Google Research

Practical and Private Federated Learning

Zheng Xu
Google Research
June, 2021

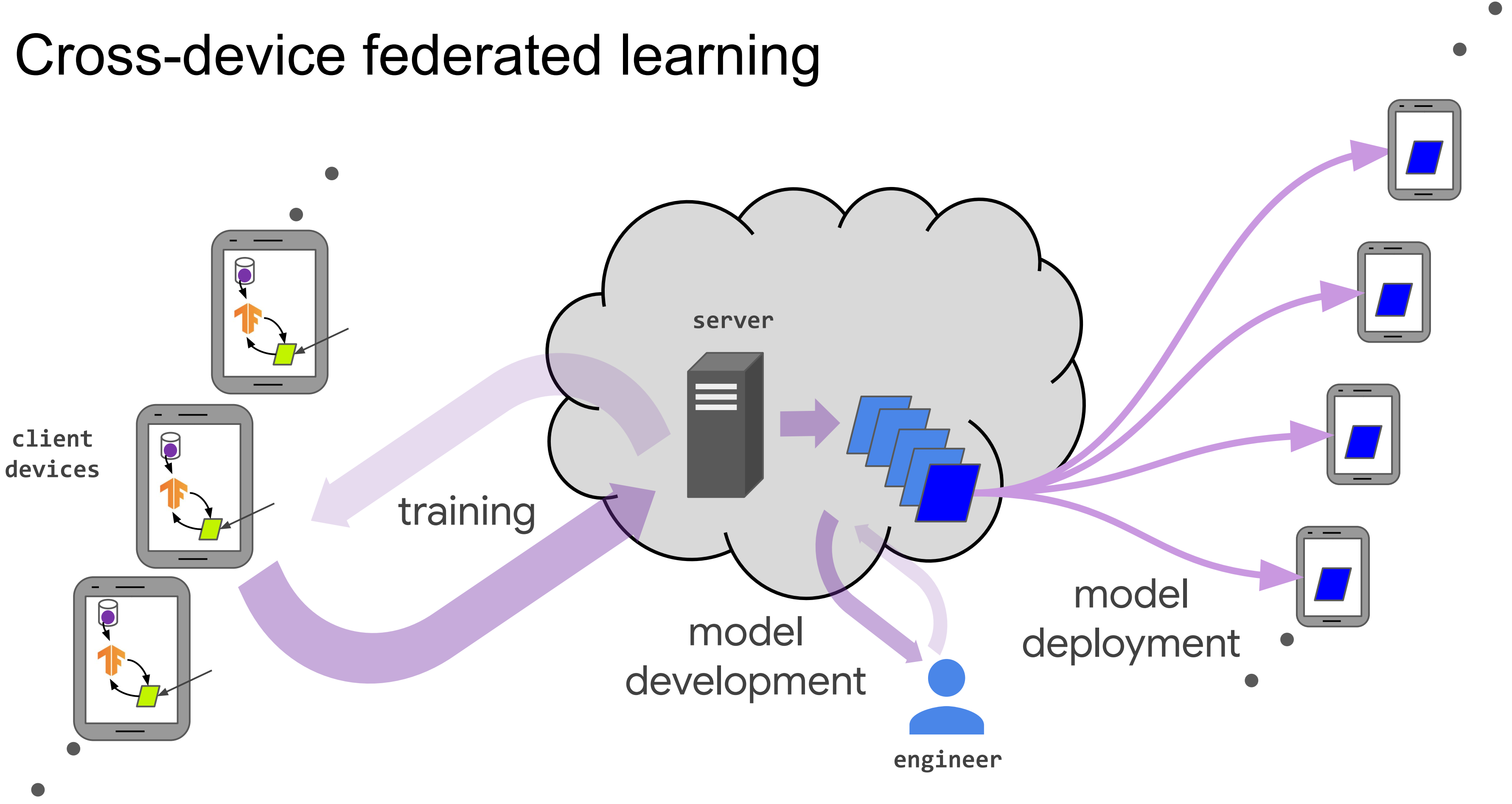
Presenting the work of many

Federated learning

Federated learning is a machine learning setting where multiple entities (clients) collaborate in solving a machine learning problem, under the coordination of a central server or service provider. Each client's raw data is stored locally and not exchanged or transferred; instead, focused updates intended for immediate aggregation are used to achieve the learning objective.

