2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) | 978-1-6654-0004-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CISP-BMEI53629.2021.9624434

FedACS: an Efficient Federated Learning Method Among Multiple Medical Institutions with Adaptive Client Sampling

Yunchao Gu, Quanquan Hu, Xinliang Wang , Zhong Zhou
School of Computer Science
Beihang University
Beijing, China

Sixu Lu
School of Artificial Intelligence
Beijing Normal University
Beijing, China

***Abstract*—With the development of deep learning, the neural network model trained by massive data is widely used in various fields, which improves production and living efficiency. However, the construction of a large open-source dataset is difficult due to commercial or privacy reasons, which limits the performance of the deep learning model. In this paper, we focus on medical image analysis and explore the possibility of federated training for medical image classification in different hospitals. We propose an adaptive client sampling algorithm, which creatively applies the curriculum learning strategy to federated learning. Our proposed method can effectively reduce the communication overhead in federated learning, and provide technical support for deep learning training in cross-institutional and non-data sharing scenarios. Comparative experiments on CIFAR-10, CIFAR-100, and a chest X-Ray classification dataset show the effectiveness of the proposed algorithm.**

***Index Terms*—Deep learning; federated learning; curriculum learning; medical image analysis; classification**

I. Introduction

With the development of deep learning, a large number of related technologies have been applied to medical image processing tasks. Computer-aided diagnosis can reduce the burden of medical practitioners. Medical imaging, which covers X-ray, CT, MRI, ultrasound, etc., is an interdisciplinary science integrating mathematics, computer science, information processing and other disciplines. Modern information processing tools and deep learning technology can be used to process images produced by different medical imaging equipment according to actual needs. However, due to privacy and legal reasons, medical institutions cannot share data directly. It is difficult to set up large-scale datasets such as ImageNet in the medical field, and there is no effective cooperation between medical institutions. At the same time, the limited amount of data owned by a single institution limits the performance of the deep learning model.

Federated learning is a way to solve the above problems, which means that multiple clients (such as mobile devices or the whole organization) cooperatively train the machine learning settings of the model under a central server (such as service providers). This setting also ensures the decentralization of training data. The term federated learning was first proposed by McMahan in 2016 [1], [2], [2], [3]. But before that, there has been a lot of related research work devoted to data privacy protection. For example, the encryption method of computing encrypted data appeared in 1980s. At first, federated learning only emphasized the application of mobile and edge devices, and researchers called these two settings cross-device and cross-silo respectively. However, due to the need to transmit the model and the calculated vectors, the communication overhead in the process of federated learning is often large.

In this paper, federated learning technology is adopted to achieve cooperation among medical institutions through aggregation of model parameters, thus solving the problem that data is difficult to share. Based on Federated Averaging Algorithm (FedAvg) [1], an improved adaptive equipment selection algorithm is proposed, which is shown in Fig. 1. Applying the curriculum learning strategy to the equipment

Corresponding author: Xinliang Wang, Email address: wangxinliang@buaa.edu.cn.

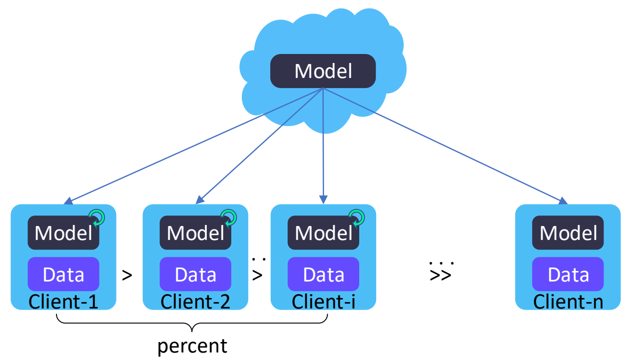

Fig. 1. Our proposed FedACS applies curriculum learning to federated learning. The scoring function discriminates the difficulty of the whole data on the client and obtains the equipment difficulty ranking. The designed pacing function makes the equipment selection ratio increase with global round, giving priority to the equipment with higher data difficulty. This method can effectively reduce the communication overhead of equipment, accelerate the convergence of models, and complete the deep learning and efficient training of non-data sharing scenes.

selection process in federated learning can effectively reduce the communication overhead of FedAvg and reduce the frequent communication between institutions and servers. Our contributions mainly include the following aspects:

- To our best knowledge, the method we proposed is the first method to apply curriculum learning strategies to federated learning. In this paper, a communication-efficient device selection algorithm, FedAvg with adaptive client sampling (FedACS) is designed.
- We apply FedACS to medical image processing, simulate the federated training of many institutions, and solve the problem that it is difficult for a single medical institution to train the model effectively.
- Experimental comparisons on CIFAR-10, CIFAR-100, and a lung image dataset show that FedACS can effectively reduce the communication cost of federated training, and provides a technical basis for cross-device and cross-institutional training.

