RESULTS OF THE FIRST INTERNATIONAL CONTEST ON EVOLUTIONARY OPTIMISATION (1st ICEO)

The Organising Committee: H. Bersini, M. Dorigo, S. Langerman, G. Seront, L. Gambardella.

IRIDIA
CP 194/6
Universite Libre de Bruxelles
50, av. Franklin Roosevelt
1050 Bruxelles - Belgium

Abstract– This paper is dedicated to the presentation of the results of the first ICEO contest. Two types of optimisation problems were the subject of the contest: real function optimisation and the well-known TSP. 8 participants tested their algorithm on the real function benchmark and 3 on the TSP problems.

I. INTRODUCTION

We believe the time has come to organize a competition on evolutionary algorithms to take place every year during the ICEC conference. Results should be presented during the conference. The reasons to believe so are manifold. First it is becoming urgent to help people find their way in the jungle of new algorithms that are introduced at an increasing rate. A lot of them amount to slight variations on existing previous algorithms leading thus to increasing performance with respect to these previous algorithms. Unfortunately often these previous algorithms are themselves of discutable quality so that we attend a kind of recurrent and circular race where each algorithm serves as comparison to another one but completely loosing trace of the progression. Also numerous algorithms cannot compare with each other (since they don't know about each other existence and performance) and are obliged to make comparison with a common canonical algorithm which could itself be poorly performing. A common benchmark appears as a straightforward way of comparing a new method with a huge number of existing ones without having to re-implement and re-test all of them.

Secondly, a competition can naturally provide a kind of selection mechanism for new algorithms to be proposed. A very natural Darwinism scheme we are all familiar with can here again drive the selection and evolution of the methods. Suppose that the competition results are given according to some criteria like: best-solutions, number-of-function-evaluations, computer-speed, etc.... If the new algorithm to be proposed does not improve on the best current ones with respect to at least one criterion, then this method will not "survive" and the author should wait for the next ICEC before to submit the paper again. Since a lot of the new methods will result from hybridization and small mutation of existing ones, a meta-GA will take place to naturally trace the way to new promising methods. At least, this selection mechanism should make the task of referees much easier in the future.

Obviously, they are and they will be numerous discussions and controversies on the right selection of the criteria upon which to base the comparison. Quality of the solution, number of function evaluations, computational speed, computational load (memory,...), universality and robustness, ease-to-use, ease-to-understand (simplicity) are all acceptable criteria. Often an opportunistic defense of a new method argues for the consideration of a so far neglected criterion. For this contest, we have proposed several indexes depending on the two categories of problems. There is nothing definitive in this choice. Better or other indexes could emerge from discussions with the competitors which feel frustrate because their algorithm propose some kind of advantages which are not reflected in these indexes. As a matter of fact, the competition rules will have to evolve with the years. Criteria can evolve and the size of the test bed must certainly increase with new challenging problems getting in.

Thirdly, there is no better way to understand how another algorithm is running than by observing how it behaves on a problem you are familiar with. This is facilitated by a common benchmark which helps you to realize how another algorithm avoids some traps that yours were unable to avoid. Indeed this common benchmark provides a natural forum for fruitful interactions. Also, we have no doubt (and these

0-7803-2902-3/96/$4.00©1996 IEEE

competitions results seem to confirm this view) that a good algorithm will often be a hybrid between various heuristic searches coming from different optimisation schemes: population and recombination taken from GA, local heuristics problem or not problem dependent (like the LK local search for the TSP problem and the simplex for real function optimisation), SA annealing mutation, Tabu memory, etc.. So this competition, by facilitating the interaction between the method designers, should entail the production of new offspring methods.

Fourth, it makes no sense to suppose the existence of an universal best algorithm which could beat all others according to all possible optimisation indexes. Instead, what is much more likely to emerge with time is a clustering of problems based on the type of algorithms which work well (and according to some criteria) on these problems. The more problems constituting the common benchmark, the more reliable and usable this clustering will be. By usable we mean that the selection of the adequate algorithm for any new problem should become easier and easier just by some form of nearest neighbor classification (or any form of prototype-based classification) within the space of problems.

Fifth, a new competition is nothing original since it has a long history in the optimisation community but also in the fields of control, classification, time series prediction, etc. It is a well-established tradition and we just wanted to pursue this tradition in a systematic and convenient way allowing participants to present their results every year to a same conference and also to have an easy access to all other participants results. Also during the conference, the exchange of algorithm (computer codes) should become frequent and natural moves.

Finally, although it sounds somewhat paradoxical, we would like this competition to be done not in a competitive but in a fully collaborative spirit: No losers all winners, that is researchers happy to collect new exciting ideas to be tested in the future. There is a lot to learn in any algorithm whatever performance it shows on the benchmark.

