Supplementary Materials: Spatio-temporal Heterogeneous Federated Learning for Time Series Classification with Multi-view Orthogonal Training

Anonymous Authors

1 EXPERIMENTAL DETAILS

1.1 Datasets Details

We provide the introduction and statistics of the adopted benchmark datasets:

- **HAR** [5]: The Human Activity Recognition database was constructed from recordings of 30 volunteers performing activities of daily living while carrying waist-mounted smartphones with embedded inertial sensors. The raw data captures 3-axis linear acceleration and 3-axis angular velocity at a constant rate of 50Hz. There are 6 types of activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying down.
- **HHAR** [13]: This dataset consists of 3-axis accelerometer measurements from 30 participants. Measurements were captured at 50 Hz. Non-overlapping segments of 128 time steps were used for classification purposes. The dataset includes six activity labels: biking, sitting, standing, walking, going upstairs, and going downstairs.
- **WISDM** [9]: WISDM is an open dataset widely used for human motion recognition and behavior analysis from 30 paticipants. The sensor measures the acceleration values of the phone in three axes (x, y, z) and samples them at fixed time intervals. The data is mesured at 20 Hz. The dataset contains six activities: walking, jogging, sitting, standing, walking upstairs, and walking downstairs.
- **Sleep-EDF** [6]: The Sleep-EDF dataset consists of single-channel EEG signals sampled at 100Hz. Sleep recording covers five different sleep stages, namely awake (W), non-rapid eye movement (N1, N2, N3) and rapid eye movement (REM).
- **Epilepsy** [2]: Epilepsy is an epileptic seizure recognition dataset. It includes EEG signals from 500 subjects. we follow the processing of previous letearets [4, 17] for classification, dividing the entire dataset into two classes.

Table 1: Summary of datasets used in the experiments.

	HAR	HHAR	WISDM	Sleep-EDF	Epilepsy
# Train	7352	12716	1350	35503	9200
# Test	2947	5218	720	6805	2300
Length	128	128	128	3000	178
# Subjects	30	9	30	20	500
Channel	9	3	3	1	1
# Class	6	6	6	5	2

1.2 Baseline Details

We provide the detailed introduction and hyper-parameters for the four groups of baselines we adopted:

- **FedAvg** [12]: Fedavg stands as a benchmark for all FL algorithms with local update and global aggregation phases. This typical FL algorithm does not need extra hyperparameters. We set the learning rate as 0.01, and communication rounds at 100 as a standard for other baselines.
- **FedProx** [10]: FedProx adds a local penalty constant by the divergence between the local model and the global, to realize personalization. The local penalty constant μ is selected from $\{0.001, 0.1, 1\}$. And 0.001 is set as the proximal term.
- **FedDyn** [1]: In each round, this method adds a penalty term sent by the server to the learning goal of each client, so that the model of each device converges toward the global optimal direction. There is a new learning weight parameter α to control model update. We apply $\alpha = 0.01$ as set in the original literature.
- **FedRoD** [3]: The gap between the generalized model and the personalized model is explored in this work. They decouple the feature extractor and prediction head. Feature extractors are fused as usual, prediction heads are divided into local heads and global heads. The output prediction is the mixture of two heads. They also add a balanced softmax loss to combat label non-iid. We follow this work to set the temperature factor as $\gamma = 1$.
- **FedProto** [15]: FedProto consider to exchange prototypes instead of models. They add a contrastive loss to minimize the distance between local prototypes and global prototypes. The weight λ of contrastive loss is set as 0.1.
- **FedBN** [11]: the batch normalization layers are preserved locally trained while the other part of the model is globally shared. We add batch normalization layers after the convolutional layers for the CNN model architecture we adopted.
- **AlignFed** [19]: AlignFed performed as a combination with FedProto and FedRoD for feature skew. They decouple the feature extractors into global ones and local ones. Prototypes are also applied to pull the features toward the global feature centres of their corresponding classes.
- **FedFA** [18]: FedFA modelS feature statistics via a Gaussian distribution. The mean of a Gaussian distribution represents the original statistic, while the variance represents the enhancement range. Sampling from a Gaussian distribution to synthesize new features. Based on the original work, we set the momentum coefficient $\alpha = 0.5$ and probability $p = 1$.
- **FedTHE** [8]: following the training architecture in FedRoD, FedTHE adds two modules to enhance the test-time adaption ability in pFL. They adopt entropy minimization and feature space alignment. We adopt the default hyperparameter $\lambda_s = 0.5$ to balance the weight of losses.
- **FedICON** [14]: FedICON focus on feature skew in the training phase and TTA challenge in the test phase. They propose

contrastive learning to extract invariant information within the same class for each client. In the test time, they consider a consistency of augment data.

- **FLAMES2Graph** [16]: They extract and visualize those input subsequences that are highly activated by convolutional neural networks. Additionally, an evolution graph is created to capture the temporal dependencies between the different extracted subsequences. We adopted the best hyperparameter described in the paper where activation threshold is 0.1, cluster size selection is ($l_1 = 25, l_2 = 18, l_3 = 13$), graph embedding size is 256, segment length is 15.