II. Related Work

A. Federated Learning

Some previous studies obtained pre-training models on datasets like ImageNet and then fine-tuned them on medical image datasets. This transfer learning method can make the model have better performance by using only a small amount of medical image data. However, this scheme can't share the data information between hospitals and related institutions, and can only carry out relevant analysis independently on data islands. In recent years, distributed decentralized federated learning technology has gradually emerged, and has been preliminarily applied in banks, hospitals, and other inter-institutions, which are generally difficult to collect data due to privacy protection, business competition, and other reasons. Federated learning can realize collaborative and decentralized neural network training without sharing patient or user data [4], [5], [6]. In federated learning, each institution is regarded as an independent node, and each node is responsible for training its own local model and submitting it to the parameter server regularly. The server continuously accumulates and aggregates the contributions of each node, and then creates a global model, which is shared with all nodes.

B. Federated Learning in Medical Images Analysis

Sheller et al. [7] applied federated learning to medical image semantic segmentation and compared it with two other cross-institutional training methods. In addition, it analyzed the influence of some hyperparameters of federated learning (such as EpR, number of institutions) on convergence, which showed the feasibility and effectiveness of federated learning. Although federated learning can guarantee high privacy security, data can still be reproduced through model inversion. To help improve the security, Wenqi Li et al. [8] studied and tested the feasibility of using ε-differential privacy framework. Beaulieu-Jones et al. [6] have done similar work, using differential privacy and periodic weight transmission to enhance the security of clinical data federated learning. In addition to differential privacy to enhance the security, there are some works that use the method of adding noise. Chang et al. [9] used the federated learning method to classify two datasets (retinal fundus images and breast X-ray images) in four simulation institutions, and added noise by adjusting the resolution and class imbalance.

Algorithm 1 FedAvg with adaptive client sampling, FedACS

Input: initial global model w_0
Output: trained model W

Server(w_0):
  for each global round r **do**
    p = pacing func(r)
    r_k = (the rank list of K clients)
    m = max($K \times p$, 1)
    s_r = (the top m of r_k)
    for each client $k \in s_r$ in parallel **do**
      w^k_{r+1} = ClientUpdate(k, w_r)
      $w_{r+1} = \sum_{k=1}^{K} \frac{n_k}{n} w^k_{r+1}$
  return W

ClientUpdate(k,w):
  B = (split n_k into batches of size b)
  for each local epoch i **do**
    for batch $b \in B$ **do**
      w = w - η $\nabla l(w, b)$
  return w

Most of the previous work simply applied federated learning technology to the medical image processing field to strengthen the cooperation between medical structures, to achieve the model effect that a single institution could not achieve. However, few related studies have solved some inherent problems in federated optimization at the same time, such as not independent and identically distributed data of different institutions, communication efficiency, system heterogeneity, and so on.

III. Methodology

A. Federated Learning Baseline

In this study, cross-device method is employed to complete the federated learning of multi-source medical data. We distribute open-source datasets to different clients to simulate different data collection centers. The original data of each client is stored locally and cannot be exchanged or migrated. Federated learning uses local updating to complete model training, and the central server accumulates and aggregates their respective contributions, and then creates a global model, which is shared with all nodes.

FedAvg [1] and its related improved algorithms [10], [11], [12], [13] are popular algorithms used in federated learning, which are robust to class imbalance and non-independent data distribution. FedAvg randomly selects a certain proportion of clients from a set of clients with a fixed number of k to participate in the training. Clients selected in each global round

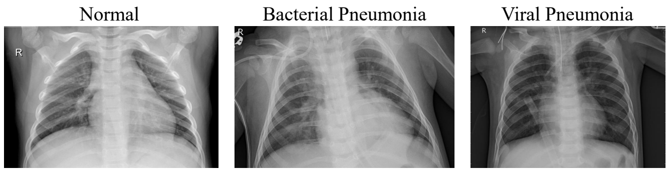

Fig. 2. The used chest X-Ray (Pneumonia) dataset with three classes: normal, bacterial pneumonia and viral pneumonia.

send model updates to the server for aggregation, in which the aggregation weight of different equipment models is set as the proportion of their data, n_k/n, where n is the number of total samples and n_k is the amount of data distributed to the k-th device. This is repeated for several rounds to converge. The optimization function is defined as:

$$f(w) = \frac{1}{n} \sum_{k=0}^{K} \sum_{i=1}^{n_k} f_i(w), \tag{1}$$

where w is the parameters of the model.

