Masterarbeit

Karlsruher Institut für Technologie

Institut für Programmstrukturen und Datenorganisation (IPD)
Lehrstuhl für Systeme der Informationsverwaltung, Prof. Böhm

Masterarbeit

Anytime Tradeoff Strategies with Multiple Targets

Anytime algorithms are a class of algorithms that can return a valid solution to a problem, even if interrupted before completion. In general, the solutions obtained via early interruption are of lower quality than the final solution, but require less resources -- in terms of number of iterations, for example. This class of algorithms is particularly handy in settings where resources are limited. For instance, in predictive maintenance, results are typically required in real-time, and the data is available as an infinite, ever-evolving stream. When the flow of the stream increases, resources need to scale, otherwise traditional algorithms may not return results in time. Anytime algorithms are a natural choice in this case, as they deliver approximate solutions of increasing quality with respect to computation time. Thus, they offer a tradeoff mechanism between runtime and quality of the solution.

Anytime algorithms have become an ubiquitous paradigm of machine learning: Various approaches exist for clustering, classification, outlier detection, dependency estimation, etc. The existing contributions vary in the depth of their analysis: while some algorithms are based on heuristics, other provide a model of the result quality one can expect for a given computational investment. Nonetheless, an unaddressed question is the optimization of the output quality of anytime algorithms w.r.t. multiple targets: Given an anytime algorithm A and a set of problems $P = \{p_1, \dots, p_n\}$, how to allocate resources to the problems to maximize the global quality of the set of solutions $S = \{s_1, \dots, s_n\}$?

The focus of this thesis is the development of efficient meta-strategies for anytime algorithms with multiple target problems. In particular, the following aspects are of interest:

- Given a global computational budget, how should one dispatch resources between problems to optimize the global quality of the set of solutions? Vice versa: Given a target quality, how to meet this quality criterion globally under parsimonious resource consumption?
- The inherent “difficulty” of the problems in P may vary and is unknown a priori. How to take into account the apparent difficulty to solve the problems in the allocation strategies?
- Simultaneously, the “interestingness” of the solutions in S may vary. For example, when estimating correlation between many pairs, one is typically more interested in strong correlations than weak correlations. How to bias strategies towards solutions with an apparently more interesting outcome?

This results in the following tasks:

- Exploratory analysis of anytime algorithms, approximate computing and multi-objective optimization.
- Development of a framework with a set of meta-strategies for the allocation of resource to optimize quality and vice versa, considering the difficulty of problems and the interestingness of solutions.
- Implementation and validation of this framework w.r.t. existing anytime algorithms via systematic experiments.

Throughout this work, the student will acquire a deep knowledge about machine learning under parsimonious resource consumption and quality/runtime tradeoffs. The student will produce useful pieces of open-source software that will be used in the research community. The student will learn how to conduct controlled experiments to compare results of her/his work to the state of art.

Ansprechpartner

Edouard Fouché, M. Sc.	edouard.fouche@kit.edu	+49 721 608-47337	Raum: 342
Am Fasanengarten 5	76131 Karlsruhe	Gebäude: 50.34	