Applied Soft Computing Journal 93 (2020) 106502

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Editorial

Special Issue on Benchmarking of Computational Intelligence Algorithms in the Applied Soft Computing Journal

Computational Intelligence (CI) is a huge and expanding field, which is rapidly gaining importance, attracting more and more interest from both academia and industry. It includes a wide and ever-growing variety of optimization and machine learning algorithms, which, in turn, are applied to an even wider and faster growing range of different problem domains. For all of these domains and application scenarios, researchers want to pick the best algorithms. Actually, they want to do more: They want to improve upon the best algorithm. This requires a deep understanding of the problem at hand, the performance of the existing algorithms for that problem, the features that make instances of the problem hard for these algorithms, and the parameter settings for which the algorithms perform the best. Such knowledge can only be obtained empirically, by collecting data from experiments, by analyzing this data statistically, and by mining new information from it.

Benchmarking is the engine driving research in the fields of optimization and machine learning for decades, while its potential has not yet been fully explored. Benchmarking can go beyond simple statistics: It can be an application of Computational Intelligence itself! During the recent years, the importance of the field is realized by more and more researchers. Workshops on the topic have become a regular and important part of leading international conferences such as GECCO [1–4], PPSN [5,6], CEC [7–11], or ICACI [12].

In our special issue on Benchmarking of Computational Intelligence Algorithms in the Applied Soft Computing journal, we aimed to collect novel contributions from this domain. We were very positively surprised by the strong feedback and many submissions we received, from which we finally could select 14 for publication. The works in this issue can roughly be divided into four categories:

1. new benchmarks and benchmark generators,
2. new visualization and evaluation methods,
3. new tools, and
4. studies on general topics.

1. New benchmarks and benchmark generators

In our special issue, several new benchmarks and benchmark generators for different application fields of Computational Intelligence are introduced, with a focus on multi-objective optimization and pattern recognition, but also for specific application areas such as circuit design. With seven contributions, this is the largest field of our special issue.

Meneghini et al. [13] propose a parameterized generator of scalable and customizable benchmark problems for many-objective optimization problems. Their Generalized Position-Distance (GPD) generator allows for constructing infinitely many problems by tuning parameters that control features such as the number of variables and objectives, deceptiveness, multimodality, the existence of robust solutions, the shape of the Pareto front, and constraints. The resulting functions have known optimal solutions, are easy to understand, visualize, and implement, as well as fast to compute.

Tanabe and Ishibuchi [14] point out that synthetic problems for multi-objective optimization may have unrealistic properties. They, therefore, present a problem suite with 16 bound-constrained and 8 constrained real-world problems, four of which are multi-objective mixed-integer optimization problems. They provide implementations in Java, C, and Matlab and use their benchmark to investigate the performance of six evolutionary multi-objective optimization algorithms.

Wu et al. [15] present the *Multi-Type Aircraft Remote Sensing Images* (MTARSI) benchmark set for aircraft type recognition. They mitigate the problem that most of the state-of-the-art algorithms in the field have been evaluated on different (and often not publicly available) data sets, making it hard to understand and compare their performance. MTARSI contains 9,385 images of 20 aircraft types, with complex backgrounds, different spatial resolutions, and complicated variations in pose, spatial location, illumination, and time period. A comprehensive performance analysis of state-of-the-art aircraft type recognition and deep learning approaches on MTARSI is provided, too.

Jian et al. [16] follow a similar goal with their *Marine Underwater Environment Database* (MUED) benchmark, which contains 8,600 underwater images of 430 individual groups of conspicuous objects with complex backgrounds, multiple salient objects, complicated variations in pose, spatial location, illumination, and turbidity of water. They include manually labeled ground-truth information. Using this benchmark, they compare nine different state-of-the-art saliency-detection algorithms.

Müller de Souza et al. [17] propose a set of representative problems for benchmarking metaheuristics for combinational logic circuit design along with a set of performance measurements and descriptive statistics to analyze their results. The benefits of this benchmark are highlighted by a case study investigating Cartesian Genetic Programming variants.

Last but not least, Fischbach and Bartz-Beielstein [18] point out two weaknesses of current benchmark instance sets: they

https://doi.org/10.1016/j.asoc.2020.106502

may either not be real-world problems or are limited in size, creating the potential problem of overfitting our algorithms to them. Additionally, the statistical tools for comparing several algorithms over several problems are often complex. They aim to overcome these problems by combining ideas from problem generation and statistical analysis of experiments, utilizing ANOVA as a standard statistical tool.

2. New visualization and evaluation methods

During the past decade, researchers have begun to recognize that comparing benchmark results is more than just comparing average result qualities or runtimes. If the goal is to gain a deep understanding of algorithm performance and problem hardness, specialized visualization and evaluation approaches are needed [19]. Four articles in our special issue make contributions to this area.