B. Curriculum Learning

Curriculum learning was proposed by Bengio et al. [14] that inspired by the natural learning process of human beings. This method makes the model come into contact with samples with higher complexity in the training process, similar to human learning more complex knowledge and concepts gradually. Such a strategy of training samples in a specific order can make the model converge faster than random order training, which can reduce the communication overhead of client-server interaction in federated learning. In the practical use of curriculum learning, it is necessary to judge the difficulty of data by certain methods, and how to sort the difficulty of data is the key and difficult point of applied curriculum learning [15], [16], [17].

In curriculum Learning, a scoring function is needed to sort the data difficulty [18], and a pacing planning function is needed to control the growth of the data exposed by the model from less to more.

C. FedACS

In this paper, the idea of curriculum learning is applied to federated learning, and an adaptive device selection algorithm is designed, which can effectively reduce the communication overhead in the process of federated training. The method of controlling the number of devices selected in different rounds of federated training and the method of calculating data difficulty and prioritizing data in different rounds are explored. We find that different training rounds in federated learning have different contributions to the convergence of the model. The more equipment selected in the later rounds, the better the final effect of the model and the stronger the robustness of the model.

We use the pacing function in the curriculum learning to control the growth of the data quantity that the model comes into contact with from less to more. In the training process,

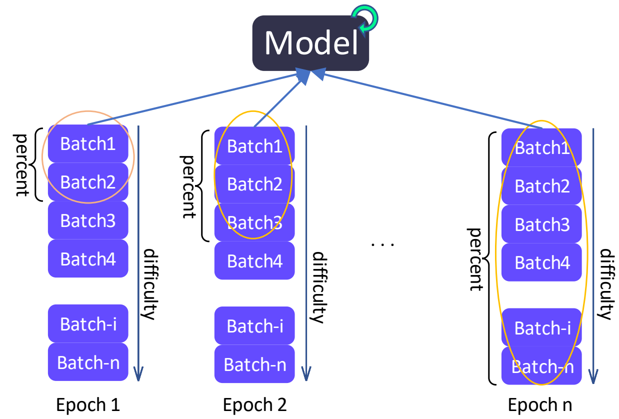

Fig. 3. Distributed curriculum learning with batch-level control granularity.

assuming that the data quantity selected by the i-th global round is p, the data quantity selected by the $i+1$-th global round is defined as:

$$p_{i+1} = p_i + \lambda \times r, \tag{2}$$

where λ is a hyperparameter and r is the number of current global round.

The scoring function designed by us utilizes the global model. In each global round of federated training, the equipment needs to send the loss of local data on the global model to the server, and the server can rank all the equipment on the data difficulty.

The specific algorithm flow is shown in Algorithm 1, where η is the learning rate and $\nabla l(x)$ is the loss function. We determine the training data amount to be selected by the global round through the pacing function. All devices test the average loss of local data on the global model and send it to the server. Then, the service completes the ranking of equipment according to the loss. In this process, the equipment with high loss or data difficulty in the current global model will be preferred, because the data on these equipment often contains more information, which is more helpful to accelerate the convergence of the global model. The global model is continuously learned and updated in the whole process, and the data on each device is not always in the global model loss ranking, so our client sampling strategy is adaptive.

IV. Experiments

In this section, we take image classification as the proxy task and conduct extensive experiments to verify the effectiveness of our proposed FedACS. The datasets we used include CIFAR-10, CIFAR-100, and chest X-Ray (Pneumonia) [19]. CIFAR-10 contains 60,000 images of 10 categories; CIFAR-100 contains 60,000 images of 100 categories. We use the chest X-Ray dataset to simulate the scene of multiple medical institutions. This dataset contains 5,863 images in three categories, as shown in Fig. 2. In all experiments, we employ ResNet [20] as the feature extractor and use the accuracy as

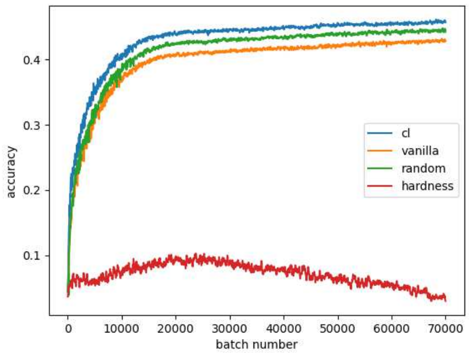

Fig. 4. The performance of distributed curriculum learning based on batch-level strategy on CIFAR-100 dataset. cl: distributed batch-level curriculum learning; vanilla: federated learning; random: only pacing function works; hardness: only scoring function works. None that the batch size on all devices is same.