Advances and Open Problems in Federated Learning, arxiv 2019

Cross-device federated learning



Key issues

Federated learning

- Communication efficiency
- (Data) heterogeneity
- Computational constraints
- Privacy and security
- System complexity

Cross-device settings

- Number of clients
- Client availability
- Connection topology
- Computation and communication
- Client states

Key issues

Federated learning

- Communication efficiency
- (Data) heterogeneity
- Computational constraints
- Privacy and security
- System complexity

Cross-device settings

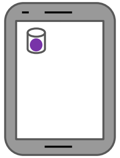
- Number of clients
- Client availability
- Connection topology
- Computation and communication
- Client states

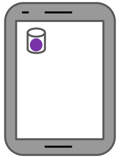
Practical and Private (Deep) Learning without Sampling or Shuffling, ICML 2021

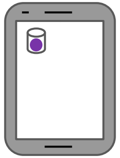
Joint work with Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta

User-level differential privacy

client
devices
("database")

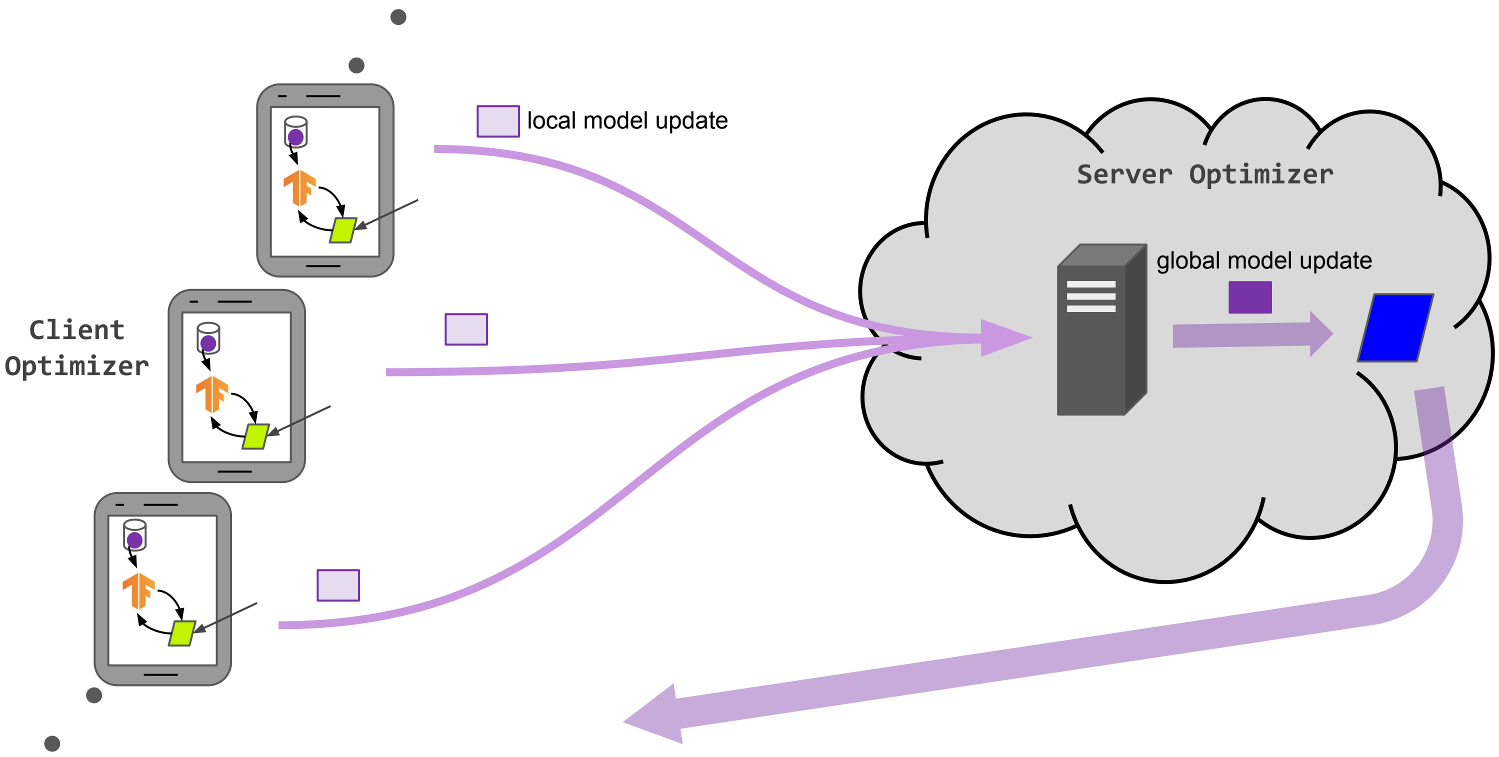




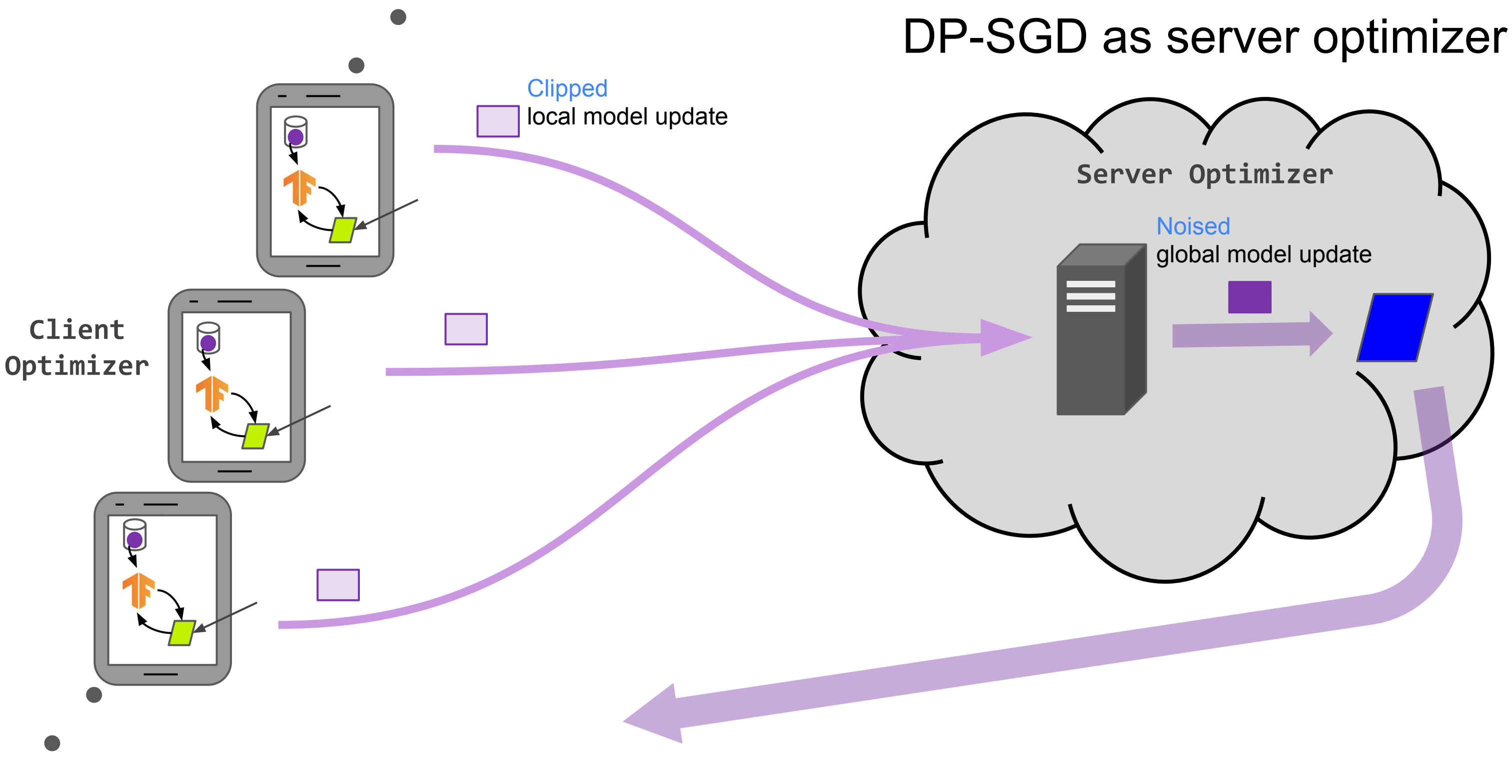


User	Example
NORFOLK	Stay, my lord,
NORFOLK	And let your reason with your choler question
NORFOLK	What 'tis you go about: to climb steep hills
NORFOLK	Requires slow pace at first: anger is like
NORFOLK	A full-hot horse, who being allow'd his way,
NORFOLK	Self-mettle tires him. Not a man in England
NORFOLK	Can advise me like you: be to yourself
NORFOLK	As you would to your friend.
BUCKINGHAM	Come on you target for faraway laughter,
BUCKINGHAM	Come on you stranger, you legend, you martyr,
BUCKINGHAM	You reached for the secret too soon, you
BUCKINGHAM	cried for the moon.
HENRY VIII	My life itself, and the best heart of it,
HENRY VIII	Thanks you for this great care: I stood i' the level
HENRY VIII	Of a full-charged confederacy, and give thanks
HENRY VIII	To you that choked it. Let be call'd before us
...	...

Generalized federated averaging (FedOpt)

