A Federated Unsupervised Personalisation for Cognitive Workload Estimation

Dario Fenoglio
Università della Svizzera Italiana
Lugano, Switzerland
dario.fenoglio@usi.ch

Martin Gjoreski
Università della Svizzera Italiana
Lugano, Switzerland
martin.gjoreski@usi.ch

Marc Langheinrich
Università della Svizzera Italiana
Lugano, Switzerland
marc.langheinrich@usi.ch

ABSTRACT

Accurate Cognitive Workload (CW) estimation, crucial in mobile healthcare and human-machine interaction, is impeded by client heterogeneity, data limitations, and privacy concerns, especially in the presence of Out-of-Distribution (OoD) clients. This study proposes a robust framework that is based on Federated Learning to protect data privacy, and utilizes context-based STRNet to enable joint cross-user learning on heterogeneous datasets, enhancing model generalisability. The framework includes a novel Unsupervised Client Personalisation strategy that prevents accuracy loss in OoD clients. We tested our framework on two publicly available CW datasets, COLET and ADABase. The framework improved the accuracy of centralized approaches while preserving data privacy. The framework is model-agnostic, efficient, and enables unsupervised personalisation for each client, bolstering the quality and robustness of the end-to-end deep learning models.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Security and privacy → Distributed systems security**; • **Computer systems organization** → *Client-server architectures.*

KEYWORDS

Cognitive Workload, Federated Learning, Privacy, Deep Learning, Unsupervised Personalisation.

1 INTRODUCTION

Accurate Cognitive Workload (CW) estimation, which quantifies the cognitive effort exerted by individuals during a task, can prevent burnout, reduce medical errors, and enhance human-machine interfaces [3, 11, 12]. CW estimation commonly exploits physiological signals, such as respiratory, ocular, neural, muscular, cardiac, and electrodermal activities. [5, 7, 11]. Nonetheless, CW estimation faces several challenges. Resource-intensive data collection often yields limited or heterogeneous datasets, necessitating training on single datasets, which restricts deep learning model generalization. In real-world scenarios, unseen clients might diverge from the training distribution, termed Out-of-Distribution (OoD) clients, leading to reduced performance [1, 10]. Furthermore, privacy protection remains a significant concern, since contemporary state-of-the-art sensors used for CW estimation collect sensitive data, e.g., cameras, eye trackers, and biomedical sensors [7, 8, 11]. In response to these challenges, we propose: i) a unified framework that utilizes joint Federated Learning (FL) across two heterogeneous datasets, thereby expanding the training data without requiring client data transfer; ii) a context-based version of the STRNet [4] that processes sensor signals from diverse heterogeneous datasets, and exploits temporal and frequency signal components, and iii) an Unsupervised Client Personalisation mechanism, designed to tailor the model to each unseen user, especially effective in bolstering performance on OoD clients. To our knowledge, this study is the first exploration of Federated joint training and Unsupervised Client Personalisation for the development of privacy-aware CW estimation pipelines.

2 DATASETS AND PREPROCESSING

In this study, we utilize two public datasets, COLET [7] and ADABase [11], for CW estimation. COLET provides eye-tracking data from 46 participants across four tasks (A1/A2/A3/A4). Each task designed to induce different CW levels. We utilized seven signals, focusing on classifying two CW levels: high (tasks A1 and A2) and low (tasks A3 and A4). Differently, ADABase involved 29 participants using multimodal sensor data. They captured physiological metrics during a driving simulation and a memory test, designed to modulate CW. In our analysis, eight signals were utilized to classify between high and low CW levels. Data preprocessing, executable locally to align with FL methodologies, involved: 1) applying subject-wise normalization for uniformity, comparability and enhanced CW differentiation across datasets; 2) resampling signals to a consistent 50Hz; 3) segmenting into 10-second windows with a 75% overlap, and 4) structuring input windows of 15 signals (seven from COLET, eight from ADABase), with one shared signal (detected blinks). The first model layer had 14 inputs - six from COLET, seven from ADABase, and one shared, that enables cross-dataset knowledge sharing. Each client then populates the input with the signals it possesses, setting the others to zero, while a context vector indicated signal presence.

3 PROPOSED FRAMEWORK FOR UNSUPERVISED PERSONALISATION

Informed by the experimental outcomes presented in [2] and [4], we utilized the spectro-temporal ResNet (STRNet) architecture, designed with 14 inputs to exploit both temporal and frequency domain information. Differing from the original implementation, our adaptation permits a variable number of input signals by introducing a context vector to mask activations from absent inputs. This modification enables learning a joint model from multimodal heterogeneous datasets. Our STRNet can be generalized into an Encoder and two distinct Heads. As illustrated in Figure 1, the Encoder and global Head (Head_g) are simultaneously trained across both datasets via Global Federated Learning (GFL), while the local Head (Head_l), specific to each dataset, is trained on a single dataset using Federated Fine-Tuning (FFT). Specifically, inspired by [6], our training process for each client unfolds in two subsequent phases: GFL and FFT. First, in the GFL phase, clients receive a global model

