GECCO Black-Box Optimization Competitions: Progress from 2009 to 2018

Urban Škvorc
Jožef Stefan Institute and
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
urban.skvorc@ijs.si

Tome Eftimov
Jožef Stefan Institute
Ljubljana, Slovenia
tome.eftimov@ijs.si

Peter Korošec
Jožef Stefan Institute
Ljubljana, Slovenia
Faculty of Mathematics, Natural
Sciences and Information
Technologies, University of Primorska
Koper, Slovenia
peter.korosec@ijs.si

ABSTRACT

The GECCO Workshop on Real-Parameter Black-Box Optimization Benchmarking Series is a series of benchmarking workshops held every year since 2009 that evaluates the performance of new optimization algorithms.

Originally, the workshop organizers provided results for every year the workshop took place. In this article, we directly compare algorithms from every workshop held from 2009 to 2018. We analyze the compared algorithms on the noiseless benchmark function set using the competition's official empirical approach, and with a statistical approach called Deep Statistical Comparison.

Our goal is to investigate how algorithm performance has evolved throughout the years, and to show differences between empirical and statistical approaches to evaluating results.

CCS CONCEPTS

• **Computing methodologies → Continuous space search**; • **Mathematics of computing →** *Nonparametric statistics*;

KEYWORDS

Competitions, Optimization, Statistics, Comparison, Benchmarking

ACM Reference Format:

Urban Škvorc, Tome Eftimov, and Peter Korošec. 2019. GECCO Black-Box Optimization Competitions: Progress from 2009 to 2018. In *Proceedings of the Genetic and Evolutionary Computation Conference 2019 (GECCO '19).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3321996

1 INTRODUCTION

The GECCO Workshop on Real-Parameter Black-Box Optimization Benchmarking Series (BBOB workshops) [4] is a series of yearly benchmarking workshops in the field of numerical optimization that evaluate algorithm performance on a constant set of benchmark functions. Since every year the workshops use the same set of benchmark functions, the results can be compared between years. However, as far as we are aware, the organizers of the workshops only provide comparisons between algorithms from a specific year.

In this article, we present collected results that compare algorithms from every year of BBOB workshops. We compare algorithms using the same empirical approach used by the organizers, as well as using a statistical approach called Deep Statistical Comparison [5]. Using this comparison, we aim to show the overall improvement (or lack thereof) in algorithm performance from the first year of the workshop in 2009 till today, as well as the difference in these two approaches.

2 RELATED WORK

For each year of the GECCO Workshop on Real-Parameter Black-Box Optimization Benchmarking series, its organizers publish the complete results for all submitted algorithms, which are available in [1]. These results are used to create a comparative empirical comparison of the algorithms using the methods described in [7]. The organizers also provide code files that can be used to perform the same empirical analysis [3].

To the best of our knowledge, no comprehensive analysis that would compare algorithms over the years exists. Molina et al. [10], provide an overview of several optimization benchmarks, but only summarize the presented results over the years.

3 EXPERIMENTS

For our experiments, we have analyzed the data of all BBOB workshops on the noiseless test function set. In order to analyze the results, we first obtained the data of every algorithm that entered the workshops between 2009 and 2018. We obtained the results for 196 algorithms. The entire list of algorithms is available in [2], and the results data is available in [1].

Every algorithm contained data about the performance on 24 different benchmark functions with 5 different dimensionalities each. In this paper, we will focus only on the highest available dimensionality of 40.

We analyzed this data in two different ways: by using the official benchmark method [7], and by using a statistical method called Deep Statistical Comparison (DSC) [5].

4 RESULTS

In this section, we provide only some interesting results of our analysis. We provide the results of both the empirical and statistical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
GECCO '19, July 13–17, 2019, Prague, Czech Republic
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3321996

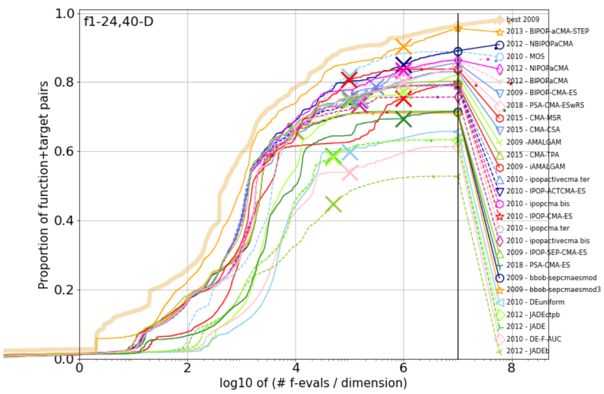

Figure 1: Empirical results for 40D functions

approaches. For readability and space concerns, we present the results of only a selected subgroup of algorithms, not on every algorithm from every year of the BBOB workshops, and only on 40 dimensional functions.