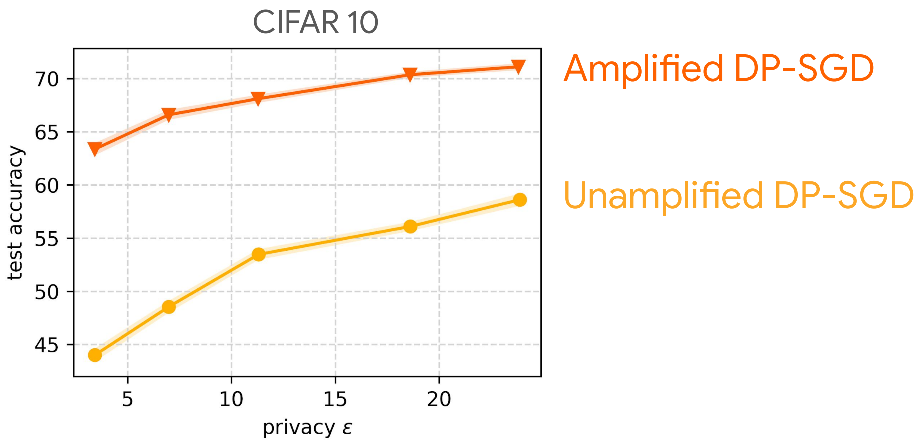


Differentially private FedOpt



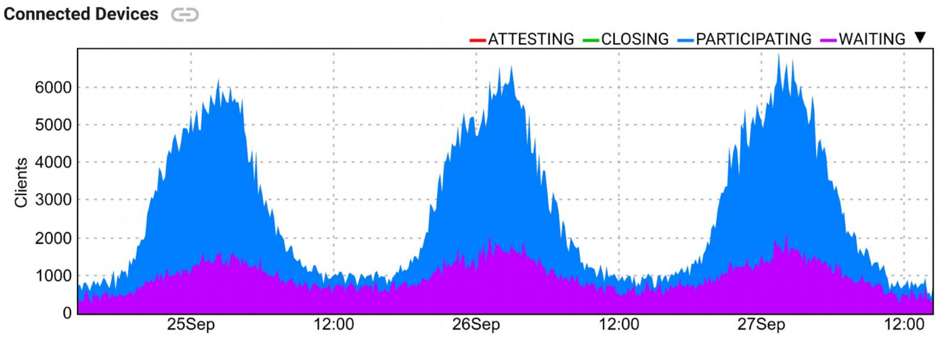
Amplification for privacy/utility trade-off

- Subsampling can amplify privacy if the minibatch is uniformly sampled.:
    - Without amplification: noise ~ √T / batch size
    - With amplification: noise ~ √T / data size



Difficulty of amplification in practice

In federated learning:



Device availability [BKMTT20]

In centralized training:

- Sampling every round is expensive
- Most deployed systems do not implement it faithfully

DP-FTRL: DP-Follow The Regularized Leader

Can we avoid sampling and achieve similar privacy/utility trade-offs?

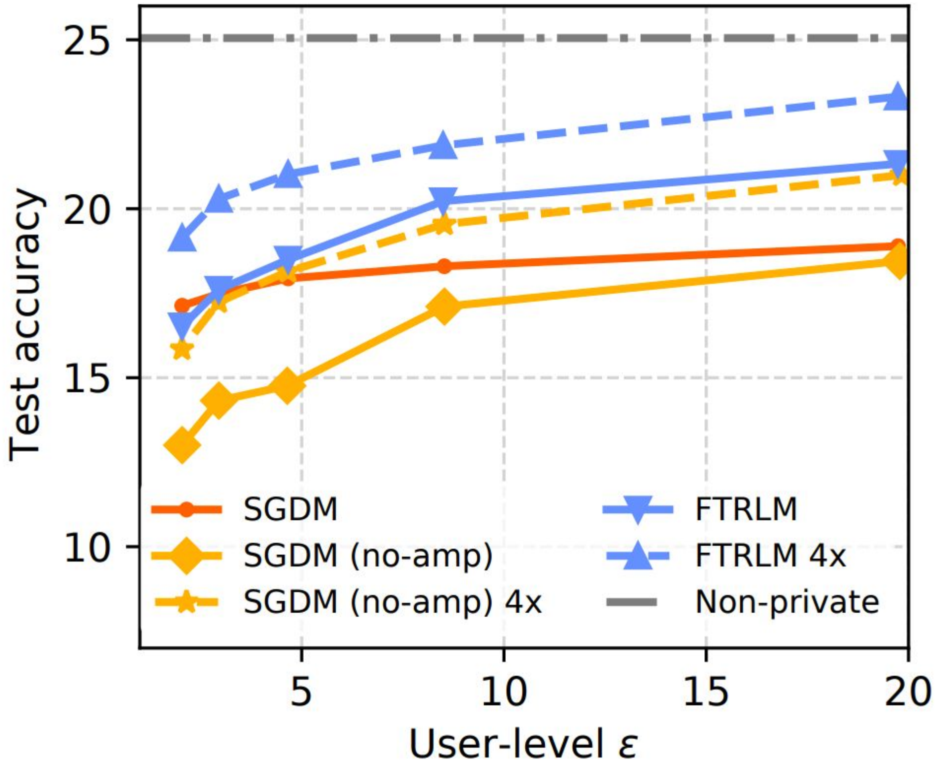
Data-dependent component in SGD: prefix sums of gradients g

$$g_1 \quad , \quad g_1 + g_2 \quad , \ldots, \quad g_1 + g_2 + \ldots + g_n$$

Main Idea: Compute prefix sums privately using Tree Aggregation Protocol (correlated noise) [CSS10,DNPR10,ST13,Honaker15]

Best known excess population risk for a single pass algorithm that *does not* rely on convexity for privacy.

