HyperUCB: Hyperparameter Optimization using Contextual Bandits

Maryam Tavakol, Sebastian Mair, Katharina Morik

maryam.tavakol@tu-dortmund.de, mair@leuphana.de, katharina.morik@tu-dortmund.de

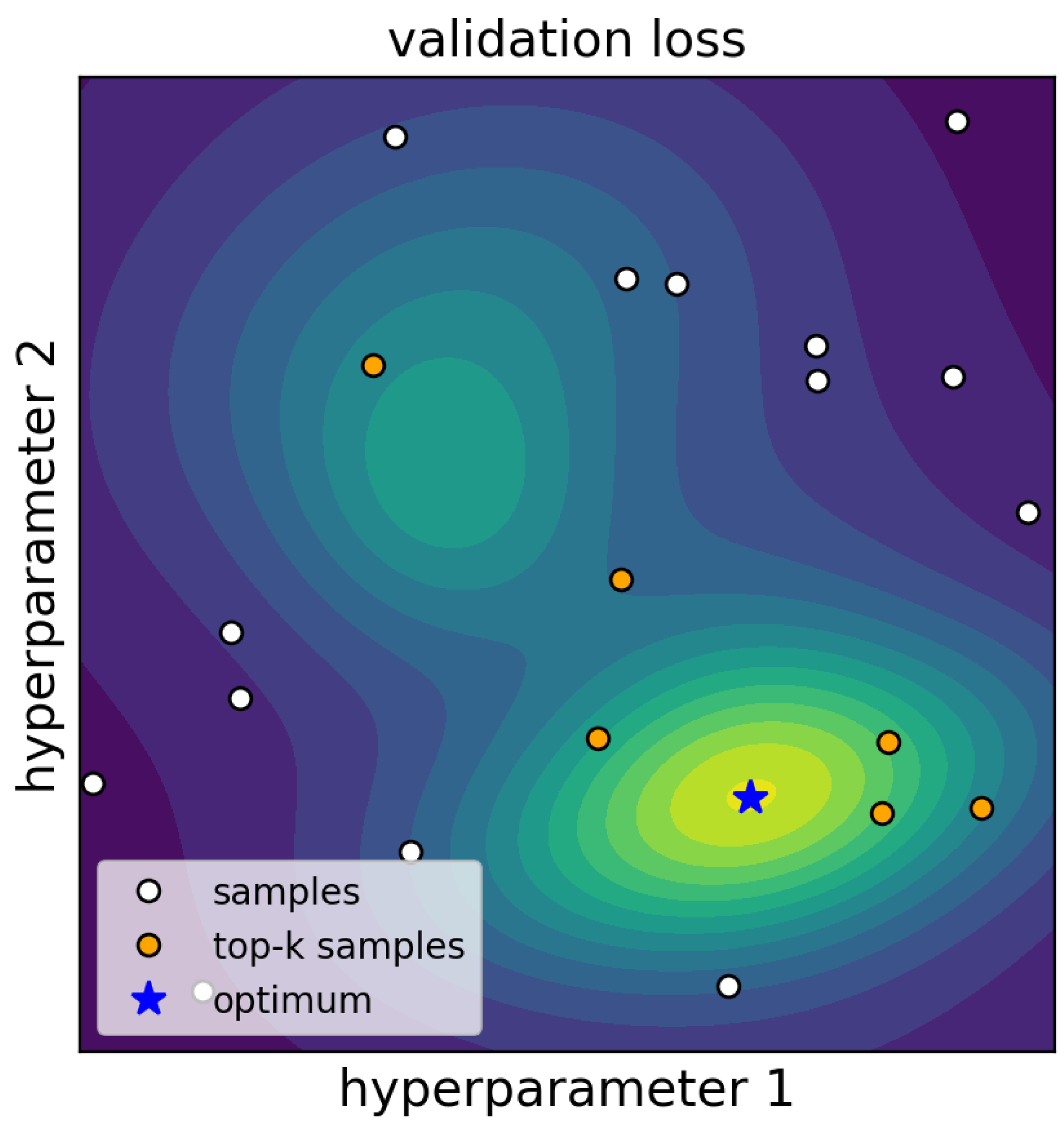
Summary

- HyperUCB is a contextual bandit extension for Hyperband
- Hyperparameters are pre-evaluated using a UCB strategy
- Only the k-best configurations are actually evaluated

