Bayesian Similarity-Weighted Aggregation for Federated Brain Tumor Segmentation

Anonymous Author(s)
Affiliation
Address
email

Abstract

We introduce Bayesian SimAgg, which is a probabilistic model that optimally combines privacy-compliant federated collaborators weights for brain lesion segmentation, adapting to data variability and uncertainty across collaborators.

1 Introduction

Multi-institutional collaboration is crucial for developing generalizable machine learning (ML) models. Federated Learning (FL) addresses this challenge by leveraging distributed computing power and data sources across various institutions (1).

2 Methods

2.1 Data

This research utilized multi-parametric magnetic resonance imaging (mpMRI) data from glioblastoma (GBM) cases, which were made publicly accessible through the Federated Tumor Segmentation (FeTS) 2022 challenge. The dataset included 1251 mpMRI scans from confirmed GBM patients distributed among 33 collaborators.

2.2 Bayesian Similarity Weighted Aggregation

2.3 Collaborator Selection Policy

To select collaborators, we used candidate profiles for each FL round. The selection process employed reinforcement learning (RL) with a multi-armed bandit approach, where each collaborator is viewed as a "bandit." Based on performance history of all collaborators across federation rounds, We used the upper confidence bounds (UCB) strategy. It balances between exploration and exploitation by selecting arms with the highest potential rewards based on their upper confidence bounds.

2.3.1 Weight Aggregation Policy

FedAvg is not suitable for non-IID data because of divergence of model parameters contributed by collaborators. To tackle this challenge, we utilize a weighted aggregation method at the server. Collaborators are assigned weights based on their similarity to the unweighted average.

During round r, the server receives the parameters p_{C^r} contributed by the collaborating entities C^r. Subsequently, the server computes the average of these parameters as follows

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

$$\hat{p} = \frac{1}{C^r} \Sigma_{i \in C^r} p_i. \tag{1}$$

Next, we proceed to determine the inverse distance (similarity) of each collaborator c within C^r from the calculated average

$$sim_c = \frac{\Sigma_{i \in C^r} p_i - \hat{p}}{p_c - \hat{p} + \epsilon}, \tag{2}$$

where $\epsilon = 1e-5$ (small positive constant). We standardize the distances to derive the "similarity weights" in the subsequent manner

$$u_c = \frac{sim_c}{\Sigma_{i \in C^r} sim_i}. \tag{3}$$

Collaborators whose parameters closely align with the average are assigned greater similarity weights, whereas those with more significant deviations receive comparatively lower weights. This methodology can effectively mitigate the influence of outliers or instances of substantial divergence, reducing their impact on the aggregation process.

To accommodate the varying influence of distinct sample sizes across each collaborator c within C^r, we employ "sample size weights" that prioritize collaborators with a greater number of samples.

$$v_c = \frac{N_c}{\Sigma_{i \in C^r} N_i}, \tag{4}$$

where N_c is the number of examples at collaborator c.

Using the weights obtained using Eqs. 3 and 4, the *aggregation weights* are computed as:

$$w_c = \frac{u_c + v_c}{\Sigma_{i \in C^r} (u_i + v_i)}, \tag{5}$$

Ultimately, the aggregation of parameters is done through the harmonic mean of the aggregation weights.

$$p^m = \frac{1}{\sum_{i \in C^r} \frac{w_i}{p_i}} \cdot \Sigma_{i \in C^r} (w_i \cdot p_i). \tag{6}$$

In the following rounds of federation, the normalized aggregated parameters p^m are extended as payout to the subsequent cohorts of collaborators.

Algorithm 1 Harmonic similarity aggregation algorithm

procedure WEIGHT AGGREGATION(C^r, p_{C^r})
  $\epsilon \leftarrow 1e-5$ ▷ C^r = set of collaborators (at round r)
  $\hat{p}$ = average(p_{C^r}) using **Eq. 1** ▷ p_{C^r} = parameters of the collaborators in C^r
  for c in C^r **do**
    Compute similarity weights u_c using **Eqs. 2** and **3**
    Compute sample weights v_c using **Eq. 4**
  for c in C^r **do**
    Compute aggregation weights w_c using **Eq. 5**
  Compute master model parameters p^m using **Eq. 6**
  return p^m

During each round of federation, a total of 318 tensors are processed. Among these, 118 tensors are directly related to the weight and bias of 95 layers of a U-Net model with 33 million parameters.

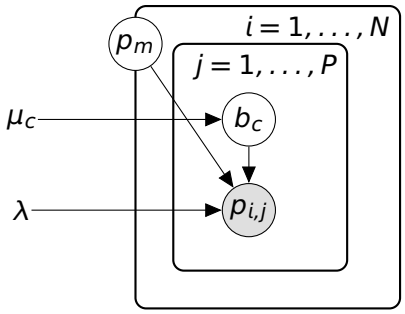


Figure 1: Plate diagram for the Bayesian model.

