Guest Editorial Special Issue on Benchmarking Sampling-Based Optimization Heuristics: Methodology and Software

Thomas Bäck, Carola Doerr, Bernhard Sendhoff, Thomas Stützle

Guest Editorial
Special Issue on Benchmarking Sampling-Based Optimization Heuristics: Methodology and Software

Thomas Bäck, *Fellow, IEEE,* Carola Doerr, *member, IEEE,*
Bernhard Sendhoff, *Fellow, IEEE,* Thomas Stützle, *Fellow, IEEE,*

Benchmarking provides an essential ground base for adequately assessing and comparing evolutionary computation methods and other optimization algorithms. It allows us to gain insights into strengths and weaknesses of different existing techniques, and consequently design more efficient optimization approaches. The need for good benchmarking practices opens up a broad range of complementary research questions, arising as a byproduct of challenges encountered when optimization methods are assessed. From the selection of representative benchmark problem instances, different algorithms and suitable performance metrics, over efficient experimentation, to a sound evaluation of the benchmark data, these research questions lie at the core of establishing a well-designed and standardized benchmarking procedure.

In recent years, a number of techniques addressing various aspects of benchmarking has become available to assist researchers and users in evolutionary computation. However, most of them are developed independently from one another, without the fully modular benchmarking pipeline in mind. This hinders knowledge transfer between the different research groups and between academic and industrial practitioners of evolutionary computation methods.

This special issue on *Benchmarking Sampling-Based Optimization Heuristics: Methodology and Software (BENCH)* aims to highlight recent advances in benchmarking evolutionary computation methods and to motivate the need for knowledge exchange in view of standardized benchmarking practices. With this issue, we provide a snapshot of the prominent efforts covering different aspects of benchmarking.

Based on a rigorous review process, 14 papers have been accepted to the BENCH special issue. These works cover a broad range of ongoing research activities centered around the benchmarking of evolutionary computation methods. The topics range from the suggestion of new benchmark problems over classical benchmarking studies to performance evaluation and visualization. Five accepted papers consider meta-algorithmic problems such as algorithm selection and configuration, demonstrating a continued interest in blending evolutionary computation methods with machine learning approaches. We briefly describe all selected papers in the following paragraphs.

Thomas Bäck (Email: t.h.w.baeck@liacs.leidenuniv.nl) is with the Leiden Institute for Advanced Computer Science, Leiden, The Netherlands.

Carola Doerr (Email: Carola.Doerr@lip6.fr) is with Sorbonne Université, CNRS, LIP6, Paris, France.

Bernhard Sendhoff (Email: bernhard.sendhoff@honda-ri.de) is with the Honda Research Institute Europe, 63073 Offenbach, Germany.

Thomas Stützle (Email: thomas.stuetzle@ulb.be) is with the IRIDIA laboratory, Université libre de Bruxelles (ULB), Belgium.

Benchmark Collections and Applications

A new framework, HAWKS, using evolutionary algorithms for the generation of synthetic datasets for clustering algorithms, is introduced in the paper by Shand *et al.* [10]. The authors illustrate how HAWKS can be used to evolve benchmark datasets consistent with predefined properties, and to evolve datasets that can emphasize performance differences between pairs of algorithms.

A real-world dataset from automotive engineering, the OSU-Honda Automobile Hood Dataset *CarHoods10k* is provided to the scientific community by Wollstadt *et al.* [13]. The authors illustrate the application of geometric deep learning, use machine learning to predict hood performance from the latent space representation, and combine this with evolutionary algorithms for a topology optimization application.

Ligot *et al.* [4] introduce an experimental protocol for the optimization-based design of robot swarms, based on a mission generator for generally applicable benchmark creation. The protocol is then illustrated by comparing the performance of two off-line fully automatic design methods.

The design of benchmark suites for multi-objective black-box optimization is taken up in the work by Yap *et al.* [14]. Using exploratory landscape analysis, standard benchmark suites are investigated with respect to their coverage of the instance space. The authors also demonstrate how to generate new benchmark problems to fill regions that are not well covered by the existing collections.

Constrained Optimization

A common challenge faced by evolutionary computation approaches in real-world problem solving are constraints. With the objective to support researchers in the analysis of different constraint-handling techniques, Sergeyev *et al.* propose a generator for constrained optimization problems in small and medium dimensions. The problems have known optima and scalable difficulty [9].

Constrained optimization is also the scope of the work by Kadavy *et al.* on the impact of boundary control methods on bound-constrained black-box optimization [3]. Using an empirical comparison of three algorithms with six different boundary control methods on the IEEE CEC competitions 2017 and 2020, it is shown that the control method has a non-negligible impact on the results and therefore needs to be clearly specified when empirical results are reported.

