Benchmarking of Federated Learning

Comparison of three algorithms

Bjarne Pfitzner, Prof. Dr. Bert Arnrich

Algorithms

- Federated learning parameters:
  - K: # of participating clients
  - C: Fraction of clients selected per global epoch
  - T: # of global epochs
- Federated Averaging (FedAvg) [1]

Algorithm 1 `FederatedAveraging (FedAvg)` [1]

Server executes:
  Initialise θ^0
  $m \leftarrow \max(C \times K, 1)$
  for $t = 1$ to T **do**
    $S^t \leftarrow$ (random set of m clients)
    for each client $k \in S^t$ **do**
      $\theta_k^t \leftarrow$ **ClientUpdate**(θ^{t-1})
    end for
    $\theta^t \leftarrow \sum_k \frac{n_k}{n} \theta_k^t$
  end for

ClientUpdate(θ): $\triangleright$ for client k
  $\theta_k \leftarrow \theta$
  for each local iteration **do**
    for each batch b_k in client's split **do**
      $\theta_k \leftarrow \theta_k - \eta \nabla L(b_k; \theta_k)$
    end for
  end for
  return local model θ_k

- LoAdaBoost [2]
  - Keep track of median client loss L_{median}
  - Train locally for E/2 epochs
  - While $L_k < L_{median}$ do:
    - Repeat local training for (E/2 - trial) epochs
    - Break if total local epochs ≥3E/2
- Co-Learning [3]
  - Increasing local epochs (ILE), using:

$$E^t = \begin{cases} E^0, & \text{if } t = 1, \\ 2 * E^{t-1}, & \text{if } t > 0 \;\&\; \frac{|\theta^t - \theta^{t-1}|}{|\theta^{t-1}|} \leq \epsilon, \\ E^{t-1}, & \text{if } t > 0 \;\&\; \frac{|\theta^t - \theta^{t-1}|}{|\theta^{t-1}|} > \epsilon \end{cases}$$

  - Cyclical learning rate using: $\eta^e = \eta^0 * \lambda^{e/E^t}$

Results

- Dataset: MNIST (handwritten digits)
- Model: CNN
- K=100,
- n_k=600,
- LR (η)=0.01,
- LR decay=0.001,
- Co-Learning parameter (ε)=0.01
- RoC: Rounds of communication to reach 97% accuracy
- **„Infinite" batch size (full local dataset) performs poorly**
- **LoAdaBoost has overall slower convergence than the other two**
- **Generally higher *E* lead to faster convergence**

Algorithm	C	B	E	RoC
FedAvg	0.1	∞	5	528
	0.1	∞	10	237
	0.1	10	5	12
	0.1	10	10	11
	0.5	∞	5	497
	0.5	∞	10	232
	0.5	10	5	11
	0.5	10	10	8
	1.0	∞	5	522
	1.0	∞	10	221
	1.0	10	5	12
	1.0	10	10	9
Co-Learning	0.1	∞	5	63
	0.1	∞	10	87
	0.1	10	5	13
	0.1	10	10	12
	0.5	∞	5	81
	0.5	∞	10	57
	0.5	10	5	22
	0.5	10	10	12
	1.0	∞	5	81
	1.0	∞	10	88
	1.0	10	5	18
	1.0	10	10	9
LoAdaBoost	0.1	∞	5	(90.96%)
	0.1	∞	10	(94.26%)
	0.1	10	5	134
	0.1	10	10	233
	0.5	∞	5	(96.60%)
	0.5	∞	10	357
	0.5	10	5	21
	0.5	10	10	17
	1.0	∞	5	(96.83%)
	1.0	∞	10	498
	1.0	10	5	27
	1.0	10	10	20

Future Work

- Increase hyperparameter search space
- Include more complex algorithms & datasets
- No independent and identically distributed data

[1] McMahan, H. Brendan, et al. „Federated learning of deep networks using model averaging." (2016).
[2] L. Huang, Y. Yin, Z. Fu, S. Zhang, H. Deng, and D. Liu. "LoAdaBoost: Loss-Based AdaBoost Federated Machine Learning on medical Data"
[3] K. Xu, H. Mi, D. Feng, H. Wang, C. Chen, Z. Zheng, and X. Lan. "Collaborative Deep Learning Across Multiple Data Centers"

Digital Health and Connected Healthcare | Prof. Dr. Bert Arnrich
Bjarne Pfitzner

Rudolf-Breitscheid-Straße 187 I D-14482 Potsdam
Tel.: +49 (0) 331 5509-1374
E-Mail: bjarne.pfitzner@hpi.de
Web: hpi.de/arnrich

Digital Health Center
by Hasso-Plattner-Institut
HPI