ISSN No:-2456-2165

Benchmarking of Problems and Solvers: A Ranking Theory Approach

Joseph Gogodze
Institute of Control System, TECHINFORMI,
Georgian Technical University
Tbilisi, Georgia

Abstract:- In this note, we propose a new approach for benchmarking computational problems and their solvers. The proposed methods are based on special construction of the paired comparison matrices of the problems and solvers; in other words, the score-matrices of the problems and the solvers. Having these matrices at our disposal, different rating/ranking methods can be used. We illustrate our approach using rating/ranking methods originated for ranking sports teams. The proposed approach is illustrated using an example to demonstrate its viability and suitability for applications.

Keywords:- *Benchmarking; Software; Solvers; Problems; Testing; Ranking Method.*

I. INTRODUCTION

In recent years, intensive studies have been conducted to evaluate the effectiveness of various solvers. Several methods for this purpose have been proposed in the literature review [1-9]. As noted in [9], most benchmarking tests utilize evaluation tables that display the performance of each solver for each problem under a specific evaluation metric (for e.g., CPU time, number of function evaluations, number of iterations, etc.). Different methods (based on suitable "statistical" quantities) are used to interpret data from these tables. The selection of a benchmarking method currently depends on the subjective tastes and individual preferences of researchers, who perform evaluations using solver/problem sets and evaluation metrics. The advantages and disadvantages of each proposed method are often a source of disagreement; however, this only stimulates further investigation in the field.

In our previous work [10], which was inspired by the work [11], based on the concept of Revealed Comparative Advantage (RSA) [12] and the simplest version of the PageRank method [13] [14], we proposed a benchmarking method which was also tested on the particular example for problems/solvers benchmarking. In this paper, we continue our investigation on problems and solvers benchmarking and propose new methods based on the special construction of paired comparison matrices (score-matrices) of the problems and the solvers. Having these matrices at our disposal, different rating/ranking methods can be used. We illustrate our approach using rating/ranking methods originated for sports teams ranking.

The approach which was introduced in [10] explores the natural relations between problems and solvers, which is determined by their evaluation tables. More specifically, we present data for benchmarking in the form of the benchmarking context, i.e., as a triple $< S, P, J >$, where S and P are the sets of solvers and problems respectively, and $J : S \times P \to \mathrm{R}$ is an assessment function (a performance or evaluation metric). It must be noted that each benchmarking context $< S, P, J >$, corresponds to special multi-objective decision making (MODM) problem.

In this paper, we will offer new methods of benchmarking which are based on the special construction of the score-matrices for problems and solvers in a benchmarking context $< S, P, J >$. Having these matrices at our disposal, well-known rating/ranking methods can be used. Of course, rating/ranking methods can be very different from one another. We illustrate our approach using rating/ranking methods originated for the rating/ranking of sports teams.

The remainder of this paper is organized as follows: section 2 describes the proposed methodology, section 3 considers the applications of the proposed tool in a selected benchmarking problem, and section 4 contains a conclusion.

II. METHOD

Throughout the paper, R^n is an n-dimensional space, with norm $\|\cdot\|$ and scalar product $<\cdot,\cdot>$.
Moreover, we use the following notations for the special sets:

$$\mathrm{R}^n_+ = \left\{\xi \in \mathrm{R}^n \mid \xi_k \geq 0, k = 1,\dots,n\right\};$$

$$\mathring{\Delta}_n = \left\{\xi \in \mathrm{int}\,\mathrm{R}^n_+ \mid \sum_{k=1}^{n} \xi_k = 1\right\}.$$

A. Benchmarking Context

Consider a set P of problems and a set of solvers S under the assumption that a function $J : S \times P \to \mathrm{R}$, henceforth referred to as the assessment function (performance metric), is given. Further assume, for definiteness, that the high and low values of J correspond to the "worst" and "best" cases respectively, and for convenience interpret $J(s, p)$ as the cost of solving the problem $p \in P$ with solver $s \in S$. Note that if $J(s, p) < J(s', p')$, it can be said that $s \in S$ solves $p \in P$ better than solver $s' \in S$ solves problem $p' \in P$

(i.e., the problem $p \in P$ was easier for solver $s \in S$ than the problem $p' \in P$ was for solver $s' \in S$).

