Group-Personalized Federated Learning for Human Activity Recognition

Ahmed A. Al-Saedi[1] and Veselka Boeva[1]

***Abstract*— Human Activity Recognition (HAR) plays a significant role in various fields including health care and well-being. Traditional centralized methods reach very high recognition rates, but they incur privacy and scalability issues. The emergence of Federated learning (FL) enables users to train a global model on distributed data in a privacy-preserving manner. However, for HAR task scenarios, the existing systems mainly focus on a unified model that cannot provide users with personalized recognition of activities. In this study, we propose a novel group-personalized FL (GP-FL) algorithm that dynamically trains several global machine learning (ML) models. The performance of the algorithm is studied and evaluated on public HAR data.**

I. PROBLEM DESCRIPTION AND MOTIVATION

Traditional ML solutions require sending a massive amount of data to a central server and training there in a centralized way. However, this introduces huge communication overhead, consumes network resources, and brings privacy concerns [1]. To solve this problem, Google proposes a FL approach, where model parameters instead of data are transferred between the central server and edge nodes (called workers hereafter) [2]. Basically, there are two main ways to model FL: a central framework where learning a single global model among a different number of workers and a personalized model for each worker to take a completely local approach. To find a trade-off between the two extreme cases, we propose grouping workers based on the similarity of their empirical class probabilities. When updating the global model, only local updates uploaded by the workers within the same group will be aggregated. In that way, our proposed GP-FL algorithm [3] is capable of training simultaneously several global models, one per each group of workers with similar activity patterns. The performance of our GP-FL algorithm is benchmarked to that of two other FL algorithms, FedAvg [2] and CFL [4], on public HAR data. FedAvg trains a single global model across all workers. Our proposed GP-FL is similar to the CFL method, which trains a set of global models, one per cluster of workers.

II. RELATED WORK AND NOVELTY

In this section, we explore some existing approaches in personalized FL in HAR related to our work. For example, in [5] the authors use local sensitivity hashing for calculating the similarity between different users in order to select a subset of the top-k most similar users for training. A novel hybrid approach for HAR that combines semi-supervised and FL is suggested in [6]. In [7] a method that relies on a semi-supervised gradient aggregation method for activity detection is introduced. In [8], the authors have proposed the FedStack framework, which supports ensemble heterogeneous architectural client models for mobile health sensor datasets. In [9], the authors have presented a novel FL framework according to the similarity of the local updates for HAR. Clustered FL (CFL) [4], which is closely related to our study, proposes hierarchical clustering to form client clusters and those in the same cluster share the same model for training.

Notice that most of the above mentioned works are aimed at the personalized training of deep learning models in a FL setup. In contrast, we introduce a lightweight model based on logistic regression that is more suitable for modern resource-constrained wearable devices.

III. PROPOSAL OUTLINE

In this section, we formally present our group-personalized FL (GP-FL) algorithm [3]. We propose to group the available workers according to their empirical class probability distributions. The workers with similar empirical probabilities are group together into the same cluster based on their similarity measured by Wasserstein distance. In addition, the built grouping is not static, but it is dynamically updated at each training round by applying cluster eccentricity analysis [10]. This approach allows to build a global model at the cluster level, overcoming the issue of personalization in traditional FL techniques. GP-FL algorithm consists of the following two phases:

Initialization Phase:

