



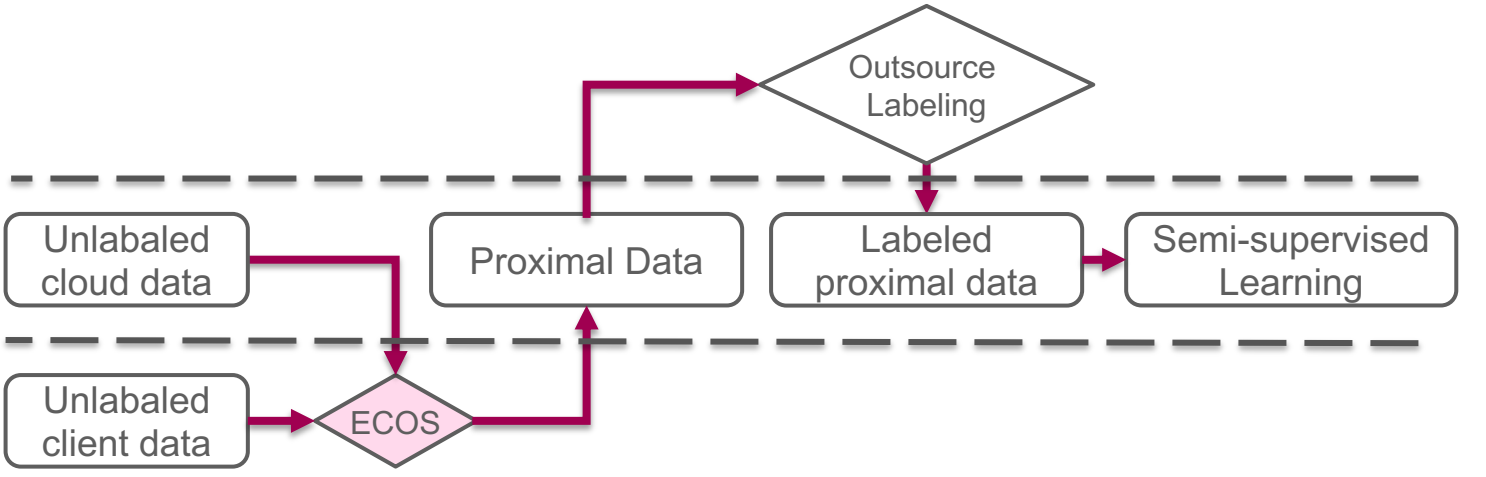
Outsourcing Training without Uploading Data via Efficient Collaborative Open-Source Sampling

