HeteroFL:
Computation and Communication Efficient Federated Learning for Heterogeneous Clients

Presenter: Enmao Diao, Duke University

Paper

Code

Enmao Diao[1] Jie Ding[2] Vahid Tarokh[1]
[1]Duke University [2]University of Minnesota

Overview

- Motivation
- Heterogeneous Federated Learning
  - Heterogenous models
  - Static Batch Normalization (sBN), Scaler, and Masked Cross-Entropy Loss
- Experiments
  - Interpolation experimental results
  - Combination of various computation complexity
- Conclusion

Motivation

- Emerging: a large set of heterogeneous IoT devices
- Existing federated learning: one global model architecture for all client devices

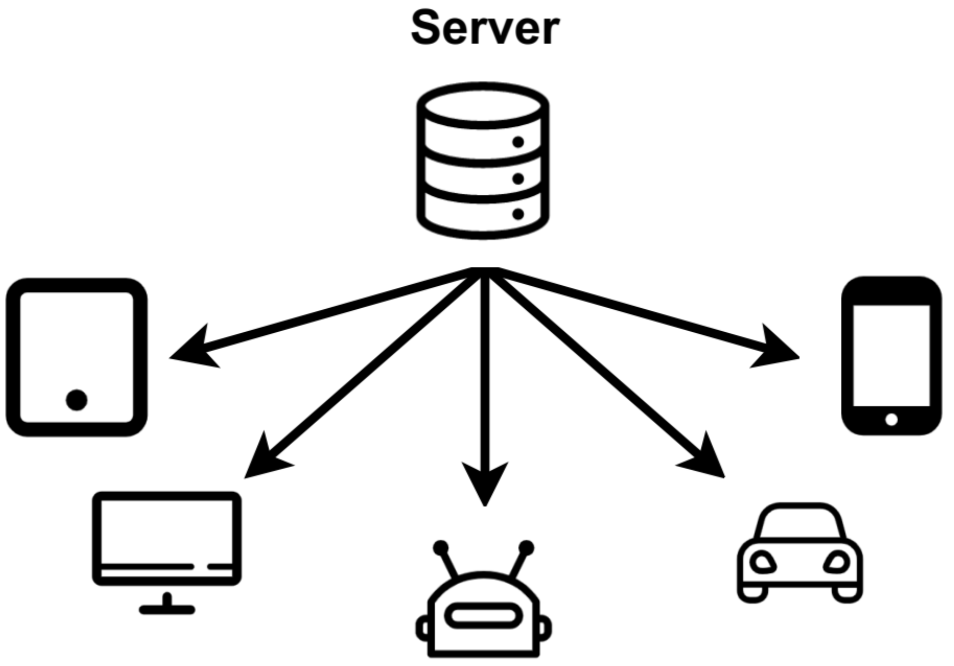

Figure 1. The computation and communication capabilities of each local client may vary significantly and even dynamically.

Motivation

Local models must share the same architecture as the global model in federated learning?

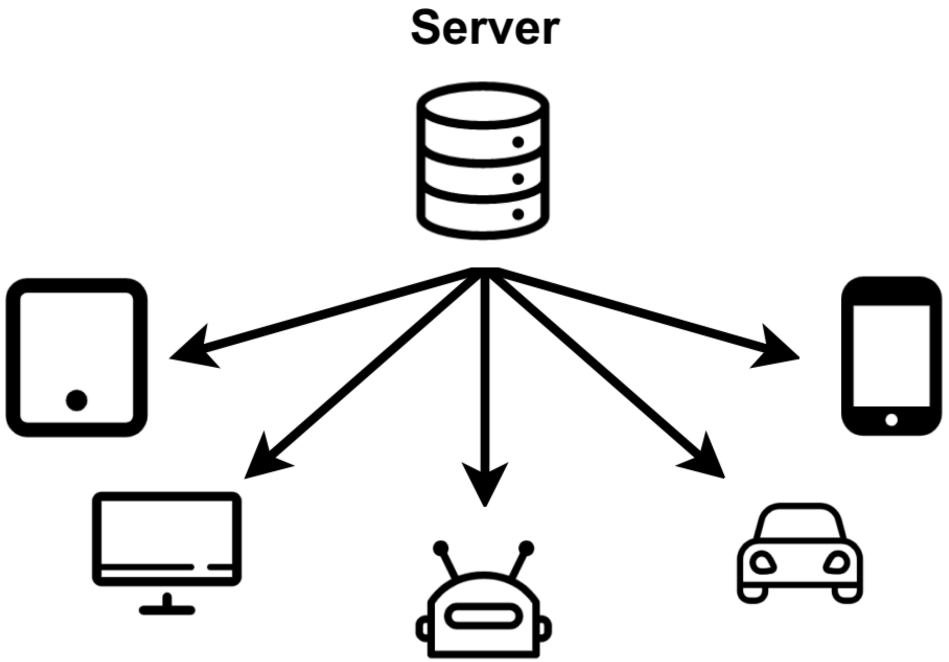

Figure 1. The computation and communication capabilities of each local client may vary significantly and even dynamically.

