Adaptive Local Training in Federated Learning

Donald Shenaj[1] Eugene Belilovsky[2,3] Pietro Zanuttigh[1]

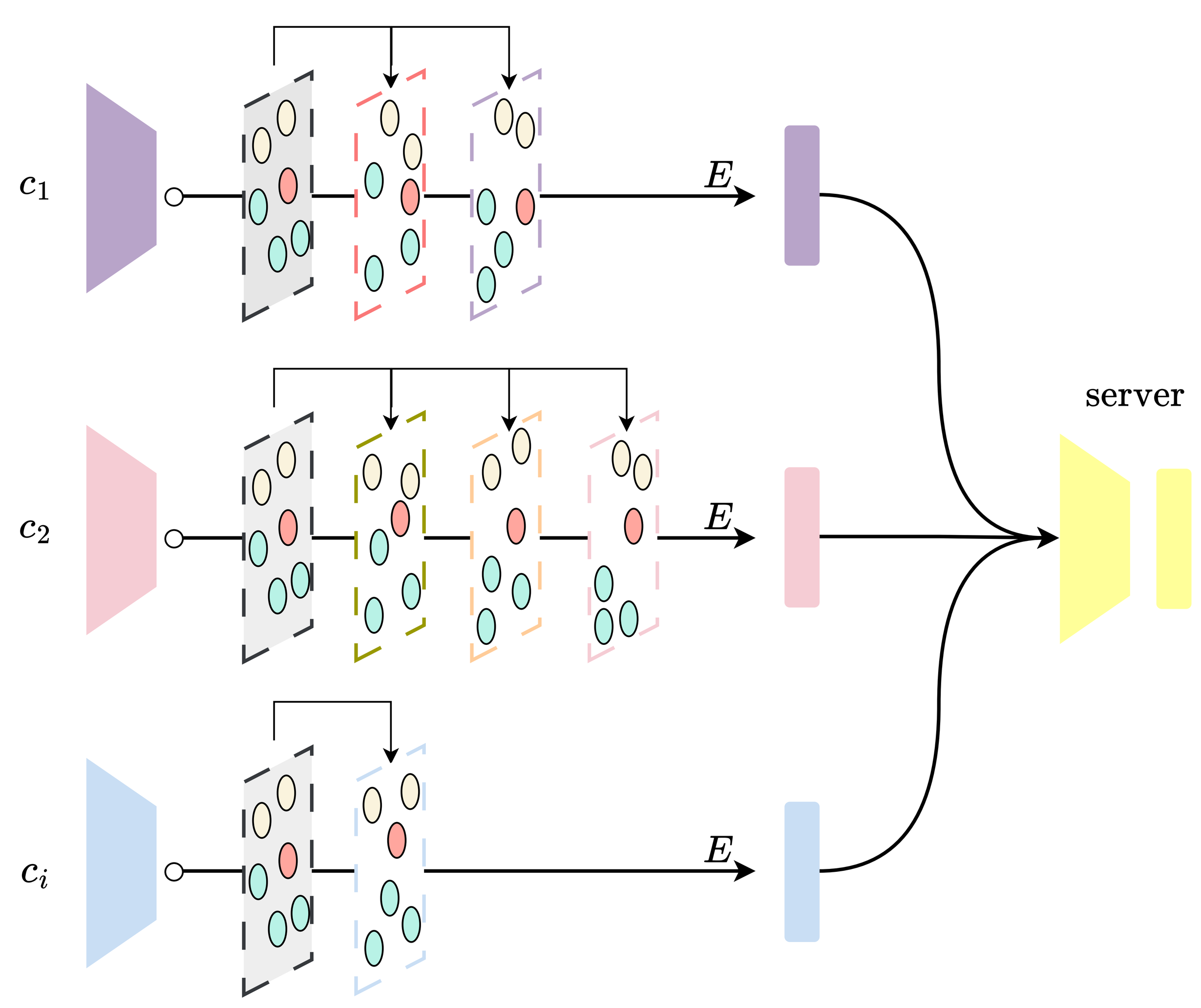
University of Padova[1] Mila[2] Concordia University[3]

Abstract

Federated learning is a machine learning paradigm where multiple clients collaboratively train a global model by exchanging their locally trained model weights instead of raw data. In the standard setting, every client trains the local model for the same number of epochs. We introduce ALT (Adaptive Local Training), a simple yet effective feedback mechanism that can be exploited at the client side to limit unnecessary and degrading computations. ALT dynamically adjusts the number of training epochs for each client based on the similarity between their local representations and the global one, ensuring that well-aligned clients can train longer without experiencing client drift. We evaluated ALT on federated partitions of the CIFAR-10 and Tiny-ImageNet datasets, demonstrating its effectiveness in improving model convergence speed and accuracy.

