Supplementary material of GSpfl: Graph Structured Personalized Federated Learning for Spatio-temporal Crime Prediction with Skewed Data

Author Name
Affiliation
email@example.com

In this supplementary material, we will provide more details on the experiment settings, implementations and more experiment results.

1 Data Collection and Description

We evaluated GSpfl on two real-world benchmarks collected in Chicago and Los Angeles by CrimeForecaster [Sun *et al.*, 2021], details are as follows:

LA city The crime data is collected across 113 regions of Los Angeles County for year 2018, and all crimes are aggregated into 8 categories. Table 1 shows all eight crime categories and the total counts of each in our dataset.

CHI city This dataset includes all reported crime events that occurred in Chicago during 2015. Table 2 shows the eight categories of crime events and their counts.

2 Focal loss

The focal loss function for a multi-label binary crime prediction problem, where each category of crime is represented by c, is given by:

$$\theta^{r+1} = \theta^r - \eta \nabla F_n(\theta_n^{(r)}) + \lambda_1 R(\theta_n^{(r)}, \Theta^{(r)}) + \lambda_2 R(\theta_n^{(r)}, \phi_n^{(r)}). \tag{1}$$

where:

- c - Represents a specific **crime category**.
- $y_c \in \{0, 1\}$ - Ground truth label for crime category c, where $y_c = 1$ if crime c occurred, otherwise $y_c = 0$.
- p_c - Model-predicted probability for crime category c.
- α_c - Class-balancing weight for category c, helping to address class imbalance.
- γ - Focusing parameter that adjusts the weight of easy vs. hard examples. A higher γ increases the focus on hard-to-classify crime categories.

The focal loss function mitigates the impact of well-classified examples, ensuring that minority and difficult-to-classify crime categories contribute more significantly to the model training.

3 CKA based similarity

The **Centered Kernel Alignment (CKA)** similarity between model layers of different clients in federated learning is given by:

$$\text{CKA}(K, L) = \frac{\text{Tr}(KL)}{\sqrt{\text{Tr}(KK)\text{Tr}(LL)}} \tag{2}$$

where:

- $K = HH^T$ - The Gram matrix for client model layer activations H.
- $L = GG^T$ - The Gram matrix for another client model layer activations G.
- $H \in R^{n \times d}$, $G \in R^{n \times d}$ - Activation matrices of two clients, where:
  - n is the number of samples.
  - d is the feature dimension.
- Tr$(\cdot)$ - The trace operator, summing the diagonal elements of a matrix.

4 Sampling and Grouping of Regions

The grouping shown in Table 3, D_s, M_s, and S_s is based on 20:30:50 percentiles of the sorted data, i.e., 20% dense, 30% moderate, and 50% sparse. The reason for taking such percentiles are due to long-tail distribution, where most of the data follow the distribution same as tail. Then we sample 10 clients from the total pool using exponential sampling.
For Los Angeles the selected 10 clients from pool of 113 are:
S-Alpha = {3, 15, 10, 58, 49, 20, 30, 32, 75, 56 },
S-Beta = {10, 58, 49, 20, 75, 22, 53, 93, 81, 68},
S-Gamma = {58, 20, 75, 53, 93, 81, 68, 45, 13, 108};
Similarly for Chicago the 10 clients from pool of 77 are:
S-Alpha = {16, 11, 34, 30, 33, 36, 19, 13, 25, 2},
S-Beta = {11, 33, 36, 25, 2, 23, 32, 18, 10, 9},
S-Gamma = {33, 45, 2, 32, 26, 18, 9, 6, 47, 28}

5 Evaluation Metrics

The metrics are defined as:

- **Precision** The ratio of correctly predicted instances of crime category TP_c to the total instances predicted as crime category.

Category	Theft	Vehicle Theft	Burglary	Fraud	Assault	Sexual Offenses	Robbery	Vandalism
Counts	66,697	17,123	14,517	15,578	32,372	6,161	8,864	17,123

Table 1: Los Angeles City crime dataset statistics.

Category	Theft	Criminal Damage	Narcotics	Robbery	Assault	Deceptive Practices	Burglary	Battery
Counts	56,695	28,589	21,607	9,632	16,692	14,085	13,103	48,824

Table 2: Chicago City crime dataset statistics.

SUBSETs	Included Regions
S-Alpha	$D_s(10)$
S-Beta	$D_s(5)$ + $M_s(5)$
S-Gamma	$D_s(3)$ + $M_s(4)$ + $S_s(3)$

Table 3: Selected Subsets from pool of total clients

- **Recall** The ratio of correctly predicted instances of crime category TP_c to the total actual instances of crime category.
- **F1-Score** The harmonic mean of precision and recall for crime category c.
- **Micro F1** The micro F1 score aggregates the contributions of all crime categories by summing the true positives (TP), false positives (FP), and false negatives (FN) across all categories before computing the F1 score. This approach gives equal weight to each instance, making it effective for datasets with imbalanced category distributions that means gives more weight to the classes with more instances.

$$\text{Micro F1} = \frac{2\sum_{c=1}^{C} \text{TP}_c}{2\sum_{c=1}^{C} \text{TP}_c + \sum_{c=1}^{C} \text{FP}_c + \sum_{c=1}^{C} \text{FN}_c} \tag{3}$$

- **Macro F1** The macro F1 score is the arithmetic mean of the F1 scores for each crime category. It treats each category equally by first computing the F1 score for each category and then averaging them. It treats all classes equally, regardless of their imbalance in the dataset. This score gives equal weight to each class, which is useful because each type of crime is equally important in the evaluation.

$$\text{Macro F1} = \frac{1}{C}\sum_{c=1}^{C} \frac{2 \cdot \text{TP}_c}{2 \cdot \text{TP}_c + \text{FP}_c + \text{FN}_c} \tag{4}$$

here C is the total number of classes, and TP_c, FP_c, and FN_c are the true positives, false positives, and false negatives for class c, respectively.