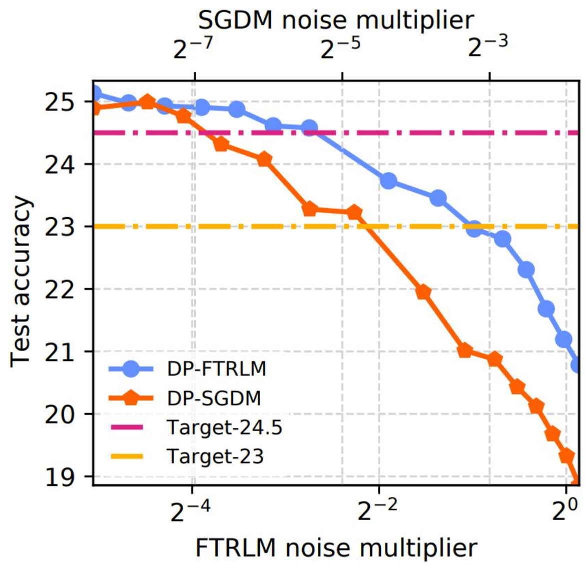
Privacy/Utility trade-offs

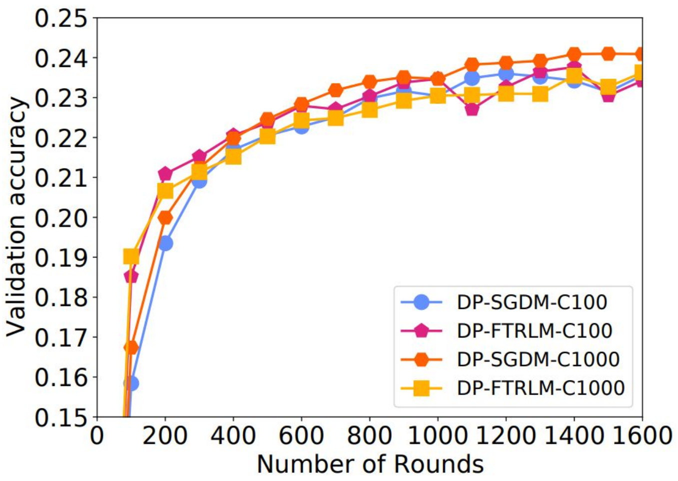


StackOverflow Next Word Prediction

- DP-FTRL outperforms DP-SGD without amplification
- DP-FTRL is competitive to/outperforms amplified DP-SGD at $\epsilon > 2$
- DP-FTRL outperforms amplified DP-SGD with modest increase in clients per round

Privacy/Computation trade-offs for targeted utility



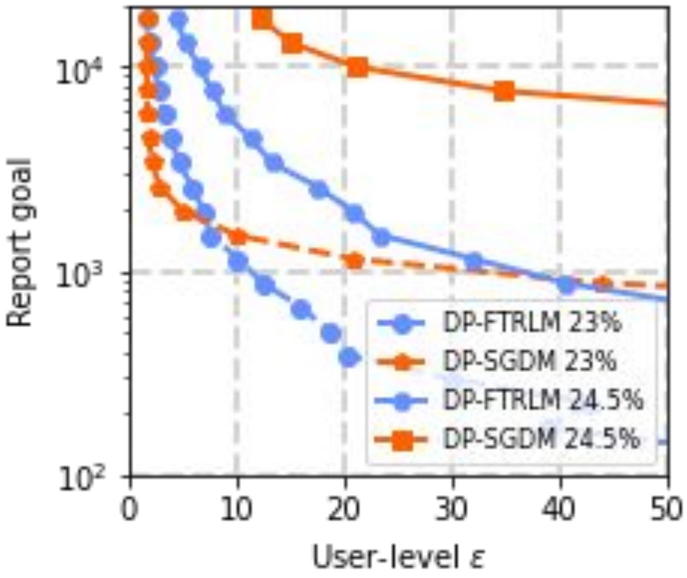


Hypothesis verified

In practice, targeted utility can be met by increasing computation based on hypothesis:

For sufficiently large data, the utility accuracy will not drop if noise multiplier and clients per round proportionally increase. [MRTZ'17]

Privacy/Computation Trade-offs for Targeted Utility



DP-FTRL provides similar/better privacy-computation trade-offs than DP-SGD

Key issues

Federated learning

- Communication efficiency
- (Data) heterogeneity
- Computational constraints
- Privacy and security
- System complexity

Cross-device settings

- Number of clients
- Client availability
- Connection topology
- Computation and communication
- Client states

