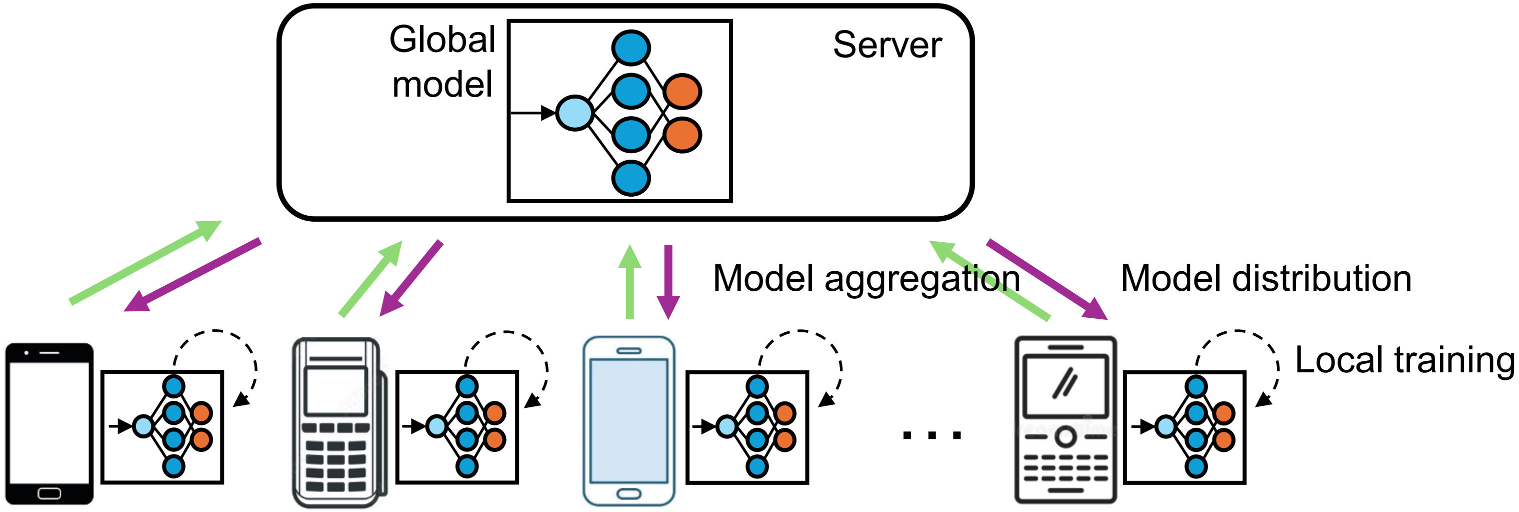
Layer-wise Update Aggregation with Recycling for Communication-Efficient Federated Learning

Jisoo Kim[1], Sungmin Kang[2], Sunwoo Lee[1*]

1. Inha University, Incheon, Republic of Korea
2. University of Southern California, Los Angeles, CA, USA

Problem Statement

- **Expensive model aggregation cost in Federated Learning**
  - The Larger the model the more expensive the communication cost.



Related Work

▪ Existing Communication-efficient Methods for FL

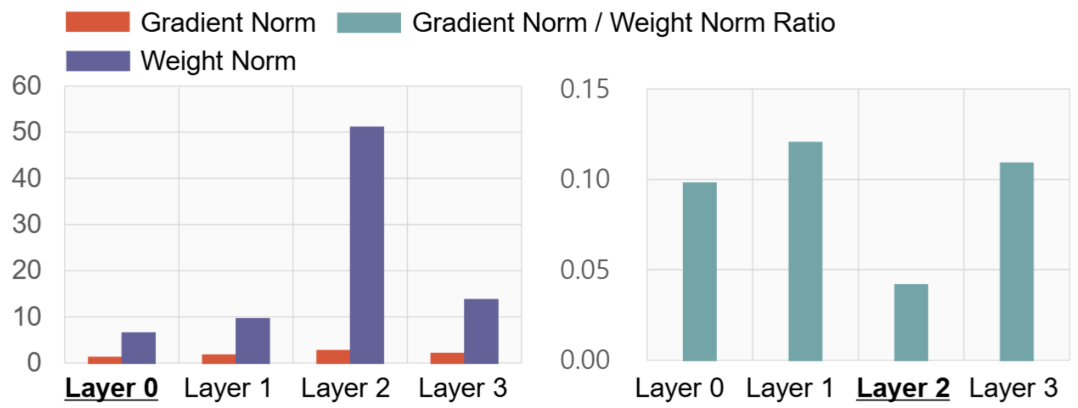
- Quantization (Reisizadeh et al., AISTATS 2020, FedPAQ)
- Sparsification (Jiang et al., TNNLS 2022, PruneFL.)
- Binarization (Li et al., ICML 2024, FedBAT.)
- Model Reparameterization (Nam et al., ICLR 2022, FedPara.)

➡ Commonly drop a part of model parameters or updates.