Motivation

HeteroFL: to enable the training of heterogeneous local models with dynamically-varying computation complexities, while still producing a single global inference model!

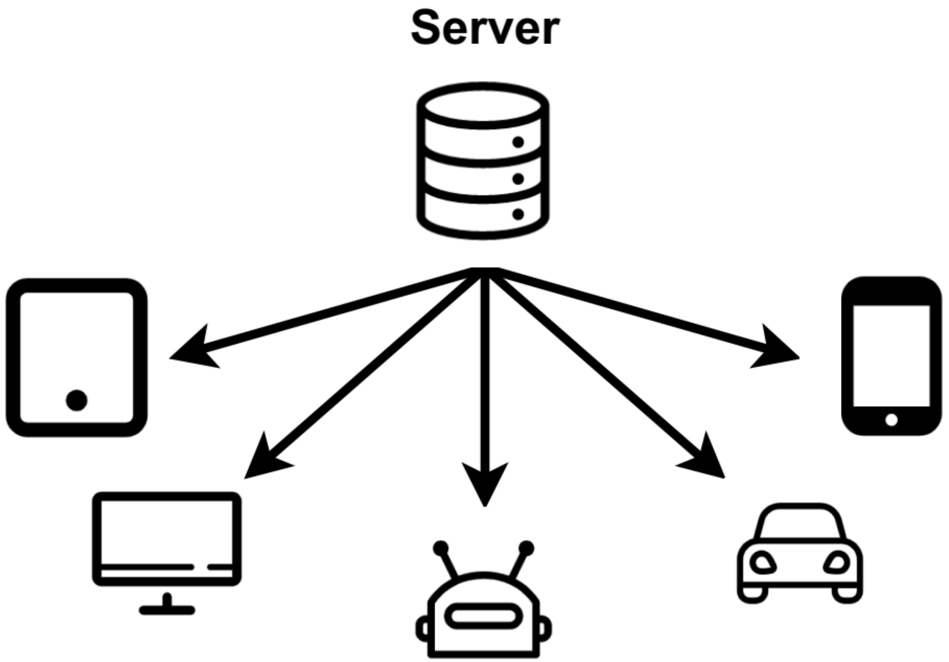

Figure 1. The computation and communication capabilities of each local client may vary significantly and even dynamically.

Heterogeneous Federated Learning (HeteroFL)

- Small local models can benefit more from FL by performing global aggregation on a part of larger local model parameters.

