FedSampling: A Better Sampling Strategy for Federated Learning

Tao Qi[1], Fangzhao Wu[2], Lingjuan Lyu[3], Yongfeng Huang[1,4,5], Xing Xie[2]

[1]Department of Electronic Engineering & BNRist, Tsinghua University, Beijing 100084, China

[2]Microsoft Research Asia, Beijing 100080, China

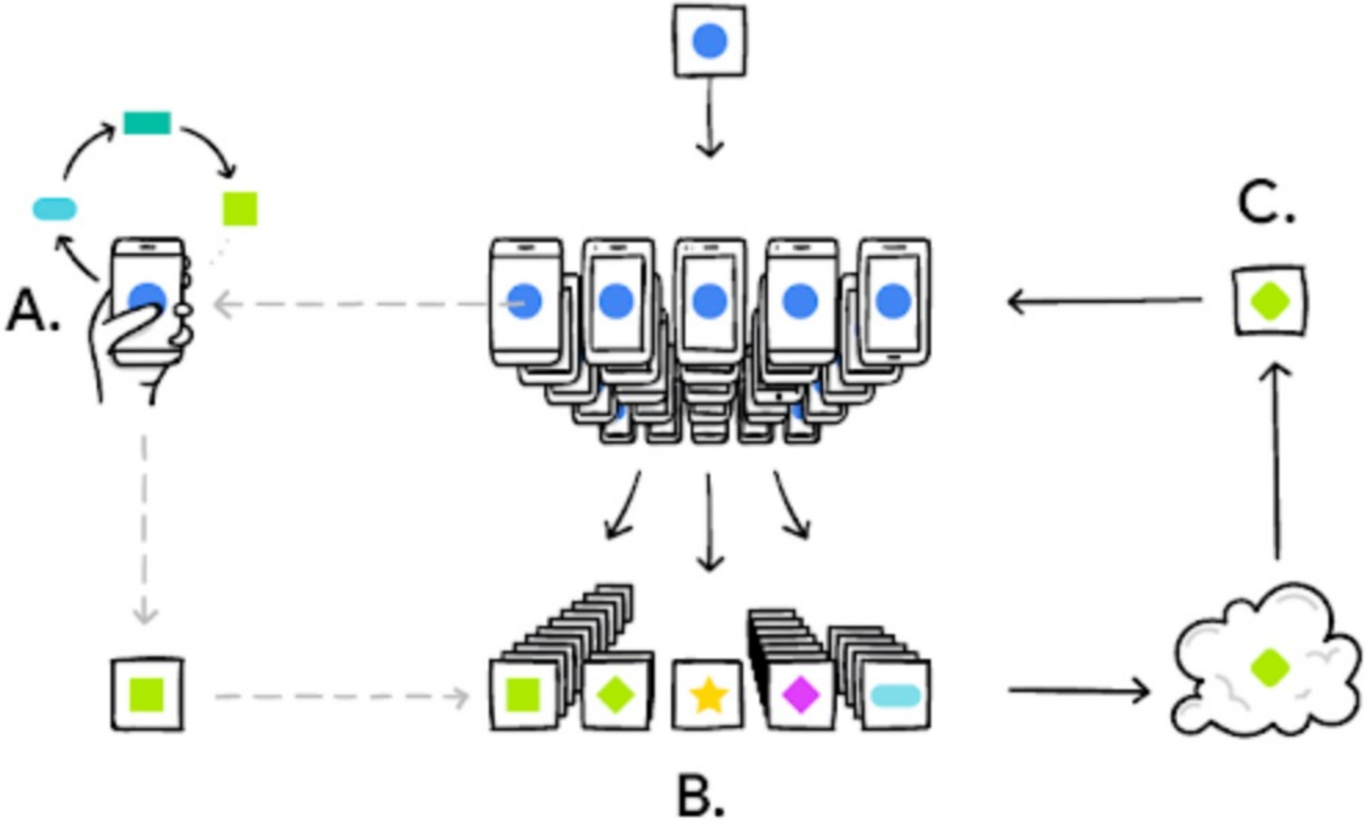
[3]Sony AI, 1-7-1 Konan Minato-ku Tokyo 108-0075, Japan

[4]Zhongguancun Laboratory, Beijing 100094, China

[5]Institute for Precision Medicine of Tsinghua University, Beijing 102218, China