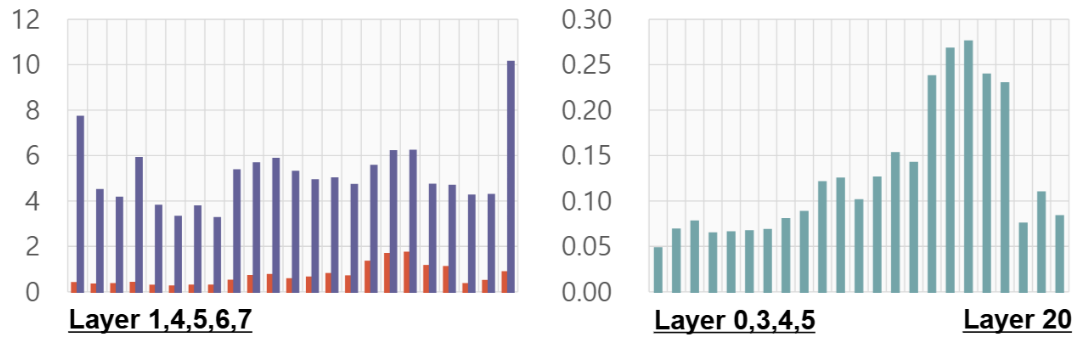
Is update dropping the best way to reduce the communication cost?

Motivation

- **Limitations of gradient norm as a parameter importance metric**



(a) FEMNIST (CNN)

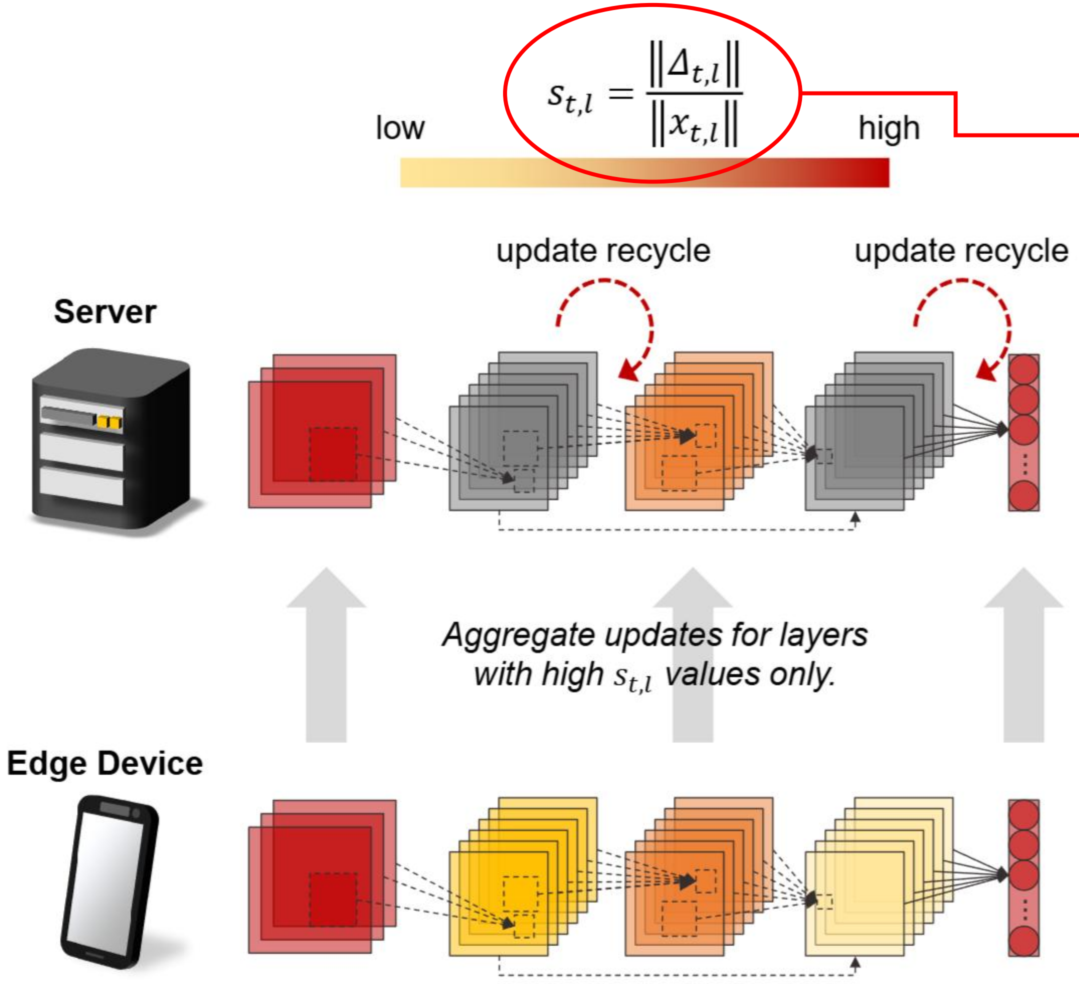


(b) CIFAR-10 (ResNet20)

The gradient norm does not account for parameter magnitude, so it does not effectively represent importance in terms of parameter changes.