1) At time $t=0$, the Server initializes the model $\mathcal{M}_0$, set of workers W_t, and number of iterations T.
2) The Server transmits the initial global model $\mathcal{M}_t$ to the subset of workers W_t ($W_t \subset W$).
3) Each $w_i \in W_t$ receives the global model $\mathcal{M}_t$ and produces updated parameters, i.e. $\mathcal{M}_t^i$, alongside with a empirical probability vector $\hat{p}_t(w_i)$. These are sent back to the Server.
4) The Server performs the following operations:
   a) Laplace smoothing is applied to vector $\hat{p}_t(w_i)$ of each worker $w_i \in W_t$.
   b) The smoothed vectors $\hat{p}_t(w_i)$, for $w_i \in W_t$, are used to create a distance matrix. It is then passed as a parameter in a Markov clustering algorithm. Groups of workers with similar empirical probability vectors are produced, i.e. $C_t = \{C_{t1}, C_{t2}, ..., C_{tk}\}$.
   c) For each $C_{tj} \in C_t$, $(j = 1,2,...,k)$, a global group model $\mathcal{M}_t^j$, is built by averaging over the model parameters of the workers assigned to C_{tj}.
   d) For each cluster $C_{tj} \in C_t$ mean data vector μ_i^j and aggregated variance σ_i^j are also calculated.
5) The Server aggregates the parameters $\{\mathcal{M}_t^i \mid w_i \in W_t\}$

[1]Ahmed A. Al-Saedi and Veselka Boeva are with Department of Computer Science, Blekinge Institute of Technology, E-371 79 Karlskrona, Sweden

uploaded by the selected workers W_t to update the overall global model $\mathcal{M}_t$ through the FedAvg.

Iteration Phase:

1) The Server sends each group global model $\mathcal{M}_t^j$, ($j=1,2,...,k$) to its group of workers C_{tj}.
2) Each $w_i \in C_{tj}$ receives $\mathcal{M}_t^j$ and optimizes its parameters, i.e. $\mathcal{M}_{t+1}^i$ local update and the empirical probability vector $\hat{p}_{t+1}(w_i)$ are produced.
3) The Server updates the existing empirical probability vector $\hat{p}_{t+1}(w_i)$ by taking the average of it with the vector from the previous data batch, i.e. $\hat{p}_t(w_i)$.
4) The Server applies Laplace smoothing to each vector $\hat{p}_{t+1}(w_i)$, for $i=1,2,...,|W_t|$.
5) The Server adapts the grouping C_t to the current vectors $\hat{p}_{t+1}(w_i)$, for $i=1,2,...,|W_t|$, by invoking eccentricity score $\xi^j(w_i)$, $(j = 1, 2, \ldots, k)$ assessing whether each $w_i \in C_{tj}$ is still well tight with its current cluster.
   a) If $\xi^j(w_i)$ is below the threshold $\upsilon(t)$ the worker does not change its cluster C_{tj}.
   b) If $\xi^j(w_i) > \upsilon(t)$ then we calculate $\xi^l(w_i)$ for each cluster $C_{tl} \in C_t \setminus C_{tj}$, and will assign the worker w_i to cluster C_{tl} for each $\xi^l(w_i) < \upsilon(t)$. If this is true for more than one cluster the worker is assigned to the cluster with the lowest score.
   c) If $\xi^j(w_i) > \upsilon(t)$ for all the clusters in $C_t \setminus C_{tk}$ then this worker w_i will give the start of a new singleton cluster. Note that $k_{(t+1)} \geq k_t$, where $k_{(t+1)} = |C_{t+1}|$.
6) For each cluster $C_{t+1j} \in C_{t+1}$, mean data vector μ_i^j and aggregated variance σ_i^j are calculated, considering the current grouping of the workers and also using the current empirical probability vectors $\hat{p}_{t+1}(w_i)$, for $i=1,2,...,|W_{t+1}|$.
7) The updated C_{t+1} is produced, and the clusters in C_{t+1} may contain different workers from ones in C_t.

Steps 1–7 of the *iteration* phase are repeated until a certain number of training rounds T is reached.