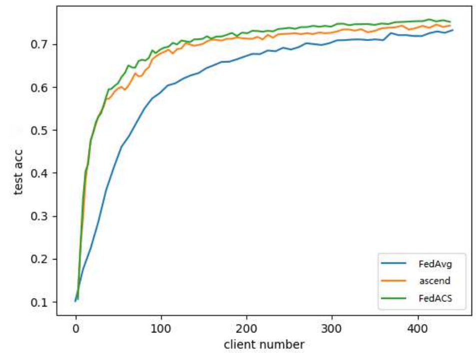

Fig. 5. Curriculum learning based on client-level control granularity. FedACS has achieved better performance. The ascend method is to select equipment from less to more, but it is randomly selected, which reflects the role of scoring function in FedACS.

the evaluation metric that can directly reflect the performance of the model.

A. Distributed Curriculum Learning (DCL)

In this section, we apply the conventional curriculum learning methods to federated learning, and verify the functions of pacing function and scoring function through comparative experiments. The conventional curriculum learning method trains batch composed of data samples, from less to more in quantity and from easy to difficult in difficulty. In the federated learning scenario, using this strategy to train and plan devices is equivalent to treating the data of one device as a relatively large batch, and the number of selected devices increases with the global communication round.

In the experiment, we keep the same amount of data on each device. As shown in the Fig. 3, after sorting the difficulty, each device selects the data with less difficulty in each global round to form batch to participate in the training. The value of percent increases with global round , and the percentage of selected data on each device is the same and increases synchronously.

As shown in Fig. 4, in the curriculum learning of distributed batch-level, we can see that the curriculum learning strategy brings convergence acceleration. The random method does not sort the data with difficulty, but only makes the data samples contacted by the model change from less to more. The hardness method gradually trains the model on the more difficult data batch in the form of a sliding window after sorting the data with difficulty. In the experiment, this method does not converge, which indicates that the scoring function has a negative impact on the convergence of the model when used alone (we will show its joint effect with pacing function in subsequent experiments), and the vanilla method is conventional federated learning.

B. Curriculum Learning in Federated Learning

In the federated learning scene, the distributed curriculum learning method increases the communication overhead (scoring function transmission, equipment data batch difficulty vector) compared with the conventional curriculum learning. We consider making adaptive modifications to the curriculum learning in the federated learning scene. The granularity of curriculum learning control is raised to client-level so that we can choose the equipment participating in federated training according to the curriculum learning strategy, which is expected to reduce the communication overhead in the federated learning scenario.

In this section, we realize the client-level curriculum learning in the federated learning scenario, and verify the improvement of communication efficiency through experiments. Distributed curriculum learning is batch-level. In the federated learning scenario, we modify it to client-level and judge the difficulty of the whole data on the client to get the equipment difficulty ranking. The most important thing is to select the equipment according to the cl strategy. Compared with conventional curriculum learning, when the same accuracy is achieved, the number of devices involved in the curriculum learning of client-level is less, and the information interaction with the server is also less, so compared with FedAvg random selection, the communication overhead can be significantly reduced, and the comparison results are shown in the TABLE I.

As shown in Fig. 5, FedACS has achieved the best performance in convergence speed and final accuracy. In addition, the ascend scheme is to select more devices from less, but the devices are randomly selected, which reflects the role of the scoring function in FedACS.

In Fig. 6, the ascend scheme is to select equipment from less to more, but randomly; The cl scheme is to give priority to the equipment with low data loss; It can be seen that our proposed FedACS performs best in convergence speed and accuracy. In addition, the histogram shows the optimized ratio of communication efficiency, and FedACS can effectively

TABLE I
COMPARISON OF OUR PROPOSED FEDACS AND DCL (DISTRIBUTED CURRICULUM LEARNING) IN TERMS OF CONTROL GRANULARITY, STRATEGY AND COMMUNICATION EFFICIENCY.