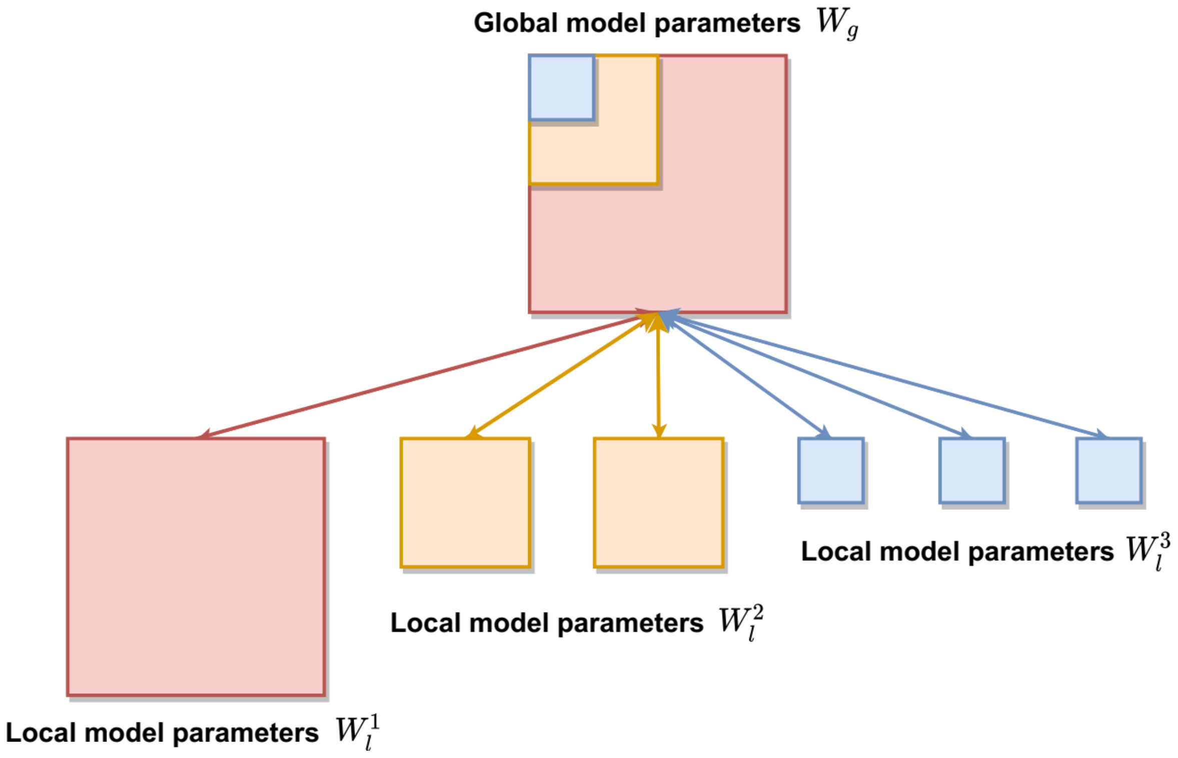

Figure 2. Global model parameters are distributed to 6 local clients with 3 computation complexity levels.

Heterogeneous Federated Learning (HeteroFL)

- **Static Batch Normalization (sBN)** does not track running estimates and cumulatively update global BN statistics after training is finished.

- **Scaler** scales representations during the training phase and the global model can be directly used for inference without scaling.

- **Masked Cross-Entropy Loss** replaces the last layer outputs that are not associated with local labels.