We assume further that the following assumptions hold (n_P, n_S below are given natural numbers):

(A0) $P = \{p_1, \ldots, p_{n_P}\}$ and $S = \{s_1, \ldots, s_{n_S}\}$;

(A1) $J(s, p) \geq 0 \quad \forall (s, p) \in S \times P$;

(A2) $I_P(s) = \sum_{p \in P} J(s, p) > 0 \; \forall s \in S$ and

$$I_P(s) = \sum_{p \in P} J(s, p) > 0 \; \forall s \in S.$$

Note that assumption (A2) can be interpreted as a "no triviality condition" of the assessments (i.e., each solver and each problem should be tested with at least one problem and one solver respectively, without zero). The triple $< S, P, J >$, which satisfied assumptions (A0), (A1), and (A2) is henceforth referred to as the benchmarking context.

Note also that for a given benchmarking context $< S, P, J >$, numerous new quantities can be defined. Most of them have a "statistical" nature and, as was noted in the introduction, such "statistical" quantities are frequently encountered in benchmarking research, e.g. see [1,4,5,7,8,9].

B. Relation with MODM Problem

Let us assume now that $< S, P, J >$ is a given benchmarking context. Define the set of alternatives as $A \overset{def}{=} S = \{s_1, \ldots, s_{n_S}\}$. At the same time, define the criteria set $C = \{c_1, \ldots, c_{n_P}\}$ as follows:

$$c_j(\cdot) = J(\cdot, p_j) : S \to \mathrm{R},$$
$$p_j \in P = \{p_1, \ldots, p_{n_P}\},$$

for $j = 1, \ldots, n_P$ (i.e., we can assume that $C = P$).

Hence, we obtain an MODM problem as a pair $< A, C >$ where A is a set of alternatives and C is a set of criteria. Note also that in the case under consideration, we can consider criteria as normalized and as nonbeneficial (in the sense that low values correspond to the "worst" and the high values correspond to "best" case).

Let's recall now the basic concepts of multi-objective optimization theory [15, 16]. For a given MODM problem $< A, C >$ where $A = \{a_1, \ldots, a_m\}$ is set of alternatives and $C = \{c_1, \ldots, c_n\}$ is a set of criteria, let's introduce the following notations and definitions: we say that A is the set of admissible alternatives and map $\vec{c} = (c_1, \ldots, c_n) : A \to \mathrm{R}^n$ is the criteria map (correspondingly $\vec{c}(A) \subset \mathrm{R}^n$ is the set of admissible values of criteria). Alternative $a_* \in A$ is weakly Pareto optimal (weakly efficient) if there is no $a \in A$ with $c_j(a) < c_j(a_*)$ for all $j = 1, \ldots, n.$ Alternative $a_* \in A$ is Pareto optimal (efficient) if there is no $a \in A$ with $c_j(a) \leq c_j(a_*)$ for all $j = 1, \ldots, n$ and with at least one strict inequality. The set of all (weakly) efficient alternatives is denoted by $(A_{we})\, A_e$ and called the (weakly) Pareto set. Correspondingly, $(\vec{c}(A_{we}))\, \vec{c}(A_e)$ is called the (weakly) efficient front.

Pareto optimality is an appropriate concept for MODM. However, the set $(A_{we})\, A_e$ of (weakly) Pareto-optimal alternatives is generally large and all alternatives from $(A_{we})\, A_e$ must be considered mathematically equal (or "equally good"). Hence, additional factors must be considered for selecting specific (or more appropriate) alternatives from the set $(A_{we})\, A_e$. A possible approach in that direction is based on the concepts of score-matrices for alternatives and criteria. We shall discuss these concepts in the following subsection.

C. Comparison of Alternatives and Criteria

Let's assume that a MODM problem $< A, C >$ is given (as was mentioned above, we are assuming that criteria are normalized, and each criterion is nonbeneficial, hence the goal of the decision-making is to minimize criteria simultaneously).

Now we will offer the score-matrix S^A. for alternatives. To emphasize the "sporting context" of this consideration, let's also note that we can interpret elements of as athletes (e.g. chess players) who play the matches among themselves and for each pair of athletes, $a, a' \in A$, their joint match $\mathrm{M}(a, a')$ include m games.

