Bayesian Neural Network For Personalized Federated Learning Parameter Selection

Mengen Luo* and Ercan E. Kuruoglu*
*Tsinghua-Berkley Shenzhen Institute, Institute of Data and Information,
Shenzhen International Graduate School, Tsinghua University, 518055, China

***Abstract*—Federated learning's poor performance in the presence of heterogeneous data remains one of the most pressing issues in the field. Personalized federated learning strives to discover an individualized model for each client to address the heterogeneity in the data. One of such approach involves personalizing specific layers of neural networks. However, previous endeavors have not provided a dependable rationale, and some have selected personalized layers that are entirely distinct and conflicting. In this work, we take a step further by proposing personalization at the elemental level, rather than traditional layer-level personalization. To select personalized parameters, we introduce Bayesian neural networks and rely on the uncertainty they offer to guide our selection of personalized parameters. Finally, we validate our algorithm's efficacy on several datasets and neural networks architecture, demonstrating that our proposed approach outperforms existing baselines.**

***Index Terms*—Federated Learning, Bayesian Neural Network, Distributed Learning**

I. Introduction

To facilitate collaborative learning while safeguarding client privacy, federated learning (FL) has emerged. Due to its simplicity and high communication efficiency, FedAvg [1] is the most popular algorithm in FL. However, recent experiments have revealed that FedAvg performance experiences a sharp decline [2] in heterogeneous data.

To address this issue, Personalized Federated Learning (PFL) seeks to provide each client with a personalized model to excel in their local data context. Given the complexity of real-world data and the diverse situations faced by individual clients, PFL aligns more closely with reality. Algorithms such as FedPer [3] and LG-FedAVG [4] personalize the model's feature extractor and classifier layers. These two algorithms have divergent opinions on which layers should be personalized, emphasizing the urgent need for an interpretable personalized parameter selection scheme.

To this end, We offer a more interpretable scheme for the automated selection of personalized parameters. We propose a simple yet effective approach that leverages the uncertainty of Bayesian neural networks (BNN) to select personalized parameters. Parameters with higher uncertainty indicate a greater scope for improvement, with changes in these parameters exerting minimal impact on the outcome compared to others.

Through experiments carried out on various benchmark datasets, we have demonstrated the superiority of our method over FedPer and LG-FedAvg. Furthermore, we conducted experiments tailored to diverse datasets and network architectures to ascertain their impact on the algorithm. This underscores the rationality of our approach to personalized parameters selection, which outperforms previous layer-based personalization techniques.

Corresponding author: Ercan E. Kuruoglu. This work is supported by Shenzhen Science and Technology Innovation Commission under Grant JCYJ20220530143002005, Shenzhen Ubiquitous Data Enabling Key Lab under Grant ZDSYS20220527171406015, and Tsinghua Shenzhen International Graduate School Start-up fund under Grant QD2022024C.

II. Proposed Method

In this section, we will discuss the method we propose, namely Federated BNN for Parameter Selection (FedBPS).

A. Bayesian Neural Network

BNN treat parameters as distributions rather than single fixed values. we employ the Laplace approximation method [5] to estimate posterior distribution rather than Variational Inference (VI) and Markov Chain Monte Carlo (MCMC). Because we do not require precise posterior results but rather their relative magnitudes. Laplace approximation utilizes second-order derivative information at a specific point to approximate the distribution as a Gaussian distribution, with the distribution depicted as follows.

$$w \sim \mathcal{N}(w_0, (\nabla_w^2 \mathcal{L}(\mathcal{D}; w)|_{w_0})^{-1}) \tag{1}$$

In our case, we assume no correlation between elements; hence, the posterior of parameters will follow a diagonal Gaussian distribution.

B. BNN For Personalized Parameters Selection

When a parameter exhibits a high variance, it implies that the parameter's value is uncertain, and changes in this parameter have a smaller impact on model performance compared to other parameters. Therefore, such parameters are well-suited for use as personalized parameters, as they can accommodate personalized tasks without significantly affecting the global subnetwork performance.

$$w_i = w_i \odot M + w_g \odot (1 - M), \tag{2}$$

where M indicates whether the parameter is personalized or not. In PFL, each client has their own personalized model. Consequently, selecting personalized parameters can lead to

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00057

TABLE I
THE FINAL TEST ACCURACY ON DIFFERENT DATASETS UNDER NON-IID SETTINGS.

