Karlsruher Institut für Technologie

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

Bachelor/Master Thesis

Selecting SAT Instances to Evaluate Solvers

SAT solving is one of the fundamental problems in computer science, with many real-world applications. Due to the NP-hardness of this problem, there is active research on efficient search strategies, e.g., using conflict-driven clause learning (CDCL) approaches. Solver portfolios even combine different solvers and flexibly decide which solver to use. To evaluate solvers, international competitions like the SAT Competition 2020 use a variety of SAT instances from different application areas. Solver runtime is the key performance measure, with a cut-off time and a penalty for long-running solvers. As in many areas of computer science, there is no single algorithm that performs best for all problem instances. However, getting insights into the reasons for good or bad solver performance is of vital interest for the community. Besides determining the relationship of solvers to individual instances, this also raises the question which SAT instances to use for a representative benchmark.

Techniques from data science and machine learning can help to analyze SAT instances. For example, one can cluster instances to find groups with similar properties. Using classification and regression techniques, one can relate properties of SAT instances to the performance of solvers. With such performance estimates, one can decide 1) if a solver should be included into a portfolio and 2) when it should be used within the portfolio. As benchmarking solvers on instances is expensive, it is important to keep a benchmark set as small as possible. At the same time, the benchmark should still be able to discriminate between solvers with high confidence. Besides the portfolio use case, improved benchmark sets are also interesting for SAT competitions.

The goal of this thesis is to develop an approach to create a small, yet discriminative set of SAT instances. The following questions are particularly interesting:

- How can one obtain a representative set of SAT instances in an automated manner?
- Which features are most important in instance selection?
- How does a clustering of instances relate to a clustering of solvers?

The following steps are part of your thesis:

- Review literature about performance prediction and instance selection for SAT solving.
- Develop an approach to create a representative benchmark of SAT instances.
- Evaluate your approach experimentally. You can use our server infrastructure for that.

During your work on this thesis, you will gain insights into current SAT solving research. You will acquire knowledge about state-of-the-art machine learning libraries. You will gain experience in running and evaluating large-scale scientific experiments. The thesis is part of an ongoing research project, and you will get supervision from the SAT perspective (Markus Iser, chair of Prof. Sanders) as well as the ML perspective (Jakob Bach, chair of Prof. Böhm).

You can write the thesis in English or German. Prior experience with programming in Python is beneficial, but not necessary.

Contact
Jakob Bach jakob.bach@kit.edu +49 721 608-43911 Room: 363
Am Fasanengarten 5 76131 Karlsruhe Building: 50.34

Bachelor/Master Thesis