Škvorc et al. [20] develop a generalized method of visualizing a set of optimization problems based on exploratory landscape analysis. With it, it becomes possible to determine the distribution of problems within a benchmark set visually by using exploratory landscape analysis combined with clustering and t-SNE visualization. This method places similar problems closer together. It is applied to the CEC Special Sessions and Competitions on Real-Parameter Single Objective optimization [8–11] and the GECCO Black-Box Optimization Benchmark workshops [1]. Interestingly, the authors find that a number of features derived on the above problems by state-of-the-art exploratory landscape analysis libraries are redundant or not invariant to transformations like scaling and shifting.

Walker and Craven [21] introduce a technique for visualizing the relative performance of algorithms optimizing the same multi-objective problem. Well-known performance indicators from literature are used to characterize the behaviour of the algorithms. The approach is validated by investigating the different parameterizations of NSGA-II and NSGA-III on the DTLZ and CEC 2009 [7] problem suites.

Torabi and Wahde [22] develop an approach for assessing the performance of optimization methods in cases where the global optimum of the objective function is unknown. The idea is to discretize the search space and then finding the optimum of the discretized space with brute force. Using this method, the performance of a GA applied to the speed profile optimization for heavy-duty vehicles is investigated.

Bossek et al. [23] discuss a multi-objective view of performance measurement of single-objective algorithms. Based on the Traveling Salesperson Problem (TSP) as example, the trade-off between the fraction of failed runs and the mean runtime of successful runs is investigated for state-of-the-art solvers. The hypervolume indicator (HV) is then used within per-instance algorithm selection models. The work also offers insights into behavior of inexact TSP solvers.

3. New tools

More sophisticated approaches to data analysis in benchmarking are often complex and hard to implement correctly. Providing tools that both make benchmarking easier [24] and provide standardized processes is a great support for the community. We are happy that two new such tools have been introduced in our special issue.

Doerr et al. [25] present the *IOHprofiler*, a software for creating detailed performance comparisons between iterative optimization heuristics, The IOHprofiler offers a selection of 23 discrete optimization problems with different types of fitness landscapes and has both an experiment executor as well as an analyzer component. A new module for IOHprofiler is discussed, which extents fixed-target and fixed-budget results for individual problems by ECDF results, which allows one to derive aggregated performance statistics for groups of problems. It is used to compare the performance of 12 different discrete heuristics on each problem.

With the *DSCTool*, Eftimov et al. [26] develop a statistical software for comparing the performance of stochastic optimization algorithms on one or multiple benchmark functions. They implement the concept of Deep Statistical Comparison (DSC), which ranks optimization algorithms by comparing the distribution of their result qualities. The DSCTool is provided as REST web service, which means all its functionalities can be accessed from any programming language.

4. General results

The field of benchmarking itself has evolved into a research domain. Besides new problems, new approaches, and new tools, it is also important to expand our basic understanding of fundamental concepts. Two articles of our special issue make such contributions.

Based on a very comprehensive review of the literature on forecasting, Oprea [27] develops a general framework for benchmarking and a set of guidelines for selecting the best algorithm for a specific problem in the domain. The idea is to integrate knowledge and software engineering best practices into CI benchmarking: It is proposed to improve the benchmarking process by using two knowledge bases, one for the application domain and one for CI algorithms. The use of the derived knowledge from an application domain-oriented survey into the general benchmarking framework, together with guidelines for CI algorithm selection, can improve the accuracy and response time of forecasting.

Tangherloni et al. [28] remind us that the performance of algorithms on artificial benchmarks can drastically differ from their performance on a real-world problem. They compare the performance relationship of state-of-the-art metaheuristics on the Parameter Estimation (PE) problem of biochemical systems with what is observed on common benchmark problems. They find that algorithms performing better on the benchmarks may perform worse on the PE problem. One important lesson is the necessity of knowledge about how the algorithms work and how to represent a problem: By applying a transformation of the PE problem to a more suitable representation, the state-of-the-art algorithms become competitive again. Only relying on benchmark results without considering algorithm features may not lead to expected results.

5. Concluding words from the guest editors

At this point, we want to sincerely thank our authors and reviewers. During the past two years, our reviewers and authors have put very much work into making this issue a success. The accepted papers underwent 2.5 revisions on average and up to 4 revisions at most, during which we together tried our best to make them perfect in every aspect. This high number of revisions posed a big workload on everyone involved – but we think the result was worth it. Thank you.

