FEDERATED LEARNING FOR PERSONALIZED SITUATION RECOGNITION USING MULTIMODAL WEARABLE AND CONTEXTUAL DATA

Zia Ur Rehman, zrehman@unisa.it

ABSTRACT

This project proposes a personalized federated learning (FL) framework for situation recognition using multimodal data from wearable devices. The system uses sensor and contextual features from 60 users to infer real life situations such as "Watching TV", "Walking-Talking", or "Brushing". To preserve user privacy and improve generalization, we implement four FL strategies: FedAvg, FedProx, FedPer, and Graph Total Variation Minimization (GTVMin). We build a similarity graph using cosine similarity between user level feature distributions and apply GTVMin to enforce model smoothness across similar clients. Experiments show that FedAvg, FedProx and FedPer achieve less F1-scores (35%), but struggle with underrepresented classes like Standing-Talking or walking-talking. In contrast, GTVMin significantly improves performance, achieving the best per client F1-score of (55%) and a weighted average F1-score of 80 % across all situations. The proposed per situation evaluation reveals that GTVMin enhances complex activity recognition, such as Running-Exercise, Standing-Talking, and Watching-TV, while maintaining high accuracy for common classes like Sleeping and Computer Work. This confirms that incorporating graph-based personalization leads to better robust multi-situation recognition.The framework balances accuracy, personalization, and communication cost, making it well-suited for privacy-preserving wearable systems in healthcare, fitness, and behavior monitoring.

Keywords: Federated learning, networks, personalized machine learning, Human activity recognition, Situation recognition

1. INTRODUCTION

In recent years, the proliferation of smartphones and smartwatches has enabled continuous monitoring of human behavior in real-life settings[1]. These wearable devices generate large amount of sensor and contextual data, which can be used not only to recognize physical activities (e.g., running, standing) but also to infer more complex, high level situations (e.g., "Sitting-WatchingTV", "Lyingdown", "Sleeping"). Recognizing such situations plays a critical role in applications like personalized health monitoring, behavior tracking, and smart assistance systems [2],[3]. The previous work introduced a centralized Situation-Aware Wearable Computing System (SA-WCS) that follows Endsley's three-level model of situation awareness [1][4]. The system included a perception phase to detect user activities using only sensor data, and in a comprehension phase to infer the user's situation by combining recognized activities with context features (e.g., location, diameter, time of the day). The approach use techniques like Context Space Theory (CST) and knowledge-driven context modeling, achieving high recognition accuracy in a centralized setting.

However, this centralized approach poses major challenges in real-world settings. Transmitting raw user data (e.g, UserId, context) to a central server raises serious privacy concerns, limits scalability, and fails to account for user-specific variations [5][6]. Different peoples exhibit diverse routines and environmental conditions, making a one-size-fits-all model suboptimal for personalized situation recognition.

To address these limitations, this project introduces Federated Learning Situation aware Wearable Computing System (FL-SAWCS). FL enables each user (client) to train models locally on their own data(e.g, sensor and contextual information of each user), preserving privacy while still benefiting from shared learning across clients. No raw data ever leaves the user's device; only model updates are communicated.In this context, our goal is to explore whether federated learning can: 1) Accurately recognize personalized situations across heterogeneous users (e.g, 60 users), 2) Reduce the generalization gap caused by non-IID data distributions, 3) Support efficient model training with minimal communication overhead, 4) Provide a privacy-preserving (FL-SAWCS) and scalable solution for real-time situation awareness.

Recent studies on context-aware systems, activity recognition and situation recognition have focused primarily on centralized machine learning pipelines. Prior work has explored manual rule-based techniques for context recognition (e.g, Context space theory), and more recent approaches have proposed KDE-based automatic context modeling. While these techniques are effective in controlled environments, they do not generalize well across users and are computationally demanding when scaled.

On the other hand, Federated Learning has gained popularity for human activity recognition [3] tasks but has not yet been fully adapted to situation identification, which requires context fusion with activity recognition and deeper personalization. Existing FL methods like FedAvg, FedProx, and FedPer address data heterogeneity and personalization to varying degrees . However, they do not explicitly consider inter-client similarity, which could improve knowledge transfer between behaviorally similar users.

This work extends the SA-WCS framework[1] by integrating federated learning with graph-based personalization to enhance multi-user situation recognition (FL-SAWCS). Specifically, we make the following contributions:

- We transform the centralized SA-WCS architecture into a fully decentralized FL system capable of local training on user devices using relevant sensor and context features.
- We implement and compare four FL algorithms: FedAvg, FedProx, FedPer, and Graph Total Variation Minimization (GTVMin).
- We model the FL network as a client similarity graph, where edge weights reflect behavioral similarity (based on cosine similarity of user sensor and context features), and GTVMin promotes parameter smoothness across connected clients.