Performance Analysis and Visualization

Benchmarking algorithms goes hand in hand with sound statistical analysis of the obtained experimental results. Rojas-Delgado *et al.* propose to switch the focus from classical null hypothesis tests to those based on Bayesian statistics when analyzing algorithm performances [7]. They introduce a Bayesian-based framework that assumes algorithm performances (i.e., rankings) are generated by a probability distribution; this fresh analytical perspective is then applied to study questions such as probability of an algorithm being ranked first or having the same relative ranking as another algorithm.

Performance metrics are also the scope of the work by Hansen *et al.* [2], which describes the rationales behind the performance metrics selected for one of the most widely used benchmarking environments in evolutionary computation, the COCO (Comparing Continuous Optimizers) platform. The work summarizes common requirements for solution quality indicators in single- and multiobjective optimization, with consideration of possible constraints or the presence of noise.

Performance analysis is always linked to the problems that the algorithms have been evaluated on. Visualizations of these problem instances are very handy not only for algorithm analysis, but also for the selection of problems to be included in a benchmark collection, as well as for the design of landscape features. In [8], Schaepermeier *et al.* compare a number of visualization techniques for multi-objective optimization problems. Their newly created moPLOT-dashboard integrates these tools and makes them available for an interactive analysis.

Meta-Algorithmics

Algorithm selection aims to determine the most suitable algorithm for an unseen problem out of an algorithm portfolio. In the paper on benchmarking feature-based algorithm selection models in numerical black-box optimization [11], Ryoji Tanabe compares existing algorithm selectors and outlines general insights revealed via benchmarking them. Among other contributions, this work highlights a performance measure more reliable than the expected runtime, and demonstrates that the difficulty of outperforming the single-best solver depends on various factors, such as algorithm portfolios, cross-validation methods, and dimensionalities.

Designing an automated hyperparameter optimization (HPO) algorithm is usually still an unsystematic and manual process, due to the black-box nature of the meta-optimization problem and the complex search space of hyperparameter values which make the problem computationally expensive. Moosbauer *et al.* in their work [5] introduce a framework to benchmark-driven automated design and apply it to multi-fidelity (MF-) HPO. They formalize the search space of MF-HPO algorithms via a configurable optimization framework and search for the best candidate in an automatic and systematic way; their findings show that the benchmark-driven approach is on par with (and in some cases even outperforms) widely used HPO techniques.

Van der Blom *et al.* introduce the Sparkle platform [1] to facilitate the use of meta-algorithmic techniques for non-expert users. A key contribution of this work is the support it offers for best practices that make it easier to correctly and effectively use algorithm selection and configuration techniques, while standardized reports help document the process and results.

Search heuristics such as evolutionary algorithms and surrogate-based optimization techniques can often be conveniently parallelized. Determining suitable batch sizes remains, however, a challenging task, especially for settings in which individual evaluations are very costly. The work by Rehbach *et al.* [6] proposes an automated batch size configuration technique for surrogate-based optimization algorithms. The configurator is partially trained on a novel benchmark generation technique that is based on Gaussian process simulation.

The work by Vermetten *et al.* introduces BIAS, a toolbox based on a large ensemble of statistical tests to detect the existence of structural bias in heuristic optimization algorithms [12]. The BIAS toolbox can be used during the algorithm design phase, to benchmark and classify algorithmic behavior in terms of structural bias, and can thus be used to improve existing algorithms.

We sincerely thank the Editor-in-Chief, Prof. Carlos A. Coello Coello, for his constant and prompt support throughout all phases of this special issue. We also gratefully acknowledge the help of the Editorial Assistant, Dr. Gregorio Toscano, with all technical matters. This BENCH special issue would not have been possible without the help of the numerous reviewers who assessed the quality of the submitted papers and who provided constructive feedback and suggestions.

We hope that the readers will enjoy this IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION special issue on benchmarking.

THOMAS BÄCK, Leiden, The Netherlands
CAROLA DOERR, Paris, France
BERNHARD SENDHOFF, Offenbach, Germany
THOMAS STÜTZLE, Brussels, Belgium