6 Baseline and State-of-the-art

We categorize the methods into three parts: (i) Standard Baselines: FedAvg, Krum (ii) and Non-iid methods: methods FedPer, Ditto (iii) FedProx, FedAvgM, Scaffold (iv) Graph structure methods: SFL, FBLG. Details of all these methods are given in Table 4, which describes the location of modification, personalization focus, Non-iid focus, underlying technique and knowledge sharing aspect.

We evaluate our proposed GSpfl method against several baselines and state-of-the-art federated learning approaches: (i) FedAvg [McMahan *et al.*, 2017] utilizes weight aggregation; (ii) Krum [Blanchard *et al.*, 2017] filter malicious or out of distribution contributions, (iii) FedAvgM [Hsu *et al.*, 2019] integrates momentum-based regularization, (iv) Fedper [Arivazhagan *et al.*, 2019] based on base + personalization layer approach, (v) FedProx [Li *et al.*, 2020], incorporate a proximal term to address client drift, (vi) SCAFFOLD [Karimireddy *et al.*, 2020] employ control variates to mitigate drift, (vii) Ditto [Li *et al.*, 2021b] based on dual-model approach where each client maintains both a personalized model and a globally consistent model, (Viii) MOON [Li *et al.*, 2021a], leverage model contrastive loss, (ix) SFL [Chen *et al.*, 2022] model the client relations with graph neural network at server, (x) FBLG [Xu *et al.*, 2024] based on losses and Jensen-Shannon (JS) divergence graph to identify similar clients for aggregation.

7 Visualization

Ethical Statement

Our research on distributed crime prediction prioritizes ethical considerations to ensure fairness, privacy, and societal impact. We acknowledge the potential risks of bias in crime data, for ensuring fairness we didn't consider protective attributes and community information for training. By leveraging federated learning, we uphold data privacy by ensuring sensitive information remains decentralized, preventing unauthorized access or misuse.

References

[Arivazhagan *et al.*, 2019] Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.

[Blanchard *et al.*, 2017] Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, and Julien Stainer.

Method	Location	Personalization	Non-IID	Technique	Knowledge sharing
FedAvg	Standard	No	Limited	Standard SGD	Weighted Averaging
Krum	Server	No	No	Distance-based Selection	Selects Closest Model Updates
FedProx	Client	No	Yes	Proximal Term in Loss	Weighted Averaging
FedAvgM	Server	No	Yes	Uses Server Momentum	Momentum-based Aggregation
Moon	Client	No	Yes	Contrastive Loss	Weighted Averaging
FedPer	Client	Yes	No	Local Adaptation of Layers	Partial Model Aggregation
Ditto	Client	Yes	No	Dual Model Training	Separate Model Per Client
SFL	S + C	Yes	No	Graph-based Model Updates	Structural Similarity Aggregation
FBLG	Server	No	Yes	Graph-based Updates	Graph based aggregation
Scaffold	Client	No	Yes	Control Variate-based Correction	Corrects Model Drift

Table 4: Comparison of Federated Learning Approaches

Byzantine-tolerant machine learning. *arXiv preprint arXiv:1703.02757*, 2017.

[Chen *et al.*, 2022] Fengwen Chen, Guodong Long, Zonghan Wu, Tianyi Zhou, and Jing Jiang. Personalized federated learning with graph. *arXiv preprint arXiv:2203.00829*, 2022.

[Hsu *et al.*, 2019] Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv preprint arXiv:1909.06335*, 2019.

[Karimireddy *et al.*, 2020] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.

[Li *et al.*, 2020] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

[Li *et al.*, 2021a] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10713–10722, 2021.

[Li *et al.*, 2021b] Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *International conference on machine learning*, pages 6357–6368. PMLR, 2021.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[Sun *et al.*, 2021] Jiao Sun, Mingxuan Yue, Zongyu Lin, Xiaochen Yang, Luciano Nocera, Gabriel Kahn, and Cyrus Shahabi. Crimeforecaster: Crime prediction by exploiting the geographical neighborhoods' spatiotemporal dependencies. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V*, pages 52–67. Springer, 2021.

[Xu *et al.*, 2024] Yi Xu, Ying Li, Haoyu Luo, Xiaoliang Fan, and Xiao Liu. Fblg: A local graph based approach for handling dual skewed non-iid data in federated learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24, K. Larson, Ed. International Joint Conferences on Artificial Intelligence Organization*, volume 8, pages 5289–5297, 2024.