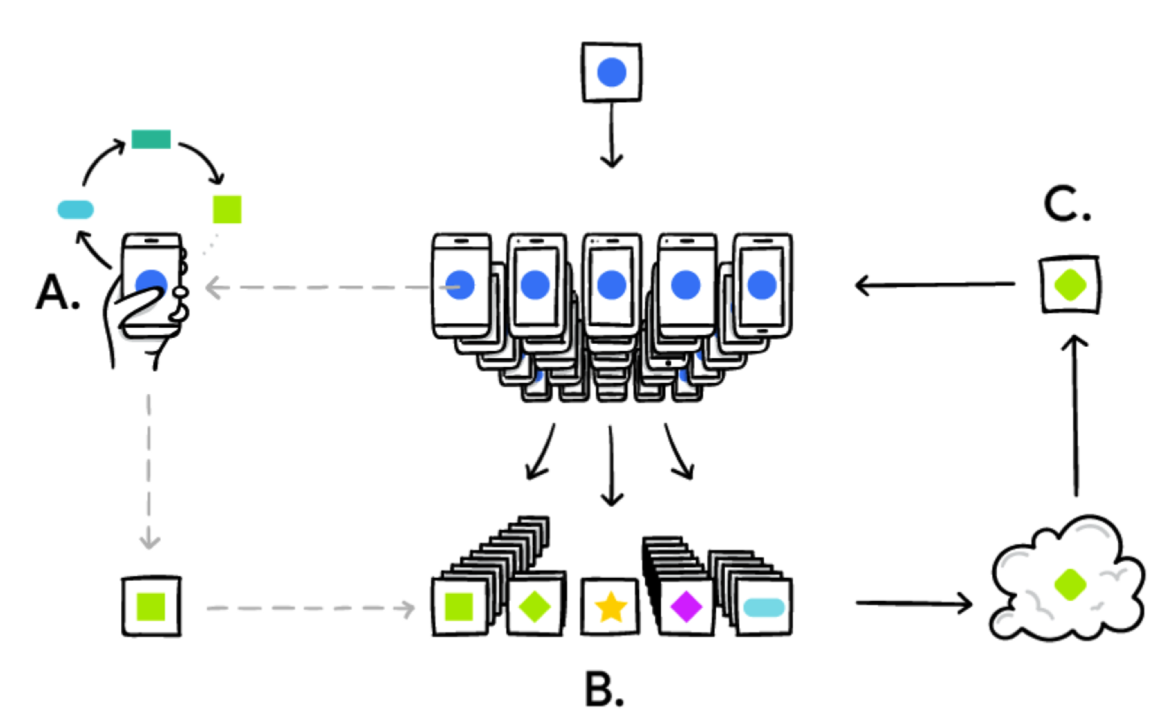
Summary

- Caveat: federated learning under *arbitrary* data and system heterogeneity is not PAC-learnable (Probably Approximately Correct) in the worst case.
- Solution: using server-side auxiliary dataset as a control knob to revive PAC learnability of FL.
- A baseline algorithm, named as SAFARI, is proposed by designed coordination between server and clients.

Federated Learning

Federated Learning achieves the best of both worlds: learning and preserving privacy.

- Collaboration from many clients.
- Keep training data local → data privacy protection.



Google Federated Learning Demo

https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

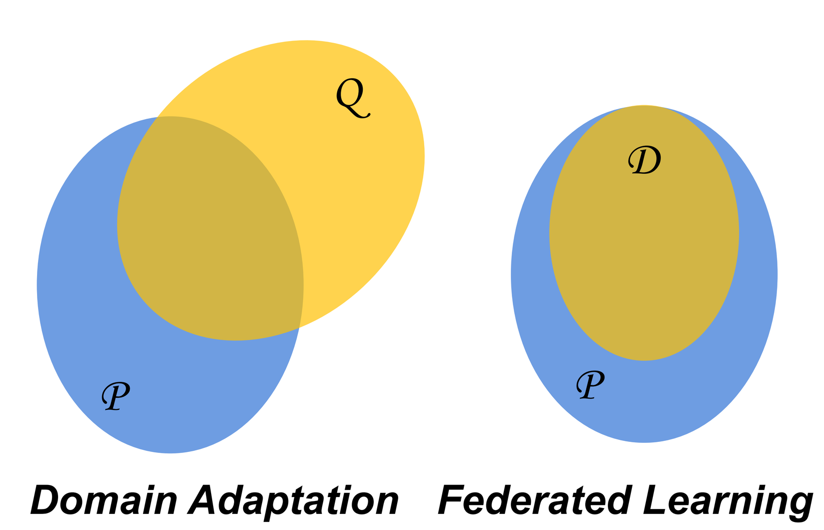
However, federated learning introduces two challenges:

- *data heterogeneity*: Non-IID datasets.
- *system heterogeneity*: unpredictable/uncontrollable clients

→ **Fact: incomplete client participation (ICP).** Some clients may never participate in the FL training.
→ **Question:** what if Federated Learning with ICP?
→ **Solution:** not PAC-learnable → adding auxiliary dataset in server's side → sufficient condition for PAC learnability

