technische universität dortmund

SFB 876 Providing Information by Resource-Constrained Data Analysis

# Project A3

## Methods for Efficient Resource Utilization in Machine Learning Algorithms

**Prof. Dr. Jian-Jia Chen, Prof. Dr. Jörg Rahnenführer**

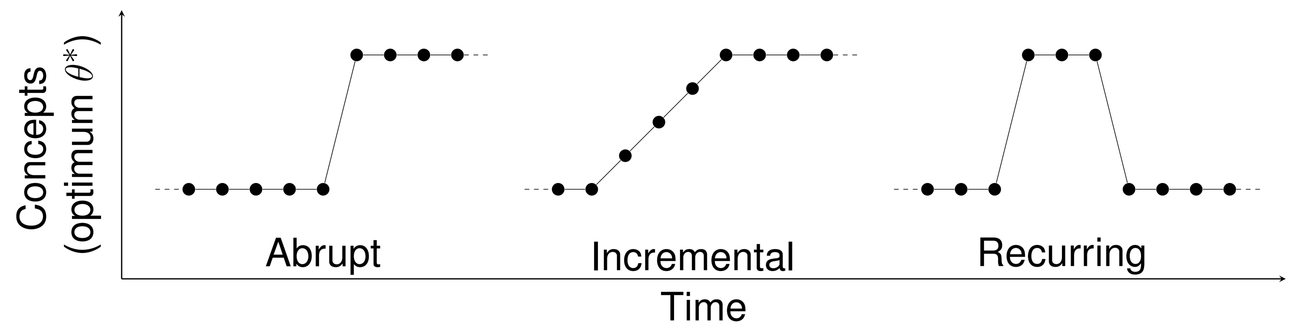
# Problem

### Optimization Scenarios for MBO (Model-Based Optimization)

- Efficiently optimize black box problems
  - Efficiently select the best algorithms for a specific problem
  - Efficiently optimize the performance of machine learning algorithms

### Challenges

- Data streams with concept drift: Relationship between features and labels changes over time



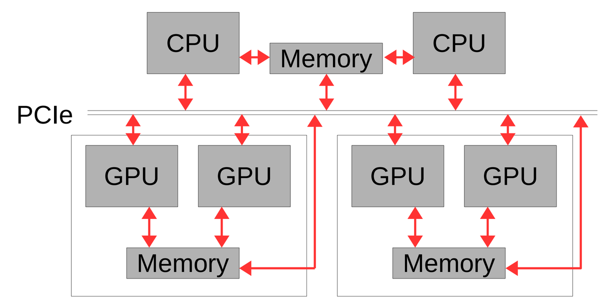

- Small sample sizes: Data sets are too small to obtain reliable predictions and to reliably estimate performance of models
- Expensive labels: High costs of additional labels to increase size of data set
- Available resources: Resources are abundant and heterogeneous and should be used efficiently

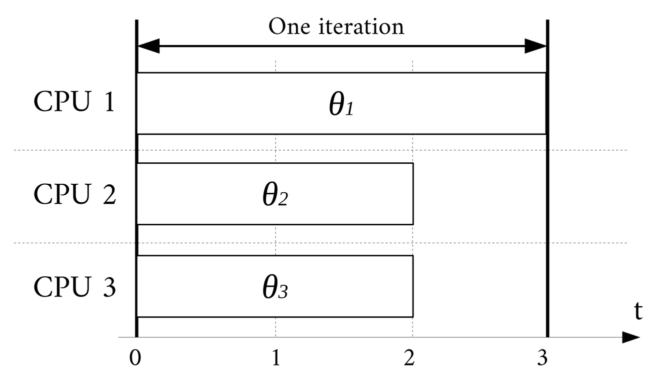
### Problems that Fit the Black-Box Paradigm

- Optimizing analysis pipeline for virus detection with the PAMONO sensor, with optimization regarding energy consumption and runtime (B2)
- Tuning deep neural networks (B2)
- Mapping of TensorFlow pipeline to improve efficiency (B2)
- Tuning the prediction of data rate of mobile devices in cellular networks (A4)
- Industrial production processes: Optimizing the prediction of product quality (B3)
- Optimizing the prediction of travel-times by guiding the selection of the best algorithm that considers environmental conditions (B4)
- Survival prediction of heterogeneous cancer cohorts (A3)

### Challenges of Executing MBO Efficiently

- Minimization of idle times
- Multiple computation units (i.e. multicore systems)
- Heterogeneous computation units (i.e. GPU, processors)
- Different memory and communication overheads
- Parallelization and concurrent execution



