Fair Comparison of Population-based Heuristic Approaches
The Evils of Competitive Testing

Anikó Csébfalvi and György Csébfalvi
University of Pécs, Pécs, Hungary

Keywords: Heuristic Algorithm, Population-based Metaheuristic Algorithm, Statistical Test, Nonparametric Test, Statistical Comparison, Sampling Theorem, Sample Design and Analysis Kolmogorov-Smirnov Test.

Abstract: 17 years ago, Hooker (1995) presented a pioneering work with the following title: "Testing Heuristics: We Have It All Wrong". If we ask the question now: "Do we have it all wrong?" the answer will be undoubtedly yes. The problem of the fair comparison remained essentially the same in the heuristic community. When we use stochastic methods in the optimization (namely heuristics or metaheuristics with several tunable parameters and starting seeds) then the usual presentation practice: "one problem - one result" is extremely far from the fair comparison. From statistical point of view, the minimal requirement is a so-called "small-sample" which is a set of results generated by independent runs and an appropriate "small-sample-test" according to the theory of the experimental design and evaluation and the protocol used for example, in the drug development processes. The viability and efficiency of the proposed statistically correct "bias-free" nonparametric methodology is demonstrated using a well-known nonlinear structural optimization example on the set of state-of-the-art heuristics. In the motivating example we used the presented solutions as a small-sample generated by a "hyperheuristic" and we test its quality against ANGEL, where the "supernatural" hybrid metaheuristic ANGEL combines ant colony optimization (AN), genetic algorithm (GE) and a gradient-based local search (L) strategy. ANGEL is an "essence of the different but at the same time similar heuristic approaches". The extremely simple and practically tuning-free ANGEL presents a number of interesting aspects such as extremely good adaptability and the ability to cope with totally different large real applications from the highly nonlinear structural optimization to the long-term optimization of the geothermal energy utilization.

1 INTRODUCTION

The problem of fair comparison, as a fundamental requirement of the evaluation of the real progress is a general problem of the heuristic community (Hooker, 1995).

When we use stochastic algorithms, the usual presentation practice: "one problem - one result", which is probable the first (most promising) element of a larger ordered list, is extremely far from the fair comparison. From statistical point of view, the minimal requirement is a so-called small-sample which is a set of results generated by 10-30 independent runs and an appropriate small-sample-test according to the theory of the experimental design and evaluation and the protocol used for example, in the drug development processes. We have to mention it, that even the usual mean or standard deviation parameter may be misleading or wrong when the distribution function is far from the "normality". When the sample size is small, then the nonparametric version of the Kolmogorov-Smirnov test (NKST) or any other appropriate nonparametric test may be the correct solution of the fair comparison problem (Csébfalvi, 2012).

The measuring of computational efficiency is generally a more complicated task. In this case, we have to define an appropriate "timeless" measure, which is invariant to progress of the optimization methodology and computational technology and able to characterize efficiency of a given approach as a whole. From this point of view the solution time is one of the worst from such measures, because (1) we have to replace the real running times with hypothetical but comparable times, and (2) we have to eliminate the effect of "polished code - readable code" like conflicts somehow to sure the fair and bias-free comparison. In our opinion is simple: we have to replace the solution time with a measure

Csébfalvi A. and Csébfalvi G..
Fair Comparison of Population-based Heuristic Approaches - The Evils of Competitive Testing.
DOI: 10.5220/0004168403060309
In *Proceedings of the 4th International Joint Conference on Computational Intelligence* (ECTA-2012), pages 306-309
ISBN: 978-989-8565-33-4

which is invariant to the environmental factors and problem types and able to characterize the computational efforts of the problem solving process as a whole. A good and generally usable measure may be the sum of times each variable has obtained a value (the total number of variable settings) divided by the number of variables. Naturally, we have to assume, that the total number of variable settings contains the number of settings of the parameter-setting or fine-tuning phase also. We note, when the approach has several not necessarily independent parameters, than the preliminary fine-tuning may be more complicated than the real problem solving process. This phase may mean, for example, a complete "experimental design and analysis" like time-consuming step in the problem solving process.