⇒ The sampling is guided towards more promising area

Motivation

- Performance in ML highly depends on the hyperparameters
- Hyperparameters are usually tuned via grid/random search
- Hyperband (HB) speeds up random search while optimizing computational resources
- but HB does not leverage the information of previous runs



Contextual HyperUCB

HyperUCB extends HB to contextual bandit setting:

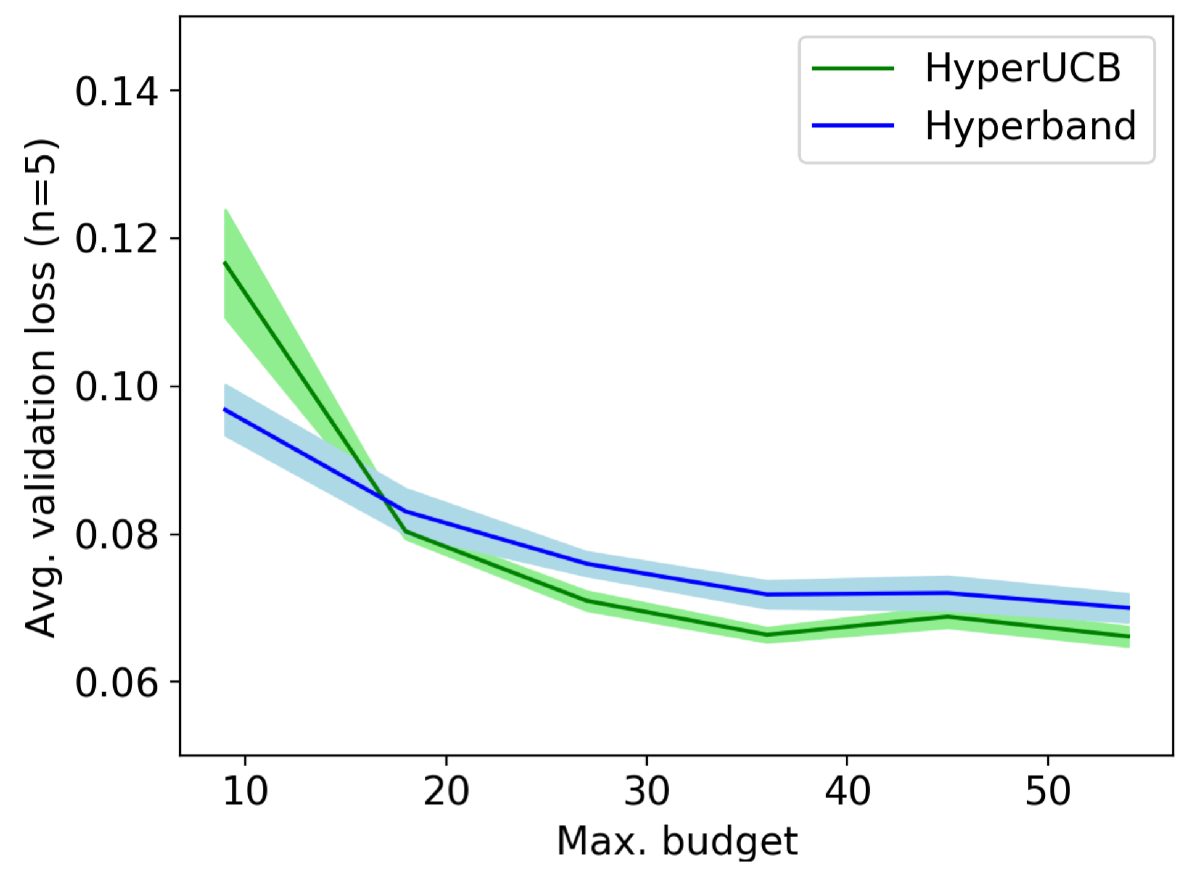
- Given a fixed budget B
- allocate B to s_{max} number of iterations
- in every iteration s:
  - ~~sample n_s configurations~~
  - sample n configurations and choose n_s with highest UCB
  - compute validation loss for configurations until B is exhausted
    - ~~run successive halving to choose top-k~~
    - run contextual UCB to choose top-k

⇒ Learn a model to keep track of previous evaluations
⇒ Only execute promising hyperparameter configurations