Evaluation and Preliminary Results: We have compared the performance of workers' personal (local) models with that of both the built federated learning (global) model and global group models trained by our GP-FL algorithm. For each experiment, the three models (local, global and group) associated with each worker are run on its test data at each round. Overall, the group global models built by the GP-FL algorithm have produced accuracy scores that are higher or at least compatible to those generated by the global model. We also compare the performance of our GP-FL algorithm with that of FedAvg and CFL algorithms. In Figure 1, we consider the first 20 global rounds for the three algorithms and display the accuracy averaged over all the workers versus the number of communication rounds for Non-IID label skew data (30%) for two datasets. Within 10 communication rounds in case of REALWORLD dataset, CFL and FedAvg reach 87%, and 85% accuracy, respectively, while our GP-FL algorithm achieves an accuracy of 89%. In case of HHAR Non-IID data, the GP-FL algorithm has obtained an accuracy of 94% in 8 communication rounds, while CFL and FedAvg have reached 93% and 92%, respectively. The proposed GP-FL algorithm demonstrates the best performance on all benchmarks across all datasets.

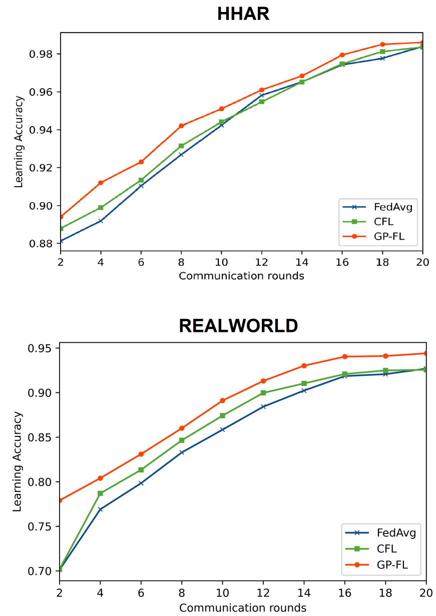

Fig. 1. Comparison of the achieved accuracy versus the number of communication rounds for Non-IID data (30%) of the three FL algorithms: FedAvg, CFL and GP-FL on HHAR (top) and REALWORLD (bottom) datasets, respectively.

IV. CONCLUSIONS

In this paper, we have discussed a new approach for building a set of group personalized FL models in case of Non-IID data, proposed in [3]. The performance of our proposed GP-FL algorithm has been evaluated and compared with that of two other baseline FL algorithms, FedAvg and CFL, on public HAR data. The GP-FL has outperformed the both algorithms in the conducted experiments with respect to the achieved accuracy. Our future plans include further evaluation the GP-FL algorithm properties and performance in other applied FL scenarios.

REFERENCES

[1] Q. Zheng et al. Research on hierarchical response recovery method of distribution network fault based on topology analysis. *Int. J. Crit. Infrastructures*, 17:216–236, 2021.
[2] H. B. McMahan et al. Communication-efficient learning of deep networks from decentralized data. *ArXiv*, 1602.05629, 2016.
[3] Ahmed A. Al-Saedi and Veselka Boeva. Group-personalized federated learning for human activity recognition through cluster eccentricity analysis. In *24th EANN/EAAAI 2023*, 2023. accepted.
[4] F. Sattler et al. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Trans. on Neural Networks and Learning Systems*, 32(8):3710–3722, 2021.
[5] S. Liu et al. Federated personalized random forest for human activity recognition. *Math. Biosciences & Engineering*, 19(1):953–971, 2022.
[6] C. Bettini et al. Personalized semi-supervised federated learning for human activity recognition. *ArXiv*, abs/2104.08094, 2021.
[7] H. Yu et al. Fedhar: Semi-supervised online learning for personalized federated human activity recognition. *IEEE T. on Mob. Comp.*, 2021.
[8] T. B. Shaik et al. Fedstack: Personalized activity monitoring using stacked federated learning. *Knowl. Based Syst.*, 257:109929, 2022.
[9] R. Presotto et al. Fedclar: Federated clustering for personalized sensor-based human activity recognition. In *2022 IEEE PerCom*, pages 227–236, 2022.
[10] P. Angelov. Anomaly detection based on eccentricity analysis. In *2014 IEEE Symposium on EALS*, pages 1–8, 2014.