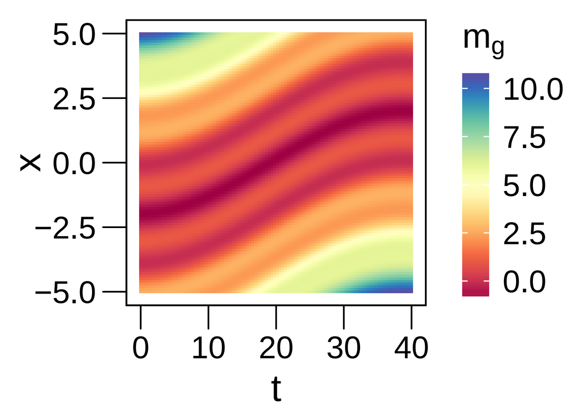

# Methodology

### Methods for Including Concept Drift in MBO

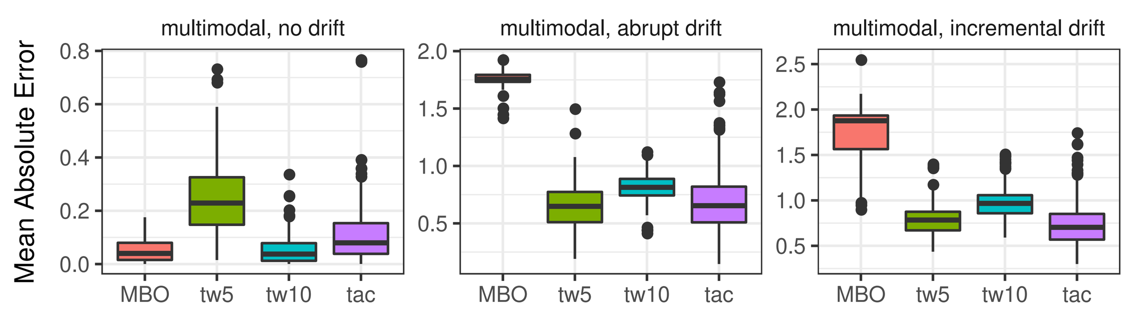
- **Naïve approach:** If concept drift detected: Discard previous evaluations, refit surrogate model
- **Historical approach:** If concept drift detected: Keep best previous settings and refit surrogate model
- **Fixed window approach:** Consider only evaluations observed in the last $t_\Delta$ time units
- **Weighted window approach:** Fixed Window + newer evaluations are weighted higher for surrogate model
- **Flexible window approach:** Window approach with adaptive $t_\Delta$
- **Time-as-covariate:** Include time as additional dimension $t$ in the surrogate model
- **Window and time-as-covariate:** Window + Time-as-covariate

### Preliminary Study

- Synthetic functions with Concept Drift:
  - One global minimum
  - Many local minima
  - Forms of drift: no-drift, abrupt, incremental

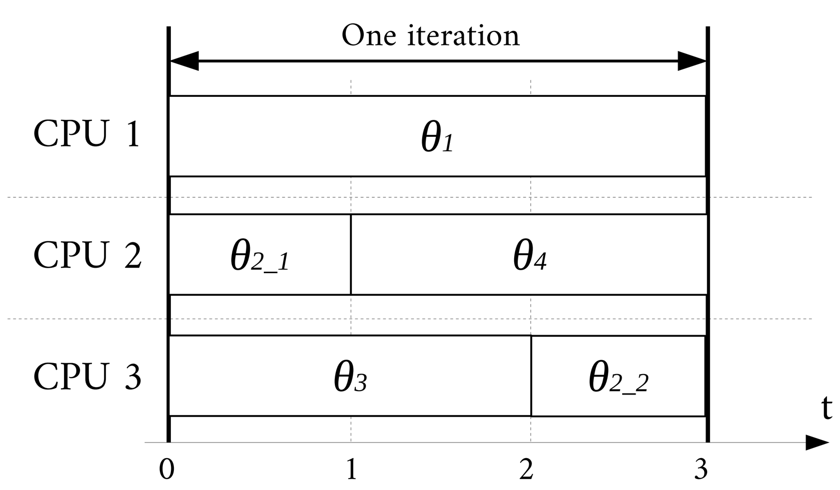



- Comparison of implementations:
  - `MBO`: ignoring the concept drift
  - `tw5`/`tw10`: MBO with Fixed Window Approach
  - `tac`: MBO with Time-as-covariate
- Results:
  - Ignoring the drift deteriorates performance
  - `tw10`, `tac`: also good performance on no-drift
  - `tac` general good performance




### Utilize Resources Efficiently

- Migration mechanism between different CPUs will be established, so that the pieces of idle time of the CPUs can be utilized efficiently