Experiments

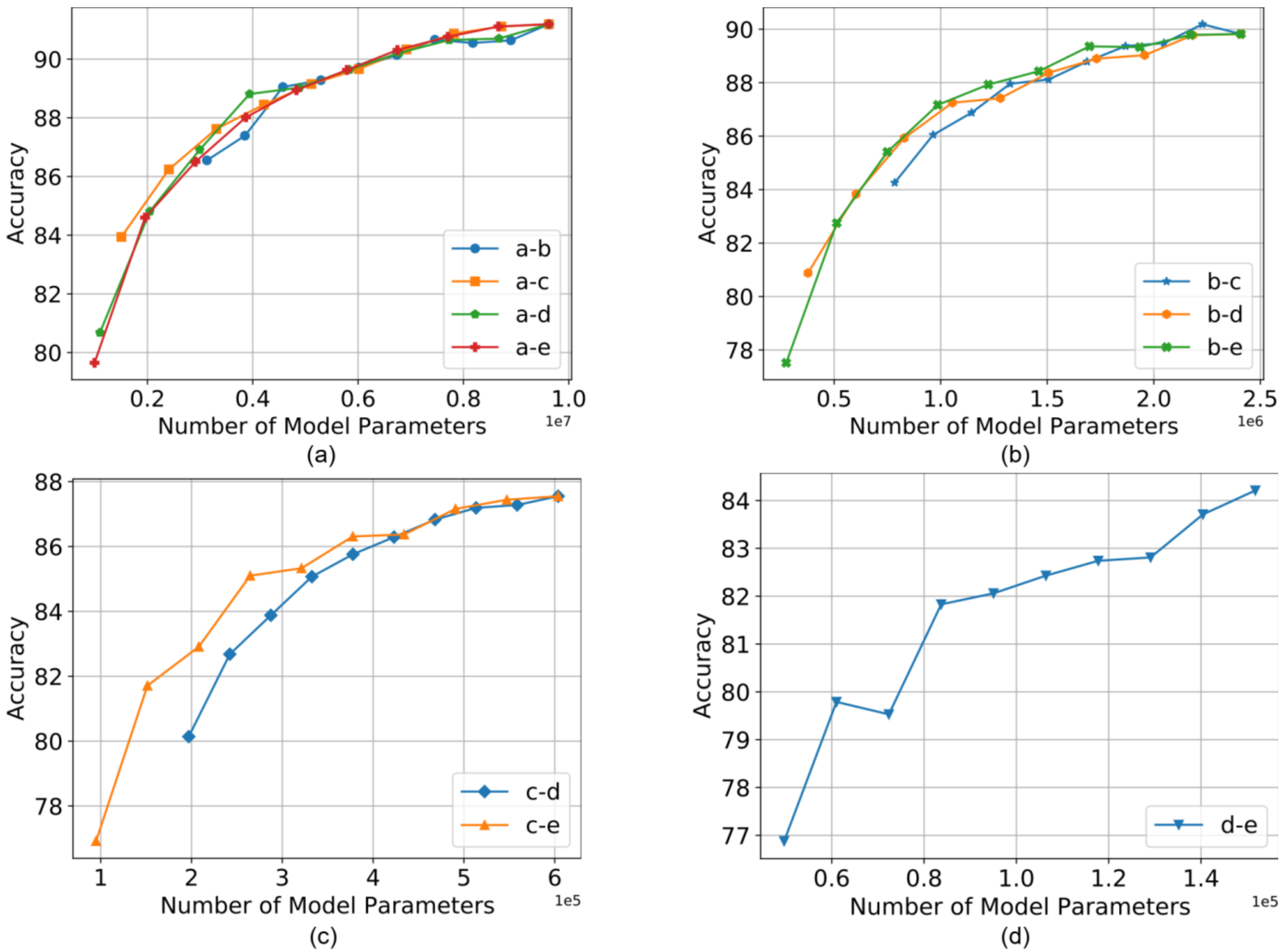

Figure 3. Interpolation experimental results for CIFAR10 (IID) dataset between global model complexity and various smaller model complexities (denoted by a, b, c, d).

Experiments

Model	Ratio	Parameters	FLOPs	Space (MB)	Accuracy		
					IID	Non-IID	
						Local	Global
a	1.00	9.6 M	330.2 M	36.71	91.19	92.38	56.88
a-e	0.50	4.8 M	165.9 M	18.43	90.29	92.10	59.11
a-b-c-d-e	0.27	2.6 M	88.4 M	9.78	88.83	92.49	61.64
b	1.00	2.4 M	83.3 M	9.19	89.82	93.83	55.45
b-e	0.51	1.2 M	42.4 M	4.67	89.10	90.68	59.81
b-c-d-e	0.33	801 K	27.9 M	3.05	87.92	91.90	59.10
c	1.00	604 K	21.2 M	2.30	87.55	91.09	55.12
c-e	0.53	321 K	11.3 M	1.22	86.88	91.83	63.47
c-d-e	0.44	265 K	9.4 M	1.01	85.79	91.49	55.42
d	1.00	152 K	5.5 M	0.58	84.21	90.77	61.13
d-e	0.63	95 K	3.5 M	0.36	82.93	90.89	56.16
e	1.00	38 K	1.5 M	0.15	77.09	89.62	54.16
Standalone (Liang et al., 2020)	1.00	1.8 M	3.6 M	6.88	16.90	87.93	10.03
FedAvg (Liang et al., 2020)	1.00	1.8 M	3.6 M	6.88	67.74	58.99	58.99
LG-FedAvg (Liang et al., 2020)	1.00	1.8 M	3.6 M	6.88	69.76	91.77	60.79

Table 1. Results of combination of various computation complexity levels for CIFAR10 dataset.

Experiments

Model	Normalization	Scaler	Masked CrossEntropy	Accuracy Non-IID			
				MNIST		CIFAR10	
				Local	Global	Local	Global
a	None	N/A	✗	97.4	97.4	42.6	42.8
	sBN			99.4	**99.4**	53.4	53.7
	None	N/A	✓	99.7	95.6	91.7	58.5
	IN			99.8	98.7	88.4	43.7
	GN			99.7	98.3	91.2	58.2
	LN			99.8	98.3	89.9	54.2
	sBN			**99.9**	98.6	**92.1**	**59.2**
e	None	N/A	✗	96.2	96.0	38.9	38.2
	sBN			90.1	90.1	40.7	40.4
	None	N/A	✓	**99.5**	**96.5**	86.6	48.9
	IN			98.5	89.8	83.7	37.0
	GN			99.2	92.2	83.5	36.7
	LN			99.3	94.0	82.6	40.0
	sBN			99.3	94.2	**90.1**	**52.9**
a-e	None	✗	✗	96.8	96.9	37.8	37.4
	sBN			99.2	99.2	41.0	41.3
	None	✗	✓	99.4	95.7	89.1	52.8
	sBN			99.8	98.0	90.7	57.7
	None	✓	✗	97.3	97.3	34.6	34.4
	sBN			99.3	**99.3**	46.0	46.7
	None	✓	✓	99.5	95.8	90.3	55.6
	IN			99.8	98.7	87.0	34.4
	GN			99.5	96.2	88.7	49.7
	LN			99.5	96.2	78.5	25.0
	sBN			**99.8**	98.2	**92.8**	**60.4**

Table 2. Ablation Study of Non-IID scenarios. Our methods significantly improve the local performance and moderately improves global performance.

Conclusion

- We identify the possibility of model heterogeneity and propose an easy-to-implement framework HeteroFL that can train heterogeneous local models.
- Our proposed solution addresses various heterogeneous settings where different proportions of clients have distinct computation and communication capabilities.
- We introduce several strategies including static Batch Normalization (sBN), scaler, and Masked Cross-Entropy Loss for significantly improving HeteroFL against the non-IID statistical heterogeneity.

Thank you!