Are All Users Treated Fairly in Federated Learning Systems?

Umberto Michieli[1,2*] and Mete Ozay[1]

[1] Samsung Research UK   [2] University of Padova

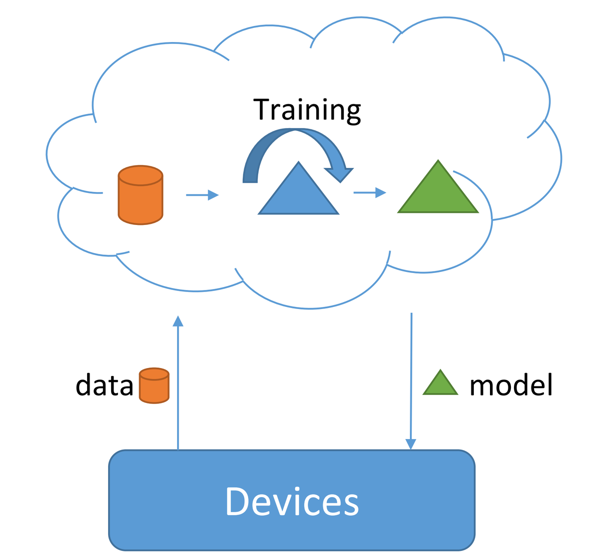
Abstract

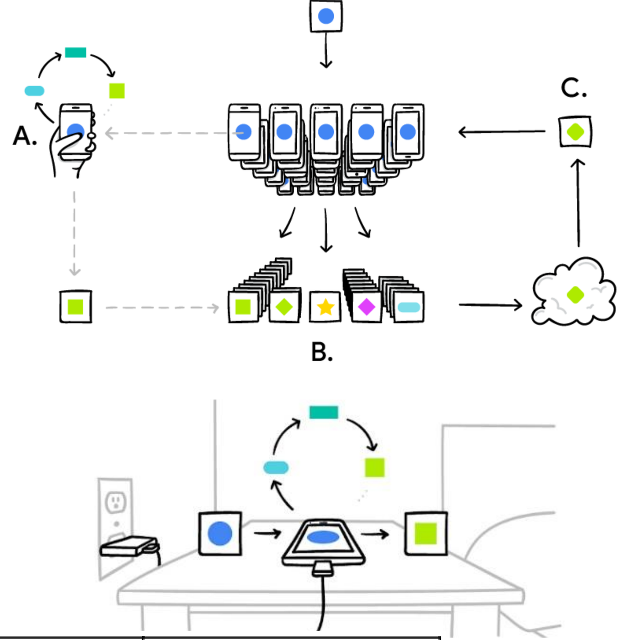
Federated Learning (FL) systems target distributed model training on decentralized and private local training data belonging to users. Current methods aggregate models with importance proportional to the frequency of local samples. However, this leads to **unfair** aggregation with respect to users. Indeed, users with few local samples are considered less during aggregation and do not offer a real contribution to optimization of the models. In the real-world, statistical heterogeneity (e.g., highly imbalanced and non-i.i.d. data) is diffused and seriously harms model training.

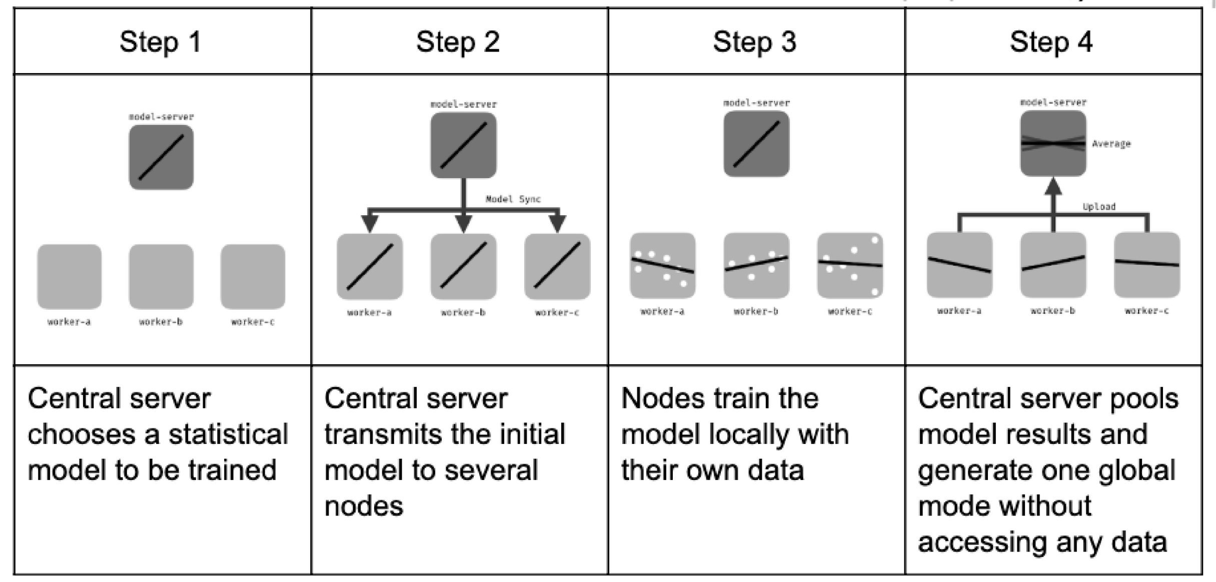
We empirically analyze the relationship between fairness of aggregation of user models, accuracy of aggregated models and convergence rate of FL methods. We compare the standard FedAvg against a fair (uniform) scheme, i.e., **FairAvg** on benchmark datasets. Experimentally, we show that fair aggregation can be beneficial for accuracy and convergence rate, whilst reducing at the same time fluctuations of accuracy of the aggregate model when clients observe non-i.i.d. data.