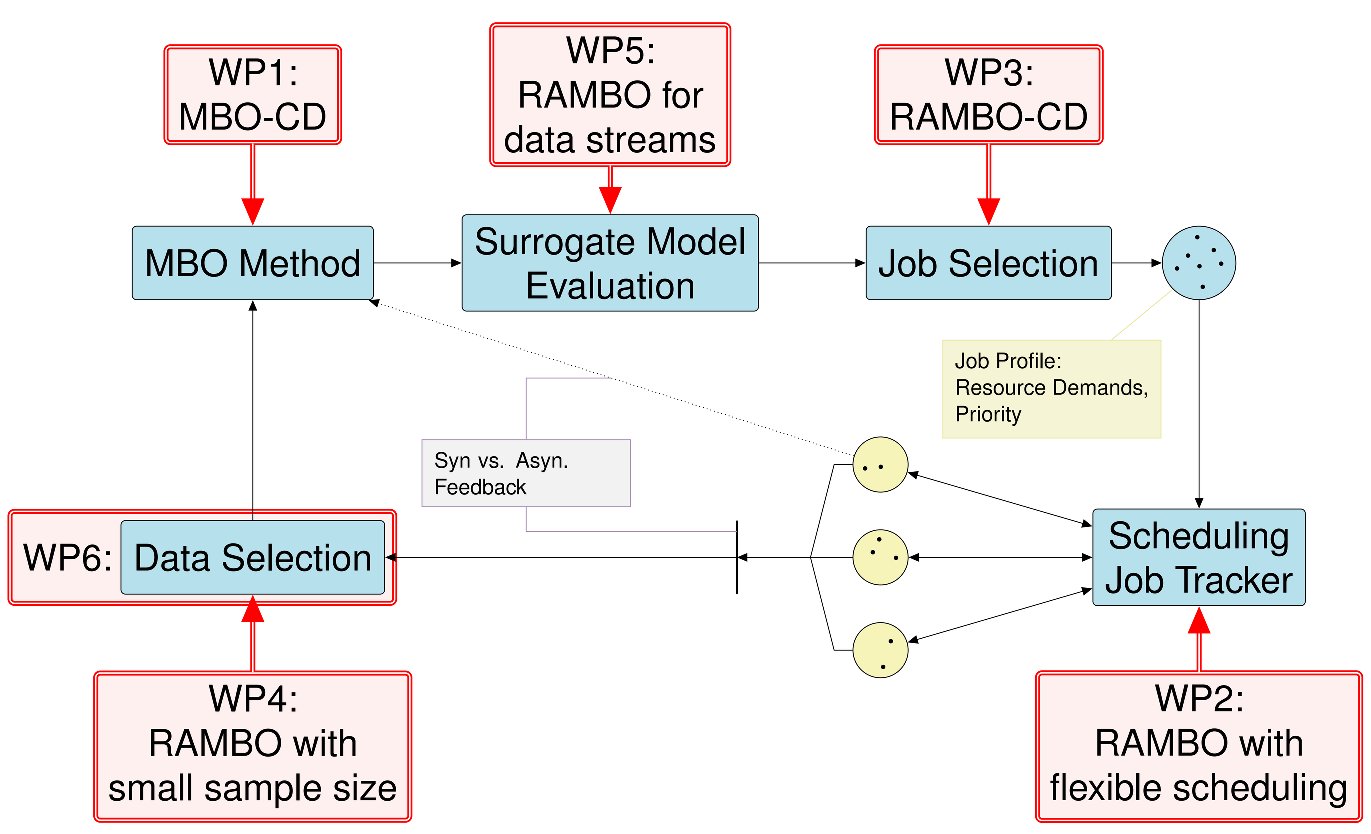

- Cancellation (CEL) of *Expected Improvement* (EI) will be explored, e.g., $EI\{\theta_1, \theta_2\} = EI(\theta_1) + EI(\theta_2) - CEL$ (w.r.t. $dis(\theta_1, \theta_2)$)
- Relationship between candidates has to be considered to maximize the total EI
- EI under concept drift has to be carefully managed, e.g., whether an unfinished evaluation of a point should be aborted or continued

# Planned Research

**WP1: MBO with concept drift (MBO-CD)**
- Implement different methods for CD in MBO
- Evaluate methods on benchmarks, both with synthetic function and real data sets
- Find principles about relationship between data situation and best method

**WP2: RAMBO with flexible scheduling**
- Develop scheduling methods to use available resources efficiently and adaptively
- Optimize for communication and memory access including distributed computing, multicore platforms, GPUs, and accelerated machine learning hardware

**WP3: RAMBO with concept drift (RAMBO-CD)**
- Combine resource-aware MBO with CD
- Explore the impact of dynamic EI in RAMBO
- Develop and evaluate aborting strategies for jobs that were selected earlier for now outdated concepts




Scenarios and Cooperations:

- Mobile Data Rate prediction (A4)
- TensorFlow Pipeline (B2)
- PAMONO Pipeline (B2)
- Tuning Deep Neural Networks (B2)
- Industrial Processes (B3)
- Travel Time Prediction (B4)
- Survival Prediction

**WP4: RAMBO with augmented samples in case of small sample size**
- Investigate methods for combining real data with additional simulated data
- Optimize hyperparameters and parameters of the simulator

**WP5: RAMBO for data streams**
- Develop starting strategy with high exploration and ignoring CD
- Develop update strategy that does not waste time if no concept drift is present
- Predict best hyperparameters for future concepts for incremental drifts

**WP6: RAMBO-CD with interactive selection of samples**
- Augment Active Learning with an automatic algorithm configuration
- Interactively select the next data points to be labeled considering the currently best machine learning method

CS 12 computer science 12

fakultät statistik