Master's Thesis

Structure-Preserving Downsampling of Machine Learning Data Sets and Streams

Fachhochschule Südwestfalen
University of Applied Sciences

(SW)²E-LAB
SOUTH WESTPHALIA SOFTWARE ENGINEERING LAB

Motivation:

The rapid advancement of Artificial Intelligence, particularly Large Language Models (LLMs), is driving an enormous increase in computational, memory, and energy demands. New data centers are being built worldwide to train increasingly complex AI models on ever-growing datasets.

A promising approach to reducing these demands is downsampling of training data without significantly compromising model quality. However, existing methods often struggle to preserve essential data structures or require computationally expensive processing. Scalable methods are therefore needed that efficiently reduce large datasets while preserving their key structural properties.

blueAIC develops Online Machine Learning technologies aimed at reducing the resource demands of AI applications. Based on its adaptive clustering algorithm CellStream, blueAIC is developing a new approach to structure-preserving downsampling. This master's thesis will systematically evaluate the approach across relevant scenarios and benchmark it against state-of-the-art methods.

Tasks:

- **Literature review:** Identify relevant state-of-the-art downsampling methods and application scenarios for batch and streaming data processing.
- **Benchmark design:** Select representative datasets, application scenarios, reference methods, and evaluation metrics.
- **Experimental evaluation:** Benchmark CellStream-based downsampling against established methods in terms of data reduction, computational efficiency, memory consumption, structural preservation, and downstream machine learning performance.

Requirements:

- Master's student in Digital Technologies or Systems Engineering & Engineering Management
- Good programming skills in Python
- Solid knowledge of Machine Learning and data analysis
- Ideally, initial experience with ML frameworks such as TensorFlow, PyTorch, or MLPro

Department of Electrical Engineering
Prof. Dr. Andreas Wübbeke
Phone: +49 (0)2921/378-3578
E-Mail: wuebbeke.andreas@fh-swf.de
Lübecker Ring 2
59494 Soest

blueAIC
Let's innovate

Wir geben Impulse

2026-10-13