Proposed Method

- **FedLUAR: Federated Layer-wise Update Aggregation with Recycling**



Gradient-to-weight ratio
(Parameter importance metric)

Empirical Results

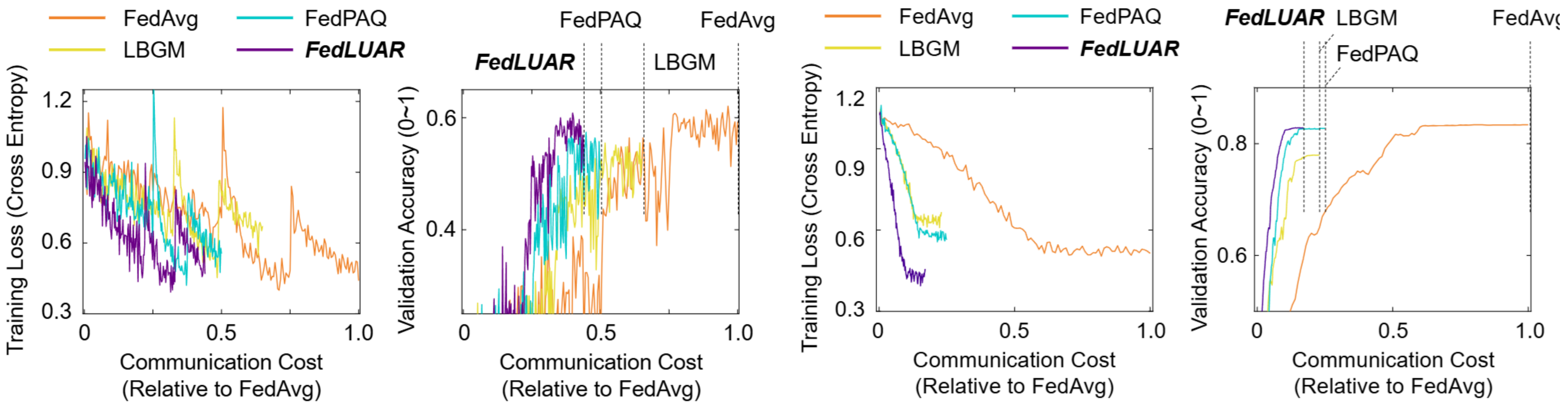
- **Classification Performance Comparison**

Method	CIFAR-10 (ResNet20) Accuracy	Comm	CIFAR-100 (WRN-28) Accuracy	Comm	FEMNIST (CNN) Accuracy	Comm	AG News (DistillBERT) Accuracy	Comm
FedAvg	61.27 ± 0.7%	1.00	59.88 ± 0.8%	1.00	71.01 ± 0.4%	1.00	82.66 ± 0.2%	1.00
LBGM	54.87 ± 0.5%	0.65	57.13 ± 0.2%	0.87	69.83 ± 1.0%	0.71	77.96 ± 0.1%	0.23
FedPAQ	57.42 ± 0.2%	0.50	36.15 ± 0.1%	0.50	71.54 ± 0.1%	0.25	82.72 ± 0.1%	0.25
FedPara	55.16 ± 0.1%	0.51	46.14 ± 0.1%	0.61	67.69 ± 0.1%	0.22	75.22 ± 0.1%	0.69
PruneFL	56.76 ± 0.1%	0.51	59.40 ± 0.1%	0.69	69.42 ± 0.4%	0.19	77.25 ± 0.1%	0.22
FDA	56.54 ± 0.3%	0.50	51.25 ± 0.1%	0.60	70.61 ± 0.1%	0.25	64.94 ± 0.1%	0.50
FedBAT	39.56 ± 0.1%	0.03	47.24 ± 0.1%	0.03	68.27 ± 0.1%	0.03	76.38 ± 0.1%	0.57
FedLUAR	**60.15 ± 0.7%**	**0.47**	**59.73 ± 0.6%**	**0.61**	**73.17 ± 0.1%**	**0.18**	**82.80 ± 0.1%**	**0.17**

- **Communication Cost Analysis**

Empirical Results

- **Learning curves w.r.t. communication cost**



(a) CIFAR-10 (b) AG News

Conclusion

1. Based on the proposed parameter importance metric (**gradient-to-weight ratio**), we demonstrate that less critical layers' updates can be safely recycled.

2. We theoretically proved that **update recycling still guarantees a convergence** for non-convex and smooth optimization problems.

3. Through various benchmarks, we empirically proved that the update recycling scheme **effectively reduces the model aggregation cost** while maintaining the model accuracy in realistic FL environments.

LMLS Lab

Large-scale Machine Learning Systems Laboratory

NEURAL INFORMATION PROCESSING SYSTEMS

Thank you for your attention!

INHA UNIVERSITY

USC University of Southern California