Persistence of Member Contribution Under Churn

Robust Decentralized Learning

EEMCS, Delft University of Technology
Author: Luka Roginić
Supervisors: Jérémie Decouchant, Bart Cox
Full paper: https://repository.tudelft.nl/

Introduction

- Decentralized learning is a paradigm to perform machine learning from multi-source data in a distributed and decentralized manner
- Churn, when members leave the network, has the most influence under non-identical and independently distributed (non-IID) data, reducing generalizability and hurting performance[1]
- Possible scenario: decentralized learing where data comes from mobile phones which often switch between being online and offline
- Existing efforts often treat the performance of decentralized learning under churn as a secondary issue, overlooking its significance

Research Question

How can we mitigate the impact of node churn in decentralized learning systems to maintain some persistence of member contributions?

Methodology

Churn:

- Permanent churn is modeled with schedule deciding in which iteration a member leaves.
- probaiblistic churn is modeled with probability to leave and rejoin the network

Joint Steps:

- Dataset condensation: generate synthetic data using distribution matching (see Figure 1)
- Exchange synthetic data with neighbors

Data Augmentation:

- augment received data to train set
- train with cross entropy loss

Synthetic Anchors (inspired by DeSA[2]):

- augment received data to synthetic set
- train with cross entropy loss + supervised contrastive loss[4] :

$$\mathcal{L} = \lambda_{\mathrm{CE}}\mathcal{L}_{\mathrm{CE}}(D_i, D^{\mathrm{syn}}, M_i) + \lambda_{\mathrm{SCL}}\mathcal{L}_{\mathrm{SCL}}(D_i, D^{\mathrm{syn}}, M_i)$$