Federated Learning

- A promising privacy-preserving machine learning framework
  - Collaborative model learning with decentralized data



Client Sampling in Federated Learning

- Client sampling is a key step for existing federated learning methods

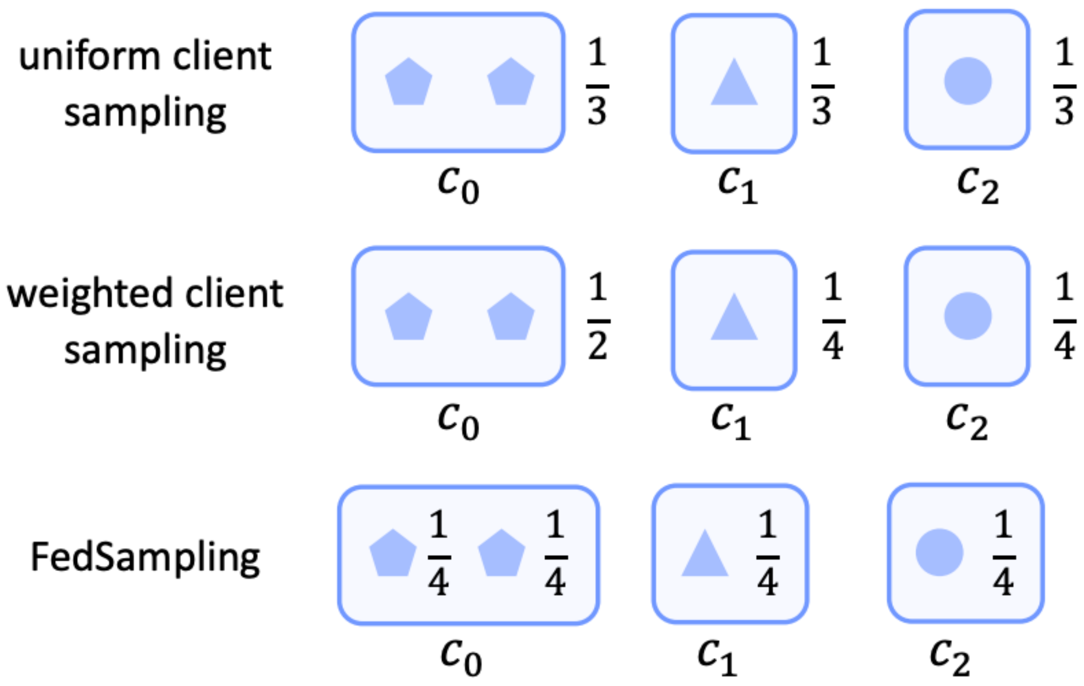
- Uniform client sampling:
  - sampling weight: $p_i = \frac{1}{M}$
  - aggregation weight: $w_i = \frac{n_i}{\sum_j^M n_j}$

- Weighted client sampling:
  - sampling weight: $p_i = \frac{n_i}{\sum_j^M n_j}$
  - aggregation weight: $w_i = \frac{1}{M}$

Algorithm 1 FedOpt

```
Input: x_0, ClientOpt, ServerOpt
for t = 0, ..., T-1 do
    Sample a subset S of clients
    x^t_{i,0} = x_t
    for each client i ∈ S in parallel do
        for k = 0, ..., K-1 do
            Compute an unbiased estimate g^t_{i,k} of ∇F_i(x^t_{i,k})
            x^t_{i,k+1} = ClientOpt(x^t_{i,k}, g^t_{i,k}, η_l, t)
        Δ^t_i = x^t_{i,K} - x_t
    Δ_t = (1/|S|) Σ_{i∈S} Δ^t_i
    x_{t+1} = ServerOpt(x_t, -Δ_t, η, t)
```