SimAgg processes 114 tensors using the harmonic mean method, and 4 'out' tensors are processed using the Bayesian method. For the Bayesian model, 4 chains are run, each with 1000 samples. During the warm-up stage, 500 samples from each chain are discarded, and every 2^{nd} sample is kept to reduce autocorrelation. This results in 2000 samples retrieved from Stan for each p_m. The averaged p_m, now referred to as p_c after sampling, is reshaped to match the expected dimensions of the final parameters. Finally, p_c is broadcasted for further rounds. The probabilistic model is formulated as follows:

$$p_c \sim \mathcal{N}(\vec{p_m} \cdot \vec{b_c}, \lambda) \tag{7}$$

$$b_c \sim \mathcal{N}_+(\mu_c, 0.1) \tag{8}$$

$$p_m \sim \mathcal{N}(0, 1) \tag{9}$$

$$\lambda = 1 \tag{10}$$

$$\mu_c = 1 \tag{11}$$

where:

- p_m: Global mean parameter for the tensors.
- b_c: Bias for each tensor.

The likelihood is modeled as:

$$p[i, j] \sim \mathcal{N}(p_m[j] + b_c[i], \lambda) \tag{12}$$

modeling each element of the tensors as a normal distribution with mean $p_m[j] + b_c[i]$ and variance λ.

The Bayesian step is designed to aggregate tensors using a probabilistic model, specifically leveraging Bayesian statistics. This approach allows for a more flexible and robust combination of the tensors, taking into account both the mean and the variability of the tensors. The R-hat statistic values close to 1 indicate good mixing of chains and convergence guarantees, along with the effective number of draws showing sufficient draws. Figure. 1 shows the plate diagram of the Bayesian model.

3 Deep Learning Experiments

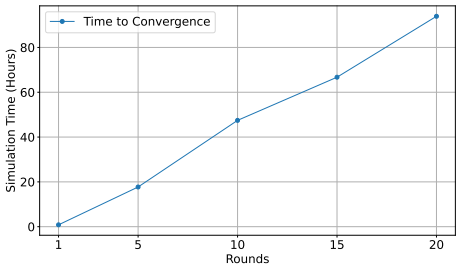
3.1 Training Setup

The experimental framework employed a 3D U-Net neural network using Intel's OpenFL platform. The performance evaluation was based on DICE similarity and Hausdorff (95%) distance metrics (2).

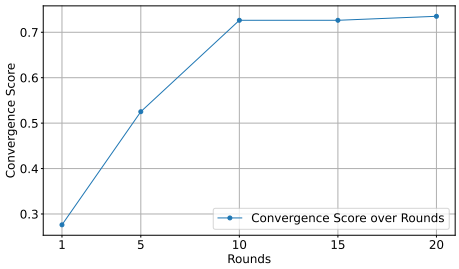
3.2 Results

3.2.1 Model training and performance using internal validation data

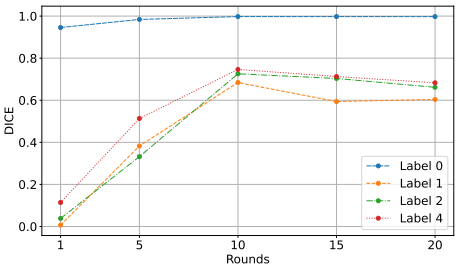
We evaluated the performance of our Bayesian SimAgg approach over 20 rounds of federated model training. Figure 2 illustrates the training performance on internal validation data, tracking simulated time, convergence score, and DICE scores.



(a) Simulation Time



(b) Projected Convergence Score



(c) DICE similarity

Figure 2: Performance metrics for model training of Bayesian SimAgg. The horizontal axis refers to the number of rounds and the vertical axis to the metrics.

Table 1: Comparison of Bayesian SimAgg, SimAgg, and RegAgg methods.

Metric	Bayesian+UCB	SimAgg	RegAgg
Simulation Time (hours)	47.45	78.14	78.13
Projected Convergence Score	0.7264	0.7273	0.7227
DICE Label 0	0.9977	0.9978	0.9980
DICE Label 1	0.6844	0.6657	0.6561
DICE Label 2	0.7257	0.6430	0.6665
DICE Label 4	0.7464	0.7603	0.7313

4 Discussion

This study highlights the efficacy of incorporating Bayesian inference, SimAgg strategies and UCB collaborator selection process into federated tumor segmentation. Our study showed that incorporating prior knowledge and stochastically aggregating the weights from collaborators leads to robust tumor or lesion segmentation in a federated setting.

5 Conclusion

Acknowledgements

References

[1] Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al.: Advances and open problems in federated learning. Foundations and Trends® in Machine Learning **14**(1–2), 1–210 (2021)

[2] Pati, S., Baid, U., Zenk, M., Edwards, B., Sheller, M., Reina, G.A., Foley, P., Gruzdev, A., Martin, J., Albarqouni, S., et al.: The federated tumor segmentation (fets) challenge. arXiv preprint arXiv:2105.05874 (2021)