We evaluate the system on a Extrasensory dataset (60 users) with 16 situation labels, reporting per-client and per-situation metrics including accuracy, precision, recall, F1-score, and communication cost. At the end, we also demonstrate that GTVMin significantly improves generalization and personalization without violating user privacy, achieving the highest weighted F1-scores across complex situation labels.

2. PROBLEM FORMULATION

In our proposed system, we model the federated learning setup as a communication graph, $G = (V, E)$, where: Each node $v \in V$ represents in Figure 1 a real user carrying a smartphone and/or smartwatch collecting multimodal sensor and context data from the Extrasensory dataset, including: Sensor features such as accelerometer, gyroscope, magnetometer and Contextual features: location, time of day, ambient data, phone usage and recognized Situations (e.g., "Sleeping").In total, $|V| = 60$ users form 60 clients in the FL network, each with their own dataset(e.g sensor and context pairs), and local behavior patterns.

Each client i trains its own local model f_i parameterized by w_i , on their personal data D_i (e.g., relevant concatenated sensor and context feature vector). The model is a multilayer perceptron (MLP) on each client.

The proposed architecture supports per-user customization and efficient training on small embedded

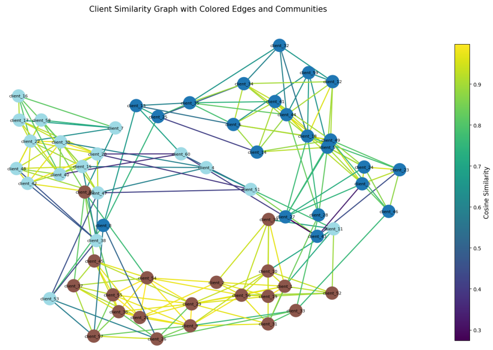

Fig. 1: Client Similarity Graph with Colored Edges and Communities

devices. Each client minimizes the weighted cross-entropy loss as shown in the equation 1 to handle imbalanced situation labels:

$$f_i(w_i) = \frac{1}{|D_i|} \sum_{(x_j, y_j) \in D_i} \alpha_{y_j} \cdot \mathcal{L}_{\mathrm{CE}}(f_i(x_j), y_j) \quad (1)$$

Where:

- α_{y_j} is the class weight for label y_j.
- $\mathcal{L}_{\mathrm{CE}}$ is the standard cross-entropy loss.
- $f_i(x_j)$ is the predicted probability vector for input x_j.

This ensures minority situations like STANDING-TALKING or STANDING-EATING" are not ignored during training.

We define edges based on behavioral similarity across clients. That is, an edge exists between clients i and j if their data distributions are similar.

Let $\mu_i \in \mathbb{R}^d$ be the mean feature vector for client i:

$$\mu_i = \frac{1}{|D_i|} \sum_{x \in D_i} x$$

Then, the edge weight between clients a_{ij}, i and j is:

$$a_{ij} = \cos(\mu_i, \mu_j) = \frac{\mu_i \cdot \mu_j}{\|\mu_i\| \, \|\mu_j\|}$$

Edges are created only for the top k most similar clients (e.g some users exercising at the same time, going office at the same time, forming a k-nearest neighbor graph (typically k=5), as shown in the equation 2

$$\min_{i \in V} \sum f_i(w_i) + \lambda \sum_{(i,j) \in E} a_{ij} \|w_i - w_j\|^2$$

Where:

- $f_i(w_i)$ is the local loss function for each client i.

- $\|w_i - w_j\|^2$ is the squared ℓ_2 distance between the models of clients i and j.
- λ is a hyperparameter that balances local accuracy against smoothness across similar clients (sharing similar information).
- a_{ij} is the cosine similarity weight between clients i and j.

This setup motivates the need for learning algorithms that not only respect client-specific behaviour but also exploit Clients with similar behavior (e.g., 2 office workers) share model traits Clients with unique behaviors. In the next section, we introduce and compare several federated learning methods tailored to this personalized, privacy-preserving scenario.

3. METHODS

In this section we describe Federated Learning (FL) algorithms applied to our situation identification task using the Extrasensory dataset [7]. Each user collects relevant sensor and contextual features and trains a local neural network (e.g, MLP) to predict real-life situations. These methods operate under a privacy preserving setting where raw data never leaves the device. Instead, models are trained locally and only model parameters are shared with other clients or users.