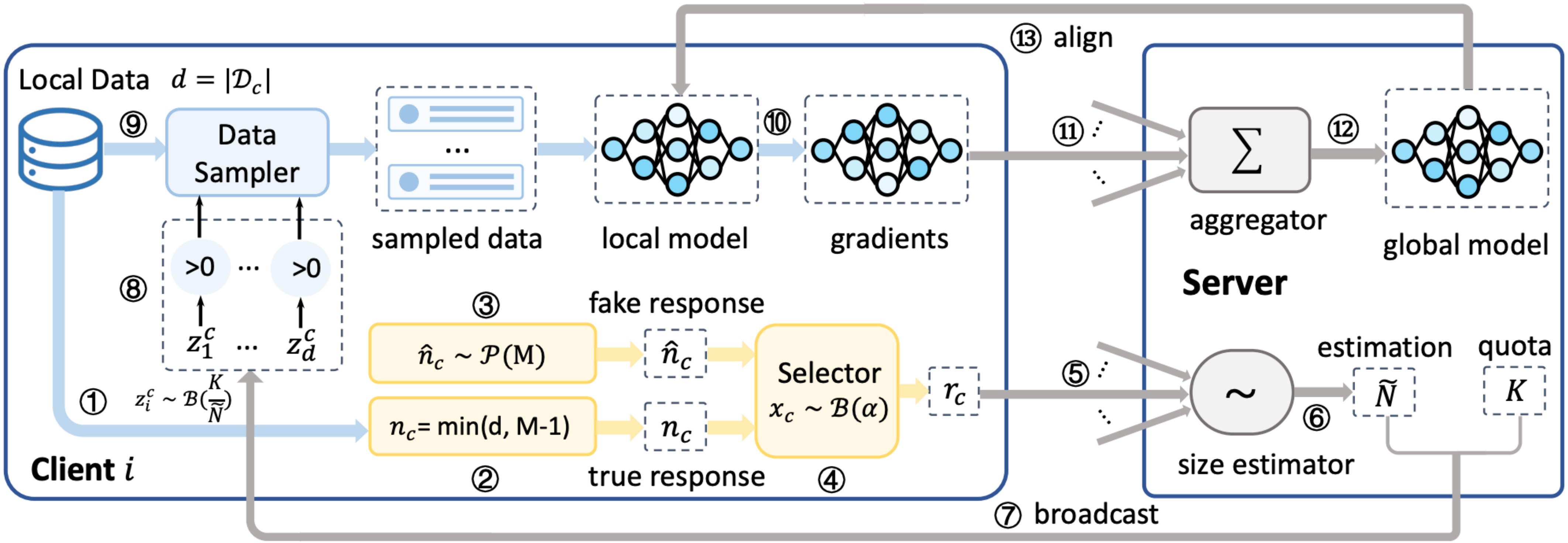
Challenges of Existing Client Sampling Methods



- Challenge:
  - Difficult to uniformly exploit decentralized samples
  - Tracking local sample sizes may also arouse privacy concerns

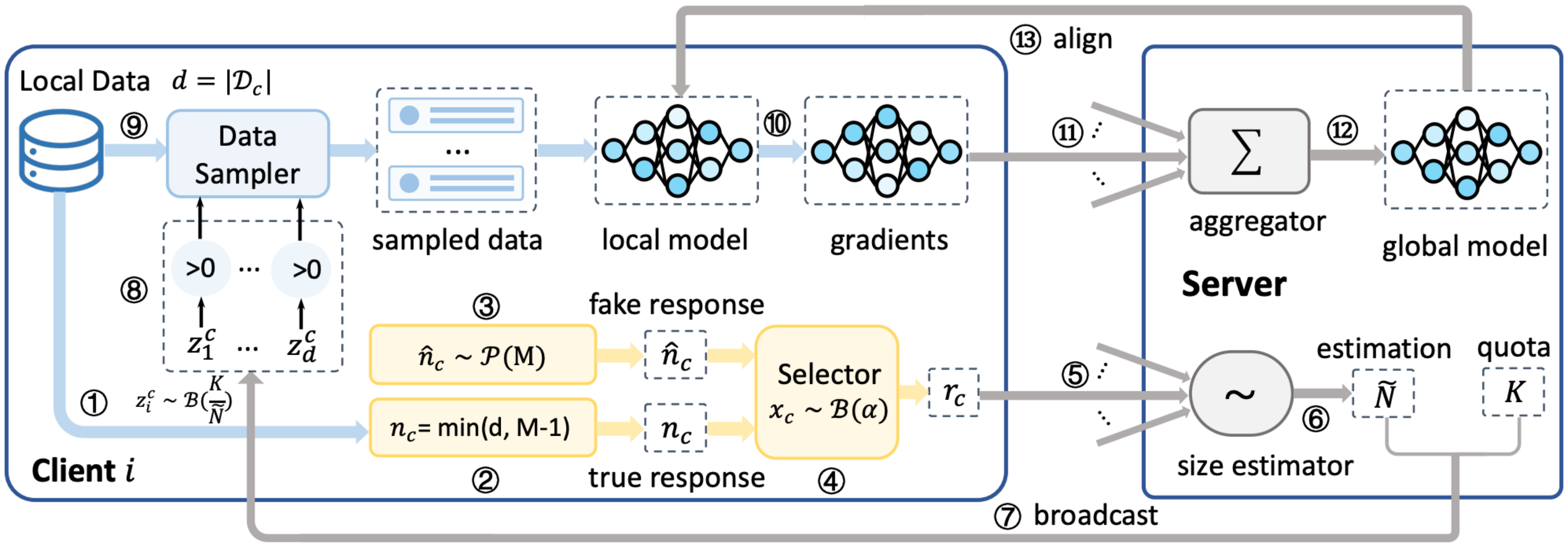
FedSampling: Uniform Data Sampling

- Independent and identical data sampling: $z_i^c \sim \mathcal{B}\left(\frac{K}{\tilde{N}}\right)$
  - K is the size of samples needed for training, $\tilde{N}$ is estimated total sample size



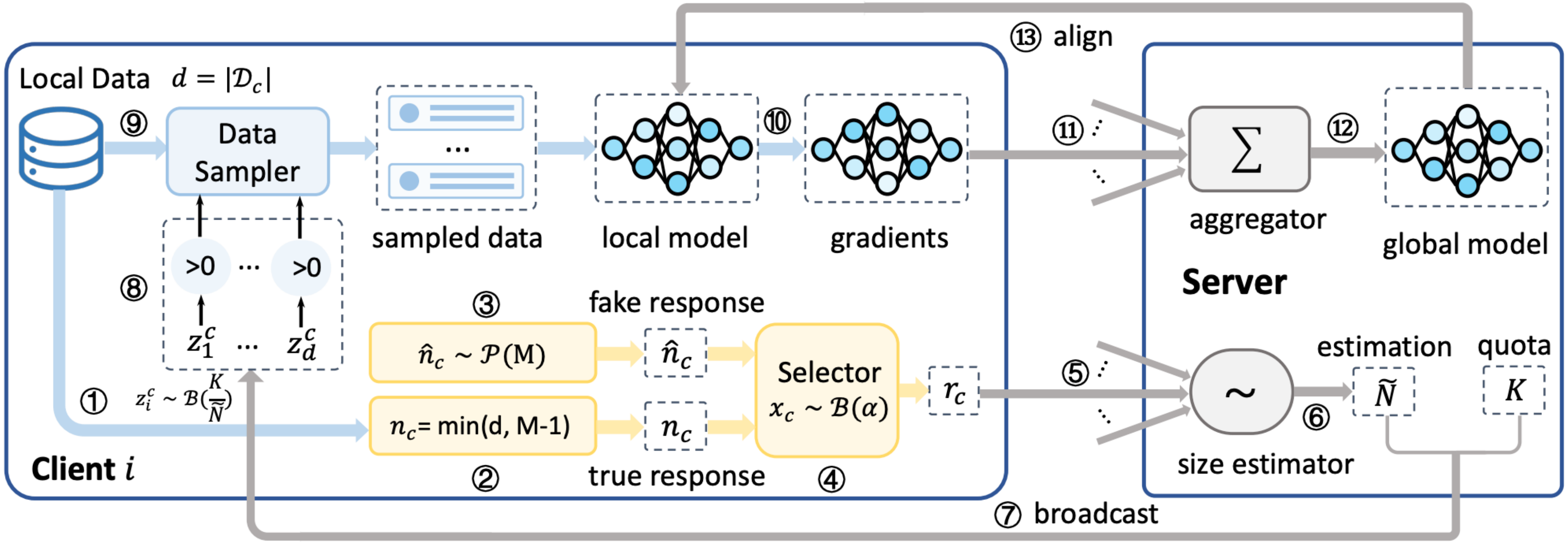
FedSampling: Privacy-Preserving Ratio Estimation

- Naive solution: Bypass the challenge by sampling data via a fixed ratio r
  - Cause privacy leakage or lead to an biased model update
- Differentially private local response: $r_c = x_c n_c + (1 - x_c)\hat{n}_c$
  - $n_c = \min(|\mathfrak{D}_c|, M), \quad x_c \sim \mathcal{B}(\alpha), \quad \hat{n}_c \sim \mathcal{P}(M-1)$
- Unbiased estimation: $\tilde{N} = \left(\sum_{c \in C} r_c - \frac{(1-\alpha)M|C|}{2}\right)/\alpha$



FedSampling: Workflow

- The workflow of FedSampling is mainly different from mainstream FL methods in data sampling



FedSampling: Discussions on Utility and Privacy

- Lemma 1: Let $p(x)$ and $\hat{p}(x)$ denote the probability of a sample x that can participate in a training step in the centralized learning and FedSampling. The MSE between $p(\cdot)$ and $\hat{p}(\cdot)$ asymptotically converges to 0
  - $\lim_{|C|\to\infty} \mathbb{E}[(p(x) - \hat{p}(x))^2] < \lim_{|C|\to\infty} \frac{Var(r_c)}{|C|\alpha^2} = 0$

- Lemma 2: FedSampling can achieve ϵ-LDP in protecting local sample sizes i.f.f. $\alpha = \frac{\exp(\epsilon)-1}{\exp(\epsilon)-2+M}$
  - $\exp(\epsilon) = \max_{c,c',y} \frac{\Pr[\mathcal{M}(n_c)=y]}{\Pr[\mathcal{M}(n_{c'})=y]} = \frac{(M-1)\alpha+1}{1-\alpha}$

Experiential Datasets and Settings

- Datasets
  - FEMNIST: A benchmark image classification datasets for federated learning
  - Amazon-Toys: A review sentiment analysis datasets in the toy domain
  - Amazon-Beauty: A review sentiment analysis datasets in the beauty domain
  - MIND: A text classification dataset based on news corpus

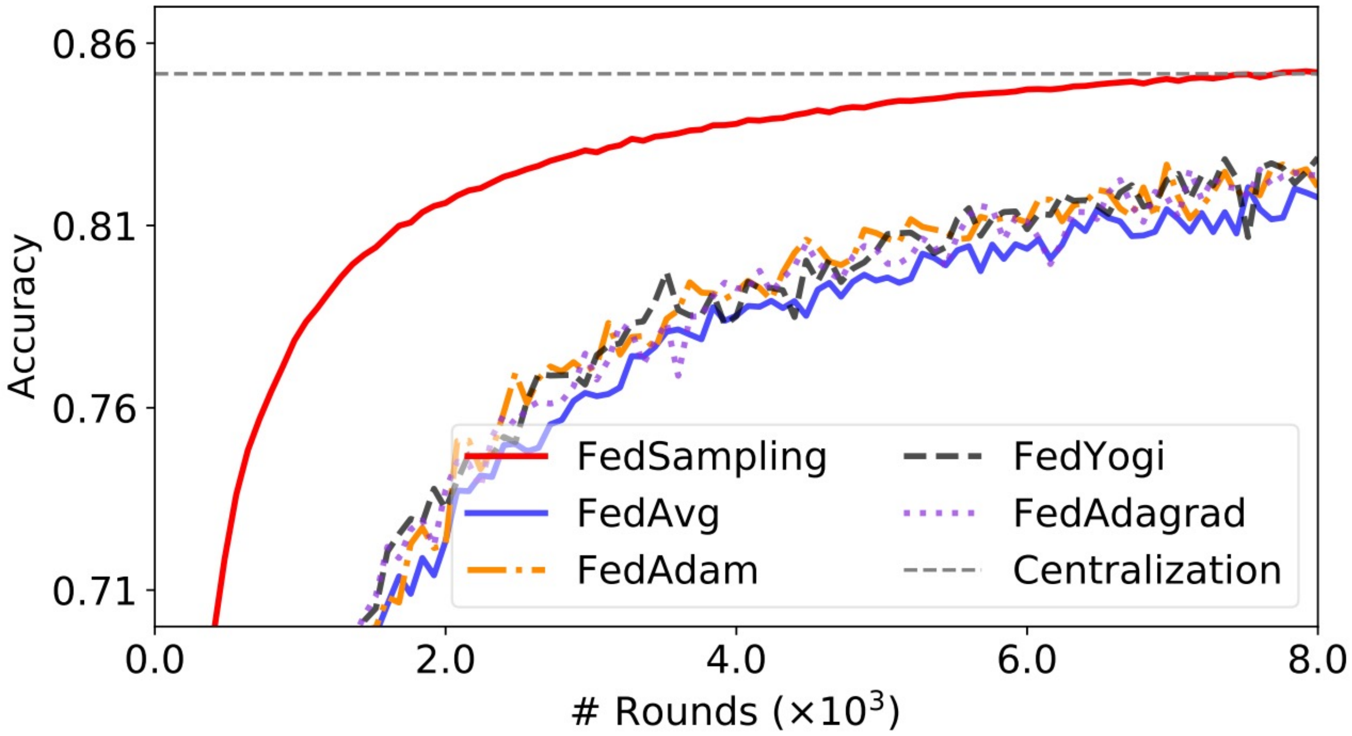
- Data patriation settings
  - Amazon datasets: Patriation data into clients based on the user ID
  - MIND: Patriation training data based on imbalanced data size distribution (log-normal)
  - FEMNIST: Patriation training data based on the class non-IID setting.

Performance Evaluation

Model	Training Algorithm	MIND		Toys		Beauty	
		Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy
Text-CNN	Centralization	51.52±0.57	71.14±0.45	39.61±1.13	63.71±0.22	43.90±0.97	62.20±0.67
	FedAvg	48.11±0.66	69.23±0.73	35.32±0.78	61.63±0.33	38.44±1.43	60.75±0.36
	FedYogi	49.12±0.71	68.92±0.40	35.62±2.34	61.22±0.39	38.77±0.89	60.35±0.91
	FedAdagrad	48.55±0.92	67.74±1.89	34.69±0.70	60.63±1.36	37.20±1.90	60.64±0.70
	FedAdam	48.54±0.65	68.22±0.50	35.27±1.59	61.35±0.32	39.09±0.80	60.43±1.05
	FedSampling	**51.33**±0.62	**71.15**±0.30	**40.15**±1.27	**63.41**±0.74	**43.04**±0.83	**62.96**±0.16
Transformer	Centralization	53.73±0.62	72.19±0.28	41.86±0.96	63.56±0.57	44.31±0.70	62.92±0.48
	FedAvg	50.46±0.99	70.74±0.52	38.68±0.93	60.30±2.06	37.82±1.36	60.41±0.27
	FedYogi	50.94±0.59	70.29±0.53	37.75±1.87	61.44±0.36	38.10±1.07	60.17±0.33
	FedAdagrad	50.99±0.68	70.65±0.48	38.06±0.61	59.69±1.60	38.59±1.56	59.87±0.51
	FedAdam	50.69±0.58	70.83±0.28	37.58±0.77	60.59±1.24	38.44±1.42	60.65±0.46
	FedSampling	**53.43**±0.57	**71.98**±0.37	**41.63**±1.12	**64.03**±0.46	**43.47**±0.94	**62.67**±0.60

Comparisons under Class Non-IID Distribution

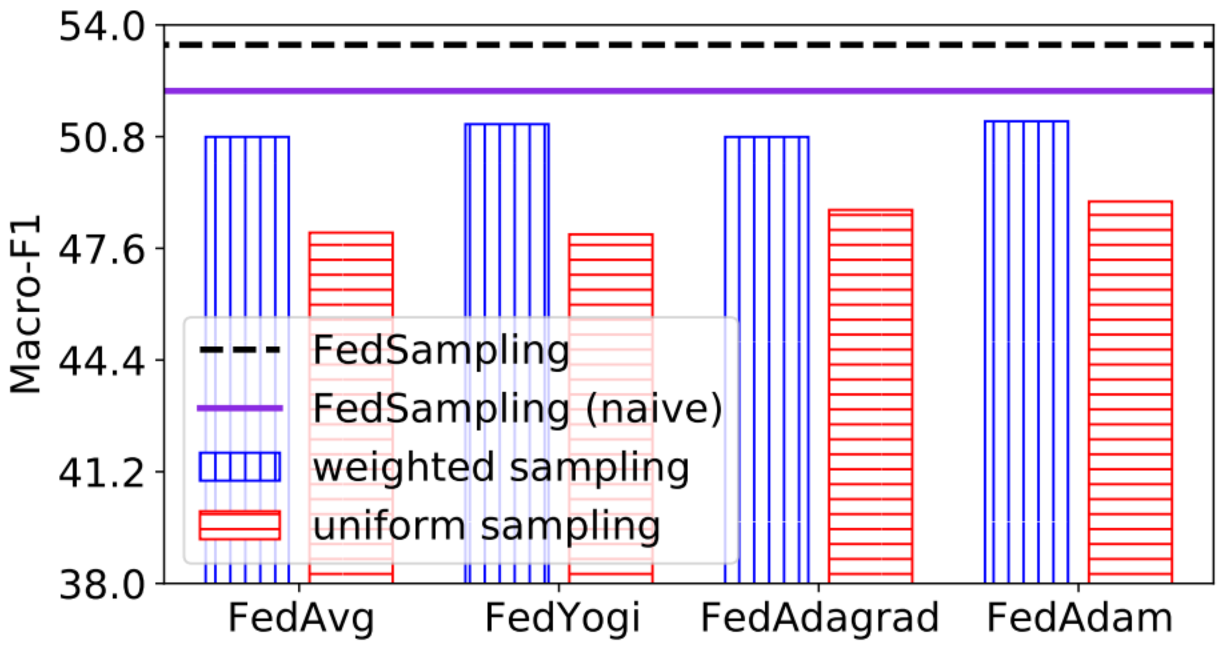
- Compare different methods on FEMNIST under the class non-IID setting



FedSampling outperforms baseline methods under class non-IID data distribution

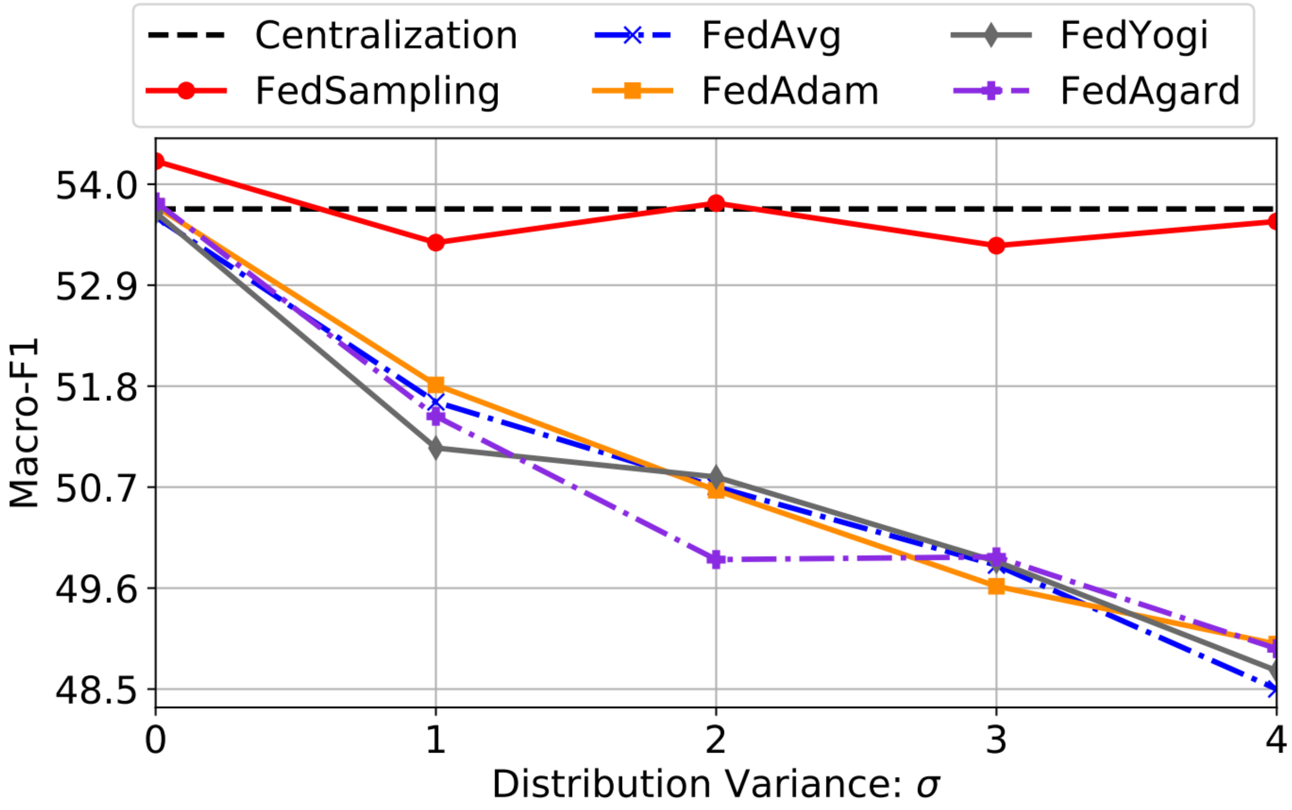
Comparisons with Weighted Sampling

- Compare FedSampling with its ablations on the text classification task



FedSampling achieves the best performance among its several ablation methods.

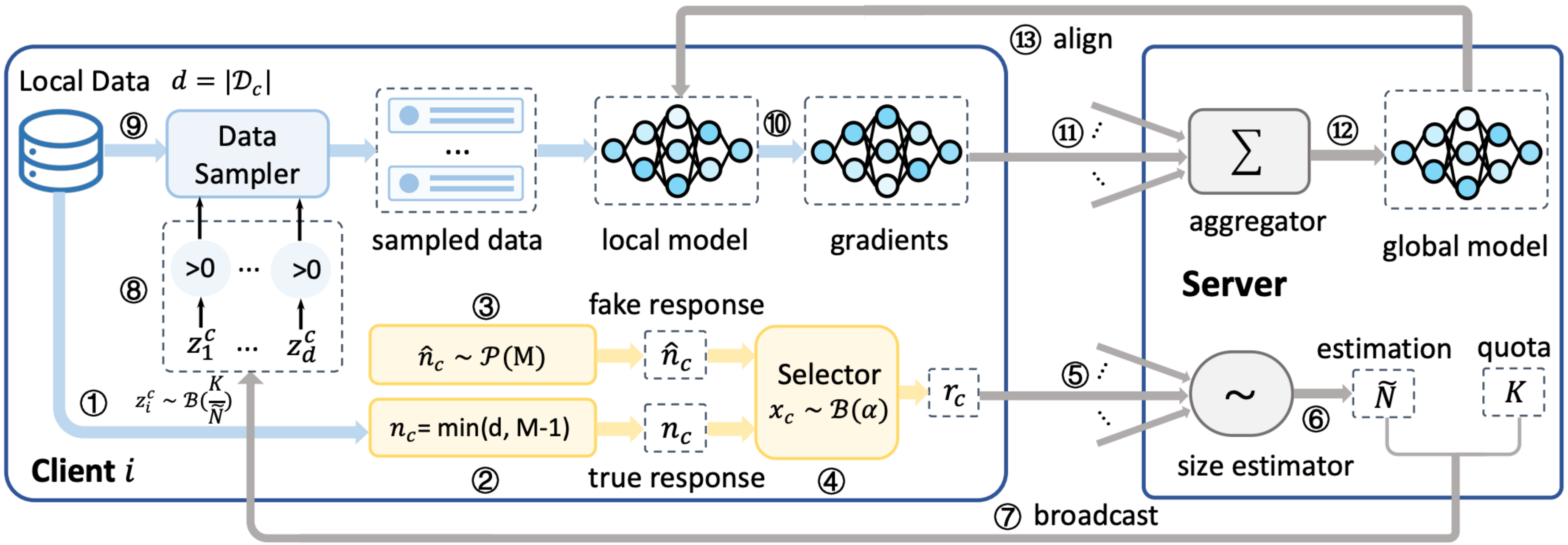
Influence of Data Size Imbalance degree



With the increasing of imbalance degree, the performance of baselines quickly degrades, while the performance of FedSampling drops slightly

Conclusion

- Propose an effective data sampling strategy for federated learning, which can achieve an uniform data exploitation in a privacy-preserving way



- Paper: https://arxiv.org/abs/2306.14245
- Code: https://github.com/taoqi98/FedSampling

Tao Qi

Tsinghua University

taoqi.qt@gmail.com