April 2020.
Thomas Weise, Markus Wagner,
Bin Li, Xingyi Zhang, and Jörg Lässig.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. Hansen, S. Finck, R. Ros, A. Auge, Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions, Research Report RR-6829, INRIA, 2009, inria-00362633v2.
[2] Good Benchmarking Practices for Evolutionary Computation (BENCHMARK@GECCO) Workshop at the Genetic and Evolutionary Computation Conference (GECCO'20), July 8–12, 2020, Cancún, Quintana Roo, Mexico, ACM, 2020.
[3] C. Doerr, P.S. Oliveto, T. Weise, B. Wróbel, A. Zamuda (Eds.), Black Box Discrete Optimization Benchmarking (BB-DOB) Workshop at the Genetic and Evolutionary Computation Conference (GECCO'19), July 13–17, 2019, Prague, Czech Republic, ACM, 2019.
[4] P.S. Oliveto, M. Wagner, T. Weise, B. Wróbel, A. Zamuda (Eds.), Black-Box Discrete Optimization Benchmarking (BB-DOB@GECCO) Workshop at the 2018 Genetic and Evolutionary Computation Conference (GECCO'18), July 15–19, 2018, Kyoto, Japan, ACM, 2018.
[5] Good Benchmarking Practices for Evolutionary Computation (BENCHMARK@PPSN) Workshop at the Sixteenth International Conference on Parallel Problem Solving from Nature (PPSN XVI), September 5–9, 2020, Leiden, The Netherlands, Springer, 2020.
[6] P.S. Oliveto, M. Wagner, T. Weise, B. Wróbel, A. Zamuda (Eds.), Black-Box Discrete Optimization Benchmarking (BB-DOB@PPSN) Workshop at the Fifteenth International Conference on Parallel Problem Solving from Nature (PPSN XV), September 8–9, 2018, Coimbra, Portugal, Springer, 2018.
[7] Q. Zhang, A. Zhou, S. Zhao, P. Suganthan, W. Liu, S. Tiwari, Multiobjective Optimization Test Instances for the CEC 2009 Special Session and Competition, Tech. Rep. CES-487, University of Essex, UK, 2008, Nanyang Technological University, Singapore; and Clemson University, USA.
[8] J. Liang, B. Qu, P. Suganthan, A. Hernández-Díaz, Problem definitions and evaluation criteria for the CEC 2013 special session on real-parameter optimization, Tech. Rep., Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China, 2013, and Technical Report, Nanyang Technological University, Singapore.
[9] J. Liang, B. Qu, P. Suganthan, Problem Definitions and Evaluation Criteria for the CEC 2014 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization, Tech. Rep., Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China, 2013, and Nanyang Technological University, Singapore.
[10] J. Liang, B. Qu, P. Suganthan, Q. Chen, Problem Definitions and Evaluation Criteria for the CEC 2015 Competition on Learning-based Real-Parameter Single Objective Optimization, Tech. Rep., Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China, 2014, and Nanyang Technological University, Singapore.
[11] N. Awad, M. Ali, J. Liang, B. Qu, P. Suganthan, Problem Definitions and Evaluation Criteria for the CEC 2017 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization, Tech. Rep., Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China, 2016, and Nanyang Technological University, Singapore.
[12] T. Weise, B. Li, M. Wagner, X. Zhang, J. Lässig (Eds.), International Workshop on Benchmarking of Computational Intelligence Algorithms (BOCIA) at the Tenth International Conference on Advanced Computational Intelligence (ICACI'18), March 29–31, 2018, Xiamen, Fujian, China, IEEE, ISBN: 978-1-5386-4362-4, 2018.
[13] I.R. Meneghini, M.A. Alves, A. Gaspar-Cunha, F.G. Guimarães, Scalable and customizable benchmark problems for many-objective optimization, Appl. Soft Comput. J. 90 (2020) 106139, http://dx.doi.org/10.1016/j.asoc.2020.106139.
[14] R. Tanabe, H. Ishibuchi, An easy-to-use real-world multi-objective optimization problem suite, Appl. Soft Comput. J. 89 (2020) 106078, http://dx.doi.org/10.1016/j.asoc.2020.106078.
[15] Z.-Z. Wu, S.-H. Wan, X.-F. Wang, M. Tan, L. Zou, X.-L. Li, Y. Chen, A benchmark data set for aircraft type recognition from remote sensing images, Appl. Soft Comput. J. 89 (2020) 106132, http://dx.doi.org/10.1016/j.asoc.2020.106132.