Now, for any $a, a' \in A$, we define:

$$S^A(a, a') = \sum_{c \in C} s_c^A(a, a'),$$

where

$$s_c^A(a, a') = \begin{cases} 1, & c(a) < c(a'); \\ 0, & c(a) \geq c(a'); \end{cases} \quad \forall c \in C.$$

Thus, the equality $s_c^A(a, a') = 1$ means that $c(a) < c(a')$ for criterion $c \in C$ and the alternative a ("athlete a") receives one score (i.e. "the athlete a wins a game $c \in C$ in the match $\mathrm{M}(a, a')$). Correspondingly, $S^A(a, a')$ indicates number of total wins of the athlete a in the match $\mathrm{M}(a, a')$. Obviously,

$$m \geq S^A(a, a') \geq 0,\; S^A(a, a) = 0, \;\; \forall a, a' \in A.$$

We say that an alternative a has defeated an alternative a', if $S^A(a, a') > S^A(a', a)$. We also say that the result of the match $\mathrm{M}(a, a')$. is "$S^A(a, a')$ wins of the alternative a (losses of alternative a'),

$S^A(a',a)$ wins of the alternative a' (losses of alternative a), and number of draws $\left(m - S^A(a,a') - S^A(a',a)\right)$".

Matrix $S^A = \left[S^A(a,a')\right]_{a,a' \in A}$ will be named as score-matrix (of alternatives) for MODM problem $< A,C >$.

For any MODM problem $< A,C >$ with the set of alternatives $A = \{a_1,\ldots,a_m\}$ and the set of criteria $C = \{c_1,\ldots,c_n\}$ we can consider also adjoining[1] MODM problem $< C, A >$. Hence, we can define score-matrix for the MODM problem $< C, A >$, which will be denoted as $S^C = \left[S^C(c,c')\right]_{c,c' \in C}$. Matrix S^C will be named as score-matrix (of criteria) for MODM problem $< A,C >$.

D. Ranking Methods

Now we will briefly introduce the basic concepts of ranking theory (see, for details, [17, 18] and the literature cited therein). For a natural N, we say that a $N \times N$-matrix $S = [S_{ij}], 1 \le i, j \le N,$ is a score-matrix if

$$S_{ij} \ge 0, S_{ii} = 0,\ 1 \le i, j \le N.$$

To emphasize the "sporting context", lets note that we can interpret elements of $\mathrm{N}(N) = \{1,\ldots,N\}$ as athletes (teams) who play the matches amongst themselves and for each pair of athletes $(i, j),\ 1 \le i, j \le N,$ their joint match $\mathrm{M}(i, j)$ includes N, games. We interpret entry $S_{ij}, 1 \le i, j \le N,$ as the number of total wins of the athlete i in the match $\mathrm{M}(i, j)$. We also say that the result of the match $\mathrm{M}(i, j)$. is S_{ij} wins of the athlete/team i, (losses of the athlete/team j), S_{ji} wins of the athlete/team j (losses of the athlete/team i), and $\left(N - S_{ij} - S_{ji}\right)$ draws.

Hence $G_{ij}(S) = S_{ij} + S_{ji},\ (G = [G_{ij}(S)], 1 \le i, j \le N),$ can be interpreted as the number of successful games (i.e. games that did not end in a draw) in the match $\mathrm{M}(i, j)$. We also introduced the number $g_i(S) = \sum_{j=1}^{N} G_{ij}(S)$ which reflects the number of successful games in all matches played by the athlete $i, 1 \le i \le N$. For a natural N, and a score-matrix $S = [S_{ij}], 1 \le i, j \le N,$ we say that a pair $\left(\mathrm{N}(N), S\right)$ is a ranking problem. We say that a weak order (i.e., transitive and complete relation) $R(N,S) \subset \mathrm{N}(N) \times \mathrm{N}(N)$ is a ranking method for a ranking problem $\left(\mathrm{N}(N), S\right)$, a vector $r = (r_1,\ldots,r_N) \in \mathrm{R}^N$ $(r_i, 1 \le i \le N)$, is a rating vector and each r_i is a measure of the performance of player $i, 1 \le i \le N,$ in the ranking problem $\left(\mathrm{N}(N), S\right)$. We say also that for a ranking problem $\left(\mathrm{N}(N), S\right)$. a ranking method $R(N,S)$ is induced by a rating vector $r \in \mathrm{R}^N$ if :