2 STATISTICAL COMPARISON

In this paper we present a theoretically correct comparison methodology, which can be used to compare two or more stochastic algorithms, or to evaluate the efficiency of a potential improvement for a given algorithm. The statistical analysis is a very important element of the evaluation of the real progress. In the heuristic community, the "fashion change" not necessarily means real improvement, and the "improved", "enhanced", "hybridized", etc. versions not necessarily give better results than the original algorithms. The development sometimes is driven by a specific problem set, on which the original algorithm is unable to produce the "expected better results" comparing with the competitors.

NKST for two samples obtained by running two competitive heuristics independently several times (10-30) the following:

$$H_0 : F_{HEURISTIC\,1}(x) = F_{HEURISTIC\,2}(x), \qquad (1)$$

that is, the two samples are from populations with the same distribution function.

3 EXAMPLE

We illustrate the essence of the methodological problems connected to the fair comparison by a popular structural optimization problem. In this nonlinear ten-bar truss (T10) weight minimization problem the design-variables are element cross-section areas and implicit functions define the response-variables, namely, the nodal displacements and the element stresses for the given load case (see Figure 1).

In the last decades, according to the challenging but sometimes frustrating nature of this problem and the progress of the optimization methodology and computational technology, it was investigated by several authors to demonstrate that their algorithm seems to be the best to date (it is robust, effective, and efficient).

The optimal solution of the problem is unknown but it is well-known that it has several more or less similar local optima according to the "hills and dales" like nature of the design space. In Table 1 we present the most important results obtained by using totally different methodological approaches.

The detailed investigation of the results can be found in Csébfalvi (2012). In this paper we only want to point out that the results are practically invariant to the year of publication which means that we have to evaluate the real progress very carefully. In other words, we can reach the area of alternative optima using different solution searching strategies and without knowledge about the real computational efficiency we cannot discriminate among the presented approaches.

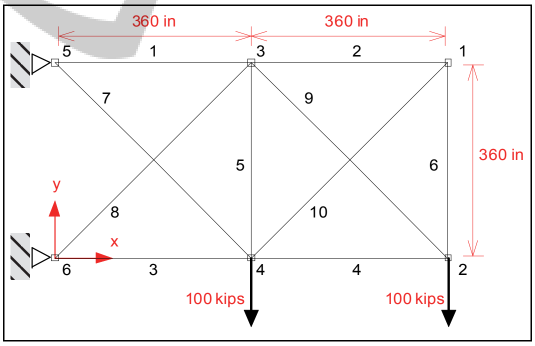

Figure 1: The benchmark-example.

In this motivating example we assume that the presented twenty solutions is a small-sample which is generated by a "hyperheuristic" and we test its quality against ANGEL developed by Csébfalvi (2007, 2011) for engineering optimization. The "supernatural" hybrid metaheuristic ANGEL combines ant colony optimization (AN), genetic algorithm (GE) and a gradient-based local search (L) strategy. We have to note, that according to the current terminology "hyperheuristic" means a metaheuristic set with a problem-specific selection mechanism.

The extremely simple and practically tuning-free ANGEL presents a number of interesting aspects such as extremely good adaptability and the ability to cope with totally different large scale real applications from the highly nonlinear structural optimization to the long-term optimization of the geothermal energy utilization (Csébfalvi and Schreiner, 2011).

Table 1: The most important results for T10.

Year	Authors	Weight (lb)
1969	Venkayya-Khot-Reddy	5084.90
1971	Gellatly-Berke	5112.00
1974	Schimit-Farshi	5089.00
1976	Rizzi	5076.66
1976	Schimit-Miura	5076.85
1976	Dobbs-Nelson	5080.00
1976	Schimit-Miura	5107.30
1979	Haug-Arora	5061.60
1979	Khan-Wilmert-Thornton	5066.98
1985	Haftka	5060.80
1991	Adeli-Kamal	5052.00
1992	Galante	4987.00
1994	Memari-Fuladgar	4981.10
1997	Ghasemi-Hinton-Wood	5095.65
1999	Lemonge	5060.92
2004	Lee-Geem	5057.88
2004	Lemonge-Barbosa	5069.09
2007	Li-Huang-Liu-Wu	5060.92
2009	Kaveh-Talatahari	5056.56
2009	Koohestani-Azad	5060.90

