Laboratory for Parallel Programming

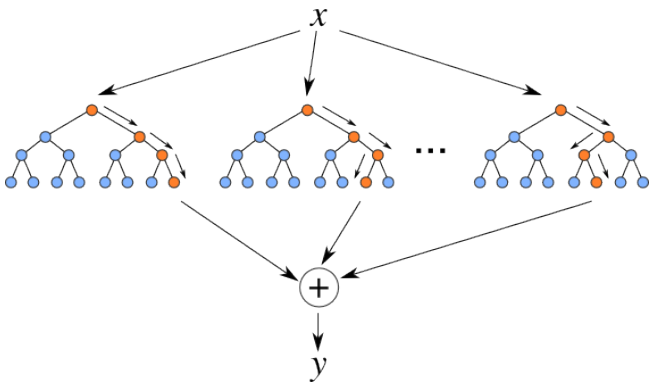
TECHNISCHE UNIVERSITÄT DARMSTADT

Bachelor's Thesis

Multi-target regression for performance modeling

Motivation

Analyzing the runtime performance and other metrics of parallel programs is crucial for their development and deployment on computing clusters. Existing solutions like ExtraP use an empiric approach to model these performance functions based on multiple measurements for different parameters. However such approaches struggle with an increasing amount of parameters.
Machine Learning solutions have gained widespread application in recent years and linear regression became one of the most common use cases of supervised learning. Though in order to construct performance functions we need to predict several not only a single target/output value. Multi-output or multi-target regression approaches generally can be separated into two different categories: problem transformation methods and algorithm adaption methods. Especially latter methods such as multi-target regression trees, multi-output SVR and rule methods are very promising for creating performance models with multiple parameters.



A random forest for regression.

Task

The target of this thesis project is to implement and evaluate a multi-target regression approach for performance modeling of parallel programs. In order to achieve this goal the student will work on the following tasks:

- Research the state of the art of multi-target regression approaches
- Identify the most promising approach and implement it, or adapt an existing solution
- Evaluate this solution with synthetic or open source data and compare it to the existing approach

Requirements

- Java, R or Python
- Parallel Programming
- Optional: Experience in Machine Learning

Contact

Marcus Ritter <ritter@cs.tu-darmstadt.de>

References

[1] Borchani, Hanen and Varando, Gherardo and Bielza, Concha and Larrañaga, Pedro. A survey on multi-output regression. 2015

Issue date: **February 2019**