1.3 Model Architecture

Encoder: Following the existing work [7, 16, 17], to ensure a fair comparison, we carefully selected an appropriate backbone model for all methods. This consideration applies to all our comparisons. To extract spatio-temporal features, we use a one-dimensional convolutional neural network (CNN) as the encoder. This configuration was kept consistent across all methods to ensure a fair comparison, where differences in prediction performance may be attributed to the adaptive algorithm itself. The 1D-CNN architecture consists of three blocks, each containing a 1D convolutional layer, a 1D batch normalization layer, a rectified linear unit (ReLU) function for nonlinearity, and finally a 1D max pooling layer.

2 DETAILED ALGORITHM

In this section, we provide a detailed training process as Algorithm 1:

Algorithm 1 FedST Training Procedure

Require: Clients K; local dataset $\mathcal{D}_k$; communication rounds T; Local epcoh E.

Initialization: global model $\boldsymbol{\theta}$,

```
for each round t = 1, ..., T do
    Distribute θ to all clients k ∈ K.
    for each client k ∈ 1, ..., K in parallel do
        (θ_k^(t), {p_{k,c}}_{c∈C}^{(t),{T,F}}) ← ClientUpdate(θ, D_k).
    end for
    Perform the global model and prototype aggregation.
end for
procedure ClientUpdate(θ, D_k)
    Compute Time and Frequency feature z_i^{T,F};
    Compute local prototypes p_{k,c}^{T,F} = 1/|D_k| Σ z_{i,c}^{T,F}, ∀c ∈ C;
    Project feature into orthogonal subspace h_i^{{T,F},{s,p}};
    Compute orthogonal loss by Eq. (11,12,13);
    Compute the overall loss from Eq. (14);
    Conduct local update θ_k^(t+1) ← θ_k^(t) − η∇L_k(θ_k^(t); D_k);
    Return , |D_k|, θ_k^(t), {p_{k,c}}_{c∈C}^{(t),{T,F}}.
end procedure
```

REFERENCES

[1] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas, Matthew Mattina, Paul Whatmough, and Venkatesh Saligrama. 2021. Federated Learning Based on Dynamic Regularization. In *Proceedings of International Conference on Learning Representations (ICLR)*.

[2] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* 64, 6 (2001), 061907.

[3] Hong-You Chen and Wei-Lun Chao. 2021. On bridging generic and personalized federated learning for image classification. In *Proceedings of International Conference on Learning Representations (ICLR)*.

[4] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. 2021. Time-Series Representation Learning via Temporal and Contextual Contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI*.

[5] Daniel Garcia-Gonzalez, Daniel Rivero, Enrique Fernandez-Blanco, and Miguel R Luaces. 2020. A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors* 20, 8 (2020), 2200.

[6] A Goldberger, L Amaral, L Glass, J Hausdorff, PC Ivanov, R Mark, HE Stanley, and PhysioToolkit PhysioBank. 2000. PhysioNet: Components of a new research resource for complex physiologic signals Components of a new research resource for complex physiologic signals. *Circulation* 101 (2000), e215–e220.

[7] Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and Marinka Zitnik. 2023. Domain adaptation for time series under feature and label shifts. In *Proceedings of International Conference on Machine Learning (ICML)*.

[8] Liangze Jiang and Tao Lin. 2023. Test-Time Robust Personalization for Federated Learning. In *Proceedings of International Conference on Learning Representations (ICLR)*.

[9] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.

[10] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated Optimization in Heterogeneous Networks. In *Proceedings of Machine Learning and Systems*.

[11] Xiaoxiao Li, Meirui JIANG, Xiaofei Zhang, Michael Kamp, and Qi Dou. 2020. FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. In *Proceedings of International Conference on Learning Representations (ICLR)*.

[12] Brendan McMahan, Eider Moore, and et al. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*.

[13] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 2015. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*.

[14] Yue Tan, Chen Chen, Weiming Zhuang, Xin Dong, Lingjuan Lyu, and Guodong Long. 2024. Is heterogeneity notorious? taming heterogeneity to handle test-time shift in federated learning. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*.

[15] Yue Tan, Guodong Long, and et al. 2022. Fedproto: Federated prototype learning across heterogeneous clients. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*.

[16] Raneen Younis, Zahra Ahmadi, Abdul Hakmeh, and Marco Fisichella. 2023. Flames2graph: An interpretable federated multivariate time series classification framework. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[17] Weiqi Zhang, Jianfeng Zhang, Jia Li, and Fugee Tsung. 2023. A co-training approach for noisy time series learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*.

[18] Tianfei Zhou and Ender Konukoglu. 2022. FedFA: Federated Feature Augmentation. In *Proceedings of International Conference on Learning Representations (ICLR)*.

[19] Guogang Zhu, Xuefeng Liu, Shaojie Tang, and Jianwei Niu. 2022. Aligning before aggregating: Enabling cross-domain federated learning via consistent feature extraction. In *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*.