Mitigating System and Statistical Heterogeneity in Federated Learning

UMEÅ UNIVERSITY

Sourasekhar Banerjee, Umeå University
Dept. of Computing Science
Supervisors: Monowar Bhuyan, Erik Elmroth
sourasb@cs.umu.se, monowar@cs.umu.se, elmroth@cs.umu.se

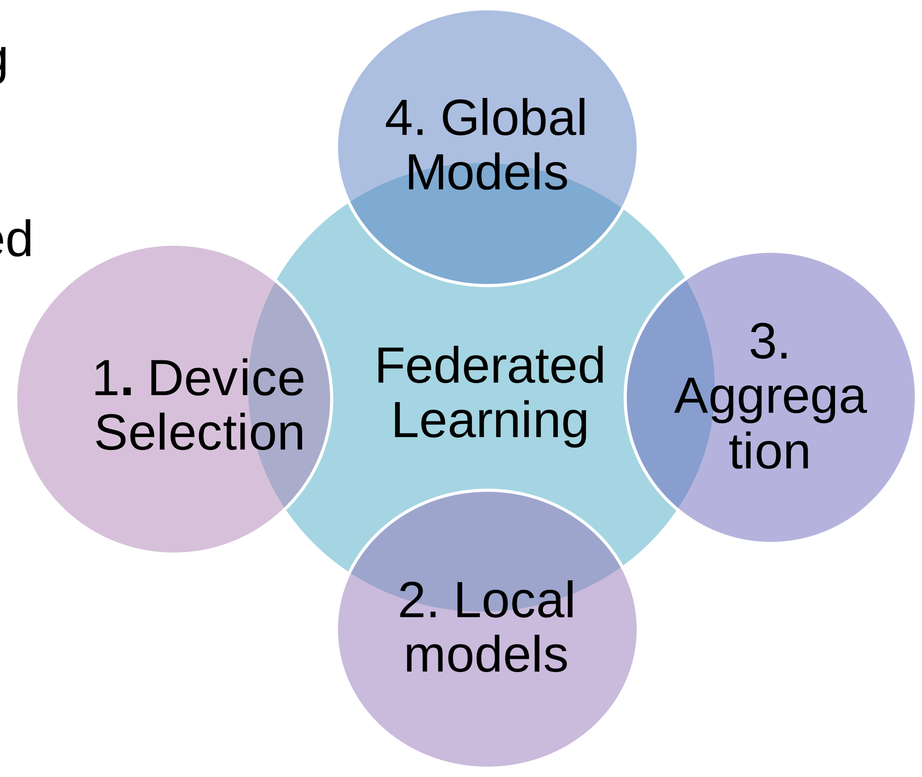
Autonomous Distributed Systems Lab

Abstract

Federated learning is a distributed learning paradigm that involves remote devices or siloed edge servers to learn a collaborative model by keeping data private. Training in a heterogeneous environment and a potentially huge network introduces unique challenges from large-scale distributed learning and optimization. In this project, we explore different characteristics and challenges (system and statistical heterogeneity) of federated learning and provide unique solutions to mitigate those challenges.

Introduction

- Federated Learning
  - Distributed
  - Collaborative
  - Privacy-preserved
- Data division
  - IID
  - Non-IID



Challenges

- **Statistical Heterogeneity**
- **System Heterogeneity (stragglers)**
- **Expensive Communications**
- Privacy

Research Questions?

- How do we select features from the dataset in a federated environment?
- **How to reduce the effect of stragglers to speed up federated learning?**
- How to mitigate the effect of statistical heterogeneity and reduce the communication cost by **personalizing** the model?
- How to learn federated learning models from small data?

Application area

- Health care
- Industrial-IoT
- Anomaly detection
- Edge cloud
- Recommender system

Current Work

Optimized and Adaptive Federated Learning for Straggler-Resilient Device Selection

Contributions

- **Fed-MOODS,** a straggler resilient **M**ulti-**O**bjective**O**ptimization based adaptive prioritized **D**evice **S**election approach.
- Fed-MOODS adaptively involve devices in Federated learning.

Objectives

- The availability of the processing capacity of each device,
- The availability of memory in devices, and
- The bandwidth capacity of the participating devices.