Federated Learning (FL)

FL: Distributed Machine Learning on Heterogeneous Data







Our Approach (FairAvg)

In general, federated aggregation can be expressed as:

$$\underbrace{\boxed{w_{t+1}}}_{\text{Central model parameters}} \leftarrow \sum_{k=1}^{\overset{\text{\# clients}}{\boxed{K}}} \underbrace{\boxed{w_t^k}}_{\text{Local model parameters}} \underbrace{a_k^t}_{\text{Attention vector}}$$

FairAvg: $a_k^t = 1/K$ (constant)

→ To show effect of data imbalance and non-i.i.d.-ness across clients

Federated Attention a_k^t

Table 1. Statistics of the employed datasets (left) and hyper-parameters (right).

Dataset	# Classes	Clients	Samples	Samples/Client Mean	Samples/Client Std.	Model	Distribution	Central. Acc. (%)	Start lr	Solver	F	Rounds	Batch size
Synthetic	10	30	9,600	320.0	1051.6	2 dense layers	Power-law	78.5	0.01	SGD	20	200	10
MNIST	10	1,000	61,676	61.7	164.7	2-layer CNN	Power-law	99.0	0.01	SGD	20	200	10
FEMNIST	10	200	16,421	82.1	143.0	2-layer CNN	Power-law	99.0	0.001	SGD	20	400	10
Sent140	2	772	40,783	53	32	Stacked-LSTM	Power-law	72.3	0.3	SGD	20	800	10
Shakespeare	80	143	517,106	3,616	6,808	Stacked-LSTM	Power-law	49.9	0.8	SGD	20	40	10