ANGEL has only three "tunable" parameters $\{P,G,I\}$, where P is the size of the population, G is the number of generations, I is the maximal number of local search iterations. Naturally, the maximal number of local search iterations means only a possibility, the procedure terminates when it reaches a size limit or a local minimum. The gradient-based L, try to make a better (lighter) feasible or a less unfeasible design from the current design obtained by AN or GE. The result of L will be the "locally best mutation".

The ANGEL sample was generated by 20 independent runs according to the number of results given by the state-of-the-art methods to date. In the investigation, the relative percent constraint tolerance was 0.001 %. We have to note, that we applied the original highly nonlinear "potential energy minimization model" without simplifications. In procedure L exact analytical derivatives were used.

In Table 2 we show an ordered ANGEL sample of 20 generated by the following settings:

$$\{P,G,I\} = \{100, 10, 10\} \tag{2}$$

Table 2: A random ordered sample of 20 for T10.

index	Weight (lb)	index	Weight (lb)
1	5063.27	11	5070.08
2	5064.80	12	5072.46
3	5065.72	13	5072.79
4	5066.11	13	5073.08
5	5067.14	15	5073.72
6	5067.70	16	5073.86
7	5068.33	17	5074.14
8	5068.52	18	5075.08
9	5068.69	19	5076.08
10	5069.71	20	5076.26

NKST (we reject the null-hypothesis) and the results of Table 2 and Figure 3 reveal that ANGEL is robust and able to produce good quality solutions within reasonable time without problem-specific preliminary investigation (fine-tuning). According to our computational experiences the range is one of the best measures of the robustness:

$$5076.26 - 5069.71 = 6.55 \tag{3}$$

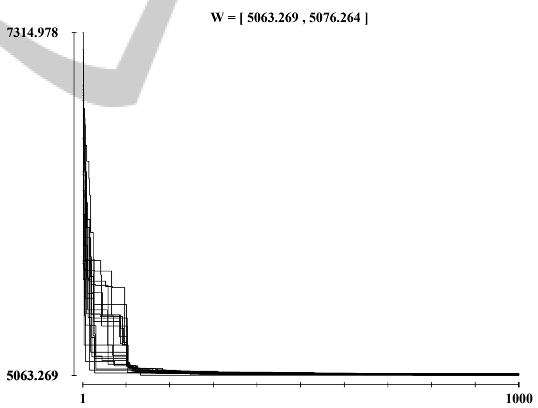

Figure 2: ANGEL searching history.

4 CONCLUSIONS

In this paper we presented a statistically correct methodology for to compare the efficiency of population-based heuristic approaches developed to generate good quality solutions within reasonable time for different optimization problems.

When we use stochastic methods to solve optimization problems, then the usual presentation practice: "one problem - one result" is extremely far from the fair comparison. From statistical point of view, the minimal requirement is the presentation of a small-sample generated by independent runs. The fair competitive testing needs an appropriate nonparametric test according to the theory of the experimental design and evaluation.

The viability and efficiency of the proposed statistically correct methodology is demonstrated using the well-known nonlinear ten-bar truss optimization example on a set of approaches developed in the last decades. In this motivating example, we assumed that the presented solutions form a small-sample generated by a "hyperheuristic" and we tested its quality against a "supernatural" hybrid metaheuristic ANGEL which combines ant colony optimization (AN), genetic algorithm (GE) and a gradient-based local search (L) strategy.

REFERENCES

Adeli, H., Kamal, O., (1991). Efficient optimization of plane trusses, *Advances in Engineering Software*; 13 (3), 116–122.

Barr, R.S., Goldenm, B. L., Kelly, J. P., Resende, M. G. C., Stewart, W. R., (1995). Design and Reporting on Computational Experiments with Heuristic Methods. *Journal of Heuristics*, 1, 9-32.