- Federated Averaging: FedAvg is a baseline FL method where each client performs (e.g all 60 users) local training and then sends its model parameters to the central server for aggregation. The global model is updated by averaging:

$$w^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} w_i^{(t)}$$

- Federated Proximal Optimization: FedProx improves upon FedAvg by introducing a proximal regularization term that limits divergence from the global model. This helps reduce client drift during local updates:

$$f_i(w^{(i)}) + \frac{\mu}{2} \|w^{(i)} - w^{(t)}\|^2$$

- Personalized Federated Learning: FedPer introduces per-client personalization by splitting the model into: A shared feature extractor (learned globally), A personalized classification head $w_{\text{head}}^{(i)}$ (kept private):$f_i(x;\, w_{\text{shared}},\, w_{\text{head}}^{(i)})$

In this work, we model our Federated Learning Situation aware Wearable Computing System (FL-SAWCS), as a graph based optimization problem, where each users $w^{(i)}$ in Extrasensory datasets has its own parametric model, and learns to minimize a combination of Local loss(based on their own sensors and context data), and Smoothness constraint that enforces similarity between neighboring users on a similarity graph. We define the variation between two clients i and j using the L2 norm of the difference between their model parameters:

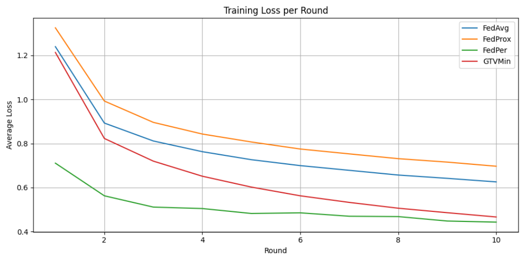

Fig. 2: Training Loss per round across all Federated Learning algorithms (FedAvg, FedProx, FedPer, GTV)

$$\phi(w^{(i)} - w^{(j)}) = \|w^{(i)} - w^{(j)}\|^2$$

This choice introduces smoothness: clients with similar behavior (e.g., daily work routine or exercise) are encouraged to have closer model parameters.

As we described client similarity graph in section 2, the optimization objective funtion is:

$$\min_{\{w^{(i)}\}_{i \in V}} \sum_{i \in V} f_i(w^{(i)}) + \lambda \sum_{(i,j) \in E} a_{ij} \cdot \|w^{(i)} - w^{(j)}\|^2$$

Where

- $f_i(w^{(i)})$: Local weighted cross-entropy loss on client i's dataset.
- a_{ij}: Cosine similarity between clients i and j, based on their mean sensor and context embeddings.
- λ: Regularization strength controlling the influence of neighboring clients (e.g, 60 clients).

To solve this optimization problem, we implement a round-based message passing algorithm that proceeds as follows: For each round $t = 1, \ldots, T$: **Local Training:** Each client i updates its model $w^{(i)}$ by minimizing $f_i(w^{(i)})$ using local data for several epochs. **Graph Based Smoothing:** Each user exchanges its model with neighbors $j \in \mathcal{N}(i)$, then updates its own model using Equation 3:

$$w^{(i)} \leftarrow w^{(i)} - \lambda \sum_{j \in \mathcal{N}(i)} a_{ij}(w^{(i)} - w^{(j)})$$

This step reduces model divergence among similar users (high a_{ij}), improving generalization.

4. NUMERICAL EXPERIMENTS

We use the ExtraSensory dataset [7], which consists of wearable sensor signals (e.g. accelerometer, gyroscope) with contextual information (e.g., location,

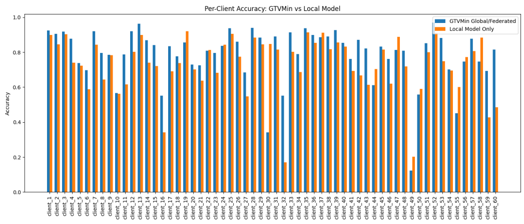

Fig. 3: Comparison of Per-Client Accuracy Comparison: Global(GTV Federated Learning) vs Local Model

Fig. 4: Confusion matrix of User1 across all Federated Learning techniques, GTV shows best performance across all 7 situations

audio, timestamp). Each user collects data independently, naturally forming a federated learning (FL) network with 60 users, where each user is treated as a node in a personalized FL graph. Our goal is to develop a Personalized Situation Recognition system (FL-SAWCS), capable of predicting complex real-world situations such as SITTING-WATCHING_TV, SITTING-TALKING, or LYING-WATCHING_TV, by fusing sensor and context features. Each client's local dataset is partitioned as follows:70% Training for local model optimization,15% Validation for early stopping and hyperparameter tuning, 15% Testing for final evaluation only.

Several federated learning methods (FedAvg, FedProx, FedPer, and GTVMin) [3] are trained for 5 communication rounds, with 5 local training epochs each round. The best performing model is selected per client based on validation F1 and accuary score and is used for test evaluation. For GTVMin, model selection occurs after the graph-based smoothing step, which incorporates information from similar neighbored.

