



HAL
open science

On-device learning for ultra-low-power wireless sensors: evaluating the effects of data subsampling

Antoine Bonneau, Fabien Mieleville, Frédéric Le Mouël

► To cite this version:

Antoine Bonneau, Fabien Mieleville, Frédéric Le Mouël. On-device learning for ultra-low-power wireless sensors: evaluating the effects of data subsampling. Journées Colloque National du GDR SoC2, GDR SOC2; INL Lyon, Jun 2023, Lyon, France. hal-04110297

HAL Id: hal-04110297

<https://inria.hal.science/hal-04110297>

Submitted on 30 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On-device learning for ultra-low-power wireless sensors: evaluating the effects of data subsampling

Antoine Bonneau^{1,2,†} Fabien Mieyeville¹ Frédéric Le Mouél²

¹ Univ Lyon, Université Claude Bernard Lyon 1, INSA Lyon, Ecole Centrale de Lyon, CNRS, Ampère, UMR5005, 69622 Villeurbanne, France

² Univ Lyon, INSA Lyon, Inria, CITI, EA3720, 69621 Villeurbanne, France

†antoine.bonneau@insa-lyon.fr

Abstract—Integrating Artificial Intelligence (AI) into embedded systems is critical for the development of viable Ambient Intelligence (AmI). However, the energy requirements of current AI computations are not compatible with the limited resources of AmI devices and global sustainability goals. To address this issue, we propose using Federated Learning across multiple intermittent embedded systems to distribute data gathering, storage, and load balancing, thus enabling the system to assess and adapt appropriately to the environment. We initially focus on relevant real-world data acquisition to check the minimum data required by a local node in a federation. Our preliminary results demonstrate a direct correlation between signal decimation, a decrease in both training time and energy usage, and a collapsing threshold limit in accuracy.

Index Terms—On-Device Learning, Intermittent Learning, Federated Learning

I. INTRODUCTION

The increasing prevalence of embedded systems has made the environment more intelligent but presents both opportunities and challenges. Ambient Intelligence (AmI) and Artificial Intelligence (AI) are intertwined, as the growing autonomy of AmI devices has spurred the deployment of connected objects that generate large amounts of data suitable for AI. However, current AI computation energy requirements are not suited to the limited resources of AmI devices and global sustainability goals.

Initially, the sensors and actuators do not require significant computing or memory capabilities. Therefore, current solutions for introducing intelligence rely on cloud-based systems. However, such solutions cannot be considered embedded in the environment. As low-end IoT devices become more powerful and energy-efficient, it may be worth considering relocating computations away from the cloud. Edge computing represents a step in this direction, but it still faces limitations that prevent the use of traditional algorithms.

To address this, we propose using Federated Learning (FL). Even though FL is mainly used in scenarios where confidentiality is crucial, this paradigm can also distribute the computational load among sensors, adapt duty cycling to energy availability, and prolong sensors' lifespan. The AmI version of FL requires the ability to synchronize with

surrounding devices and distribute the models aggregation among active participants.

Intelligent sensors must *prioritize* various tasks (e.g., monitoring, learning, inference, communication) and *adjust to energy availability*. One of these tasks is *collecting relevant data*. Our primary area of attention is this task, and we have first focused on subsampling incoming data as an effective approach to reduce computational requirements.

II. RELATED WORKS

Previous works have shown that it is possible to use advanced AI solutions despite the memory constraints of AmI low-end devices. Lin et al. [1] has successfully fine-tuned a human recognition model on an STM32F746 using less than 256 KB of RAM and 1 MB of Flash. Profentzas et al. [2] optimized a pre-trained activity-recognition neural network on an nRF-52840 SoC featuring: a 32-bit ARM Cortex-M4 with FPU at 64 MHz, 256 KB of RAM and 1 MB of Flash.

Lee et al. [3] introduced intermittent learning as a way for energy-harvested computing platforms to perform machine learning tasks effectively and efficiently. Among the frameworks for intermittent learning that later emerged, [4] developed *REHASH*, a tool helping with the design of adaptive IoT workflow heuristics. They evaluated their work by running activity recognition and greenhouse monitoring applications using an MSP430FR5994 Launchpad.

Another category of work has addressed the energy constraints imposed on intelligent sensors and the distribution of AI calculations in embedded systems. Dai et al. [5], with *DispFL*, have proposed to act on communication costs using personalized FL and decentralized learning. Costa et al. [6] designed a framework to enable the deployment of decentralized learning in resource-constrained devices. They evaluated their contribution on a STM32L4R5ZIT6 MCU over some Artificial Neural Network models.

However, to the best of our knowledge, there has been no prior attempt to combine these distinct approaches on low-capacity and ultra-low-power devices.

III. METHODOLOGY

Different approaches exist to tailor AI models and their life cycles to meet low-energy and high-constraint requirements.

Dynamic Hardware Management is a possible model-agnostic technological lever for performance/energy trade-offs. Recent *MCUs* have different operating modes balancing performance and autonomy. *ADCs* sampling duration can be decreased at the expense of measurement accuracy.

Quantization, *Pruning*, and *Encoding* are in essence memory related techniques. By simplifying or compressing the AI model’s parameters or representations you can squeeze down models to smaller sizes and induce faster computation.

Communication is the most energy-intensive task for a low-end device. As we aim to use distributed learning, optimizing exchanges with peers will be necessary. We have to evaluate specialized FL aggregators such as *Compressed Sensing* or *Adaptive Neighbor Matching* based ones.