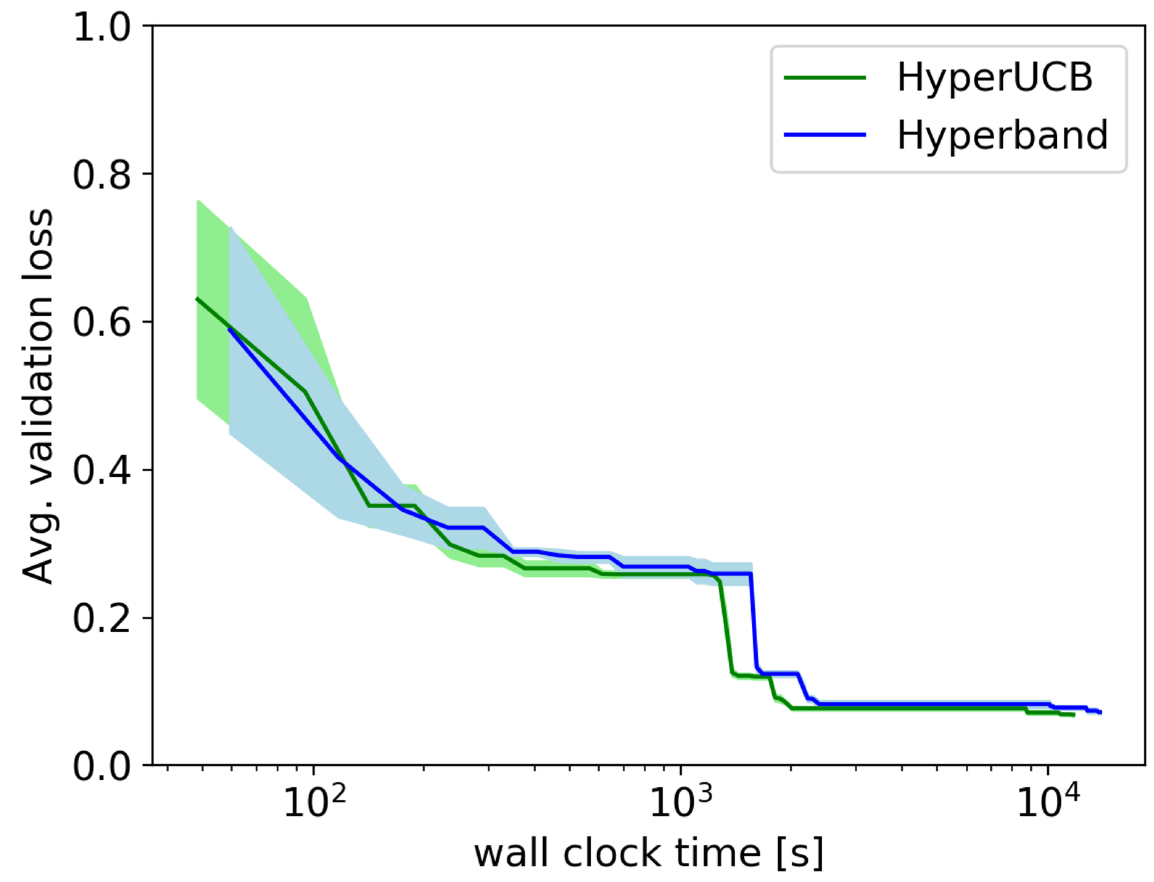
Experiments

- MNIST data of handwritten digits (60k train, 10k test)
- Multi-layer perceptron with categorical cross entropy loss
- Validation loss is evaluated on the test data
- Minimum budget corresponds to 100 mini-batches of size 100
- The model has four hyperparameters:

Hyperparameter	Range	Type
learning rate	$[0.0001, 1]$	float
# hidden layers	$\{1, 2, 3, 4, 5\}$	integer
# neurons	$\{16, 32, \ldots, 512\}$	integer
activation	{relu, tanh, sigmoid}	categorical



⇒ HyperUCB outperforms HB for a budget greater than 19



⇒ HyperUCB is as fast or faster than Hyperband

Future Work

- Use a kernalized bandit to capture non-linearity
- Derive theoretical regret bounds
- Extend the experimental setup

technische universität dortmund

LEUPHANA UNIVERSITÄT LÜNEBURG

SFB 876 Providing Information by Resource-Constrained Data Analysis