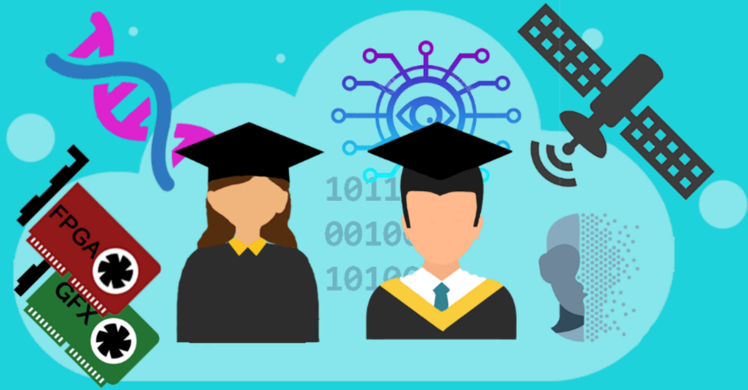
Diploma Thesis

Microproccessors and Digital Systems Laboratory



Communication Aware Model Compression for Distributed Learning

Distributed model training and federated learning have become cornerstones in Machine Learning (ML), especially as datasets and model complexities continue to grow. In distributed training, instead of relying on a single machine, the learning task is split across multiple devices or nodes. Each node processes a portion of the data and contributes to updating the shared model parameters. This parallelism enables the training of larger models and faster convergence, making it feasible to handle massive datasets and compute-heavy tasks.

A major challenge in distributed training is the communication overhead caused by the frequent exchange of model parameters and gradients between nodes, especially during the synchronization phase. As models increase in size and complexity, this exchange becomes more frequent and data-intensive, resulting in longer training times and inefficient use of network resources.

This thesis aims to design communication-aware model compression techniques tailored for distributed learning environments. The research will explore data compression methods and gradient approximation strategies to reduce the volume of exchanged data. Approaches may include quantization, sparsification, and lossy compression methods, as well as novel techniques to balance model accuracy with communication efficiency. By reducing the amount of data exchanged between devices, this work will significantly accelerate distributed training and enable more scalable machine learning systems.

Prerequisites: Python, Basic concepts of AI/ML, Distributed Computing

Desirable: Familiarity with Linux, Bash scripting

Related Material: https://doi.org/10.1109/ICDCS60910.2024.00060, https://doi.org/10.1109/TPDS.2023.3240883, https://doi.org/10.1109/TCAD.2023.3307459

Contact information:

Asst. Prof. Sotiris Xydis (sxydis@microlab.ntua.gr)

Kostis Iliakis (kiliakis@microlab.ntua.gr)

Panagiotis Eleftherakis (pelef@microlab.ntua.gr)

George Anagnostopoulos (geoanagn@microlab.ntua.gr)