AI being inherently data processing, *Data Reduction* can alleviate both memory and complexity issues in training and inference. *Adaptive Subsampling* and *Time Series Segmentation* exemplify this concept and preserve the relevant properties of the monitored phenomenon.

We selected *Data Reduction* as our primary area of compromise and measuring the influence of diverse subsampling techniques as our first task in this direction.

IV. EXPERIMENTAL SETUP

We are targeting Structural Health Monitoring (SHM) applications and have begun working with a dataset containing three mechanical physical signals from a vibrating system, built for a previous research [7] at the Ampere laboratory.

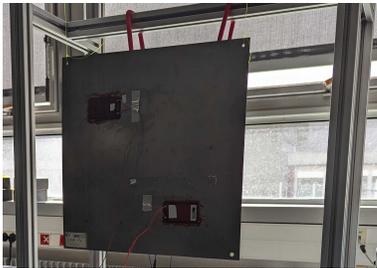


Fig. 1. Experimental vibrating testbed

The testbed, illustrated in Figure 1, consists of an aluminum plate with a modal shaker (M20/PA300E from Data Physics) attached near its center, flanked by two piezoelectric patches DuraACT P-876.A15 on either side. A laser vibrometer is used to measure the ground truth of the vibrations. Each dataset comprises recordings of an excitation sweep applied to the vibrating pot from 10 to 500 Hz at a sampling frequency of 1600 Hz. Each recording is represented by a set of timestamps, tension signals from the two piezoelectric sensors, and the vibrometer tension measured at the pot and plate contact point.

V. PRELIMINARY RESULTS

We utilized a classification model based on Fully Convolved Neural Networks (FCN) [8] to pinpoint the piezoelectric patch of signals’ origin. The classifier was trained with signal decimation (pick 1 record every 2, 4, ...) applied

on the time series data, and the experiment was repeated until the model’s accuracy decreased. To compare subsampling influence on the training, we calculated accuracy, timed it, and measured power consumption using a software wattmeter.

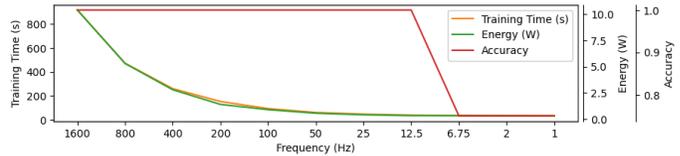


Fig. 2. Effects of signal decimation on piezoelectric classification

The experiment was conducted on an *HP Precision 5570* laptop using 15 CPUs simultaneously. A linear correlation was observed in Figure 2 between the signal decimation resultant sampling frequency and the decrease in training time and power usage down to a plateau corresponding to the laptop idle energy consumption. The accuracy decreases below 10Hz since there isn’t enough data to support effective learning.

VI. FUTURE WORKS

Future works involve comparing this data selection approach with random and variations-aware strategies. We plan to first move the training onto an STM32F302R8 device with a lightweight ML framework and then distribute it over a network.

Our current experiment is limited mainly by its size. We plan to set up a larger modular plate including a movable vibration source, additional piezoelectric patches, and the ability to simulate material changes over short periods.

Following our examination of the *Data Reduction* trade-offs, we aim to address another challenging FL task. Specifically, we plan to evaluate the costs of different synchronization strategies for distributed learning. This will involve coupling energy models from *ns3* to the *Flower* FL framework and then validating these simulated strategies with our vibrating system.

REFERENCES

- [1] J. Lin, L. Zhu, W.-M. Chen, W.-C. Wang, C. Gan, and S. Han, “On-Device Training Under 256KB Memory,” in *Advances in Neural Information Processing Systems*, vol. 35, Jul. 2022.
- [2] C. Profentzas, M. Almgren, and O. Landsiedel, “MiniLearn: On-Device Learning for Low-Power IoT Devices,” in *Proc. of the International Conf. On Embedded Wireless Systems And Networks*, Sep. 2022.
- [3] S. Lee, B. Islam, Y. Luo, and S. Nirjon, “Intermittent Learning: On-Device Machine Learning on Intermittently Powered System,” in *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiqu. Tech.*, Dec. 2019.
- [4] A. Bakar, A. G. Ross, K. S. Yildirim, and J. Hester, “REHASH: A Flexible, Developer Focused, Heuristic Adaptation Platform for Intermittently Powered Computing,” in *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiqu. Tech.*, Sep. 2021.
- [5] R. Dai, L. Shen, F. He, X. Tian, and D. Tao, “DisPFL: Towards communication-efficient personalized federated learning via decentralized sparse training,” in *Proc. of the 39th ICML*, Jul. 2022.
- [6] D. Costa, M. Costa, and S. Pinto, “Train Me If You Can: Decentralized Learning on the Deep Edge,” *Applied Sciences*, May 2022.
- [7] M. Zieliński, Piotr, “Système distribué actif sans fil basse consommation pour l’amortissement des vibrations,” PhD, Université de Lyon, Université de Lyon, Oct. 2015, <https://www.theses.fr/2015ECDL0029>.
- [8] Z. Wang, W. Yan, and T. Oates, “Time series classification from scratch with deep neural networks: A strong baseline,” in *International Joint Conference on Neural Networks (IJCNN)*, May 2017.