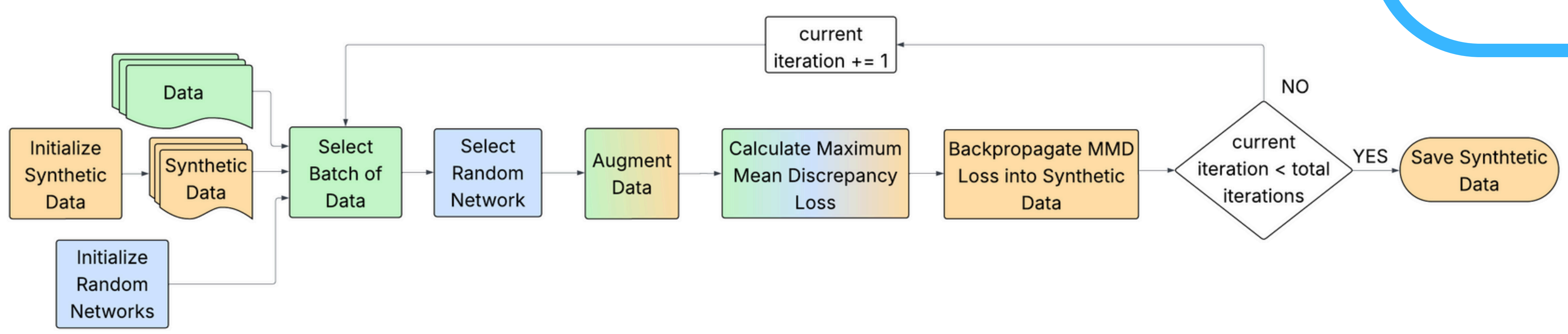

Figure 1: Data Synthesis

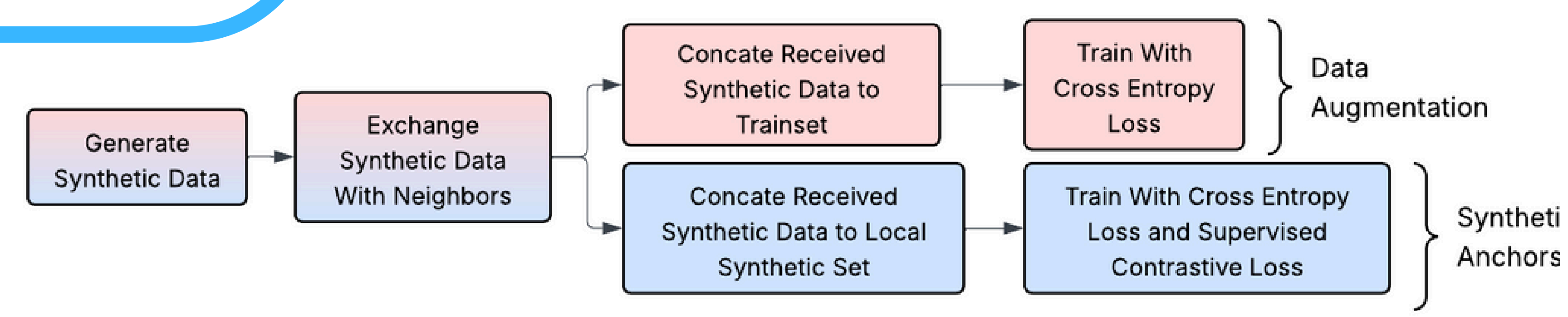

Figure 2: Data Augmentation and Synthethic Anchors Overview

Results

Setting	Degree 3 Test Acc. (%)	Degree 3 MMD Acc. (%)	Degree 3 Δ DPSGD	Degree 5 Test Acc. (%)	Degree 5 MMD Acc. (%)	Degree 5 Δ DPSGD
DPSGD Augmented	55.94 ± 2.76	–	–	56.84 ± 4.39	–	–
DPSGD - 3 members leaving	41.43 ± 4.20	39.26 ± 7.76	-14.51	46.89 ± 4.73	45.31 ± 4.53	-9.94
Data Augmentation	45.95 ± 6.94	44.08 ± 10.75	-10.00	50.85 ± 3.43	53.22 ± 4.80	-5.99
Synthetic Anchors	46.60 ± 2.61	41.95 ± 5.38	-9.34	48.45 ± 3.16	44.80 ± 2.88	-8.38
Linear Synthetic Anchors	50.82 ± 1.75	46.01 ± 4.27	-5.12	50.68 ± 2.88	49.72 ± 1.74	-6.15
DPSGD - 5 members leaving	40.37 ± 9.18	39.64 ± 7.23	-15.57	46.19 ± 4.18	33.08 ± 9.43	-10.65
Data Augmentation	45.69 ± 10.36	40.41 ± 12.12	-10.25	48.66 ± 5.37	38.11 ± 7.98	-8.17
Synthetic Anchors	42.80 ± 9.64	40.35 ± 12.35	-21.31	48.95 ± 2.26	36.98 ± 6.14	-7.88
Linear Synthetic Anchors	47.29 ± 10.90	41.82 ± 11.32	-8.65	50.79 ± 2.08	40.97 ± 5.51	-6.05
DPSGD - 8 members leaving	44.20 ± 6.67	32.91 ± 8.03	-11.74	44.77 ± 6.15	32.60 ± 6.34	-12.07
Data Augmentation	47.32 ± 5.92	37.76 ± 6.84	-9.52	49.72 ± 5.56	37.34 ± 8.62	-7.11
Synthetic Anchors	43.22 ± 4.17	31.86 ± 4.72	-12.72	47.72 ± 3.85	35.69 ± 4.01	-9.12
Linear Synthetic Anchors	46.46 ± 4.18	35.33 ± 5.07	-9.48	45.89 ± 3.68	35.53 ± 3.99	-10.95

Setting	Degree 3	Degree 5
DPSGD Augmented	55.94 ± 2.76	56.84 ± 4.39
DPSGD - p_{active}=0.80	52.07 ± 3.27	54.67 ± 3.12
Data Augmentation	52.32 ± 4.74	55.83 ± 2.14
Synthetic Anchors	51.86 ± 4.25	55.14 ± 2.65
Linear Synthetic Anchors	53.79 ± 2.71	54.68 ± 2.72
DPSGD - p_{active}=0.90	50.64 ± 4.13	52.33 ± 2.52
Data Augmentation	51.41 ± 5.26	53.26 ± 2.38
Synthetic Anchors	49.70 ± 2.61	54.67 ± 1.56
Linear Synthetic Anchors	52.63 ± 3.56	55.41 ± 1.62
DPSGD - p_{active}=0.95	47.83 ± 4.66	50.02 ± 3.20
Data Augmentation	48.86 ± 3.57	51.82 ± 3.12
Synthetic Anchors	46.26 ± 3.63	51.29 ± 2.82
Linear Synthetic Anchors	49.42 ± 2.95	52.04 ± 2.77

Setup

- 16 members in the network
- 3, 5, or 8 members leave the network
- probability of being active : 80%, 90%, or 95%
- test accuracy and accuracy on testing on missing members data (MMD)
- CIFAR10 dataset
- Linear Synthetic Anchors: dynamic weights defined by a linear function of the current iteration.

- Data augmentation improves test accuracy by up to 5.32% and MMD accuracy by up to 7.89%, while also reducing standard deviation.
- Linear synthetic anchors outperform both data augmentation and static synthetic anchors, with gains reaching 9.29% in test accuracy and 6.75% in MMD accuracy when 3 or 5 members leave. They also offer the highest stability, achieving lower standard deviation in 21 out of 24 test and MMD accuracy comparisons. Furthermore, they achieve the best performance in 5 out of 6 probabilistic churn scenarios.

Discussion

- Data augmentation as well as synthetic anchors help mitigate the impact of churn and preserve member contribution
- Static synthetic anchors perform worse than data augmentation because overrealying on synthetic data, which although is more information dense carries less total information
- Linear synthetic anchors perform the best balancing quick learning with synthetic data with model tuning on regular samples
- Limitations: small network size, simplified churn and ML models, and the computational cost of dataset condensation
- Future work: larger network, more complex churn and machine learning model. Different dataset condensation method, different weight balance between losses in synthetic anchors, and propagating synthetic data further

REFRENCES

- [1] T. Liu, Y. Cui, X. Hu, Y. Xu, and B. Liu. On the Convergence of Gossip Learning in the Presence of Node Inaccessibility. ICCC 2022, pp. 4497–4502
- [2] C. -Y. Huang, K. Srinivas, X. Zhang, and X. Li. “Overcoming Data and Model heterogeneities in Decentralized Federated Learning via Synthetic Anchors”. In: Proceedings of the 41st International Conference on Machine Learning. Vol. 235. Proceedings of Machine Learning Research. PMLR, 21–27Jul 2024, pp. 20111–20133.
- [3] B. Zhao and H. Bilen. Dataset Condensation with Distribution Matching. 2022.
- [4] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan. “Supervised Contrastive Learning”. In: CoRR abs/2004.11362 (2020).

TU Delft