

Proposal for a M2 Research Internship 2019: Automated Algorithm Configuration

October 11, 2018

Supervisors: The student will be supervised by

- Carola Doerr (CNRS and LIP6, Sorbonne Université) and
- Manuel López-Ibáñez (Alliance Manchester Business School, University of Manchester, UK)

Contact:

Carola: <http://www-ia.lip6.fr/~doerr/>, *E-mail:* Carola.Doerr@mpi-inf.mpg.de

Manuel: <http://lopez-ibanez.eu/>, *E-mail:* manuel.lopez-ibanez@manchester.ac.uk

Keywords: Optimization, Algorithm Configuration, Algorithm Selection

General Information: The student will be primarily hosted by LIP6, on the Jussieu campus of Sorbonne University, research visits to Manchester are advisable, but not mandatory; the travel cost will be reimbursed.

We aim to achieve results that are publishable in high-quality international conferences and journals.

We particularly welcome applications from students interested in continuing their internship with a PhD under our supervision.

Prerequisites: The student should bring some basic background about algorithms and probability theory, courses in optimization would be a plus, but do not constitute a formal requirement.

Some programming will be required in the course of the project. The *irace* package is implemented in R; some basic knowledge of R is therefore advisable, but not strictly required. We do require, however, *some* programming experience, as well as willingness to conduct empirical work.

Our working language is English, we therefore expect fluency in written and spoken English.

Description of the Topic: State-of-the-art optimization algorithms typically come with a set of parameters that need to be set by the user in order to configure the algorithm to the problem that he/she wants to solve. This parametrization typically offers a great flexibility in the algorithmic behavior, making the underlying approach applicable to broad ranges of problems or problem instances. As a downside, the user is left with the task of choosing these parameters, and it is not unusual that a good algorithmic performance requires a solid understanding of how different parameters choices impact the behavior of the algorithm. Obtaining such expert knowledge is a tedious task, which often requires a significant amount of time and experience, a prerequisite that is often not met in practice.

Automated algorithm configuration tools have been developed to aid the users in choosing suitable parameter settings for their applications, thus trading expert's time and knowledge against computational resources and a less person-biased, more scientific approach. Numerous applications in various industrial and academic contexts have shown significant advantages of such methods. Among the most widely used configuration tools are *irace* [LDC⁺16], *ParamILS* [HHLBS09], *SMAC* [HHLB11], and *GGA* [AMS⁺15]. Our project will focus on the first-mentioned tool, the *irace*

R package, for which Manuel is a co-developer. The main working principle of *irace* is an iterative racing approach: the search for good algorithmic configurations is divided into several steps. In each step a set of potential parameter configurations is tested on a small set of instances. Statistical tests indicate non-promising or poor-performing configurations, which are removed from the race. At the end of each step the surviving configurations are used to sample new configurations, which are added to the race of the next step. The race ends when a user-defined budget is exhausted, and the surviving configurations are suggested to the user. Statistics about the relevance of each individual parameter and dependencies between them are recorded and can be analyzed in a post-processing analysis. *irace* has been successfully applied in a broad range of contexts, ranging from machine learning over robotics to classical optimization tasks such as traveling salesperson and scheduling problems.

In this project we will address the following questions:

- In the current version of *irace* the initial parameter configurations are samples uniformly at random. *Can we obtain significant performance gains and/or other advantages such as a more robust output from replacing this initialization by a quasi-random procedure?*
- The generation of new configuration samples is based on simple unbiased sampling, thus ignoring information that *irace* has accumulated from previous steps. It seems likely that additional performance gains are possible through a better exploitation of the previous evaluations. *How do other optimization procedures compare to the unbiased generation of new candidate configurations? Is this comparison uniform across different budgets and across different dimensions?*
- *irace* has been developed with a machine learning setting of training and test instances in mind. There is, however, an increasing demand in tuning (hyper-)parameters for single instances, e.g., to understand the dependence of suitable parameter choices on the dimension, cf. [DW18] for an example. Such settings are particularly common in setups where theoretical performance bounds are sought [DD18]. *How well does irace perform on such single-instance settings, and how can we potentially adjust it to cope better with this task?*

References

- [AMS⁺15] Carlos Ansótegui, Yuri Malitsky, Horst Samulowitz, Meinolf Sellmann, and Kevin Tierney. Model-based genetic algorithms for algorithm configuration. In *Proc. of International Conference on Artificial Intelligence (IJCAI'15)*, pages 733–739. AAAI Press, 2015.
- [DD18] Benjamin Doerr and Carola Doerr. Theory of parameter control for discrete black-box optimization: Provable performance gains through dynamic parameter choices. *CoRR*, abs/1804.05650, 2018.
- [DW18] Carola Doerr and Markus Wagner. Simple on-the-fly parameter selection mechanisms for two classical discrete black-box optimization benchmark problems. In *Proc. of Genetic and Evolutionary Computation Conference (GECCO'18)*, pages 943–950. ACM, 2018.
- [HHLB11] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *Proc. of Learning and Intelligent Optimization (LION'11)*, pages 507–523. Springer, 2011.

- [HHLBS09] Frank Hutter, Holger H. Hoos, Kevin Leyton-Brown, and Thomas Stützle. ParamILS: An automatic algorithm configuration framework. *Journal of Artificial Intelligence Research*, 36:267–306, 2009.
- [LDC⁺16] Manuel López-Ibáñez, Jérémie Dubois-Lacoste, Leslie Pérez Cáceres, Mauro Birattari, and Thomas Stützle. The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives*, 3:43–58, 2016.