Practical and Private (Deep) Learning without Sampling or Shuffling, ICML 2021

Key issues

Federated learning

- Communication efficiency
- (Data) heterogeneity
- Computational constraints
- Privacy and security
- System complexity

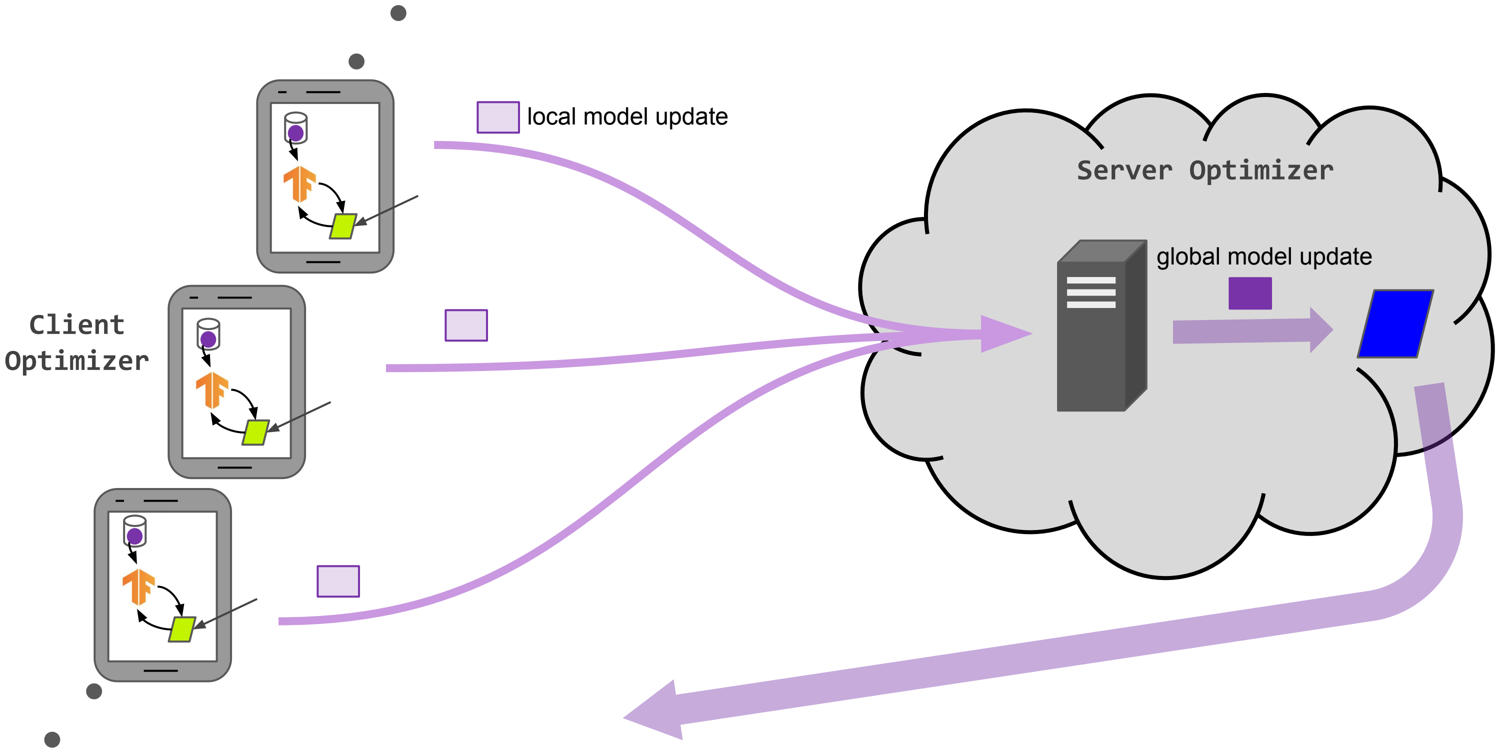
Cross-device settings

- Number of clients
- Client availability
- Connection topology
- Computation and communication
- Client states

Local Adaptivity in Federated Learning: Convergence and Consistency, arxiv 2021
Joint work with Jianyu Wang, Zachary Garrett, Zachary Charles, Luyang Liu, Gauri Joshi