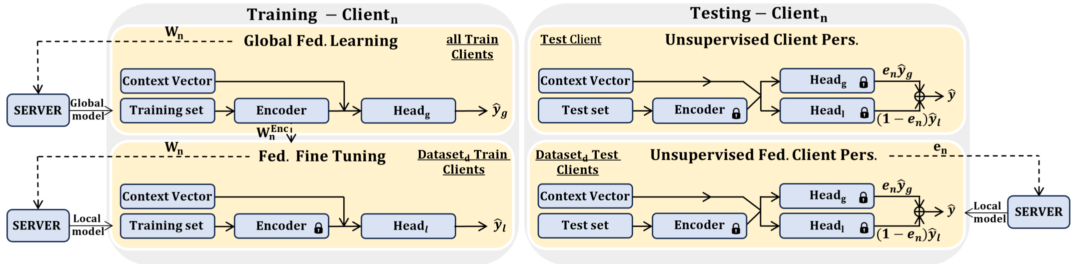

Figure 1: Training and Unsupervised Client-Personalisation for a client n and dataset d.

from the server and train it using their respective local datasets, subsequently returning an encrypted version of the updated weights to the server, where the global model is updated through weighted averaging [9]. Subsequently, two FFTs are executed using the frozen pre-trained Encoder, aggregating on the server only the weights from clients of the same dataset. Both training phases employ a binary cross-entropy loss function and utilize Adam for optimization. We opted for this implementation over client-specific fine-tuning (not federated) for two key reasons: first, the necessity of developing and testing a client-independent model; and second, client datasets may be limited, potentially resulting in unstable training. Post-training, we implement our proposed Unsupervised Client Personalisation (UCP), whereby a unique model for each client is constructed utilizing the global Encoder and $\mathrm{Head_g}$ received from the GFL, alongside the $\mathrm{Head_l}$ from the FFT (Figure 1). Predictions generated by both Heads are combined using a learnable parameter e as follow: $\hat{y} = e \cdot \hat{y}_g + (1-e) \cdot \hat{y}_l$ where $\hat{y}_g$ and $\hat{y}_l$ represent predictions from $\mathrm{Head_g}$ and $\mathrm{Head_l}$, respectively. Unsupervised learning is applied to each unseen client, simulating real-world scenarios, prior to delivering final inferences. To optimize the e parameter, Equation 1 is minimized:

$$L = \lambda \cdot \left[- \sum_i p_i(\hat{y}) \cdot \log p_i(\hat{y}) \right] + \left[e \cdot ||h_n^{EMA} - h_g||_2 + (1-e) \cdot ||h_n^{EMA} - h_l^d||_2 \right] \quad (1)$$

Here, λ is a hyperparameter, and the first term denotes the Shannon Entropy of the softmax probability for the i-th class on $\hat{y}$, which augments the model's predictive confidence. The second term enforces latent space alignment with h_g and h_l^d, adjusting the balance between the models to prioritize the one trained on the closer dataset (global or local). Specifically, h_l^d is the average latent space of clients related to dataset d, while h_g is the mean for all clients involved in the FL. h_n^{EMA} is the instantaneous latent space average for the current client, computed using Exponential Moving Average. Additionally, we also explored an Unsupervised Federated Client Personalisation (UFCP) to obtain a unique personalized model per dataset. As shown in Figure 1, clients from the same dataset undergo federated training, exclusively sharing the learnable e parameter, leading to a unified personalisation model for testing clients.

4 RESULTS AND DISCUSSION

Our results, validated using person-independent testing with six clients from ADABase and nine from COLET, are showcased in Table 1. They demonstrated that our proposed UFCP and UCP frameworks outperform traditional Centralized Training and Dataset-Specific FL on two distinct datasets. Even compared to the privacy-invasive Centralized approach, UCP improved accuracy from 91.8% to 92.5% on ADABase and from 83.7% to 91.2% on COLET. Our ablation study revealed that the ADABase dataset, when used in our GFL, notably boosts network performance for COLET clients, enhancing accuracy by 0.4% and 3.7% on ADABase and COLET respectively. This improvement suggests that joint training enhances the network's capacity to extract pertinent information from the data. Although FFT does not overtly enhance performance, it creates two local models, each specific to its training distribution, providing distinct contributions to the OoD clients. Essentially, our UCP framework enhances the Encoder's generalization capabilities through global training, and then identifies the optimal combination of two Heads trained on diverse distributions (e.g., two separate datasets), to leverage the benefits of both distributions per client. In summary, this study addresses several challenges in estimating CW by: 1) augmenting dataset size by merging two heterogeneous datasets (COLET and ADABase) to leverage common, task-related information through our STRNet with context vector; 2) safeguarding user data privacy through federated approaches; and 3) enhancing test client performance, including those outside the training distribution, via our UCP. Notably, our framework surpasses even the privacy-intrusive centralized approach. Its model-agnostic nature permits the use of any Encoder architecture and theoretically enables exploitation of numerous datasets with at least one common signal. Consequently, this study paves the way towards leveraging client-personalized deep learning models to extract high-level knowledge from data, even in the absence of a large, homogeneous dataset.