Fed-MOODS

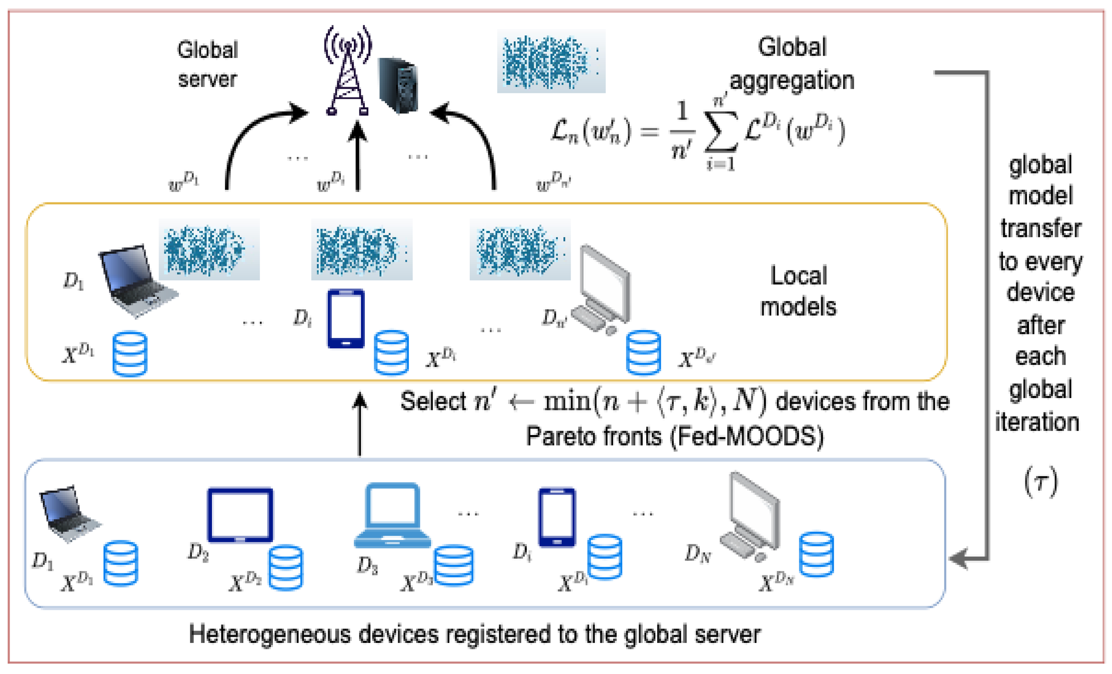
Phase 1 (Device rank)

1. The server collects meta-data and computes the objectives for each device.
2. Maximize these objective functions using Multi-objective optimization.
3. Rank each device based on the Pareto fronts.

Phase 2

1. Select n' devices from the Pareto front.
2. Learn the global model collaboratively.
3. Add another set of devices from the Pareto fronts and learn the global model.
4. Continue steps 2 and 3 until the model converges.

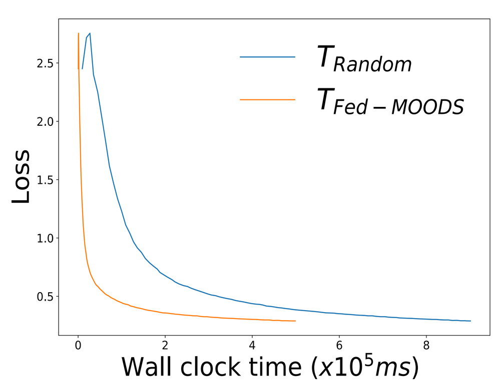
Framework

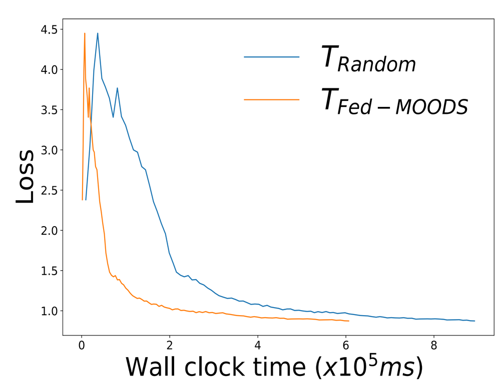


TOTAL AND AVERAGE WALL CLOCK TIME COMPARISON BETWEEN FED-MOODS AND BENCHMARK MODEL WITH RANDOM DEVICE SELECTION AT PRESENCE OF 90% STRAGGLERS ON NON-IID DATA.