Method	Control Granularity	Strategy	Communication Efficiency
DCL	Batch-level	With the increase of global round, the batch number and difficulty of feeding model increase	Reduce
FedACS	Client-level	With the increase of global round, the number of client participating in aggregation increases and the difficulty increases	Promote

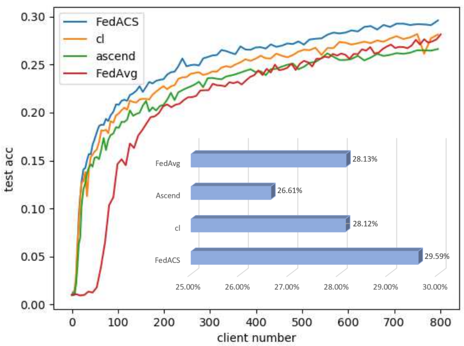

Fig. 6. Comparison of accuracy and communication efficiency on the CIFAR-100 dataset.

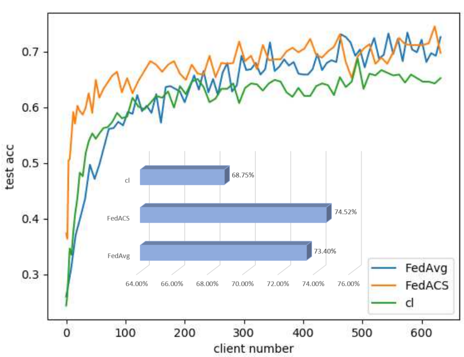

Fig. 7. Comparison of accuracy and communication efficiency on the pneumonia classification dataset.

reduce communication consumption.

Due to privacy, law, and other issues, many medical institutions cannot concentrate data for training. We simulated the federated training among many institutions through experiments, and the results are shown in Fig. 7. Experiments show that our algorithm can significantly reduce the communication overhead in federated training. Under the condition of the same device-server communication times, our method can achieve better performance.

V. CONCLUSION

In the medical field, due to the law, privacy, and other reasons, it is difficult to collect data, thus forming a data island, and various medical institutions cannot exchange information effectively. We simulated the data interaction and model training between hospitals by using federated learning, so as to open the connection between data islands. As far as we know, we are the first to apply curriculum learning strategies to federated learning. In this paper, we propose a client sampling algorithm, which can effectively reduce the communication times between the server and the client, accelerate the convergence speed of the model, and achieve better accuracy. This method can provide help for deep learning training in cross-institutional and data sharing scenarios. In future work, we will verify this method in more detail, including medical image lesion segmentation and detection.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.
[2] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.
[3] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging," *arXiv preprint arXiv:1602.05629*, 2016.
[4] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 311–320.
[5] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1310–1321.
[6] B. K. Beaulieu-Jones, W. Yuan, S. G. Finlayson, and Z. S. Wu, "Privacy-preserving distributed deep learning for clinical data," *arXiv preprint arXiv:1812.01484*, 2018.
[7] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 92–104.
[8] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso *et al.*, "Privacy-preserving federated brain tumour segmentation," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 133–141.
[9] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. L. Rubin, and J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 945–954, 2018.
[10] M. Mohri, G. Sivek, and A. T. Suresh, "Agnostic federated learning," in *International Conference on Machine Learning*. PMLR, 2019, pp. 4615–4625.
[11] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," *arXiv preprint arXiv:1905.10497*, 2019.
[12] T.-M. H. Hsu, H. Qi, and M. Brown, "Measuring the effects of non-identical data distribution for federated visual classification," *arXiv preprint arXiv:1909.06335*, 2019.
[13] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konečnỳ, S. Kumar, and H. B. McMahan, "Adaptive federated optimization," *arXiv preprint arXiv:2003.00295*, 2020.
[14] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 41–48.

[15] J. Wei, A. Suriawinata, B. Ren, X. Liu, M. Lisovsky, L. Vaickus, C. Brown, M. Baker, M. Nasir-Moin, N. Tomita *et al.*, "Learn like a pathologist: Curriculum learning by annotator agreement for histopathology image classification," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2473–2483.

[16] B. Korbar, D. Tran, and L. Torresani, "Cooperative learning of audio and video models from self-supervised synchronization," *arXiv preprint arXiv:1807.00230*, 2018.

[17] D. Weinshall, G. Cohen, and D. Amir, "Curriculum learning by transfer learning: Theory and experiments with deep networks," in *International Conference on Machine Learning*. PMLR, 2018, pp. 5238–5246.

[18] G. Hacohen and D. Weinshall, "On the power of curriculum learning in training deep networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 2535–2544.

[19] "Chest x-ray images (pneumonia)." https://www.kaggle.com/paultimothy-mooney/chest-xray-pneumonia.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.