Samples are distributed to clients according to power-law distribution

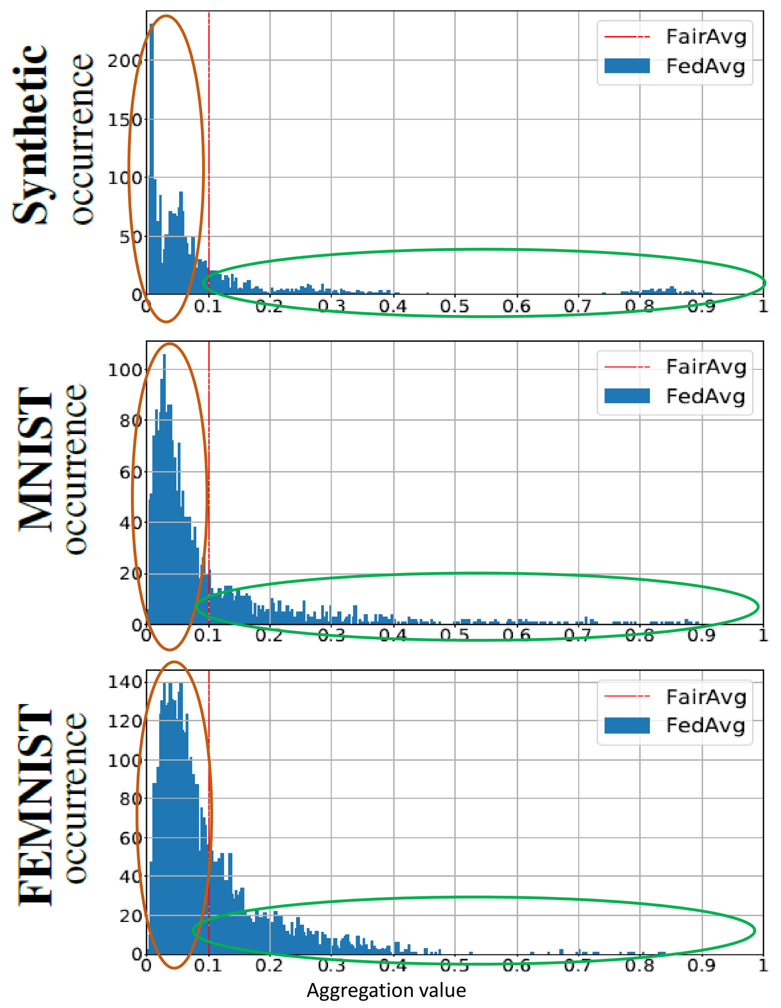
Note1: we subsample K=10 clients

Note2: $\boldsymbol{a}$ computed by FedAvg is $\boldsymbol{a}_k \propto n_k$

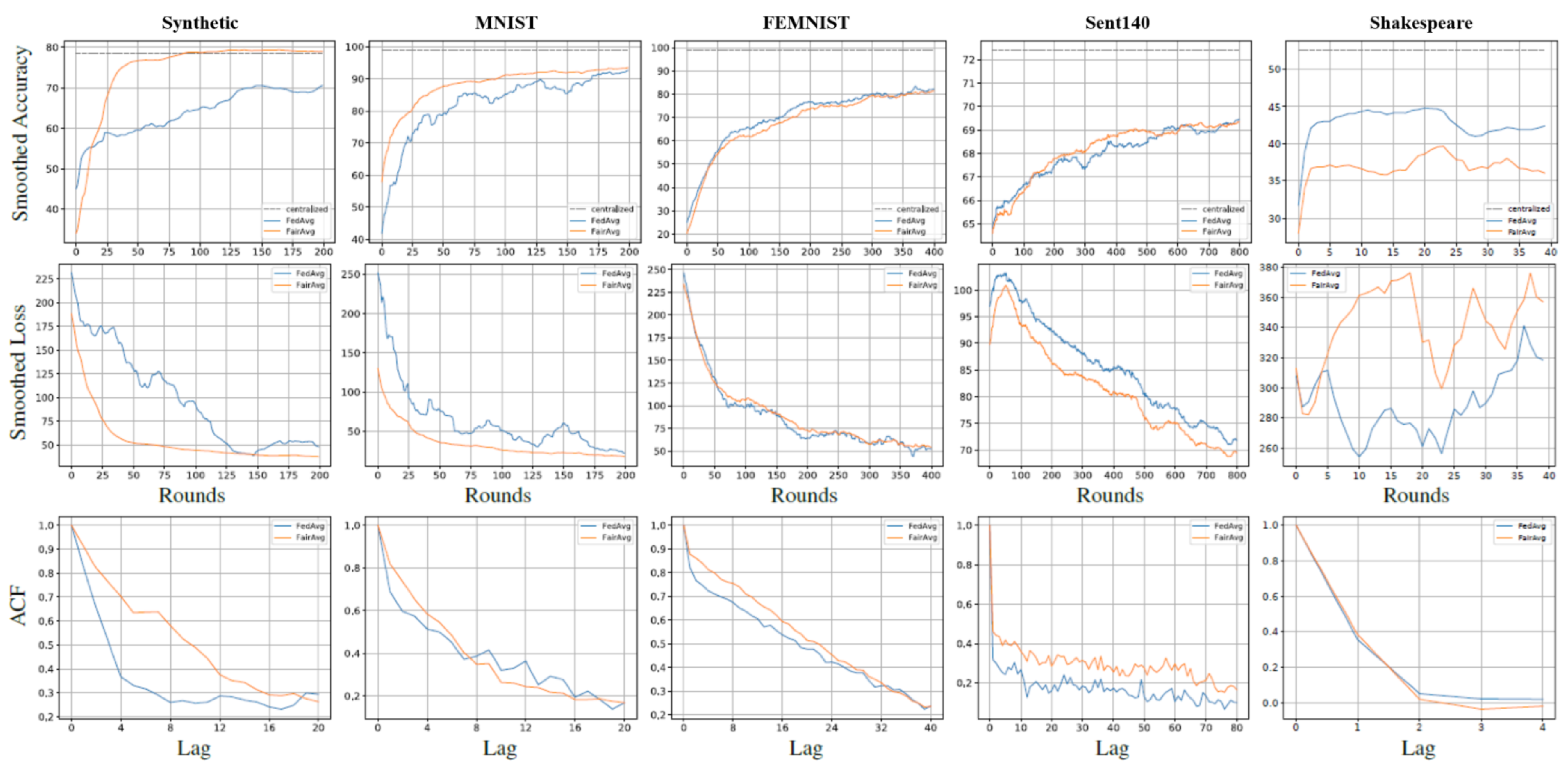
❑ **Many clients** *contribute little*

❑ **Few clients** *dominate the scene*

→If data is highly non-i.i.d. this represents a problem for convergence



Results



ACF: AutoCorrelation Function

→ FairAvg improves accuracy wrt FedAvg (when data are distributed non-i.i.d.)

→ FairAvg reduces fluctuation towards convergence values wrt FedAvg (when data are distributed non-i.i.d.)

Conclusion

✓ We explore the relationship between fairness of aggregation schemes, accuracy of aggregated models and convergence rate of federated optimization methods on non-i.i.d. data.

✓ FairAvg is beneficial compared to FedAvg for both final accuracy and convergence rate, whilst reducing at the same time fluctuations of accuracy of the aggregate model.

✓ We believe that FL models could employ federated aggregation values centered around the value employed by FairAvg for uniform treatment of user contributions.

* Researched during internship at Samsung Research UK