$$(i, j) \in R(N,S) \text{ if end only if } r_i \ge r_j.$$

In this article, for illustrative purposes, we consider only a few ranking methods from the many that are known in literature[2]. In particular, for a score-matrix $S = [S_{ij}], 1 \le i, j \le N,$

we consider the following ranking methods:-

- *Score Method*

Rating vector $r^s = (r_1^s,\ldots,r_N^s) \in \mathrm{R}^N$, for the score method, defined as an average score

$$r_i^s = \sum_{j=1}^{N} S_{ij} / g_i(S), 1 \le i \le N;$$

- *Neustadtl's Method*

The Neustadtl's rating vector $r^N \in \mathrm{R}^N$, defined by equality $r^N = \bar{S} r^s$, where

$$\bar{S} = [\bar{S}_{ij}],\ \bar{S}_{ij} = S_{ij} / g_i(S), 1 \le i, j \le N.$$

- *Buchholz's Method*

The Buchholz's rating vector $r^B \in \mathrm{R}^N$, defined by equality $r^B = \left[\bar{G}(S) + E_N\right] r^s$, where

$$\bar{G}(S) = [\bar{G}_{ij}(S)],\ \bar{G}_{ij}(S) = G_{ij}(S) / g_i(S), 1 \le i, j \le N.$$

- *Fair-Bets Method*

The rating vector $r^{fb} \in \overset{\circ}{\Delta}$, for fair-bet method defined as unique solution of the following system of linear equations:

$$\sum_{j=1}^{N} S_{ij} r_j^{fb} - \left(\sum_{j=1}^{N} S_{ji}\right) r_i^{fb} = 0,\ 1 \le i \le N.$$

[1] Note that when interpreting the results for the adjoined MODM problem, we must remember the accepted agreement regarding "worst"/"best" values of criteria. If this causes inconvenience, then it is better to carry out the appropriate normalization

[2] Note that the ranking methods considered here originated from the ranking problems for chess tournaments and go back to the investigations of H.Neustadtl, E.Zermelo, and B.Buckholdz. For a detailed explanation, see e.g. [18].

➢ *Maximum Likelihood Method:*

The rating vector $r^{ml} \in \mathrm{R}^N$, for the maximum likelihood method, is defined by equality

$$r_i^{ml} = \ln(\pi_i),\ 1 \le i \le N,$$

where vector $\pi = (\pi_1, \dots, \pi_N) \in \overset{\circ}{\Delta}_N$, as unique solution of following non-linear equations system:

$$\pi_i \left(\sum_{\substack{j=1 \\ j \ne i}}^{N} \frac{G_{ij}(S)}{\pi_i + \pi_j} \right) = r_i^s g_i(S),\ 1 \le i \le N.$$

E. Proposed approach

Based on considerations from previous subsections, we are now ready to propose the following procedure for the given benchmarking context $< S, P, J >$

- ➢ The benchmarking context $< S, P, J >$ transformed in corresponding MODM problem (see subsection II.B.);
- ➢ For the MODM problem $< A, C >$, The score-matrixes $S^A = \left[S^A(a, a') \right]_{a,a' \in A}$ and $S^C = \left[S^C(c, c') \right]_{c,c' \in C}$ is constructed (see subsection II.C.);
- ➢ Using the score-matrixes S^A, S^C ranking of the alternatives from the sets A, C, respectively, is conducted by the ranking methods R^A, R^C (e.g. one of the described in the subsection II.D.) respectively;
- ➢ The alternative (solver) best by ranking method R^A is declared as R^A-best solver;
- ➢ The criterion (problem), best by ranking method R^C is declared as R^C-best problem;

III. BENCHMARKING OF DIFFERENTIAL EVOLUTION ALGORITHMS

Recently, researchers [19] have conducted a performance analysis of differential evolution (DE) algorithms using a well-known set of test functions. In this section, we use these results to illustrate the proposed benchmarking method. All necessary calculations were conducted in the MATLAB environment.

