

Experimental Comparison of Clustering Approaches for Personalized Federated Learning

Seohee Choi, Minjung Park, and Sangmi Chai(✉)

Ewha Womans University, Seoul, Republic of Korea
seohee.choi@ewhain.net, smchai@ewha.ac.kr

Abstract. Federated learning allows clients to collaboratively train a global model while preserving data privacy. However, personalized the global model is necessary for each device to prevent performance degradation caused by the heterogeneity of each client's local data. In this paper, we discuss previously studied client clustering techniques for personalized federated learning. We also evaluate the performance of personalized models generated by various federated learning and client clustering algorithms. Our research aims to bridge the research gap by evaluating the effectiveness of personalized models when applied to users.

Keywords: Clustering · Federated Learning · Personalized Federated Learning

1 Introduction

Federated learning is a method for training a unified global model in a distributed system without transferring data belonging to each client. It saves the computing resources of edge devices to train local models without requiring direct access to client-owned data, enabling model training while preserving data privacy. It has become increasingly important with the introduction of data privacy laws such as General Data Protection Regulation (GDPR).

In federated learning, heterogeneous local data distributions compared to the distributional level of the global model can lead to poor prediction performance. In a typical federated learning approach, local models trained on edge devices with highly distributed and diverse data sets are averaged to fit a global model. However, optimizing only the accuracy of the global model can lead to a degradation in personalization capabilities [1], resulting in significant performance degradation on end-user devices. It means that prediction accuracy may suffer if there is insufficient data available for the outcome being predicted during the global model training. This problem is similar to the sampling bias that occurs during the training of machine learning models. When training a model, if the number of samples between classes of data becomes too large, the recall for classes with fewer samples will be smaller, causing the model to overfit to certain classes. However, solving the sampling bias problem in a federated learning process is challenging as the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14112, pp. 187–193, 2023.
https://doi.org/10.1007/978-3-031-37129-5_16

class of data is not easy to determine. This is due to the fact that training participants cannot be identified and the client's data is difficult to access. To overcome this challenge, previous research in federated learning has suggested various techniques such as using metadata, applying noise to data, or applying contrastive learning for sampling the data without direct access to it. Due to the existence of heterogeneous learning environments and federated learning settings tailored to the purpose of various models, there is a need for research that can be personalized for different federated learning models.

Personalization is a relevant approach to address non-IID challenges and statistical heterogeneity present in each client. However, many studies only evaluate the accuracy of global models. In this paper, we aim to mitigate this research gap from the perspective of client clustering techniques, which is an intermediate approach between a single global model and local models [2] in federated learning personalization. Clustering is a similarity-based approach that considers group-level client relationships to model client relationships [3]. It provides a trade-off between generalization and distribution mismatch caused by heterogeneity in the data distribution across clients. This technique is simple to interpret and implement, especially when the number of clients is very large, and it can help improve non-IID handling.

2 Related Works

2.1 Federated Learning

The Federated averaging (FedAvg) algorithm is one of the most well-known federated learning algorithms [4]. FedAvg aggregates local model weights which are sent from clients to generate a global model. However, when the datasets produced by different clients are non-IID, the FedAvg algorithm achieve accuracy and system performance degradation [5]. Specifically, the non-IID setting can cause the global model to become biased towards the data of certain clients, resulting in poor generalization to other clients. As a result, various extensions to the FedAvg algorithm has been proposed to address the non-IID problem. For instance, Li et al. [6] demonstrated that strong convexity in client functions can improve the convergence rate in FedAvg. The authors argued that the divergence issue of averaging could be mitigated by not involving all clients in each round. Instead, clients could be selected with a certain probability or with a probability proportional to the size of their local dataset. Furthermore, in Federated Optimization in Heterogeneous Networks (FedProx) [7], to address the issue of local models deviating significantly from the global optima in federated learning, FedProx introduced a regularization term into the local objective function of the FedAvg algorithm. This term adjusts for the impact of local updates. This led to better convergence on heterogeneous data by restricting the impact of local updates on the global model. However, FedProx assigns the same weight to all devices in the aggregation step, similar to FedAvg, and does not account for device heterogeneity [8]. MOON [9] aimed to tackle the problem of weight deviation in non-IID data environments, which results in a widening of the representation gap between local and global models, by performing contrastive learning at the model level. By comparing the representations of different models, MOON accelerates convergence and maximizes the agreement between the local model learned by the client and the representation learned by the global model. SCAFFOLD [5] proposed

a framework that utilizes additional control variables to facilitate the communication between the central server and participating clients, as well as estimate the update direction for both the global and local models. FedProx, MOON, and SCAFFOLD all aim to reduce the variance by introducing a function that compensates for local updates.