We compute the following evaluation metrics across each client: Accuracy, Precision, Recall, F1-score (see Figure 3). We also evaluate the performance Per-label: Macro and weighted F1-scores (see Figure 6, and Confusion Matrix (Figure 4), Loss curves: Training, validation, and test losses per client (Figure 2). Among all methods, GTVMin consistently outperformed others, especially for rare or noisy situation

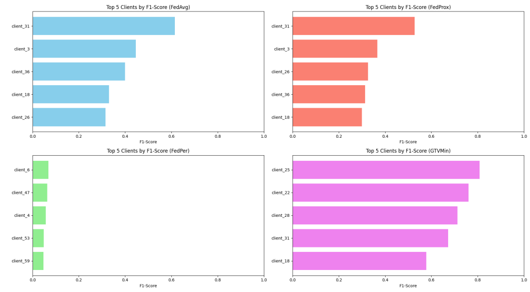

Fig. 5: Visualize Top 5 Clients per Method by F1-Score

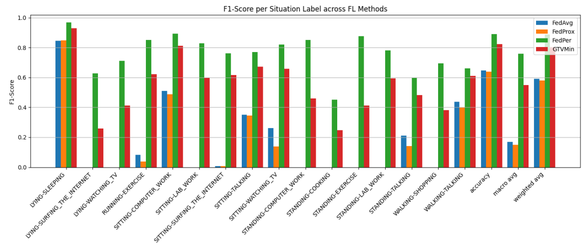

Fig. 6: F1-Score per Situation Label across FL Methods

labels (see Figure 6). This is attributed to its graph-based regularization, which enforces local smoothness across behaviorally similar clients.

A comparison of federated vs. local model accuracy reveals that in Figure 3: FedAvg improves modestly over local models, FedProx provides better stability with non-IID data,FedPer under performs on low-data clients due to limited training for personalized heads,GTVMin achieves the best trade-off between personalization and generalization, reaching an F1-score of up to 55% (Table 1).

Despite higher per-round computation, GTVMin remains communication-efficient by only sharing model parameters among top-5 most similar clients (see Figure 5), avoiding full global aggregation and enabling scalable, privacy-preserving personalization.

Table 1: Comparison of Federated Learning Methods (Best Client per Method) for Situation Recognition

Method	Personalized	Handles Non-IID	Graph-Based	Best Client	Best F1
FedAvg	No	No	No	client_52	0.3549
FedProx	No	Yes	No	client_52	0.3506
FedPer	Partial	Yes	No	client_55	0.0835
GTVMin	Full	Yes	Yes	client_25	**0.5501**

5. CONCLUSION

In conclusion, we proposed a Federated learning situation aware wearable computing system (FL-SAWCS) for situation recogniton using Extrasensory datasetusing . The propsoed system using GTVMin outperforms baseline FL algorithms (FedAvg, FedProx, FedPer) in personalized situation recognition . It achieved the best overall F1score of 0.55, with notable improvements in recognizing rare situations like STANDING-LABWORK and RUNNING-EXERCISE, where FedAvg and FedProx scored less tahn 20%. GTVMin graph based regularization enabled efficient knowledge sharing among similar and diverse range of users, leading to higher accuracy and better generalization, especially for low-data clients. Despite this, challenges like graph scalability and class imbalance persist. Future improvements may include dynamic graph updates and adaptive architectures to further enhance performance.

6. REFERENCES

[1] G. D'Aniello, M. Gaeta, R. Gravina, Q. Li, Z. U. Rehman, and G. Fortino, "Situation identification in smart wearable computing systems based on machine learning and context space theory," *Information Fusion*, vol. 104, p. 102197, 2024.

[2] D. Sachin, B. Annappa, and S. Ambesenge, "Federated learning for wearable sensor-based human activity recognition," in *International Conference on Intelligent Technologies*. Springer, 2022, pp. 131–139.

[3] O. Aouedi, A. Sacco, L. U. Khan, D. C. Nguyen, and M. Guizani, "Federated learning for human activity recognition: Overview, advances, and challenges," *IEEE Open Journal of the Communications Society*, 2024.

[4] G. D'Aniello, G. Fortino, M. Gaeta, and Z. U. Rehman, "A data-driven approach for context definition in situation-aware wearable computing systems," in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2024, pp. 1525–1531.

[5] P. F. Afandy, P. C. Ng, and K. N. Plataniotis, "Federated learning for hierarchical fall detection and human activity recognition," in *2024 IEEE 10th World Forum on Internet of Things (WF-IoT)*. IEEE, 2024, pp. 1–6.

[6] T. Gönül, O. D. Incel, and G. Isiklar Alptekin, "Human activity recognition with smart watches using federated learning," in *International Conference on Intelligent and Fuzzy Systems*. Springer, 2022, pp. 77–85.

[7] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, "Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–12.