Junyuan Hong¹, Lingjuan Lyu², Jiayu Zhou¹, Micheal Spranger²

¹ Michigan State University, ² Sony AI

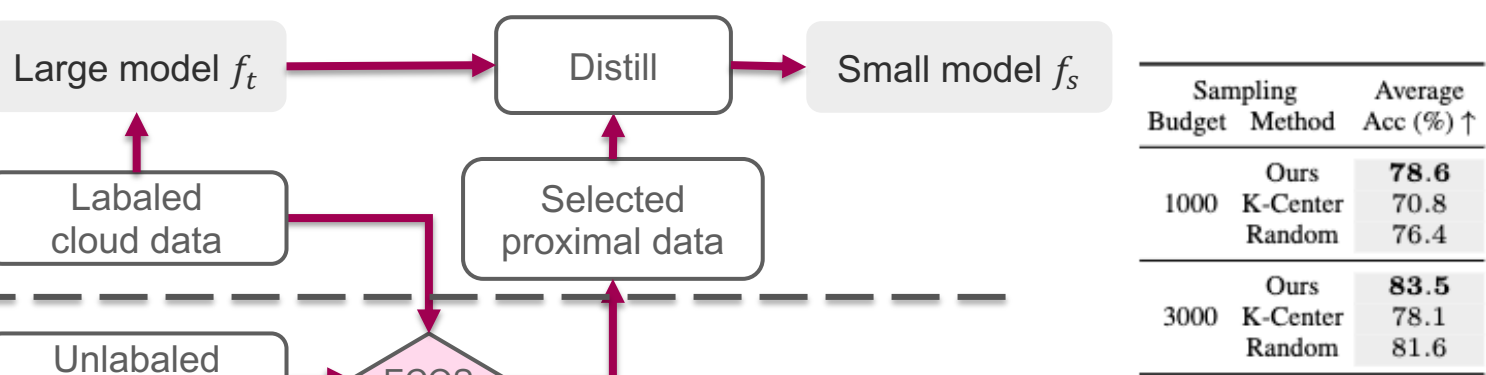
Applications

(a) Selective labeling



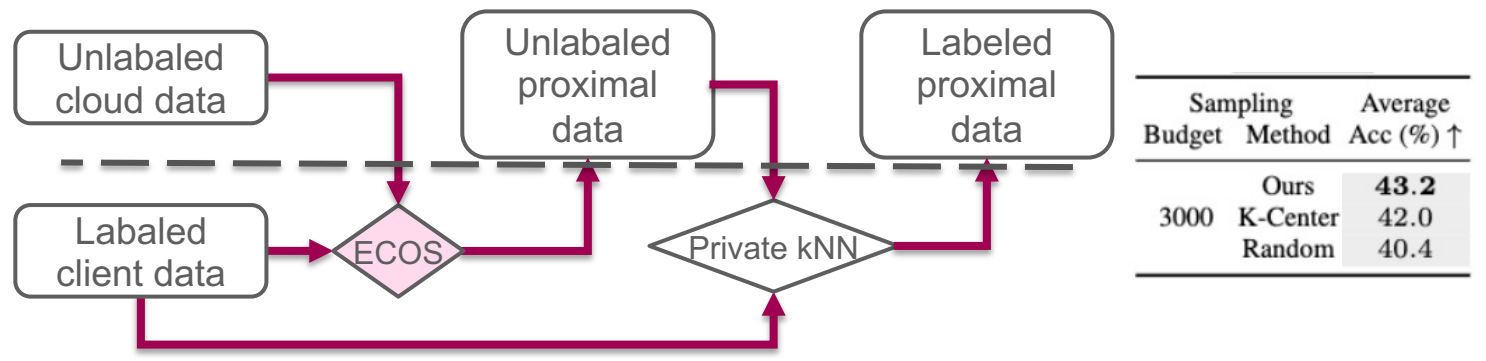
Sampling Budget	Method	Clipart		Infograph		Painting		Quickdraw		Real		Sketch		Average Acc (%) ↑
		Acc (%) ↑	ε ↓	Acc (%) ↑	ε ↓	Acc (%) ↑	ε ↓	Acc (%) ↑	ε ↓	Acc (%) ↑	ε ↓	Acc (%) ↑	ε ↓	
1000	Ours	88.4±1.5	0.58	52.6±0.9	0.58	90.4±1.7	0.58	84.3±1.6	0.58	92.1±1.2	0.58	87.2±0.5	0.58	82.5
	K-Center	86.8±0.3	0.00	50.5±0.9	0.00	89.1±1.4	0.00	27.2±1.8	0.00	92.5±0.1	0.00	85.6±1.4	0.00	72.0
	Random	86.9±0.8	0.00	47.4±2.7	0.00	88.6±0.1	0.00	77.9±2.4	0.00	91.4±0.3	0.00	86.9±0.5	0.00	79.9
3000	Ours	93.2±0.4	0.58	58.1±0.6	0.58	92.5±1.1	0.58	89.2±0.9	0.58	94.4±0.2	0.58	92.8±0.2	0.58	86.7
	K-Center	93.5±1.1	0.00	56.3±0.3	0.00	92.9±0.3	0.00	60.5±8.7	0.00	94.1±0.2	0.00	92.1±0.7	0.00	81.6
	Random	92.5±0.6	0.00	53.6±1.4	0.00	91.7±0.8	0.00	86.1±0.4	0.00	93.5±0.3	0.00	93.0±0.2	0.00	85.1
-	Local	87.0±0.3	0.00	51.7±0.7	0.00	85.9±0.4	0.00	83.5±0.3	0.00	93.5±0.1	0.00	81.7±0.6	0.00	80.6