2.2 Personalized Federated Learning

The degree to which data is distributed differs for every client that possesses it, and the same is true for the devices of end-users who will be using the federated model. To mitigate the performance degradation caused by device heterogeneity, statistical heterogeneity, and data heterogeneity, which are inevitable during the federated learning process, personalization is necessary at the device, model, and data levels. Traditional machine learning approaches that can contribute to personalization by addressing data imbalances are not easily applicable to a federated learning mechanism that does not have direct access to local data owned by the client [10]. Therefore, additional processing is required to incorporate these techniques into the personalization process of federated learning.

Previous research [1] has suggested that for personalization in federated learning to be useful, three goals must be achieved simultaneously: (1) an improved personalization model for the majority of customers, (2) a robust initial model for customers with limited or no data for personalization, and (3) a fast convergence rate that allows high quality to be reached with fewer training rounds. Also, In [3], metrics were proposed to evaluate personalized federated learning. They consider three dimensions: (1) model performance; (2) system performance; and (3) trustworthy AI.

Google Research [2] has proposed three approaches to personalization: (1) user clustering – which involves grouping users into clusters and training a separate model for each group; (2) data interpolation – where some local data is combined with global data and used to train a model with the interpolated data combination; and (3) model interpolation – which first trains a global model and then trains a local model separately, combining each model afterwards. [2] states that personalized models serve as an intermediate step between purely local and global models. In [3], personalization techniques for federated learning are classified into two categories: Global Model Personalization and Learning Personalized Models. Previous research has attempted to personalize the global model, which closely replicates the training procedure of general federated learning for a single model. Since the generalization performance of the global model directly impacts the personalized model's performance, they optimized the global model to enhance subsequent personalization performance [3]. This involves additional training on each client's local dataset before utilizing the trained global model in the inference phase [10]. Representative algorithms included in this approach are Meta-learning. Meta-learning is a personalization technique for federated learning, which uses a small number of examples to train a model and then fine-tunes it for new tasks, allowing it to learn quickly. One meta-learning method, Model-Agnostic Meta-Learning (MAML) [11], has been used in personalization research. MAML has similarities between the personalization goals of MAML and federated learning (FL), and FedAvg can also be interpreted as a meta-learning algorithm [1]. And the latter approach, referred to as Learning Personalized Models, focuses on training models that are personalized to each individual

client, instead of optimizing the global model for personalization. This trains a personalized model around the solution for each user's prediction, instead of training a single global model for personalization. Included in this approach are Multi-task learning and Client-clustering. Multi-task learning [12] is a learning paradigm used in various fields of machine learning, where multiple tasks are jointly learned to avoid the data sparsity problem when each task has less labeled data, and this knowledge can also be utilized by other tasks. From this research, we discuss clustering techniques that classify clients based on their similarities or preferences to improve accuracy by training personalized models.

2.3 Client Clustering Techniques

Clustering techniques refer to clustering and sampling that have been previously proposed in the machine learning. However, applying these algorithms to federated learning requires a lot of considerations due to the heterogeneity of the federated training environment. Client clustering in federated learning is a technique that assumes a common dataset among participating clients and their locally trained models. Client clustering techniques group different local data and the local models trained on them, i.e., clients, according to some characteristic or tendency. A group of users is a subset of the entire client group, and they are trained on the same task within the group. The idea is to learn the features of the data more efficiently by grouping similar users together, and increase the accuracy of global models trained on non-IID data.

While federated learning generally has the advantage that the identity of each participant is not disclosed and their data is protected, it also assumes significant challenges to personalization using clustering: (1) the identity of participating clients is unknown and they must be clustered without access to their data; (2) the classification criteria of clients and the subset to which they belong are disclosed, which can pose a security threat. In addition, if learning is required through additional communication, communication overhead can be a problem, and optimization of the learning process should be considered.

2.4 Federated Learning Personalization Based on Client Clustering

Client clustering techniques have been proposed as a solution to address the challenge of data heterogeneity in the personalization phase. This method enables efficient personalization by organizing de-identified data into a clustered structure.