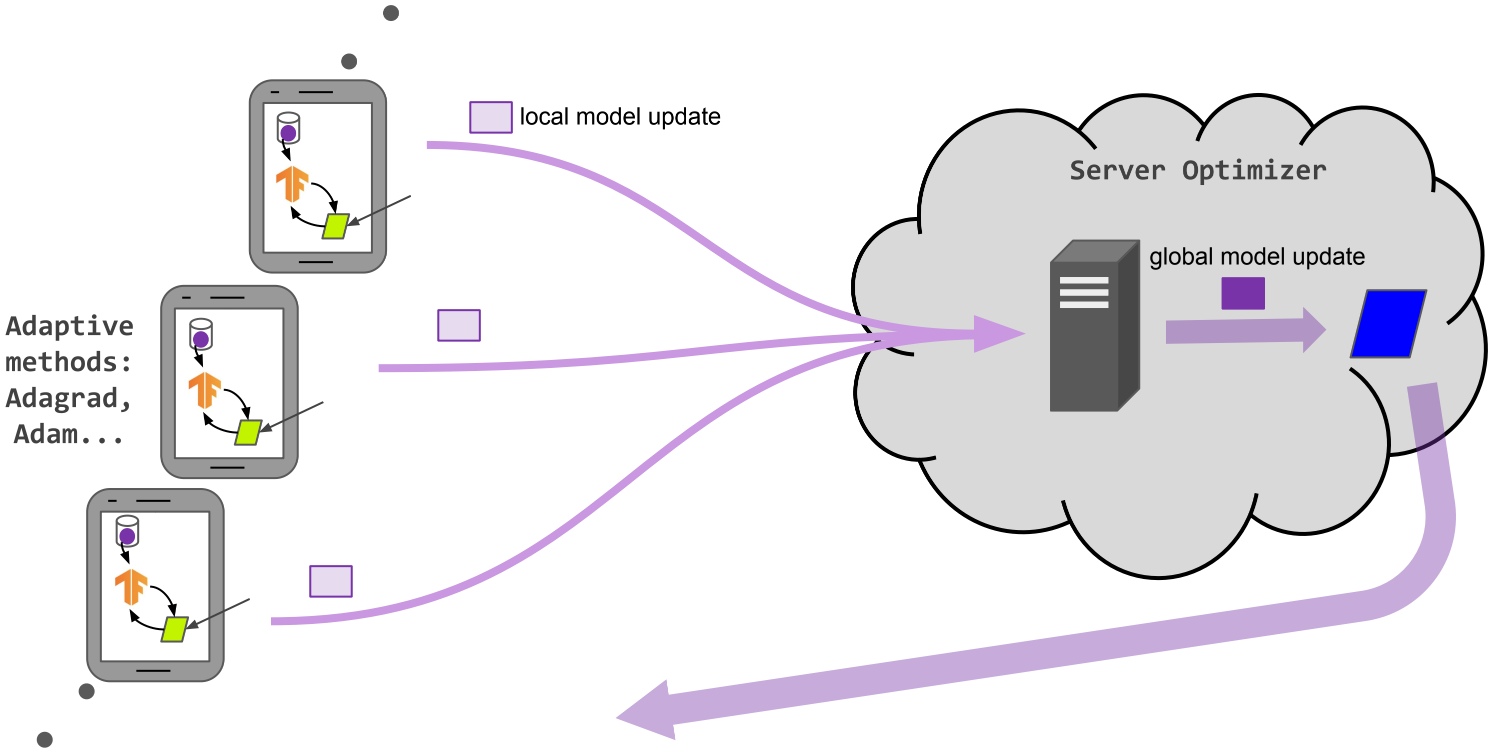
Generalized federated averaging (FedOpt)

How can we choose optimizer for heterogeneous clients?



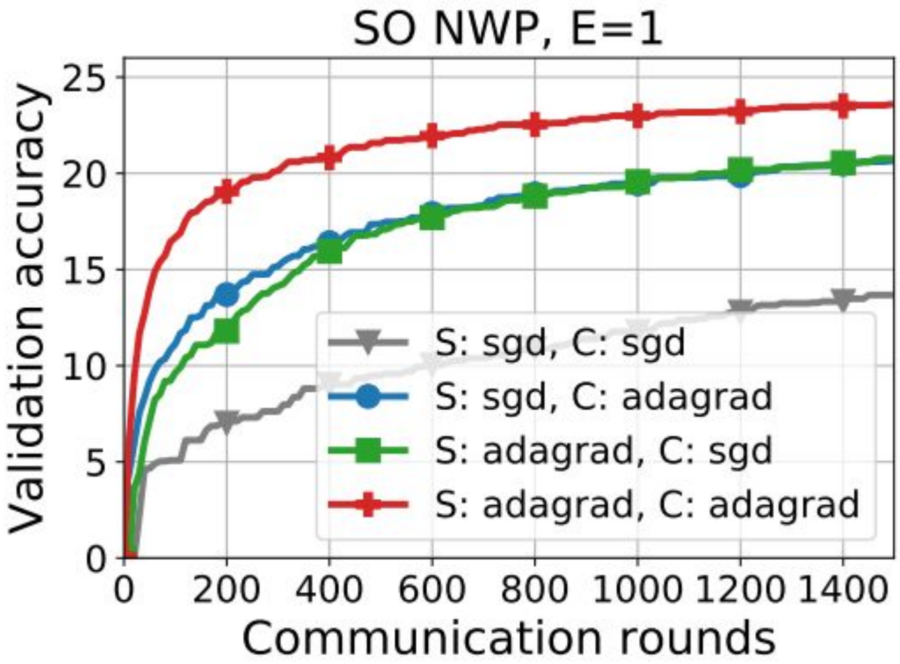
Local adaptivity in federated learning

How can we choose optimizer for heterogeneous clients?



