



Adaptive Distillation for Decentralized Learning from Heterogeneous Clients

Jiaxin Ma (OMRON SINIC X)



*Ryo Yonetani (OMRON SINIC X)



Zahid Iqbal (University Sains Malaysia)



Scenario: Learning from Decentralized Private Data

- You want to train a deep neural network for classification tasks ... but you don't have labeled data!
- You found people storing labeled data locally ... which you cannot access directly due to privacy concerns
- How can we leverage such decentralized labeled data for training a neural net?











- 1. Server distributes global model
- 2. Clients update the global model with their own data
- 3. And send them back to the server
- 4. Server aggregates (averages) the client models to update the global one





Our Focus

	Client Model Architectures	Server-client communications
Federated Learning	Homogeneous All clients need to train the same model	Multi-rounds Server & clients exhange models iteratively
Our Focus	Heterogeneous Model architectures may vary by clients	Single-round Clients share their own models only once

Our Focus

	Client Model Architectures	Server-client communications
		Multi-rounds Server & clients exhange models iteratively
Our Focus	Heterogeneous Model architectures may vary by clients	Single-round Clients share their own models only once

Decentralied Learning via Adaptive Distillation (DLAD)

- Key idea: network distillation [Hinton+, 2015] + adaptive aggregation
 - Given unlabeled data source (distillation data), training global model to imitate aggregated outputs of client models
 - Aggregating outputs adaptively to emphasize outputs from confident clients



Decentralied Learning via Adaptive Distillation (DLAD) - cont'd

- Computing confidence-based aggregation weights
 - Each client trains an additional classifier to distinguish their own data from distillation one
 - Using classifier scores for aggregation weights: higher scores means clients are more confident about the outputs



Selected Results (CIFAR-10)

Different types of client data distributions



Different numbers of clients



Number of clients increases up to 30

Summary

- DLAD learned global model from decentralized data using adaptive distillation
- It outperformed a baseline method with naïve aggregation strategy
- Model-free RL extension available at IJCAI'20 -> Search by "MULTIPOLAR transfer RL"

