Adapted Weighted Aggregation in Federated Learning

Yitong Tang
University of British Columbia, 2329 West Mall
Trusted and Efficient AI (TEA) Lab
yitongta@student.ubc.ca

Abstract

This study introduces FedAW, a novel federated learning algorithm that uses a weighted aggregation mechanism sensitive to the quality of client datasets, leading to better model performance and faster convergence on diverse datasets, validated using Colored MNIST.

Introduction

Research has demonstrated that conventional federated approaches like FedAvg (McMahan et al. 2017), which do not specifically account for non-iid data distribution, can experience substantial performance drops or even convergence issues under such conditions (Kairouz et al. 2019; Karimireddy et al. 2019). This highlights the necessity for developing more robust federated learning algorithms that can effectively handle the intrinsic data variability across diverse clients.

In this study, we introduce **FedAw**, an innovative federated optimization algorithm that tackles the inherent challenges of data heterogeneity from. Our approach is underpinned by a set of meticulously devised evaluation metrics aimed at assessing the quality of datasets held by individual clients. These metrics inform a weighted aggregation scheme, wherein clients possessing more representative samples exert greater influence on the aggregation of the global model.

Code is avaialble (https://github.com/Yitong999/FedAW)

Background

Federated learning's privacy is challenged by data disparity, impacting global model efficacy. Previous solutions like q-FFL (Li et al. 2020) partially address this; however, FedAW innovatively evaluates local data quality via biased models trained by Generalized Cross Entropy Loss (Nam et al. 2020), enhancing model fairness and effectiveness.

Dataset Setup

Colored MNIST, has 10 digits thus $y \in [10], where[m] = \{0, 1, ..., m-1\}$. We also have the sensitive attribute A (which is digit color in our case) , For an image x, it contains

Figure 1: Illustration of bias-aligned samples for Colored MNIST

a digit y, and it has a unprivileged color $a_y \in A$, which is the major injected color of digit y, and it has privileged colors $A_{/a_y}$ which is all the other colors. For example, $y = 1$, it has unprivileged color which is $a_1 = yellow$ as shown in Figure 1, and it has privileged colors $a_{\neg y}$ which is a set containing all other colors except yellow. For each digit y class, we consider correct predictions as positive outcomes. We use (x, y, a) to denote each training example.

In the training dataset, we denote **ratio** $r = \frac{|\{X:(x,y,a_y)\}|}{|X|}$. In Federated Learning setting, 4 clients have **r** = 0.005, 2 clients have **r** = 0.01. 2 clients have **r** = 0.02, and 2 clients have **r** = 0.05.

Approach

Inspired by an observation from Chang et al. (Chang and Shokri 2023), we propose an adaptive reweighing algorithm to prevent the performance degradation of a well-trained global model by local models trained on biased datasets. From Table 1, without reweighing on models in aggregation, models with $r = 0.005$ will have a strong negative impact on the global model. Initially, we train two separate local models to unsupervisedly identify those trained on biased datasets (Section 4.1). Subsequently, we focus on reassigning weights during the model aggregation step in communication. To maximize the effectiveness of this weight reassignment, we schedule it after the local biased models are adequately trained.

Identifying Well Trained Client's Models

$$score \leftarrow \frac{CE(x^i, y^i; \phi_k)}{CE(x^i, y^i, \phi_k^t) + CE(x^i, y^i; \phi_k^t)} \tag{1}$$

This involved training a biased model, f_B , and then focusing on samples that f_B misclassified to train a debiased

Algorithm 1: Adapted Weighted Aggregation in Federated Learning

Input: $T, T_w, K, D = \{D_1, D_2, ..., D_K\}$
Parameter: debiased models $\{\phi_1, ..., \phi_K\}$, biased models $\{\phi'_1, ..., \phi'_K\}$, learning rate η

1: **for** $t = 0, 1, ..., T-1$ **do**
2: scores $\leftarrow [1]$
3: **for** $k = 1, ..., K$ **in parallel do**
4: $\phi_k^t, score \leftarrow$ **LocalUpdate**(θ^t)
5: scores $\leftarrow$ score
6: **end for**
7: $\theta^{t+1} \leftarrow \sum_{k=1}^{K} \frac{score_k}{\sum scores} \phi_k^t$
8: **end for**
LocalUpdate(θ^t):
9: $\phi_k \leftarrow \theta^t$
10: **for** $e = 1, 2, ..., E_f$ **do**
11: Draw a mini-batch $B = \{x_i^{y_i}\}_{i=1}^{B}$ from D_k
12: score $\leftarrow CE(x_i^{y_i}; \phi_k^t) + CE(x_i^{y_i}; \phi_k^{t'})$
13: $R_g(\phi_k^t) \leftarrow \frac{1}{B} \sum_{i=1}^{B} GCE(x_i^{y_i}; \phi_k)$
14: $\phi_k^t \leftarrow \phi_k^t - \eta \nabla R_g(\phi_k^t)$
15: $R_f(\theta^t) \leftarrow$ local debiased model training loss
16: $\phi_k^t \leftarrow \phi_k^t - \eta \nabla R_f(\phi_k^t)$
17: **end for**
18: **return** $\theta^t, score$

