Note

Xin Huang

2025 年 3 月 17 日

0.1 Paper and notes

[1] FedAvg开山之作

[2] solves the GAN application in Federated Learning scenario.

[3] propose Dual-Adapter Teacher (DAT) to address data heterogeneity by regularizing the client local updates and applying Mutual Knowledge Distillation (MKD) for an efficient knowledge transfer. Parameter-Efficient Finetuning (PEFT) for Federated Learning is similar to LoRA which use two matriex to replace a big matrix. Adapter is similar to LoRA. This paper uses Mutual Knowledge Distillation (MKD) which is quite famliar in FL. This work is in a scenario of Mutil-Modal trainning.

[4] reduce the communication overhead in SCAFFFOLD[5].

[6] add extra dynamic regularizer which uses gradient of last trainning epoch based on FedProx[7]. But I cannot Reproduce the results. The formulars appearing in the Code are not consistent with that in the paper. They are totally different.

[8] FedACG add momentum on the server side, sending to the model with momentum to clinets avoiding the stroage overhead on the client side, and accelerate the convergence rate. It's motivation is reduce the inconsistency between the local models and the divergence of the resulting global model. It add two components: momentum on the server side and momentum based regularization on the client side. The most similar work is FedAvgM [9]. Also, [10]demostrates that the momentum benefits FL.

Looking ahead is a common strategy. We can look ahead in server side and in client side with one more epoch trainning

FedNova[11] is a normalized averaging method that eliminates objective inconsistency while preserving fast error convergence.

t-SNE, weight divergence, CKA similarity

Knowledge Distillation:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_{CE}(q^S, y) + \lambda\mathcal{L}_{KL}(q^S_\tau, q^T_\tau) \tag{1}$$

$q^S$is the softlabel predicted by the Student, y is the groud truth. q^T_τ is the label predicted by the Teacher. τ is the smooth control variable.

FedDistill [12]: participants compute the average soft label on own private datasets and send to the server. The server averages the global soft label. High efficience the communication. FedDistill [12] is similar to FedProto [13]

FedGKD [14] averages the historical global buffer or distributes the historical global models. The rest of the work is similar to FedDistill. In this method, I noticed that distributing historical global models is kind of like making serilization just like the git rebase.

FedCOG [15] generates data to complement original data and use knowledge Distillation.

Inco aggregation[16] observers a decline in model accuracy and layer-wise similarity (layer similarity) as measured by Centered Kernel Alignment. The deeper layers share lower layer similarity across the clients, while the shallower layers exhibit greater alignment.

参考文献

[1] H. B. McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *International Conference on Artificial Intelligence and Statistics*, 2016.

[2] Zuobin Xiong, Wei Li, and Zhipeng Cai. Federated generative model on multi-source heterogeneous data in iot. In *AAAI Conference on Artificial Intelligence*, 2023.

[3] Haokun Chen, Yao Zhang, Denis Krompass, Jindong Gu, and Volker Tresp. Feddat: An approach for foundation model finetuning in multi-modal heterogeneous federated learning. In *AAAI Conference on Artificial Intelligence*, 2023.

[4] Xinmeng Huang, Ping Li, and Xiaoyun Li. Stochastic controlled averaging for federated learning with communication compression. *ArXiv*, abs/2308.08165, 2023.

[5] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank J. Reddi, Sebastian U. Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for on-device federated learning. *ArXiv*, abs/1910.06378, 2019.

[6] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N. Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. *ArXiv*, abs/2111.04263, 2021.

[7] Anit Kumar Sahu, Tian Li, Maziar Sanjabi, Manzil Zaheer, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *arXiv: Learning*, 2018.

[8] Geeho Kim, Jinkyu Kim, and Bohyung Han. Communication-efficient federated learning with accelerated client gradient. In *CVPR*, 2024.

[9] Tzu-Ming Harry Hsu, Qi, and Matthew Brown. Measuring the effects of non-identical data distribution for federated visual classification. *ArXiv*, abs/1909.06335, 2019.

[10] Ziheng Cheng, Xinmeng Huang, and K. Yuan. Momentum benefits non-iid federated learning simply and provably. *ArXiv*, abs/2306.16504, 2023.

[11] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H. Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *ArXiv*, abs/2007.07481, 2020.

[12] Eunjeong Jeong, Seungeun Oh, Hyesung Kim, Jihong Park, Mehdi Bennis, and Seong-Lyun Kim. Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data. *ArXiv*, abs/1811.11479, 2018.

[13] Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *AAAI Conference on Artificial Intelligence*, 2021.

[14] Dezhong Yao, Wanning Pan, Yutong Dai, Yao Wan, Xiaofeng Ding, Chen Yu, Hai Jin, Zheng Xu, and Lichao Sun. Fedgkd: Toward heterogeneous federated learning via global knowledge distillation. *IEEE Transactions on Computers*, 73(1):3–17, 2024.

[15] Rui Ye, Yaxin Du, Zhenyang Ni, Siheng Chen, and Yanfeng Wang. Fake it till make it: Federated learning with consensus-oriented generation. *ArXiv*, abs/2312.05966, 2023.

[16] Yun-Hin Chan, Rui Zhou, Running Zhao, Zhihan Jiang, and Edith C. H. Ngai. Internal cross-layer gradients for extending homogeneity to heterogeneity in federated learning. *ArXiv*, abs/2308.11464, 2023.