(b) Adaptive model compression



Sampling Budget	Method	Average Acc (%) ↑
1000	Ours	78.6
	K-Center	70.8
	Random	76.4
3000	Ours	83.5
	K-Center	78.1
	Random	81.6

(c) Automated client labeling



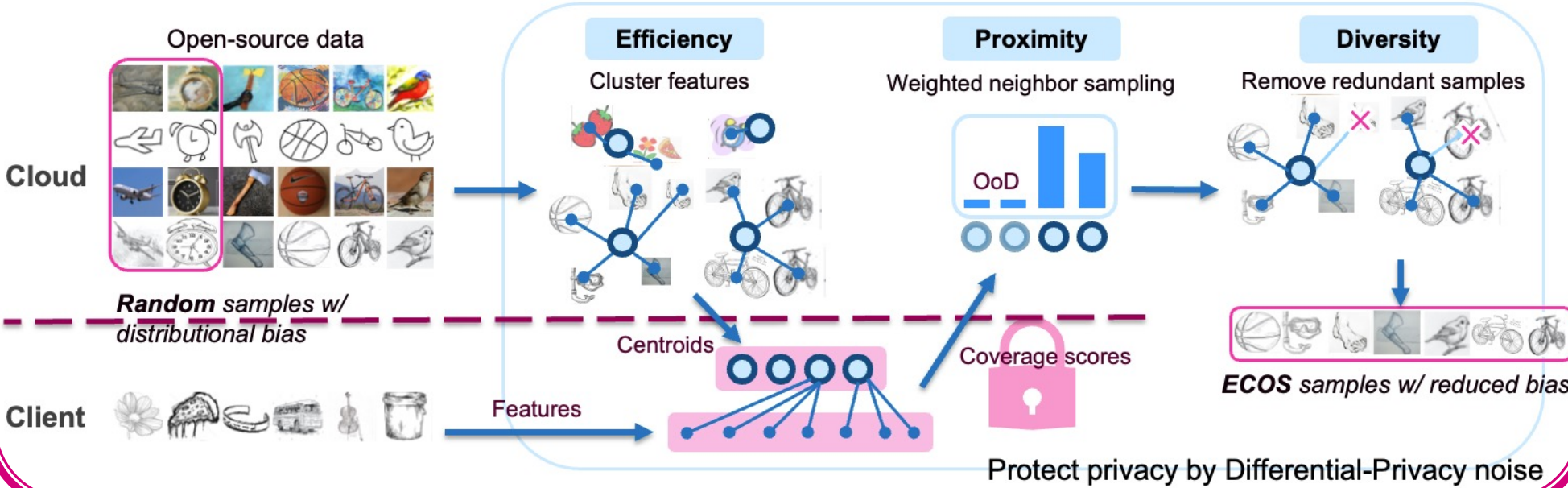
Sampling Budget	Method	Average Acc (%) ↑
3000	Ours	43.2
	K-Center	42.0
	Random	40.4

New Problem: Outsourcing Training without Uploading Data



New Idea: ECOS Efficient Collaborative Open-Source Sampling

- ❑ **Efficiency:** Open-source data could be enormous.
- ❑ **Bias:** Distributional gaps between private and open-source data.
- ❑ **Redundancy:** Duplicated samples from non-curated sources.
- ❑ **Privacy:** Communication may expose private information.



Efficiency



Conclusions

- ❑ **New problem:** Outsourcing training without uploading data.
- ❑ **New sampling paradigm:** ECOS.
- ❑ ECOS improves model accuracy.
- ❑ ECOS is widely applicable with different supervisions.
- ❑ ECOS is efficient in multiple dimensions.

Acknowledgement

This research was funded by Sony AI. This material is based in part upon work supported by the National Science Foundation under Grant IIS-2212174, IIS-1749940, Office of Naval Research N00014-20-1-2382, and National Institute on Aging (NIA) RF1AG072449.

