EuroMLSys
Paper ID: 10

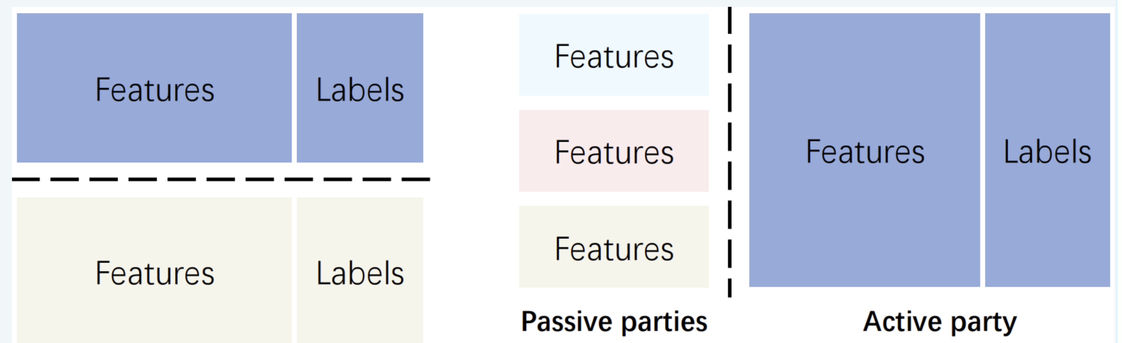
Gradient-less Federated Gradient Boosting Trees with Learnable Learning Rates

Chenyang Ma, Xinchi Qiu, Daniel J. Beutel, Nicholas D. Lane
Computer Science and Technology, University of Cambridge, United Kingdom

Problem

- Horizontal federated XGBoost relies on the sharing of gradients because finding the optimal split condition of a single tree depends on the order of the data samples.

- The sharing of gradients causes:
1) Per-node level communication frequency
2) Serious privacy concerns



(a) Horizontal Setting (b) Vertical Setting

Contributions

- We propose a novel privacy-preserving framework, **FedXGBllr**, a **fed**erated **XGB**oost with **l**earnable **l**earning **r**ates in the horizontal setting which do not rely on the sharing of gradients and hessians.

- Our framework disentangles the per-node level communication frequency when training a federated XGBoost.

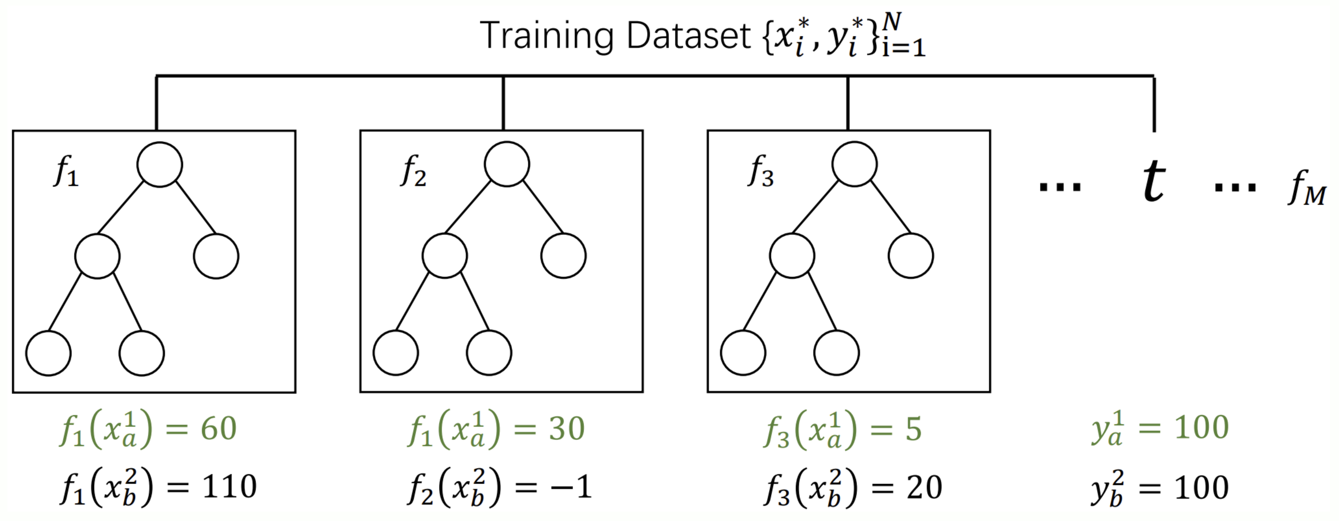
- The total communication overhead of our framework is independent of the dataset size and is significantly lower (by factors ranging from 25x to 700x) than previous methods.

- We show that FedXGBllr is interpretable with carefully framed reasoning and analysis.