4.1 Empirical results

The empirical results are available in Figure 1. We present the results using the graph form, which shows the number of successful runs (runs where the error of the results was less than 10^{-8} across all benchmark functions), depending on the number of function evaluations. The BBOB benchmark tool-set allows for other forms of result presentations (such as tables) that allow a deeper look at the results, but here we focus only on the graphical version.

From the empirical results on the entire function set, we see that the algorithm BIPOP-aCMA-STEP [9] from 2013 appears to be the best, even when compared to algorithms from following years.

Another interesting result is the high amount of test functions that were easily solved by algorithms. When we looked at individual function results, 15 out of 24 test functions were solved correctly when looking at the results for dimensionality of 40.

4.2 Statistical Results

The statistical analysis performs pairwise comparisons between different algorithms, and determines whether one algorithm is statistically better than another on the benchmark function set. Table 1 shows this comparison on a small subsection of algorithms that performed best using statistical comparison. The values in the cells show the p-values of these comparisons. A value smaller than 0.05 means that the algorithm in the row of the table performed statistically better than the algorithm in the column. For easier visualization, such cells have a darker background. The algorithms shown are presented in [6, 8, 9, 11].

When using the statistical approach there is no statistical significance between most of the presented algorithms. We can also see that NBIPOPaCMA is shown to outperform PSA-CMA-ESwRS, while others do not. Both of these observations were not observed through empirical analysis and show that the statistical approach can provide different results.

	09-BI	12-NB	12-NI	13-BI	18-PS
09-BIPOP-CMA-ES	-	0.98	1.00	0.99	0.54
12-NBIPOPaCMA	0.04	-	0.58	0.58	0.02
12-NIPOPaCMA	0.01	0.58	-	0.58	0.06
13-BIPOP-aCMA-STEP	0.02	0.58	0.58	-	0.21
18-PSA-CMA-ESwRS	0.54	0.99	0.96	0.88	-

Table 1: Statistical results for 40D

5 CONCLUSION

In this article, we presented a quick overview of the empirical and statistical approaches to comparison of algorithms from the GECCO Workshop on Real-Parameter Black-Box Optimization Benchmarking, and the differences in results between the two methods. Our results showed that the algorithm performance peaked in the year 2013, and has not shown any significant overall improvement since.

In the future, we would like to expand this kind of algorithm analysis to understand not only which algorithms perform better, but also the reasons why they perform better on certain benchmark functions, for example by using Landscape Feature Analysis to group together similar functions on some lower level.

ACKNOWLEDGMENTS

This work was supported by the projects from the Slovenian Research Agency (research core funding No. P2-0098 and young researcher funding No. Pr-08987)

REFERENCES

[1] 2018. BBOB algorithm data. (mar 2018). Retrieved January 22, 2019 from http://coco.gforge.inria.fr/data-archive/bbob/
[2] 2018. BBOB Algorithms. (mar 2018). Retrieved January 22, 2019 from http://coco.gforge.inria.fr/doku.php?id=algorithms-bbob
[3] 2018. BBOB downloads. (mar 2018). Retrieved January 22, 2019 from http://coco.gforge.inria.fr/doku.php?id=downloads
[4] 2018. The BBOB workshop series. (mar 2018). Retrieved January 22, 2019 from http://numbbo.github.io/workshops/index.html
[5] Tome Eftimov, Peter Korošec, and Barbara Koroušić Seljak. 2017. A Novel Approach to statistical comparison of meta-heuristic stochastic optimization algorithms using deep statistics. *Information Sciences* 417 (2017), 186–215.
[6] Nikolaus Hansen. 2009. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*. ACM, 2389–2396.
[7] Nikolaus Hansen, Anne Auger, Dimo Brockhoff, Dejan Tušar, and Tea Tušar. 2016. COCO: Performance assessment. *arXiv preprint arXiv:1605.03560* (2016).
[8] Ilya Loshchilov, Marc Schoenauer, and Michèle Sebag. 2012. Black-box optimization benchmarking of IPOP-saACM-ES and BIPOP-saACM-ES on the BBOB-2012 noiseless testbed. In *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*. ACM, 175–182.
[9] Ilya Loshchilov, Marc Schoenauer, and Michèle Sebag. 2013. Bi-population CMA-ES agorithms with surrogate models and line searches. In *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*. ACM, 1177–1184.
[10] Daniel Molina, Antonio LaTorre, and Francisco Herrera. 2018. An insight into bio-inspired and evolutionary algorithms for global optimization: Review, analysis, and lessons learnt over a decade of competitions. *Cognitive Computation* 10 (2018), 517–544.
[11] Kouhei Nishida and Youhei Akimoto. 2018. Benchmarking the PSA-CMA-ES on the BBOB noiseless testbed. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 1529–1536.