[16] M. Jian, Q. Qi, H. Yu, J. Dong, C. Cui, X. Nie, H. Zhang, Y. Yin, K.-M. Lam, The extended marine underwater environment database and baseline evaluations, Appl. Soft Comput. J. 80 (2019) 425–437, http://dx.doi.org/10.1016/j.asoc.2019.04.025.
[17] L.A. Müller de Souza, J.E.H. da Silva, L.J. Chaves, H.S. Bernardino, A benchmark suite for designing combinational logic circuits via metaheuristics, Appl. Soft Comput. J. 91 (2020) 106246, http://dx.doi.org/10.1016/j.asoc.2020.106246.
[18] A. Fischbach, T. Bartz-Beielstein, Improving the reliability of test functions generators, Appl. Soft Comput. J. 92 (2020) 106315, http://dx.doi.org/10.1016/j.asoc.2020.106315.
[19] T. Weise, Y. Chen, X. Li, Z. Wu, Selecting a diverse set of benchmark instances from a tunable model problem for black-box discrete optimization algorithms, Appl. Soft Comput. J. 92 (2020) 106269, http://dx.doi.org/10.1016/j.asoc.2020.106269.
[20] U. Škvorc, T. Eftimov, P. Korošec, Understanding the problem space in single-objective numerical optimization using exploratory landscape analysis, Appl. Soft Comput. J. 90 (2020) 106138, http://dx.doi.org/10.1016/j.asoc.2020.106138.
[21] D.J. Walker, M.J. Craven, Identifying good algorithm parameters in evolutionary multi- and many-objective optimisation: A visualisation approach, Appl. Soft Comput. J. 88 (2020) 105902, http://dx.doi.org/10.1016/j.asoc.2019.105902.
[22] S. Torabi, M. Wahde, A method for performance analysis of a genetic algorithm applied to the problem of fuel consumption minimization for heavy-duty vehicles, Appl. Soft Comput. J. 80 (2019) 735–741, http://dx.doi.org/10.1016/j.asoc.2019.04.042.
[23] J. Bossek, P. Kerschke, H. Trautmann, A multi-objective perspective on performance assessment and automated selection of single-objective optimization algorithms, Appl. Soft Comput. J. 88 (2020) 105901, http://dx.doi.org/10.1016/j.asoc.2019.105901.
[24] T. Weise, X. Wang, Q. Qi, B. Li, K. Tang, Automatically discovering clusters of algorithm and problem instance behaviors as well as their causes from experimental data, algorithm setups, and instance features, Appl. Soft Comput. J. 73 (2018) 366–382, http://dx.doi.org/10.1016/j.asoc.2018.08.030.
[25] C. Doerr, F. Ye, N. Horesh, H. Wang, O.M. Shir, T. Bäck, Benchmarking discrete optimization heuristics with IOHprofiler, Appl. Soft Comput. J. 88 (2020) 106027, http://dx.doi.org/10.1016/j.asoc.2019.106027.
[26] T. Eftimov, G. Petelin, P. Korošec, DSCTool: A web-service-based framework for statistical comparison of stochastic optimization algorithms, Appl. Soft Comput. J. 87 (2020) 105977, http://dx.doi.org/10.1016/j.asoc.2019.105977.
[27] M. Oprea, A general framework and guidelines for benchmarking computational intelligence algorithms applied to forecasting problems derived from an application domain-oriented survey, Appl. Soft Comput. J. 89 (2020) 106103, http://dx.doi.org/10.1016/j.asoc.2020.106103.
[28] A. Tangherloni, S. Spolaor, P. Cazzaniga, D. Besozzi, L. Rundo, G. Mauri, M.S. Nobile, Biochemical parameter estimation vs. benchmark functions: A comparative study of optimization performance and representation design, Appl. Soft Comput. J. 81 (2019) 105494, http://dx.doi.org/10.1016/j.asoc.2019.105494.

Thomas Weise *
Institute of Applied Optimization, School of Artificial Intelligence and Big Data, Hefei University, Hefei, Anhui, 230601, China
E-mail address: tweise@hfuu.edu.de
URL: http://iao.hfuu.edu.cn

Markus Wagner
School of Computer Science, The University of Adelaide, Adelaide, SA 5005, Australia
E-mail address: markus.wagner@adelaide.edu.au
URL: http://cs.adelaide.edu.au/~markus/

Bin Li
Department of Electronic Engineering and Information Sciences, School of Information Science and Technology, University of Science and Technology of China, Hefei, Anhui, 230027, China
E-mail address: binli@ustc.edu.cn
URL: http://staff.ustc.edu.cn/~binli/

Xingyi Zhang
School of Computer Science and Technology, Anhui University, Hefei, Anhui, China
E-mail address: xyzhanghust@gmail.com

Jörg Lässig
Enterprise Application Development Group, Department of Computer Science, Hochschule Zittau/Görlitz, D-02826 Görlitz, Germany
E-mail address: jlaessig@eadgroup.org

* Corresponding editor.