Method	MNIST	FMNIST	CIFAR10
FedAvg	**98.49**	87.28	71.53
FedPer	96.48	92.09	75.27
LG-FedAvg	96.44	91.92	76.52
FedBPS(ours)	96.87	**92.51**	**76.80**

TABLE II
THE IMPACT OF DIFFERENT DATASETS ON HYPERPARAMETERS IN A HETEROGENEOUS DATA SETTING.

Proportion	MNIST	FMNIST	CIFAR10
30%	95.44	91.13	75.27
50%	**97.74**	**93.14**	75.72
70%	96.87	92.52	**76.80**
90%	96.23	92.19	76.23

disparities. To address this, we will first aggregate the global parameter distribution $\mathcal{N}(\mu_g, \sigma_g)$ and then determine which parameters should undergo personalization by using σ_g. We opt for an aggregation scheme based on KL divergence.

$$\mu_g = \sum_{i=1}^{N} \pi_i \mu_i, \ \sigma_g = \sum_{i=1}^{N} \pi_i \sigma_i + \pi_i(\mu_i - \mu)^2 \tag{3}$$

From the formula, it is evident that the variance of the aggregated model is determined by the mean of client model variances and the variance of that mean. Consequently, the global model comprises components with low variance and minimal discrepancies among clients, while personalized parameters exhibit high variance or substantial inter-client disparities.

III. EXPERIMENTS

A. Setup

Datasets And Partition We evaluate our algorithm and the baseline on three popular datasets: MNIST, Fashion-MNIST, and CIFAR-10. To simulate each client's heterogeneous dataset, we follow the approach of [6]. Each client possesses major and minor classes.

Baselines We will compare the following three baselines. FedAvg, FedPer and LG-FedAvg. Our method, FedBPS, using 70% of the parameters as personalized parameters.

Implementation details For each method, we will adopt common settings for local training. We will use SGD with a learning rate of $\eta = 0.01$, a batch size of $B = 128$. In FL, the local epoch will be set to $E = 5$. The total communication rounds will be set to $T = 60$ for MNIST and Fashion-MNIST, and $T = 100$ for CIFAR-10. For the neural networks, we will employ LeNet-5 for MNIST and Fashion-MNIST. For CIFAR-10, we will broaden the channels of the first and second convolutional layers of LeNet to 16 and 32, respectively. To investigate the effect of network depth on hyperparameters, we will conduct experiments using ResNet architectures of varying depths on the CIFAR-10 dataset.

TABLE III
THE IMPACT OF THE DEPTH OF RESNET ON HYPERPARAMETERS IN A HETEROGENEOUS DATA SETTING.

	ResNet20	ResNet32	ResNet44	ResNet56
30%	72.13	72.47	72.17	71.75
50%	73.21	72.96	73.19	73.34
70%	74.92	74.25	73.94	73.73
90%	75.08	74.96	74.49	74.60

B. Numerical Results

We conducted experiments on three distinct datasets, and it is evident from the results in Table I that our method outperforms the baseline in most scenarios. Our algorithm has demonstrated superior performance on more complex datasets.

The proportion of personalized parameters has a profound impact on the effectiveness of the algorithm. Here, we continue to assess the influence of hyperparameters on experimental results across these datasets. We conducted experiments with personalization proportions of $30\%, 50\%, 70\%, 90\%$, and the final results are presented in Table II. It should be noted that the optimal personalization proportion varies depending on the dataset and model used. These results can only suggest that the optimal proportion may fall between 50% and 90%.

For ResNet architectures of different depths, the result are presented in Table III. The recommended personalization proportion falls within the range of 70% to 100%, which differs from the earlier LeNet architecture. However, it is evident that the depth of network with similar architecture has minimal influence on the choice of personalization proportion.

IV. CONCLUSION

In this paper, we introduced an approach for selecting personalized parameters to enable PFL, leveraging the uncertainty obtained from BNN for parameter selection. Our method offers enhanced flexibility, allowing for the selection of personalized parameters at a finer granularity, and the algorithm's performance demonstrates insensitivity to the introduction of additional hyperparameters.

REFERENCES

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.

[2] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018.

[3] Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary, "Federated learning with personalization layers," *arXiv preprint arXiv:1912.00818*, 2019.

[4] Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B Allen, Randy P Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency, "Think locally, act globally: Federated learning with local and global representations," *arXiv preprint arXiv:2001.01523*, 2020.

[5] Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig, "Laplace redux-effortless bayesian deep learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20089–20103, 2021.

[6] Jian Xu, Xinyi Tong, and Shao-Lun Huang, "Personalized federated learning with feature alignment and classifier collaboration," *International Conference on Learning Representations*, 2023.