Datasets	Random Device selection		Fed-MOODS	
	$T_{Random}(ms)$	$\bar{T}_{Random}(ms)$	$T_{Fed-MOODS}(ms)$	$\bar{T}_{Fed-MOODS}(ms)$
MNIST	9×10^5	9×10^3	$\mathbf{4.9 \times 10^5}$	$\mathbf{4.9 \times 10^3}$
FMNIST	8.9×10^5	8.9×10^3	$\mathbf{6 \times 10^5}$	$\mathbf{6 \times 10^3}$

Wall-clock time comparison





Results

Test accuracy

Dataset		SF %	Fed-MOODS + FedAvg	Fed-MOODS + FedProx	Random + FedAvg	Random + FedProx
MNIST	IID	90	**97.2**	96.31	97.2	96.89
		70	97.54	97.49	97.61	97.5
		50	**97.94**	97.76	97.74	97.61
		10	98.11	**98.39**	98.05	98.11
	Non-IID	90	92.31	91.93	92.04	. 91.47
		70	**93.91**	92.79	89.18	93.43
		50	**94.69**	93.47	93.05	89.61
		10	**95.74**	93.59	93.17	93.86
CIFAR-10	IID	90	**53.43**	50.20	49.15	48.86
		70	46.3	47.15	48.62	47.17
		50	43.59	**49.42**	46.25	48.9
		10	46.71	**47.33**	45.48	44.72
	Non-IID	90	49.23	**49.55**	15.84	10
		70	**48.75**	47.68	33.99	29.75
		50	**46.56**	45.93	24.98	38.44
		10	45.86	**47.81**	33.75	34.0
FMNIST	IID	90	78.66	78.48	80.48	79.44
		70	82.63	82.81	83.04	77.63
		50	83.32	83.89	85.17	82.59
		10	**85.39**	85.22	84.44	84.68
	Non-IID	90	63.22	**65.33**	50.26	58.18
		70	**67.16**	65.54	56.92	64.07
		50	70.0	**70.97**	55.56	61.81
		10	**71.76**	67.58	58.18	59.26

Analysis

1. Fed-MOODS produces better performance random.
2. Convergence of Fed-MOODS is quicker than random device selection.
3. Fed-MOODS is 1.8× and 1.48× faster than the benchmark model (FedAvg) with random device participation on the MNIST and FMNIST non-IID datasets, respectively.

Conclusion

- In this project, we are dealing with the challenges of the system and statistical heterogeneity in federated learning.
- We raised 4 research questions that need to be solved to mitigate these challenges.
- Though we are proposing algorithms, we have identified five application areas where the proposed solutions can be applied and assessed.
- In future, we are focusing on personalization to mitigate the effect of heterogeneity and reduce the communication rounds.

References

1. Banerjee, S.; Vu, X.-S.; and Bhuyan, M. 2022. Optimized and adaptive federated learning for straggler-resilient device selection. accepted in International Joint Conference in Neural Networks (IJCNN).
2. Reisizadeh, A.; Tziotis, I.; Hassani, H.; Mokhtari, A.; and Pedarsani, R. 2020. Straggler-resilient federated learning: Leveraging the interplay between statistical accuracy and system heterogeneity. arXiv preprint arXiv:2012.14453.
3. T Dinh, C.; Tran, N.; and Nguyen, J. 2020. Personalized federated learning with moreau envelopes. Advances in Neural Information Processing Systems 33:21394–21405
4. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017a. Communication-efficient learning of deep networks from decentralized data. Artificial Intelligence and Statistics 1273–1282
5. Fallah, A.; Mokhtari, A.; and Ozdaglar, A. 2020. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. Advances in Neural Information Processing Systems 33:3557–3568.

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM