SparsyFed: Sparse Adaptive Federated Training

Adriano Guastella*[,1], **Lorenzo Sani***[,2,3], Alex Iacob [2,3], Alessio Mora [1], Paolo Bellavista [1], Nicholas D. Lane [2,3]

[*] *Equal contribution, correspondence to Adriano Guastella, adriano.guastella2@unibo.it*
[1] *Dipartimento di Informatica Scienza e Ingegneria, Università di Bologna*
[2] *Department of Computer Science and Technology, University of Cambridge*
[3] *Flower Labs, UK*

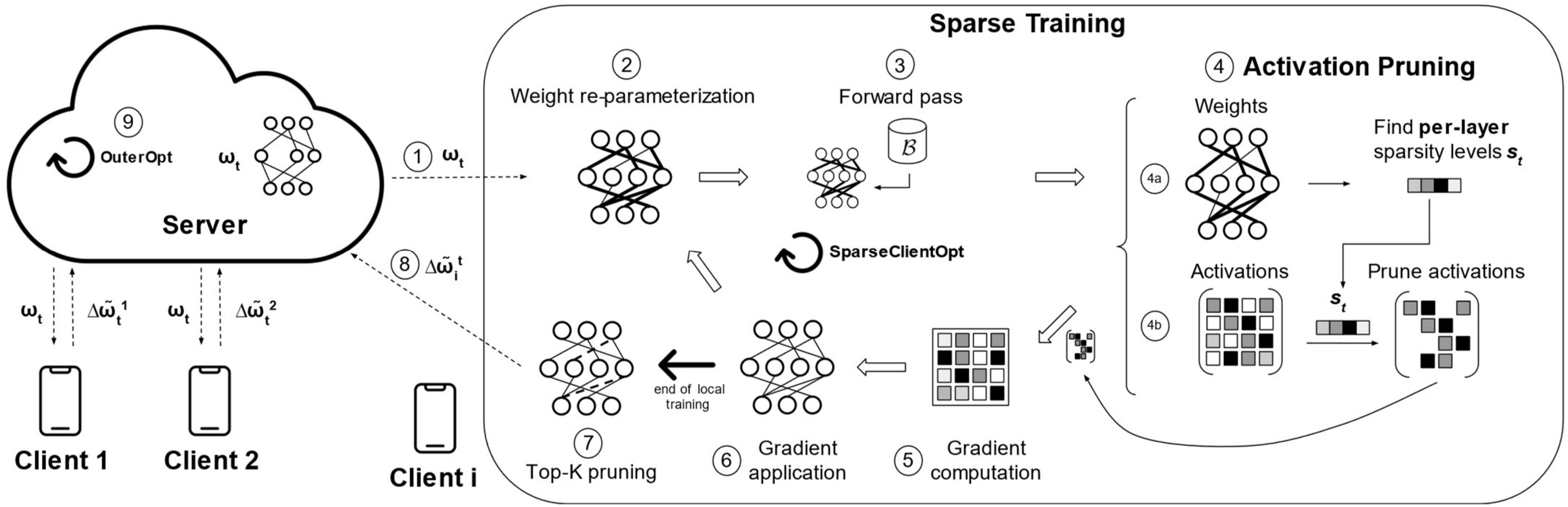
Introduction

Federated Learning (FL) enables decentralised training on client devices, enhancing privacy and reducing reliance on central storage. However, **high communication costs** and **limited device resources** pose major challenges.

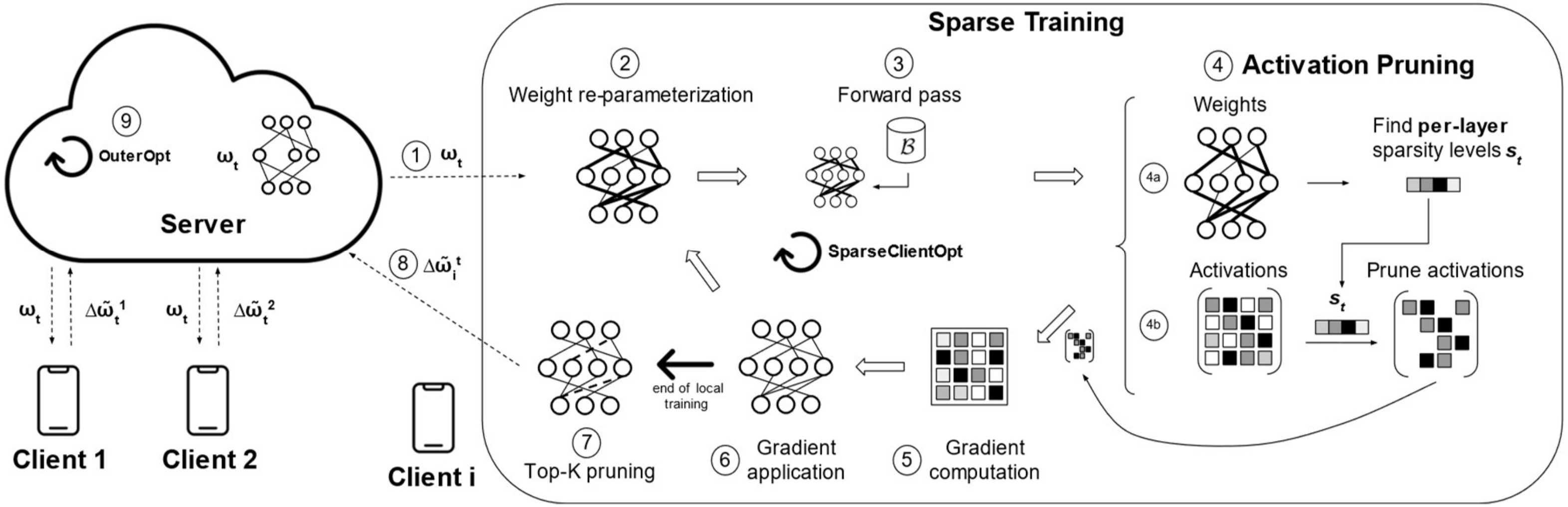
Sparse training mitigates these issues by reducing model size and computation but introduces new hurdles: achieving **consensus on sparse models**, ensuring **efficient convergence** across diverse clients, and minimizing communication overhead without **accuracy loss**.

Method

SparsyFed introduces a novel adaptive sparse training approach designed for cross-device FL. By **dynamically pruning activations** and **re-parameterizing weights**, *SparsyFed* enhances model efficiency while maintaining high accuracy, even at extreme sparsity levels.



Method



Key Components:

- **Activation Pruning**: Reduces computational overhead by selectively pruning activations before backpropagation, lowering memory usage and FLOPs while preserving critical information.
- **Weight Re-parameterization**: Uses a sparsity-inducing re-parametrization to improve resilience to pruning, enabling efficient training without excessive hyperparameter tuning.
- **Sparse Communication**: Prunes model updates before transmission, significantly reducing communication costs without degrading accuracy.

Results: Performance & Efficiency

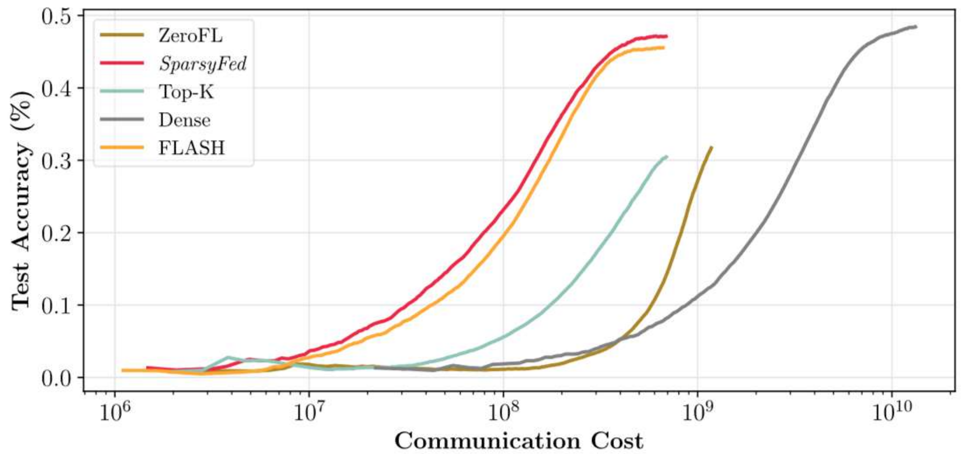
SparsyFed outperforms other baselines in accuracy across all settings, sometimes even matching dense model performance.

Dataset	Sparsity	$\alpha = 1.0$				$\alpha = 0.1$			
		ResNet-18	ZeroFL	FLASH	*SparsyFed*	ResNet-18	ZeroFL	FLASH	*SparsyFed*
CIFAR-10	dense	83.70 ± 1.70	-	-	-	73.81 ± 4.84	-	-	-
	0.9	80.56 ± 1.90	76.16 ± 1.30	81.15 ± 1.03	$\mathbf{82.13 \pm 1.53}$	69.79 ± 3.78	67.40 ± 4.11	71.87 ± 2.63	$\mathbf{75.00 \pm 2.78}$
	0.95	74.71 ± 3.29	75.53 ± 2.27	79.36 ± 1.03	$\mathbf{82.60 \pm 1.58}$	60.00 ± 4.66	61.55 ± 4.18	72.08 ± 2.09	$\mathbf{75.95 \pm 3.39}$
	0.99	66.27 ± 5.08	70.71 ± 0.15	73.45 ± 1.37	$\mathbf{77.71 \pm 1.69}$	43.96 ± 11.99	51.71 ± 3.54	56.91 ± 3.55	$\mathbf{63.69 \pm 3.90}$
	0.995	63.82 ± 2.41	56.02 ± 3.95	69.15 ± 1.60	$\mathbf{70.01 \pm 0.43}$	19.02 ± 10.77	41.33 ± 3.64	52.15 ± 3.87	$\mathbf{56.79 \pm 3.97}$
	0.999	31.79 ± 19.10	17.66 ± 8.34	36.07 ± 7.49	$\mathbf{51.39 \pm 3.19}$	11.50 ± 4.49	18.76 ± 4.28	29.31 ± 6.75	$\mathbf{43.68 \pm 7.61}$
CIFAR-100	dense	52.29 ± 1.14	-	-	-	48.34 ± 2.71	-	-	-
	0.9	46.57 ± 1.71	40.70 ± 4.72	51.99 ± 0.21	$\mathbf{53.08 \pm 0.90}$	41.96 ± 2.16	31.92 ± 7.65	45.59 ± 0.75	$\mathbf{48.37 \pm 1.73}$
	0.95	28.07 ± 23.27	38.82 ± 1.75	47.19 ± 1.88	$\mathbf{52.81 \pm 1.72}$	11.48 ± 17.51	34.21 ± 7.65	44.31 ± 2.14	$\mathbf{48.27 \pm 2.70}$
	0.99	19.65 ± 16.30	18.97 ± 2.08	42.76 ± 4.08	$\mathbf{46.64 \pm 1.59}$	0.14 ± 0.72	13.07 ± 2.26	34.75 ± 3.38	$\mathbf{41.03 \pm 2.14}$
	0.995	9.51 ± 14.81	6.01 ± 4.74	36.43 ± 4.97	$\mathbf{42.21 \pm 1.03}$	0.14 ± 0.72	7.04 ± 5.25	26.44 ± 17.35	$\mathbf{35.72 \pm 2.01}$
	0.999	3.81 ± 2.18	1.96 ± 0.66	5.80 ± 2.86	$\mathbf{15.96 \pm 0.64}$	0.14 ± 0.72	1.66 ± 0.97	3.56 ± 2.07	$\mathbf{13.84 \pm 3.69}$
Speech Commands	dense	91.49 ± 0.94	-	-	-	80.15 ± 2.69	-	-	-
	0.9	84.28 ± 0.88	87.79 ± 1.40	88.68 ± 1.72	$\mathbf{92.32 \pm 1.59}$	65.44 ± 0.97	70.35 ± 2.65	77.15 ± 0.77	$\mathbf{79.67 \pm 2.78}$
	0.95	78.58 ± 0.44	84.29 ± 1.50	84.89 ± 0.49	$\mathbf{89.14 \pm 1.15}$	57.39 ± 1.04	65.90 ± 1.88	71.28 ± 1.75	$\mathbf{75.46 \pm 2.24}$
	0.99	65.01 ± 0.84	57.79 ± 0.82	69.22 ± 1.59	$\mathbf{75.82 \pm 3.72}$	50.42 ± 6.26	41.42 ± 1.60	53.55 ± 2.00	$\mathbf{56.69 \pm 4.56}$
	0.995	56.73 ± 1.00	37.16 ± 2.71	58.23 ± 1.84	$\mathbf{68.02 \pm 3.14}$	34.20 ± 1.43	22.61 ± 3.45	43.16 ± 3.47	$\mathbf{48.30 \pm 5.39}$
	0.999	21.56 ± 12.79	10.10 ± 4.01	17.70 ± 2.58	$\mathbf{47.43 \pm 1.66}$	19.25 ± 6.01	8.85 ± 3.76	17.14 ± 2.97	$\mathbf{29.24 \pm 2.34}$