FL with incomplete client participation

- **Federated learning with ICP is not PAC-learnable.**
There exists a client participation process $\mathcal{F}$, a distribution P, and a system capacity $\alpha = \frac{m}{M}$, such that
$$\mathbb{P}_{S\sim P}\left[\mathcal{R}_P(\mathcal{L}(\mathcal{F}(S)), f) > \frac{1-\alpha}{8}\right] > \frac{1}{20}$$
- **Under mild conditions, the PAC-learnable of SA-FL is revived.**
(α, β)-positively-related: there exist constants $\alpha \geq 0$ and $\beta \geq 0$ such that
$$|\varepsilon_P(h) - \varepsilon_Q(h)| \leq \alpha[\varepsilon_Q(h)]^{\beta}, \forall h \in \mathcal{H}.$$
With probability at least $1-\delta$ for any $\delta \in (0,1)$, it holds that
$$\varepsilon_P(\hat{h}_Q^*) = \mathcal{O}\left(\left(\frac{d_{\mathcal{H}}}{n_T+n_S}\right)^{\frac{1}{2-\beta_Q}} + \left(\frac{d_{\mathcal{H}}}{n_T+n_S}\right)^{\frac{\beta}{2-\beta_Q}}\right),$$
where $d_{\mathcal{H}}$ denotes the finite VC dimension for hypotheses class $\mathcal{H}$.
✓ (α,β)-positively-related of P and Q → Generalization error: $\mathcal{O}\left(\left(\frac{d_{\mathcal{H}}}{n_T+n_S}\right)^{\frac{\beta}{2-\beta_Q}}\right)$
P is target distribution, D is the shifted distribution due to ICP, $(Q = P + D)$
- **With further conditions, FL is strictly better than centralized learning.**
If $\hat{\mathcal{R}}_P(\hat{h}_Q^*) \leq \hat{\mathcal{R}}_P(h_Q^*)$ and $\varepsilon_P(h_Q^*) = \mathcal{O}(\mathcal{A}(n_T,\delta))$, then with probability at least $1-\delta$ for any $\delta \in (0,1)$, it holds that $\varepsilon_P(\hat{h}_Q^*) = \mathcal{O}\left((d_{\mathcal{H}}/n_T)^{\frac{1}{2-\beta_P}}\right)$, $\mathcal{A}(n_T,\delta) = \frac{d_{\mathcal{H}}}{n_T}\log(\frac{n_T}{d_{\mathcal{H}}} + \frac{1}{n_T}\log(\frac{1}{\delta}))$.
- **Diagram of distribution for domain adaptation and federated learning.**



SAFARI algorithm

- **Update steps.**
  - With Prob. p: $\mathbf{x}_{t+1} = \mathbf{x}_t + \eta\left(\frac{1}{|S_t|}\sum_{i\in S_t}\Delta_t^i\right), \Delta_t^i = -\sum_{k=0}^{K-1}\nabla F_i(\mathbf{x}_{t,k}^i, \xi_{t,k}^i)$ ★ client update option
  - Otherwise: $\mathbf{x}_{t+1} = \mathbf{x}_t - \eta_s \nabla F(\mathbf{x}_t, \xi_t)$ ★ server update option
- **Convergence Guarantees.** Under mild conditions, the convergence rate is $\mathcal{O}(1/\sqrt{mKR})$, where m is the clients' number, K is the local update steps, and R is the number of communication rounds.

Experiments

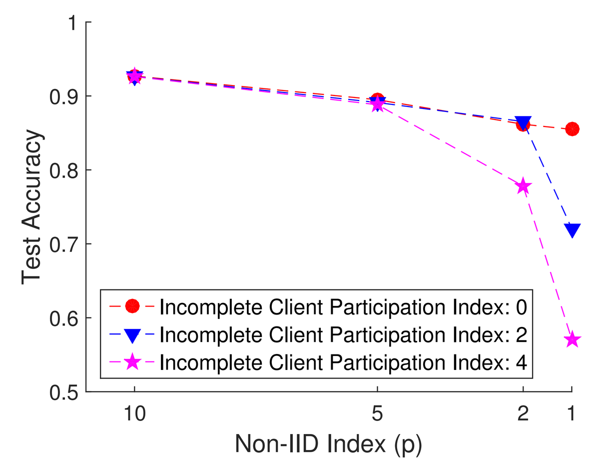


Figure 1:Test Accuracy of FedAvg on MNIST with incomplete client participation.

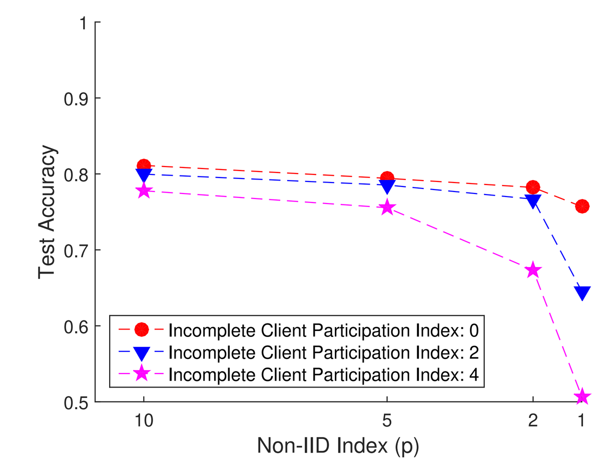


Figure 2:Test Accuracy of FedAvg on CIFAR-10 with incomplete client participation.

Table 1:Test accuracy improvement (%) with auxiliary dataset.

Server Datasize	Non-IID Index (p) 10	5	2	1
50	-	-	-	12.32
100	-	-	5.24	16.48
500	-	-	9.40	27.55
1000	-	-	10.08	28.78

Conclusion

- Auxiliary dataset in server's side help revive the learnability of federated learning with incomplete client participation.
- A new algorithm, SAFARI (server-assisted federated averaging), is proposed with the same linear speedup convergence guarantees as classic FL with ideal client participation.

Acknowledgements

This work is supported in part by AI Seed Funding and the GWBC Award at RIT, as well as NSF grants CAREER CNS-2110259, CNS-2112471, and IIS-2324052. The authors also thank Mr. Zhe Li for his assistance with the part of the experiments.

Correspondence to:

- Haibo Yang (hbycis@rit.edu)
- Jia Liu (liu@ece.osu.edu)