Local adaptivity: advantages

Fast convergence



Local adaptivity: advantages

Fast convergence

Accuracy improvement

	C:SGD	C:ADAGRAD
S:SGD	14.19	21.68
S:ADAGRAD	21.80	24.40

(a) SO NWP

Local adaptivity: advantages

Fast convergence

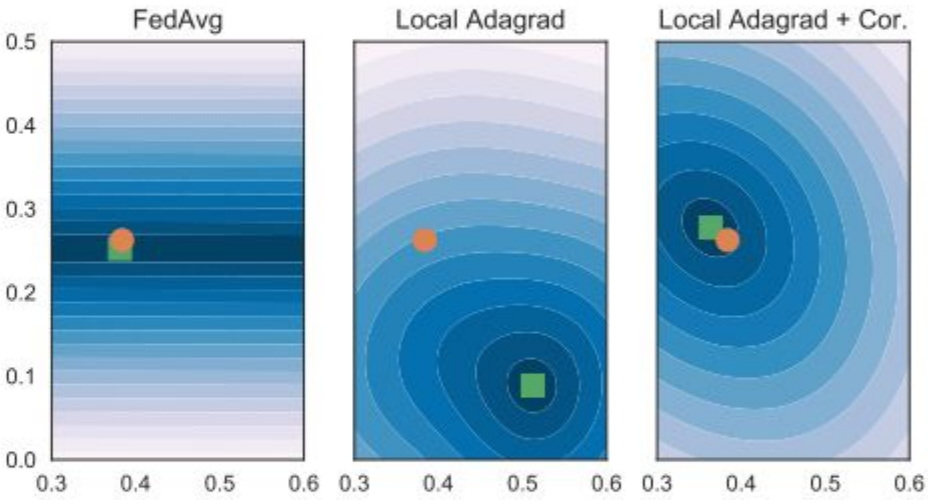
Accuracy improvement

Hyperparameter sensitivity

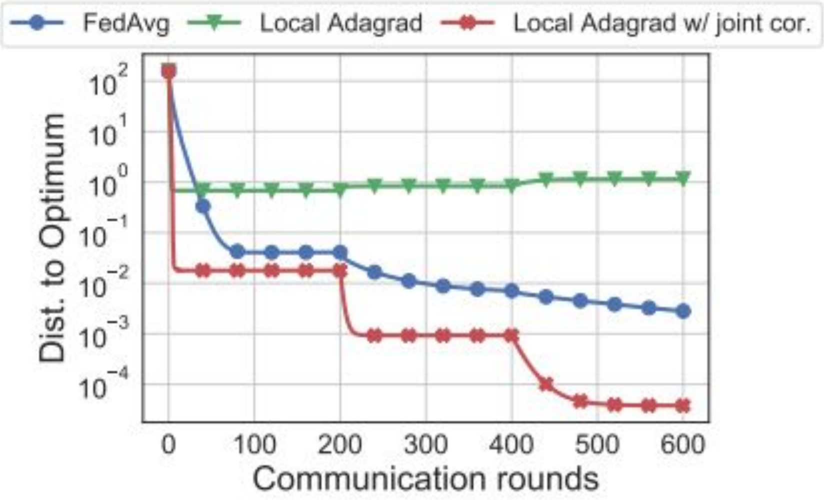
(a) S: ADAGRAD, C: ADAGRAD

(b) S: ADAGRAD, C: SGD

Local adaptivity: consistency and correction



(a)



(b)

Local adaptivity: empirical results

Training Tasks	ServerOpt	ClientOpt			
		SGD	AdaGrad	+ Local Cor.	+ Joint Cor.
SO NWP	Adam	24.40	24.70	24.81	**24.85**

ClientOpt	No Cor.	Local Cor.	Joint Cor.
Yogi [13]	24.80	25.29	**25.33**
Adam	24.86	25.15	**25.35**

Key issues

Federated learning

- Communication efficiency
- (Data) heterogeneity
- Computational constraints
- Privacy and security
- System complexity

Cross-device settings

- Number of clients
- Client availability
- Connection topology
- Computation and communication
- Client states

Practical and Private (Deep) Learning without Sampling or Shuffling, ICML 2021
Local Adaptivity in Federated Learning: Convergence and Consistency, arxiv 2021

Conclusion

- Federated learning can be practical and private
- “Constraints” of practical federated learning
  - Privacy protection and system complexity
- Interdisciplinary research with many open problems
  - Simulation for evaluation
  - Theory and practice
  - Robustness, fairness, and personalization

Advances and Open Problems in Federated Learning, arxiv 2019
Upcoming white paper on arxiv in about one week