Results: Performance & Efficiency

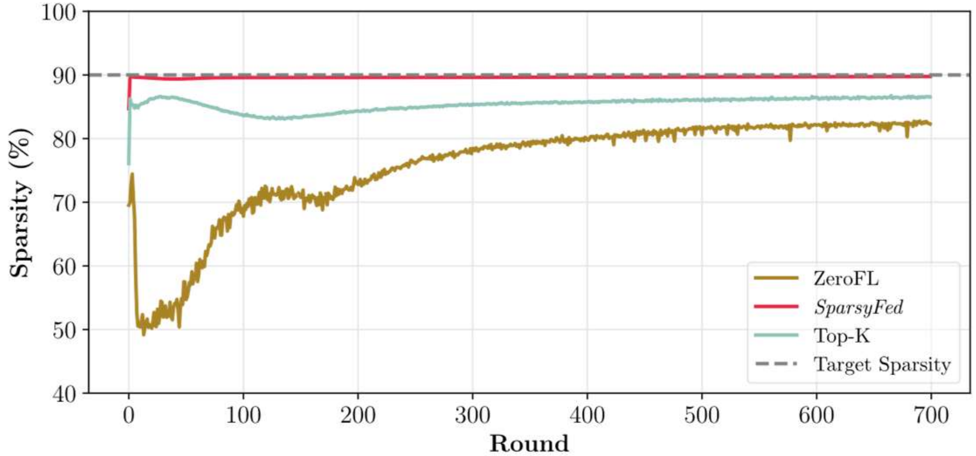
Significant Communication Savings

SparsyFed reduces communication costs by up to **19.29×** compared to dense models, minimizing both uplink and downlink overhead, without degrading model performance.



Sparsity Level Convergence

SparsyFed efficiently converges to the **target sparsity level** early in training, maintaining stable sparsity throughout the learning process, ensuring minimal deviation from the desired sparsity.



Thank you

Full paper: https://arxiv.org/abs/2504.05153

Repository: https://github.com/AGuastella/sparsyfed