FEDERATED FORGETTING-RESISTANT REPRESENTATION LEARNING (APPENDIX)

Author(s) Name(s)

Address

1. THE MEMORY OF THE FEDERATED MODEL

In this section, we observe the memory of the federated model M_g on the MIXED dataset (see section 3). Specifically, P participants perform 100 rounds of unsupervised optimization on the M_g based on the SimSiam [1] algorithm. In the first 50 rounds, each participant randomly selects 1024 samples from the Fashion-MNIST dataset [2] to optimize M_g and simulate the learned historical knowledge. Participants randomly select 1024 samples from the MIXED dataset in each subsequent round to continue optimizing M_g. We evaluate the model accuracy on the Fashion-MNIST test set at the end of each round.

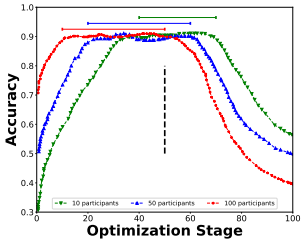

Fig. 1. *Accuracy evaluation of the federated model on the Fashion-MNIST.* The green, blue, and red denote 10, 50, and 100 participants. The vertical dotted line represents the beginning of using the MIXED dataset to optimize the model. A horizontal line identifies the highest accuracy interval, i.e., the optimal memory of the federated model.

Figure 1 depicts the accuracy of the federated model during the above optimization process. In the first 50 rounds of optimization, the accuracy has continuously improved to a peak of about 90%, and the increase in the number of participants has accelerated the performance improvement. When using the MIXED dataset to continue optimizing the model, the performance shows a downward trend, which means that the new data distribution harms the historical memory of the model. The more participants, the faster the performance decreases. Further, we replace the above image classification model with a generative adversarial network (GAN) [3] model and use the same federated training strategy to optimize the GAN. As shown in Figure 2, the scattered points represent the 2D projection (using *t-SNE* algorithm [4]) of the Fashion-MNIST training samples, and the shaded areas represent the 2D projection coverage of the generated samples from the GAN. In the first 50 optimization rounds (subgraph a to c), the GAN attempts to fit the distribution of Fashion-MNIST samples. When using the MIXED dataset to continue optimizing the GAN, the ability of GAN to fit Fashion-MNIST decreases sharply (subgraph d), which means that the model begins to forget. Figure 2 visually depicts the model from learning to forgetting and also shows that if the participant data distribution is dynamically changing, continuous optimization under the federated paradigm will inevitably lead to forgetting the historical knowledge.

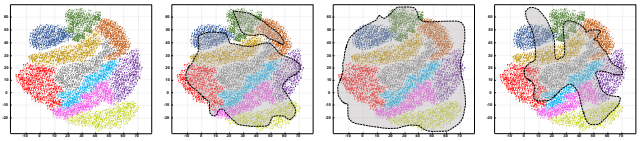

Fig. 2. *2D projection of the Fashion-MNIST samples and the generated samples from the GAN.* Different colors distinguish different categories.

2. METHODOLOGY

Our approach FedFRR can be summarized with Algorithm 1.

3. EXPERIMENTS

Datasets. **MIXED:** We mix the SVHN [5], Not-MNIST [6], Fashion-MNIST [7], FaceScrub [8], and TrafficSigns [9] to create the MIXED dataset, which consists of 280K images of 293 classes. **CMNIST:** We inject color into the digit in the MNIST dataset [10], resulting in the Colored-MNIST dataset with the combined label [*digit*, *color*], which identify the samples' essential and bias feature. **CCIFAR10:** Different textures (e.g., Gaussian Blur, Fog, and Snow) are injected into the CIFAR10 dataset [11] to create the Corrupted-CIFAR10 dataset. **FFHQ:** FFHQ is a high-quality image dataset of human faces consisting of 70K PNG images at 1024×1024 resolution and contains considerable variation in age, ethnicity, and image background. **MiniImageNet:** We

Thanks to XYZ agency for funding.

Algorithm 1 *FedFRR*

Input: P: participant number, S: stage number, C: stage capacity, T: weight truncated participant number, A: augmented sample number in each stage, $\Gamma(\cdot)$: transform function, α: forgetting penalty coefficient, η: learning rate, $\mathcal{L}$: loss function. **Output**: federated model $\mathcal{M}(\theta^S)$.

Initializing the federated model $\mathcal{M}(\theta^1)$, random $\lambda \sim Beta(8, 8)$.

for $s = 1, ..., S$ **do**
  $\{\theta_p^s\}_{p=1}^P \leftarrow \theta^s$, distribute model weights to all participants
  for $p = 1, ..., P$ **do**
    $\tilde{x}_j^i = (1 - BM_\lambda) \odot x_i + \Gamma(BM_\lambda \odot x_j), i, j \in [1..., C], i \neq j$, semantic augmented samples
    contrastive optimize θ_p^s with $\{x_i, x_j, \tilde{x}_j^i\}$ (Formula 2)
  $\theta^{s+1} \leftarrow \theta^s + \sum_{p=1}^P \Phi\left(\theta_p^s\right)$, weight truncation aggregation (Formula 3)
return θ^S

extract 120K 84x84 images in 200 classes from the ImageNet [12, 13] dataset to create the MiniImageNet dataset, which is divided into a training set of 100K samples and a test set of 20K samples.