Proposed Approach

Intuition 1: A fixed learning rate is too weak

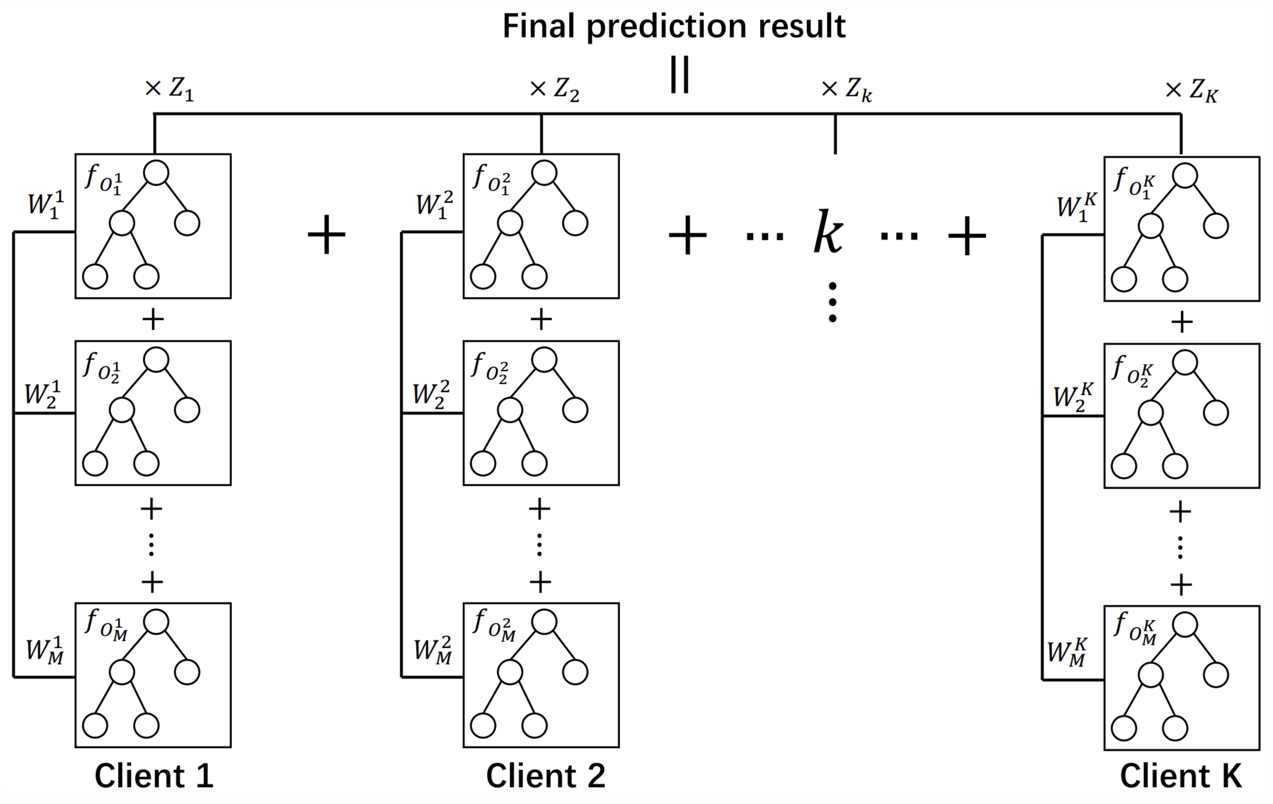
- Local client's dataset may be heterogenous
- Each tree makes different "amount of mistakes"