A. Data

In [19] considered 9 optimization algorithms listed and 25 types of test functions. In the work [19] and in the sources cited therein present detailed information on the selected algorithms and test function types. Utilizing these algorithms and test function types, the sets of 9 solvers and 50 problems were defined. Note that in [19] considered so-called ERTRSE as an assessment function for benchmarking DE algorithms, [19,20] (see also [10] for detail explanations components of the benchmarking context $< S, P, J >$ in this case-study).

B. Results

Direct calculation show that the Pareto sets are:

$$A_e^S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$A_e^P = \{1, 3, 7, 8, 9, 10, 11, 14, 17, 21, 22, 23, 24, 27, 29, 30, 32, 33, 34, 35, 36, 39, 42, 43, 44, 45, 47, 49, 50\}$$

We cannot present here the score-matrix because of its size but the score-matrix is:

$$S^A = \begin{bmatrix} 0 & 20 & 42 & 45 & 43 & 43 & 48 & 41 & 49 \\ 30 & 0 & 43 & 45 & 43 & 43 & 42 & 42 & 49 \\ 8 & 7 & 0 & 49 & 40 & 4 & 10 & 33 & 48 \\ 5 & 5 & 1 & 0 & 4 & 1 & 6 & 26 & 41 \\ 7 & 7 & 10 & 46 & 0 & 1 & 11 & 30 & 44 \\ 7 & 7 & 46 & 49 & 49 & 0 & 38 & 35 & 47 \\ 2 & 8 & 40 & 44 & 39 & 12 & 0 & 36 & 49 \\ 9 & 8 & 17 & 24 & 20 & 15 & 14 & 0 & 33 \\ 1 & 1 & 2 & 9 & 6 & 3 & 1 & 17 & 0 \end{bmatrix}$$

In Table 1, we present the ranking results of the solvers (obviously, ranking of problems can be considered analogously) obtained using different methods considered in the subsection II.D.

	rS	*rN*	*rB*	*rfb*	*rml*	*PgR*
S1	0,8275	0,3532	1,2866	0,3108	-1,1595	1,6324
S2	0,8425	0,3666	1,2997	0,3672	-1,0564	1,5966
S3	0,4975	0,1598	0,9978	0,0496	-2,9723	-0,7522
S4	0,2225	0,0694	0,7572	0,0172	-4,4527	-0,7010
S5	0,3900	0,1221	0,9038	0,0341	-3,5165	-0,7330
S6	0,6950	0,2585	1,1706	0,1117	-1,9544	-0,6340
S7	0,5750	0,1978	1,0656	0,0630	-2,5826	-0,5297
S8	0,3500	0,1439	0,8688	0,0401	-3,7258	0,5611
S9	0,1000	0,0390	0,6500	0,0062	-5,3452	-0,4403

Table 1:- Rating of Solvers by Different Methods

For comparison, Table 2 presents the correlation coefficients between rating vectors *rS*, *rN*, *rB*, *rfb*, *rml* and *PgR* (for explanation *PgR*-rating vector see [10]).

	rS	*rN*	*rB*	*rfb*	*rml*	*PgR*
rS	1					
rN	0,980	1				
rB	0,999	0,980	1			
rfb	0,855	0,929	0,855	1		
rml	0,999	0,978	0,999	0,851	1	
PgR	0,619	0,753	0,619	0,876	0,623	1

Table 2:- Correlation between Rating Vectors for Solvers

Finally, Table 3 presents the minimum and maximum points of the rating vectors *rS, rN, rB, rfb, rml, PgR* for solvers. As we can see, rating vectors *rS, rN, rB, rfb, rml* clearly recognize the solver S2 from Pareto set as best. At the same time, the rating vector *PgR* recognizes solver S1 from Pareto set as best. But it should be noted here that solver S1 is second after S2 for rating vectors *rS, rN, rB, rfb, rml*. Analogously, we can see that rating vectors *rS, rN, rB, rfb, rml, PgR* (see also footnote 1 regarding *PgR*) clearly recognize problem P14 from Pareto set as "best".

	rS	*rN*	*rB*	*rfb*	*rml*	*PgR*
Min	S9	S9	S9	S9	S9	S3
Max	S2	S2	S2	S2	S2	S1

Table 3:- Comparison of Best-Worst Solvers

IV. CONCLUSION

This short note introduced a new method of benchmarking computational problems and their solvers. The proposed method is based on a special construction of the paired comparison matrices of the problems and solvers. Having these matrices at our disposal, different rating/ranking methods can be used. We illustrate our approach using rating/ranking methods – *rS, rN, rB, rfb, rml.* Of course, other versions of the rating/ranking methods also can be used for benchmarking purposes, but we do not consider other possibilities here. Furthermore, we considered an illustrative example – benchmarking of differential evolution algorithms (solvers) and test functions (problems) – to demonstrate the viability and suitability of the proposed method for applications.