Dynamic Local Epochs

- An adaptive mechanism allows each client to dynamically determine when to stop training based on the similarity between learned representations: when the similarity is smaller than a threshold, the training halts.
- The threshold function: $T_h(r)$ is dynamic, as it is updated at every communication round r, allowing the system to progressively adjust the stopping criterion over time.
- Therefore, during each communication round, clients execute a different number of local training epochs. As an example, in the Figure below, clients c_1, c_2, c_3 perform respectively 3, 4 and 2 local epochs.



Pseudocode of FedAvg + ALT

Input: Initialize model parameters θ
Initialize maximum rounds R
Initialize threshold parameters a, b

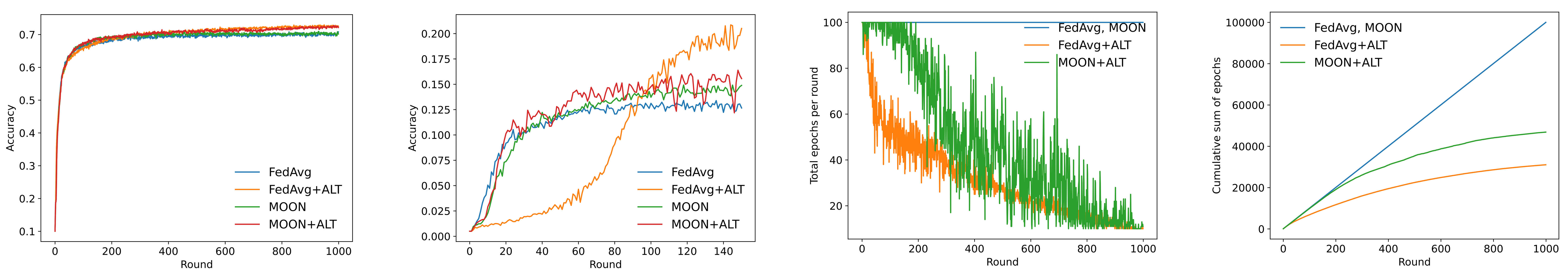
Server executes:
for $r = 1$ **to** R **do**
  $S \leftarrow$ Random subset of clients
  $T_h(r) \leftarrow a + \frac{b*r}{R}$
  for each client $i \in S$ **do**
    $\theta_i \leftarrow$ ClientUpdate$(i, \theta, r, T_h(r))$
  end for
  $\theta \leftarrow \sum_{i \in S} \frac{n_i}{n} \theta_i$
end for

ClientUpdate$(i, \theta, r, T_h(r))$**:**
if $r = 1$:
  $\theta_i \leftarrow \theta$
  $\theta_i := \{w_i, v_i\}$
$\theta_g \leftarrow \theta$
$\theta_g := \{w_g, v_g\}$
stop $\leftarrow False$
for $j = 1, 2, ..., E$ and stop $= False$ **do**
  for each batch $\mathcal{B}$ in $\mathcal{D}_i$ **do**
    $p_i \leftarrow f(w_i, \mathcal{B})$
    $p_g \leftarrow f(w_g, \mathcal{B})$
    if $\cos(p_i, p_g) < T_h(r)$: stop $\leftarrow True$
    $\theta_g \leftarrow \theta_g - \eta \nabla \mathcal{L}(\theta_g; \mathcal{B})$
$\theta_i \leftarrow \theta_g$
return θ_i

Threshold Update

- A key design decision lies in the choice of the threshold $T_h(r)$.
- Empirically, we found that a linearly increasing threshold: $T_h(r) = 0.1 + \frac{0.8 \cdot r}{R}$, yielded the most effective results.
- This formulation enables a "slow start" mechanism:
  - In the early training rounds, clients have greater flexibility in their local updates.
  - As training progresses, stricter similarity constraints are gradually enforced.

Impact of ALT on Main FL Approaches in Terms of Accuracy and Total Computations



(a) Accuracy curve (CIFAR-10)

(b) Accuracy curve (Tiny-ImageNet)

(c) Total epochs per round (CIFAR-10)

(d) Cumulative epochs per round (CIFAR-10)

Conclusion

We introduced ALT, a novel representation learning feedback mechanism designed to dynamically control the number of local epochs in federated learning. By adaptively adjusting local training duration, ALT effectively reduces both energy consumption and communication costs while maintaining model performance. Our approach is seamlessly integrable into existing FL algorithms without requiring significant modifications.

References

[1] HB McMahan, E Moore, D Ramage, S Hampson, B Agüera y Arcas, Communication-efficient learning of deep networks from decentralized data, AISTATS 2017.

[2] Q. Li, B. He, D. Song, Model-Contrastive Federated Learning, CVPR 2021.

Paper

dshenaj@gmail.com