model, f_D , as in LfF (Nam et al. 2020). From equation (1), a bias-aligned sample has a smaller value in numerator and a larger value in denominator, leading a lower score. For each client, we sum up the score for every local batch of training data. As bias-conflicting samples obtain high scores, a training dataset with more bias-conflicting samples will obtain higher scores in the summation.

Weighing Well Trained Client's Models in Aggregation

Leveraging strategies from (Li et al. 2020), we enhanced the global model's fairness by preferentially weighting less biased local models. We assign more weights to clients with higher scores in aggregation.There are two versions for assigning the weight:

v1 is more aggressive by linearly assigning the weight as indicated in line 9; **v1** is designed to exaggeratedly rely on well trained local models while in aggregation. **v2** is more gentle by softmax the list of scores. It won't lose too many features (compared to v1) from not well trained local models. There could be other weight assigning strategies (i.e. polynomial). Users have to decide weight assigning strategies case by case.

Scheduling The Weight Assigning

We must ensure that the biased model is thoroughly trained to exhibit bias before proceeding to step 2. If not, the biased model ϕ' may not effectively reduce the loss for biased samples. Consequently, Equation 1 may fail to assign lower scores to clients trained with biased datasets. The scheduling of this process is determined on a case-by-case basis. For instance, in the case of **CMNIST**, local biased models are considered well-trained after 2000 epochs.

Evaluations

$$DI(Y = y) = \frac{\Pr(outcome = +|Y = y, A = a_y)}{\Pr(outcome = +|Y = y, A \neq a_y)} \quad (2)$$

This section outlines the training dataset and federated settings utilized. Clients were trained on the Colored MNIST dataset, each with varying proportions of bias-conflicting samples. Specifically, four clients were allocated 0.5% bias-conflicting samples (the highest level of bias), two clients had 1% bias-conflicting samples, another two clients had 2% bias-conflicting samples, and the remaining two clients had 5% bias-conflicting samples (the lowest level of bias).

Dataset	Method	FedAvg	FedAw_v1	FedAw_v2
CMNIST	Vanilla	0.63	**0.65**	0.64
	LF	0.80	**0.86**	0.83
	BiasAdv	0.78	**0.80**	**0.80**

Table 1: Accuracy of image classification assessed on fair test sets from the CMNIST. The top-performing results are highlighted in bold.

Dataset	Method	FedAvg	FedAw_v1	FedAw_v2
CMNIST	Vanilla	2.1	**1.6**	1.7
	LF	**1.0**	**1.0**	0.96
	BiasAdv	1.1	**1.0**	1.6

Table 2: Disparate Impact of image classification assessed on fair test sets from the CMNIST. The top-performing results are highlighted in bold.

Baselines Our baseline is FedAvg aggregation. Comparison on test sets Table 1 shows the comparisons in accuracy with $FedAw_{v1}$, $FedAw_{v2}$, and $FedAvg$ on three unsupervised bias eliminating methods. Accuracy is evaluated on a fair colored MNIST dataset where bias attributes (color) are not related to labels.

Conclusion

This study introduces an innovative federated learning aggregation technique termed FedAW, which recalibrates the weight of individual client models based on the quality of each local model's training dataset. This strategy is designed to counter the issue of biased feature heterogeneity across diverse clients. Empirical assessments conducted on the Colored MNIST dataset illustrate that FedAW significantly elevates the convergence rate and overall performance for non-IID datasets. The study further corroborates the efficacy of FedBN, particularly in environments where a substantial proportion of clients present with highly biased training datasets.

References

Chang, H.; and Shokri, R. 2023. BIAS PROPAGATION IN FEDERATED LEARNING.

Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2019. Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977.

Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S. J.; Stich, S. U.; and Suresh, A. T. 2019. Scaffold: Stochastic controlled averaging for on-device federated learning. arXiv preprint arXiv:1910.06378.

Li, T.; Sanjabi, M.; Beirami, A.; and Smith, V. 2020. Fair resource allocation in federated learning.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data.

Nam, J.; Cha, H.; Ahn, S.; Lee, J.; and Shin, J. 2020. Learning from failure: De-biasing classifier from biased classifier.