Baselines. Our baselines consist of the supervised group (HLE [14], DER [15]), unsupervised group (SimSiam [1], RELIC [16]), and federated group (FedSimCLR [17], FedCA [18], FedWeIT [19]). The supervised group denotes the traditional centralized "training and deployment" mode, optimizing the model through limited supervision data. HLE is a rehearsal-based method that utilizes hierarchy-aware pseudo-labeling to incorporate hierarchical class information. DER is a general CL method that combines rehearsal and knowledge distillation to solve the problem of gradual or sudden changes in distribution. Similar to the supervised group, the unsupervised group uses methods such as contrastive learning to optimize the model. SimSiam is one of the state-of-the-art representational learning techniques for learning unsupervised representation. RELIC is a self-supervised representation learning method that utilizes explicit causal invariance constraints to augment the pre-training. The federated unsupervised group denotes the model optimization based on unsupervised methods in federated paradigm. FedSimCLR is a SimCLR implemented in the federated paradigm. FedCA is a federated contrastive averaging algorithm with a dictionary and alignment. FedWeIT is a federated continual learning (FCL) framework that decomposes the network weights into global federated parameters and sparse task-specific parameters, making it one of the best-performing algorithms in the FCL field.

Implementation Details: We use ResNet [20] and multi-layer CNN [21] for the PNU, the PNU number U = 10, participant number P = 20, the stage capacity C = 100, and the number of PNU with truncated weight T = 0.1U. In each stage, the augmented sample number A = 0.5C. The forgetting penalty coefficient α = 0.5, and the learning rate η=0.01. As shown in Figure 3, the layout of PNUs includes *SI*: all PNUs have the same structure. *MS*: The model's first and second half of PNUs are implemented in two different structures. *MA*: two different structures alternately implement all PNUs. The metric *Forgetting* $= MAX\{ACC - ACC', 0\}$, *ACC* is the model accuracy after pre-training, and *ACC'* is the real-time accuracy in different stages.

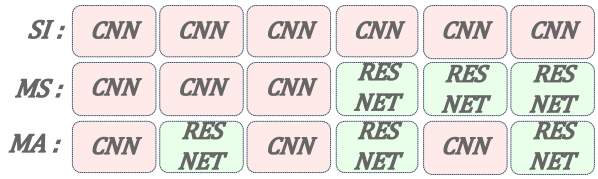

Fig. 3. *The layout of PNUs in the federated model.*

4. REFERENCES

[1] Xinlei Chen and Kaiming He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15750–15758.

[2] Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *CoRR*, vol. abs/1708.07747, 2017.

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[4] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[5] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, and Bo Wu, "Reading digits in natural images with unsupervised feature learning," 2011.

[6] Y Bulatov, "Not-mnist dataset," 2011.

[7] Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[8] Hong-Wei Ng and Stefan Winkler, "A data-driven approach to cleaning large face datasets," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 343–347.

[9] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel, "The german traffic sign recognition benchmark: a multi-class classification competition," in

The 2011 international joint conference on neural networks. IEEE, 2011, pp. 1453–1460.

[10] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] Alex Krizhevsky, Geoffrey Hinton, et al., "Learning multiple layers of features from tiny images," 2009.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[13] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al., "Matching networks for one shot learning," *Advances in neural information processing systems*, vol. 29, 2016.

[14] Byung Hyun Lee, Okchul Jung, Jonghyun Choi, and Se Young Chun, "Online continual learning on hierarchical label expansion," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[15] Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara, "Dark experience for general continual learning: a strong, simple baseline," *Advances in neural information processing systems*, vol. 33, pp. 15920–15930, 2020.

[16] Jovana Mitrovic, Brian McWilliams, Jacob C Walker, and Buesing, "Representation learning via invariant causal mechanisms," in *International Conference on Learning Representations*, 2020.

[17] Ting Chen, Simon Kornblith, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[18] Fengda Zhang, Kun Kuang, Zhaoyang You, Tao Shen, Jun Xiao, Yin Zhang, Chao Wu, Yueting Zhuang, and Xiaolin Li, "Federated unsupervised representation learning," *arXiv preprint arXiv:2010.08982*, 2020.

[19] Jaehong Yoon, Wonyong Jeong, and Sung Ju Hwang, "Federated continual learning with weighted inter-client transfer," in *International Conference on Machine Learning*. PMLR, 2021, pp. 12073–12086.

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

[21] Alex Krizhevsky and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, Eds. 2012, vol. 25, Curran Associates, Inc.