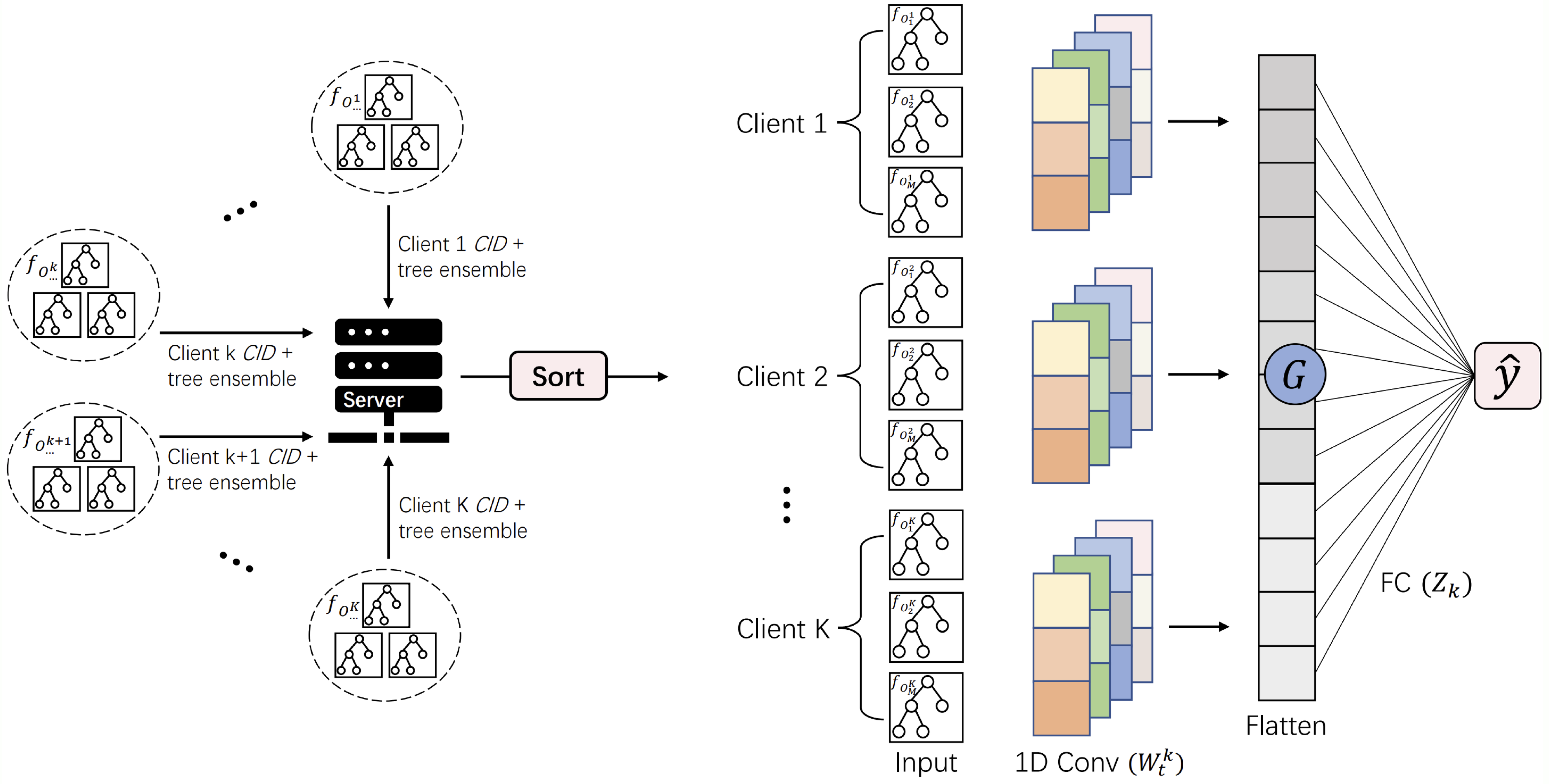
Proposed Approach cont.

Intuition 2: Moving towards the global optima

- Applying a weighted sum on the diverse prediction results given by all XGBoost tree ensembles can lead to a more accurate final prediction value



Pipeline



(a) Tree Ensembles Aggregation

(b) One-layer 1D CNN (trained by FedAvg)

Experiments

Table 1: Summary of datasets

Dataset	Task Type	Data No.	Dimension	Size
a9a	classification	32,561	123	16MB
cod-rna	classification	59,535	8	2.1MB
ijcnn1	classification	49,990	22	4.4MB
real-sim	classification	72,309	20,958	6.1GB
HIGGS	classification	1,000,000	28	112MB
SUSY	classification	1,000,000	18	72MB
abalone	regression	4,177	8	253KB
cpusmall	regression	8,192	12	684KB
space_ga	regression	3,167	6	553KB
YearPredictionMSD	regression	515,345	90	615MB

Table 2: Quantitative results of FedXGBllr compared to SimFL and centralized baseline - Accuracy ↑ (for the first six classification datasets), MSE ↓ (for the last four regression datasets).

Dataset	FedXGBllr			SimFL [16]	Centralized Baseline
	2 clients	5 clients	10 clients	2 clients	
a9a	85.1	85.1	84.7	84.9	84.9
cod-rna	97.0	96.5	95.8	94.0	93.9
ijcnn1	96.3	96.0	95.3	96.4	96.3
real-sim	93.4	93.8	92.7	92.9	93.5
HIGGS	71.5	70.9	70.3	70.7	70.7
SUSY	82.5	81.7	81.2	80.4	80.0
abalone	3.6	4.4	4.9	-	1.3
cpusmall	8.0	8.5	9.5	-	6.7
space_ga	0.024	0.033	0.034	-	0.024
YearPredictionMSD	80.3	82.7	91.6	-	80.5

Ablation Studies

Table 3: Comparison of communication overhead (MB).

Dataset	FedXGBllr	SimFL [16]
a9a	6.0	150.4 (25×)
cod-rna	6.0	249.3 (42×)
ijcnn1	6.0	218.4 (36×)
real-sim	6.0	323.1 (54×)
HIGGS	6.0	4216 (703×)
SUSY	6.0	4136 (689×)

Table 4: Model interpretability. $k = kernel\_size$, $s = stride$, $n = client\_tree\_num = 500/client\_num$.

Dataset	1-layer 1D CNN (k = s = n)	1-layer 1D CNN (k = 3, s = 1)	2-layer FCNN
a9a	85.1	83.9	82.8
cod-rna	96.5	96.3	94.7
ijcnn1	96.0	95.1	92.2
real-sim	93.8	93.2	91.6
HIGGS	70.9	70.5	67.9
SUSY	81.7	81.3	77.5
abalone	4.4	4.4	5.8
cpusmall	8.5	9.2	12.6
space_ga	0.033	0.034	0.044
YearPredictionMSD	82.7	87.5	117.7