References

[1] K. v. d. Blom, H. H. Hoos, C. Luo, and J. G. Rook. Sparkle: Towards accessible meta-algorithmics for improving the state of the art in solving challenging problems. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[2] N. Hansen, A. Auger, D. Brockhoff, and T. Tušar. Anytime performance assessment in blackbox optimization benchmarking. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[3] T. Kadavy, A. Viktorin, A. Kazikova, M. Pluhacek, and R. Senkerik. Impact of boundary control methods on bound-constrained optimization benchmarking. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[4] A. Ligot, A. Cotorruelo, E. Garone, and M. Birattari. Towards an empirical practice in off-line fully-automatic design of robot swarms. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[5] J. Moosbauer, M. Binder, L. Schneider, F. Pfisterer, M. Becker, M. Lang, L. Kotthoff, and B. Bischl. Automated benchmark-driven design and explanation of hyperparameter optimizers. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[6] F. Rehbach, M. Zaefferer, A. Fischbach, G. Rudolph, and T. Bartz-Beielstein. Benchmark-driven configuration of a parallel model-based optimization algorithm. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[7] J. Rojas-Delgado, J. Ceberio, B. Calvo, and J. A. Lozano. Bayesian performance analysis for algorithm ranking comparison. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[8] L. Schäpermeier, C. Grimme, and P. Kerschke. Plotting impossible? surveying visualization methods for continuous multi-objective benchmark problems. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[9] Y. D. Sergeyev, D. E. Kvasov, and M. S. Mukhametzhanov. A generator of multiextremal test classes with known solutions for black-box constrained global optimization. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2021.

[10] C. Shand, R. Allmendinger, J. Handl, A. Webb, and J. Keane. Hawks: Evolving challenging benchmark sets for cluster analysis. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2021.

[11] R. Tanabe. Benchmarking feature-based algorithm selection systems for black-box numerical optimization. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[12] D. Vermetten, B. van Stein, F. Caraffini, L. L. Minku, and A. V. Kononova. Bias: A toolbox for benchmarking structural bias in the continuous domain. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[13] P. Wollstadt, M. Bujny, S. Ramnath, J. J. Shah, D. Detwiler, and S. Menzel. Carhoods10k: An industry-grade data set for representation learning and design optimization in engineering applications. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

[14] E. Yap, M. A. Muñoz, and K. Smith-Miles. Informing multi-objective optimisation benchmark construction through instance space analysis. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2022.

Thomas Bäck (Fellow, IEEE) received the Ph.D. degree in computer science from the University of Dortmund, Dortmund, Germany, in 1994. He is a Professor of Computer Science with the Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands. His research interests include evolutionary computation, machine learning, and their real-world applications, especially in sustainable smart industry and health.

He is an elected member of the Royal Netherlands Academy of Arts and Sciences (KNAW, 2021), was a recipient of the IEEE COMPUTATIONAL INTELLIGENCE SOCIETY EVOLUTIONARY COMPUTATION Pioneer Award in 2015, and received the Best Ph.D. Thesis Award from the German Society of Computer Science (GI) in 1995. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and an Area Editor for the ACM Transactions on Evolutionary Learning and Optimization. He was elected as a Fellow of the International Society of Genetic and Evolutionary Computation in 2003.

Carola Doerr, formerly Winzen, (member, IEEE) is a CNRS Research Director with the LIP6 Computer Science department at Sorbonne Université in Paris, France. She received the habilitation degree (HDR) from Sorbonne Université in 2020, a Dr.-ing. degree from Saarland University, Germany, in 2011, and a Diplom in Mathematics from Kiel University in 2007. From 2007 until 2012, she was a business consultant with McKinsey & Company (on educational leave from 2010 onward). Her main research activities are in the analysis of black-box optimization algorithms, with a strong focus on their mathematical analysis. She is an Associate Editor of IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and ACM Transactions on Evolutionary Learning and Optimization, Editorial Board Member of the Evolutionary Computation journal, and Advisory Board Member of the Springer Natural Computing Book Series. She is a Founding and a Coordinating Member of the Benchmarking Network.

Bernhard Sendhoff (Fellow, IEEE) received the Ph.D. degree in applied physics from Ruhr-Universität Bochum, Bochum, Germany, in 1998. He was with Honda Research Institute Europe GmbH, Offenbach, Germany, from 2003 to 2010, as a Chief Technology Officer, and from 2011 to 2017, as a President. Since 2017, he has been an Operating Officer with Honda Research and Development Ltd., Tokyo, Japan, and the Head of the Global Operation, Honda Research Institutes. He is an Honorary Professor with the Technical University of Darmstadt, Darmstadt, Germany. He has authored or coauthored over 180 scientific publications. Dr. Sendhoff is a Senior Member of ACM and a member of SAE.

Thomas Stützle (Fellow, IEEE) received the Ph.D. degree in computer science from Technische Universität Darmstadt, Darmstadt, Germany, in 1998. He is a Research Director of the Belgian F.R.S.–FNRS with the IRIDIA Laboratory, Université Libre de Bruxelles, Brussels, Belgium. He has published extensively over 250 peer-reviewed articles in journals, conference proceedings, or edited books in the area of metaheuristics. His research interests include stochastic local search (SLS) algorithms, large-scale experimental studies, SLS algorithm engineering, and automated design of algorithms.