In [13], the Iterative Federated Clustering Algorithm (IFCA) was proposed, which employs alternating minimization to estimate the identity of a cluster and minimize the loss function via gradient descent. The algorithm broadcasts the server model parameters to k clients and collect the model parameters after local learning. The central server performs gradient descent updates to the model parameters of clients with the same cluster identity estimate. This process is repeated several times to progressively cluster. In [14], a Stochastic Expectation Maximization (SEM) technique was proposed, which adds a step of dynamically updating W_i to Expectation-Maximization (EM) algorithm to design a multi-center federated loss. In this method, each client finds the global model closest to its learned value and proceeds with clustering. In order to use these methods,

it is necessary to determine the appropriate number of clusters. In [15], a clustering structure was inferred based on the cosine similarity between clients' gradient updates, and clients were recursively separated in a top-down manner based on this clustering. This recursion stops when all clients are not far from a fixed value of the federated learning solution. Many of the previously proposed client clustering algorithms suffer from the drawback that clustering takes place over multiple rounds of communication, which can result in high communication costs. Therefore, in [16], the authors simplified the clustering process into a single step and trained a global model during a single round of communication. Clusters were generated based on the difference between the parameters of the global model and those of the local models.

3 Experiments

3.1 Experimental Setup

Datasets. We evaluate client clustering algorithms and compare their results using the MNIST [17] dataset and Federated Extended MNIST (FEMNIST), a publicly available federated benchmark dataset introduced in LEAF [18]. The MNIST dataset is a collection of handwritten digit vision images, and it is widely used to evaluate various federated learning environments. FEMNIST is suitable for testing more challenging and realistic environments, as it contains 26 handwritten 10-digit numbers, upper-and lower-case letters, and the image set is partitioned by writer. To construct datasets for clients, we select the same number of writers as the number of clients in both datasets. To implement the pathological non-IID setting in the MNIST dataset, we sort the data by label and partition the dataset so that a client has data for only two labels.

FL and PFL Algorithms. We use a pre-trained global model trained over 2000 iterations for each partitioned dataset. The global model is trained with FedAvg algorithm. A total of 100 clients participated in the training and the learning rate was 0.008. The final test accuracy was 0.82, and the final test loss was 0.63. We compare the performance of the IFCA [13], SEM [14], and CFL [15] algorithms by adjusting the percentage of clients participating in each training round, the number of clusters, and the number of communication rounds for clustering. We report the average test set accuracy for all clients, from immediately after the global model is created until the clustering task is complete, and the percentage of clients that reach the test set accuracy target at the end of the targeted communication rounds.

3.2 Experiments

We first used the IFCA algorithm to train a PFL model with client clustering. We experimented with fixing the number of clusters to $k = 3$. Also, existing research shows that the accuracy is not significantly different when $k < 3$. The experiment was run in the following environment: Batch size = 10, Participation clients in every round = 15, Learning rates = 0.03, Epoch = 1, Initial training rounds for client clustering = 30 (Figs. 1, 2 and 3).

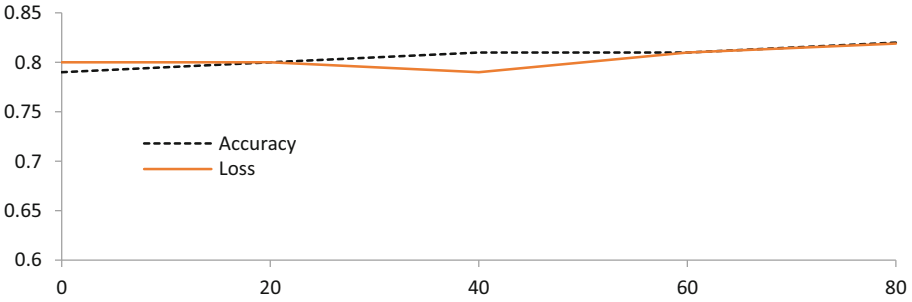

Fig. 1. Accuracy of the IFCA PFL model. FEMNIST datasets. Clients = 15.