ACKNOWLEDGMENTS

This study was funded by TRUST-ME (project 205121L_214991) and SmartCHANGE (GA No. 101080965) projects.

REFERENCES

[1] Philip Boyer, David Burns, and Cari Whyne. 2021. Out-of-Distribution Detection of Human Activity Recognition with Smartwatch Inertial Sensors. *Sensors* 21, 5 (2021). https://doi.org/10.3390/s21051669
[2] Maciej Dzieżyc, Martin Gjoreski, Przemysław Kazienko, Stanisław Saganowski, and Matjaž Gams. 2020. Can We Ditch Feature Engineering? End-to-End Deep

DOI: 10.1145/3626705.3631796
Published December 2023

Table 1: Accuracy Comparison: Centralized Training, Dataset-Specific Federated Learning (FL), Global FL (GFL), Federated Fine-Tuning (FFT), Unsupervised Federated Client Personalisation (UFCP), Unsupervised Client Personalisation (UCP).

Centralized		Data-Spec. FL		GFL		FFT		UFCP (ours)		UCP (ours)	
ADA	COLET	ADA	COLET	ADA	COLET	ADA	COLET	ADA	COLET	ADA	COLET
91.8	83.7	91.8	83	92.2	86.7	92.5	85.8	**92.6**	87.1	92.5	**91.2**

Learning for Affect Recognition from Physiological Sensor Data. *Sensors* 20, 22 (2020), 1–210. https://doi.org/10.3390/s20226535

[3] Daniel E. Ehrmann, Sara N. Gallant, Sujay Nagaraj, Sebastian D. Goodfellow, Danny Eytan, Anna Goldenberg, and Mjaye L. Mazwi. 2022. Evaluating and reducing cognitive load should be a priority for machine learning in Healthcare. *Nature Medicine* 28, 7 (2022), 1331–1333. https://doi.org/10.1038/s41591-022-01833-z

[4] Martin Gjoreski, Vito Janko, Gašper Slapničar, Miha Mlakar, Nina Reščič, Jani Bizjak, Vid Drobnič, Matej Marinko, Nejc Mlakar, Mitja Luštrek, et al. 2020. Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors. *Information Fusion* 62 (2020), 47–62. https://doi.org/10.1016/j.inffus.2020.04.004

[5] Martin Gjoreski, Tine Kolenik, Timotej Knez, Mitja Luštrek, Matjaž Gams, Hristijan Gjoreski, and Veljko Pejović. 2020. Datasets for cognitive load inference using wearable sensors and psychological traits. *Applied Sciences* 10, 11 (2020), 3843. https://doi.org/10.3390/app10113843

[6] Liangze Jiang and Tao Lin. 2023. Test-Time Robust Personalization for Federated Learning. *arXiv* (2023). https://doi.org/10.48550/arXiv.2205.10920 arXiv:2205.10920

[7] Emmanouil Ktistakis, Vasileios Skaramagkas, Dimitris Manousos, Nikolaos S. Tachos, Evanthia Tripoliti, Dimitrios I. Fotiadis, and Manolis Tsiknakis. 2022. COLET: A dataset for COgnitive workLoad estimation based on eye-tracking. *Computer Methods and Programs in Biomedicine* 224 (2022), 106989. https://doi.org/10.1016/j.cmpb.2022.106989

[8] Marc Langheinrich. 2001. Privacy by design—principles of privacy-aware ubiquitous systems. In *International conference on ubiquitous computing*. Springer, 273–291. https://doi.org/10.1007/3-540-45427-6_23

[9] H. Brendan McMahan, Eider Moore, Daniel Ramage, and Blaise Agüera y Arcas. 2016. Federated Learning of Deep Networks using Model Averaging. *CoRR* abs/1602.05629 (2016). https://doi.org/10.48550/arXiv.1602.05629 arXiv:1602.05629

[10] John P. Miller, Rohan Taori, Aditi Raghunathan, Shiori Sagawa, Pang Wei Koh, Vaishaal Shankar, Percy Liang, Yair Carmon, and Ludwig Schmidt. 2021. Accuracy on the Line: on the Strong Correlation Between Out-of-Distribution and In-Distribution Generalization. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 7721–7735. https://doi.org/10.48550/arXiv.2107.04649

[11] Maximilian P. Oppelt, Andreas Foltyn, Jessica Deuschel, Nadine R. Lang, Nina Holzer, Bjoern M. Eskofier, and Seung Hee Yang. 2022. ADABase: A Multimodal Dataset for Cognitive Load Estimation. *Sensors* 23, 1 (2022), 340. https://doi.org/10.3390/s23010340

[12] P. Ramakrishnan, B. Balasingam, and F. Biondi. 2021. Cognitive load estimation for adaptive human-machine system automation. In *Learning control*. Elsevier, 35–58. https://doi.org/10.1016/b978-0-12-822314-7.00007-9

DOI: 10.1145/3626705.3631796
Published December 2023