Csébfalvi, A., (2007). Angel method for discrete optimization problems. *Periodica Polytechnica: Civil Engineering*, 51/2 37–46, doi: 10.3311/pp.ci.2007-2.06, http://www.pp.bme.hu/ci.

Csébfalvi, A., (2009). A hybrid meta-heuristic method for continuous engineering optimization. *Periodica Polytechnica: Civil Engineering*, 53 (2), 93–100, doi: 10.3311/pp.ci.2009-2.05.

Csébfalvi, A., (2011). Multiple constrained sizing-shaping truss-optimization using ANGEL method. *Periodica Polytechnica: Civil Engineering*, 55 (1), 81–86, doi: 10.3311/pp.ci.2011-1.10, http://www.pp.bme.hu/ci.

Csébfalvi, A., Schreiner, J., (2011) A Net Present Value Oriented Hybrid Method to Optimize the Revenue of Geothermal Systems with Respect to Operation and Expansion, In: Tsompanakis, Y., Topping, B.H.V. (Eds.) *Proceedings of the Second International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering*. Chania, Greece, Civil-Comp Press, Stirling, UK.

Dobbs, M. W., Nelson, R. B., (1976). Application of optimality criteria to automated structural design. *AIAA Journal*, 14 (10), 1436–43.

El-Sayed, M. E., Jang, T. S., (1994). Structural optimization using unconstrained non-linear goal programming algorithm. *Computers and Structures*, 52 (4), 723–727.

Galante, M., (1992). Structures optimization by a simple genetic algorithm. In *Proceedings of Numerical Methods in Engineering and Applied Sciences*, CIMNE: Barcelona, Spain, 862–870.

Haftka, R. T., Kamat, M. P., (1985). *Elements of Structural Optimization*, Martinus Nighoff: Dordrecht.

Haug E. J., Arora, J. S., (1979) *Applied Optimal Design*, Wiley: New York.

Hooker, J. N., (1995). Testing Heuristics: We Have It All Wrong. *Journal of Heuristics*, 1, 33-42.

Kaveh, A., Talatahari, S., (2009) Particle swarm optimizer, ant colony strategy and harmony search scheme hybridized for optimization of truss structures. *Computers and Structures,* 87 (5-6), 267-283.

Khan, M. R., Willmert, K. D., Thornton, W. A., (1979). An optimality criterion method for large-scale structures. *AIAA Journal* 17 (7), 753—61.

Koohestani, K., Kazemzadeh Azad, S., (2009). An Adaptive Real-Coded Genetic Algorithm for Size and Shape Optimization of Truss Structures. In Topping, B.H.V., Tsompanakis Y. (Eds.), *Proceedings of the First International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering*, Civil-Comp Press, Stirlingshire, UK.

Lee, K. S., Geem, Z. W., (2004). A new structural optimization method based on the harmony search algorithm, *Computers and Structures*, 82 (9-10), 781-798.

Lemonge, A. C. C., (1999). Application of Genetic algorithms in Structural Optimization Problems. *Ph.D. Thesis, Program of Civil Engineering–COPPE*, Federal University of Rio de Janeiro, Brazil.

Lemonge, A. C. C., Barbosa, H. J. C., (2004). An adaptive penalty scheme for genetic algorithms in structural optimization. *International Journal of Numerical Methods in Engineering,* 59, 703–736.

Li, L. J., Huang, Z. B., Liu, F., Wu, Q. H., (2007). A heuristic particle swarm optimizer for optimization of pin connected structures. *Computers and Structures*, 85(7-8), 340-349.

Rizzi, P., (1976). Optimization of multiconstrained structures based on optimality criteria. In *Proceedings of 17th Structures, Structural Dynamics, and Materials Conference*, King of Prussia, PA.

Schmit, L. A. Jr, Farshi, B., (1974). Some approximation concepts for structural synthesis. *AIAA Journal*, 12 (5), 692–9.

Schmit, L. A., Miura, H., (1976). Approximation concepts for efficient structural synthesis. *NASA CR-2552*, Washington, DC: NASA.

Venkayya, V. B., (1971). Design of optimum structures. *Computers and Structures*, 1 (1–2), 265–309.