

Adaptive Distillation for Decentralized Learning from Heterogeneous Clients

Jiaxin Ma¹⁾, Ryo Yonetani¹⁾, and Zahid Igbal²⁾ 1) OMRON SINIC X 2) University Sains Malaysia



Background

• Problem: How can we train a deep neural network for classification tasks from decentralized labeled data?



Related Work

• Federated learning: server iteratively asks clients to train a global model and aggregates the client models to update the global model

Pros: no need to get direct access to client data Lim 1: clients have to train the model of the same arch. Lim 2: server should communicate with clients frequently



Our focus

- Can we allow clients to train models with different archs?
- Can we allow clients to send their model only once?

DLAD: Decentralized Learning via Adaptive Distillation

- Key idea: Network distillation [Hinton+, 2015] + adaptive distillation
- Allowing client models to take different architectures + to be sent to server only once



Non-IID #1 Distribution Client model ResNet DenseNet ResNet/DenseNet No. of clients 10 20 30 5 10 20 30 10 20 30 5 0.1696 0.1704 Client (ResNet) 0.1721 0.1697 0.1721 0.1700 0.1704 Client (DenseNet) 0.1850 0.1837 0.1847 0.1839 0.1824 0.1838 0.1818 -Global (avg.) 0.3066 0.3102 0.3590 0.3857 0.3471 0.4347 0.4719 0.4623 0.3648 0.4907 0.4771 Global (labeled) 0.7888 0.8000 0.8050 0.8160 0.8197 0.8194 0.8255 0.8264 0.8127 0.8252 0.8307 0.5748 **Global (DLAD)** 0.6376 0.5838 0.6227 0.6553 0.4725 0.5784 0.6251 0.6657 0.6379 0.6470 Non-IID #2 Distribution ResNet/DenseNet Client model ResNet DenseNet No. of clients 5 10 20 30 5 10 20 30 10 20 30 0.4331 0.4267 0.4205 0.4194 0.4331 0.4236 0.4199 Client (ResNet) Client (DenseNet) 0.4694 0.4339 0.4471 0.4533 0.3985 0.4330 0.4433 ----Global (avg.) 0.4163 0.4069 0.4185 0.4103 0.4216 0.4175 0.4182 0.4270 0.4176 0.4190 0.4199 0.7632 Global (labeled) 0.7383 0.7659 0.7728 0.7706 0.7774 0.7506 0.7987 0.7957 0.7625 0.7993 0.6904 0.7025 **Global (DLAD)** 0.6660 0.6952 0.6914 0.6151 0.6786 0.6219 0.6782 0.6772 0.6170

- Dataset: MNIST/CIFAR-10/CINIC-10 split into multile client data subsets
- Baselines: Client model | Average aggregation | Upperbound (w/ labeled distil. data)
- Model archs: ResNet / DenseNet

More information:

Full paper: https://arxiv.org/abs/2008.07948 RL extension: https://arxiv.org/abs/1909.13111

Contact:

Ryo Yonetani, Ph.D. ryo.yonetani@sinicx.com

Computing aggregation weights