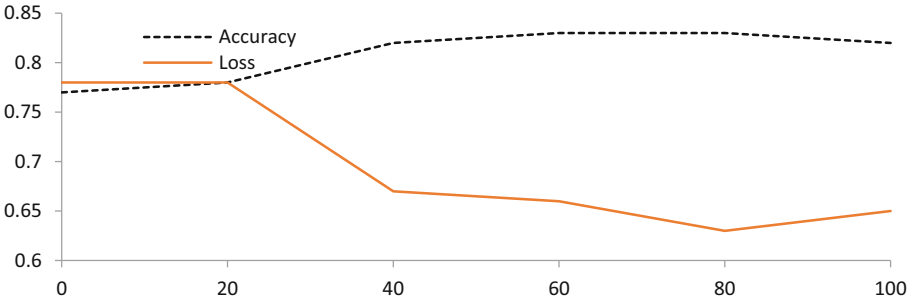

Fig. 2. Accuracy of the IFCA PFL model. FEMNIST datasets. Clients = 50

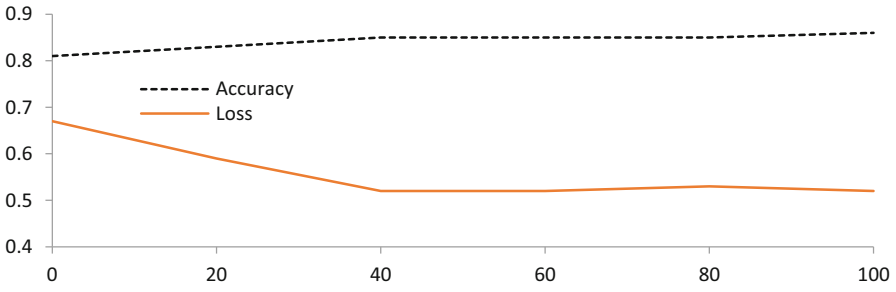

Fig. 3. Accuracy of the IFCA PFL model. FEMNIST datasets. Clients = 100

4 Conclusion

To measure the performance of the personalized federated models trained by client clustering methods, we personalized the global models generated by federated learning algorithms used in previous studies. In this personalization process, we adopted multiple client clustering methods and compared the outcomes in various experimental environments. Based on our evaluations of different client clustering techniques using MNIST and FEMNIST datasets, we present a comprehensive comparison of the performance and scalability of IFCA, SEM, and CFL algorithms in the context of personalization in

federated learning. We believe that this study can help researchers and practitioners to select appropriate client clustering methods for their federated learning personalization tasks, with leading to improve its performance.

References

1. Jiang, Y., Konečný, J., Rush, K., Kannan, S.: Improving Federated Learning Personalization via Model Agnostic Meta Learning (2019). arXiv:1909.12488
2. Mansour, Y., Mohri, M., Ro, J., Theertha Suresh, A.: Three Approaches for Personalization with Applications to Federated Learning (2020). arXiv:2002.10619
3. Tan, A.Z., Yu, H., Cui, L., Yang, Q.: Towards personalized federated learning. IEEE Trans. Neural Netw. Learn. Syst. 1–17 (2022)
4. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282. PMLR (2017)
5. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: Scaffold: stochastic controlled averaging for federated learning. In: International Conference on Machine Learning, pp. 5132–5143. PMLR (2020)
6. Li, X., Huang, K., Yang, W., Wang, S., Zhang, Z.: On the convergence of FedAvg on non-IID data (2019). arXiv preprint arXiv:1907.02189
7. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proc. Mach. Learn. Syst. **2**, 429–450 (2020)
8. Wahab, O.A., Mourad, A., Otrok, H., Taleb, T.: Federated machine learning: survey, multi-level classification, desirable criteria and future directions in communication and networking systems. IEEE Commun. Surveys Tutor. **23**, 1342–1397 (2021)
9. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10713–10722 (2021)
10. Kairouz, P., et al.: Advances and open problems in federated learning. Found. Trends® Mach. Learn. **14**, 1–210 (2021)
11. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning, pp. 1126–1135. PMLR (2017)
12. Zhang, Y., Yang, Q.: A survey on multi-task learning. IEEE Trans. Knowl. Data Eng. **34**, 5586–5609 (2022)
13. Ghosh, A., Chung, J., Yin, D., Ramchandran, K.: An efficient framework for clustered federated learning. Adv. Neural. Inf. Process. Syst. **33**, 19586–19597 (2020)
14. Long, G., Xie, M., Shen, T., Zhou, T., Wang, X., Jiang, J.: Multi-center federated learning: clients clustering for better personalization. World Wide Web **26**, 481–500 (2023)
15. Sattler, F., Müller, K.R., Samek, W.: Clustered federated learning: model-agnostic distributed multitask optimization under privacy constraints. IEEE Trans. Neural Netw. Learn. Syst. **32**, 3710–3722 (2021)
16. Briggs, C., Fan, Z., Andras, P.: Federated learning with hierarchical clustering of local updates to improve training on non-IID data. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–9. IEEE (2020)
17. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**, 2278–2324 (1998)
18. Caldas, S., et al.: Leaf: a benchmark for federated settings (2018). arXiv preprint arXiv:1812.01097