Finally, we must stress that we only provided a tool for benchmarking solvers and problems using the given benchmarking context. Note that the issues regarding selection of benchmarking context remains open in literature, i.e., we have not yet a clear and direct recommendations regarding how to select a set of solvers, set of problems, and performance metrics to obtain benchmarking results with proper justification.

REFERENCES

[1]. Benson, H.Y., Shanno, D.F., Vanderbei, R.J.: Technical Report ORFE-00-02, Princeton University, Princeton, New Jersey, (2000).

[2]. Billups, S.C., Dirkse, S.P., Ferris, M.C.: A comparison of algorithms for large-scale mixed complementarity problems. Comput. Optim. Appl., 7 (1997), pp. 3–25.

[3]. Bondarenko, A.S., Bortz, D.M., More, J.J.: COPS: Large-scale nonlinearly constrained optimization problems. Technical Memorandum ANL/MCS-TM-237, Argonne National Laboratory, Argonne, Illinois, 1998 (Revised October 1999).

[4]. Bongartz, I., Conn, A.R., Gould, N.I.M., Saunders, M.A., Toint, P.L.: A numerical comparison between the LANCELOT and MINOS packages for large-scale numerical optimization. Report 97/13, Namur University, 1997.

[5]. Conn, A.R., Gould, N.I.M., Toint, P.L.: Numerical experiments with the LANCELOT package (Release A) for large-scale nonlinear optimization. Math. Program. 73 (1996), pp. 73–110.

[6]. Mittelmann, H.: Benchmarking interior point LP/QP solvers. Optim. Methods Softw. 12 (1999), pp. 655–670.

[7]. Nash, S.G., Nocedal, J.: A numerical study of the limited memory BFGS method and the truncated Newton method for large scale optimization. SIAM J. Optim., 1 (1991), pp. 358–372

[8]. Vanderbei, R.J., Shanno, D.F.: An interior-point algorithm for nonconvex nonlinear programming. Comput. Optim. Appl., 13 (1999), pp. 231–252.

[9]. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Mathematical programming 91.2 (2002), 201–213.

[10]. Gogodze J., PageRank method for benchmarking computational Problems and their solvers, International Journal of Computer Science Issues, 15.3 (2018), pp. 1–7.

[11]. Hidalgo, C.A., Hausmann, R.: The building blocks of economic complexity. Proceedings of the National Academy of Sciences, 106 (2009), pp. 10570–10575.

[12]. Balassa, B.: Trade Liberalization and Revealed Comparative Advantage. The Manchester School of Economic and Social Studies, 3(2) (1965), pp. 99–123

[13]. Gleich, D. F.: PageRank beyond the Web. SIAM Review, 57.3 (2015), pp. 321–363.

[14]. Langville, A.N., Meyer, C.D.: Deeper inside pagerank. Internet Mathematics, 1.3 (2004), pp. 335–380.

[15]. Ehrgott, M.: Multicriteria Optimization, Springer, 2005.

[16]. Miettinen, K. M., Nonlinear Multiobjective Optimization, Kluwer, 1999.

[17]. Govan, A.Y., Ranking Theory with Application to Popular Sports, PhD thesis, North Carolina State, University, 2008.

[18]. González-Díaz, J., Hendrickx, R., Lohmann, E., Paired comparisons analysis: an axiomatic approach to ranking methods. Social Choice and Welfare, 42.1 (2014), pp.139-169.

[19]. Sala, R., Baldanzini N., Pierini M.: SQG-Differential Evolution for difficult optimization problems under a tight function evaluation budget. arXiv preprint arXiv: 1710.06770 (2017).

[20]. Auger, A., Hansen, N.: Performance evaluation of an advanced local search evolutionary algorithm. In Proceedings of the IEEE Congress on Evolutionary Computation. IEEE, 2 (2005), pp. 1777–1784.