II. THE REAL FUNCTION BENCHMARK

For the real function optimisation competition, we proposed uni-modal, highly multi-modal, separable and non-separable functions. The five functions to be minimised in their five and ten dimensional versions are:

The Sphere Model:

$$f(x) = \sum_{i=1}^{N}(x_i - 1)^2$$ with xi in [-5,5]

The Griewank's function:

$$f(x) = \frac{1}{d}\sum_{i=1}^{N}(x_i - 100)^2 - \prod_{i=1}^{N}\cos\left(\frac{x_i - 100}{\sqrt{i}}\right) + 1$$

with d=4000 and xi in [-600,600]

The Shekel's foxholes:

$$f(x) = -\sum_{i=1}^{m}\frac{1}{\|x - A(i)\|^2 + c_i}$$ with m =30 and xi in [0,10]

The Michalewicz' function:

$$f(x) = -\sum_{i=1}^{N}\sin(x_i)\sin^{2m}\left(\frac{ix_i^2}{\pi}\right)$$ with m=10 and xi in [0,π]

The Langerman's function:

$$f(x) = -\sum_{i=1}^{m}c_i\left(e^{-\frac{1}{\pi}\|x-A(i)\|^2}\cos(\pi\|x - A(i)\|^2)\right)$$

with m=5 and xi in [0,10]

We defined three performance indexes: the Expected Number of Evaluations per Success (ENES), the best value reached (BV) and the Relative Time (RT). These indexes had to be measured on every problem of the test bed for their 5 and 10 dimensional version. The ENES index represents the mean number of function evaluations needed in order to reach a certain fitness value - Value To Reach (VTR) - given with each problem. The ENES is computed by running 20 independent runs of the algorithm with the same parameters until the VTR is reached. If the VTR is not reached a stopping criterion is let to the author's appreciation but it must at least include a limit on the number of evaluated individuals.

Let's call NS the number of successes i.e. the number of runs able to reach the VTR and NE the number of evaluations i.e. the total number of individuals evaluated during the 20 runs. The ENES is defined as follows: ENES = NE/NS.

BV is the best value reached during the 20 runs. RT is the relative time taken by the algorithm to perform computations other than fitness evaluations. This time is expressed in Function Evaluation Equivalent. Let's call CT the total CPU time taken by the algorithm to perform 10000 iterations (by iteration, we mean a step that implies the evaluation of the individual), and ET the CPU time taken by 10000 fitness function evaluations.
RT is given by: (CT-ET)/ET.

The results of the real optimisation competition are given in Table 1. The participants are indicated by the first letters of their names: Bi-Pa is Bilchev and Parmee, Li is Li, Sto-Pri is

Storn and Price, Se-Be is Seront and Bersini, Van Ke is Van Kemenade, Ki-De is Kingdon and Dekker, Ka is Kargupata and FMGV is Fleurent, Glover, Michelon and Valli. All 30 values of the indexes are given for the 5 functions. We ordered the results on the basis of the ENES index that we believed to be, for this test bed, the most significant criterion on which to base this ranking.

III. RESULTS OF THE TSP AND ATSP COMPETITION

For testing the different algorithms on combinatorial problems, we proposed a set of small and large symmetric and asymmetric instances of the classical Traveling Salesman Problem. For each problem, the participants had to provide the average and best result obtained out of five experiments, each consisting of twenty random trials. Each trial had to be terminated after a number of function evaluation given by the following formula: K*n*j where K is 100 in case of TSP and 200 in case of ATSP, n is the number of cities and j varies on the set {1, 10, 25, 100}.

Results are indicated in table 2 for the 6 symmetric problems and in table 3 for the 5 asymmetric problems. The names of the problem instances can be read in the tables. Here again the names of the participants are given by the first parts of their names: Fre-Mer is Freisleben and Merz, Pfa is Pfaringher, Dor-Gam is Dorigo and Gambardella. For each problem, results for only four indexes are given: *** - b is the best solution reached after the lowest number of tour evaluations, *** - a is the average solution reached after the lowest number of tour evaluations, *** - B is the best solution obtained after the highest number of tour evaluations, *** - A is the average solution obtained after the highest number of tour evaluations. The results are ordered according to the quality of the obtained solutions.

IV CONCLUSIONS

Different conclusion can be immediately drawn from this first ICEO competition.

1. Too few participants (11) were involved in this competition. We hope that this number will increase in the future. We hope that it will become a natural endeavor for all authors presenting a new optimisation method to test it on the benchmark.

2. A competition is worth doing since the performance of the algorithms can be incredibly different. The competition results can make the future participants more aware of the minimal level of performance to be expected from a new method.

3. Hybridization seems to become an usual and key attitude for producing performant algorithms, reinforcing then the need for such an interacting forum provided by the common benchmark.

4. Several problems appeared during the competition and will need to be resolved for the next issue. For instance, do we allow a participant to tune adequately some parameters of his method for each new problem or do we constrain him to maintain a same and unique algorithm for all problems ? Are the indexes sufficient and satisfactory enough ? In real function optimisation, do we suppose to know the definitions of the functions (so as to allow strategy based on an explicit function decomposition on each axis) or just the evaluation for any new point ?

It is very likely that the rules of the ICEO competition will gradually evolve and improve so as to attract more and more participants convinced that being involved into this benchmark will be of great help for the genesis of their future ideas.

TABLE 1: Real Function Optimisation Competition Results

	Bi-Pa	Li	Sto-Pri	Van Ke	Se-Be	Ki-De	Ka	FGMV
ENES1	20	243	736	1452	326	1278.1	1522.6	12218
BV1	3.88e-15	0.0		1.5e-30			0.0	2.7e-7
RT1	2	12.7	4.67	75	0.35	49.83	0.667	3
ENES2	40	243	1892	3462	1099	2934.7	6392.2	85692
BV2	7.10e-15	0.0		4.3e-24			0.0	4.7e-7
RT2	2	13.6	4.88	32	0.35	94.83	0.2381	3.18
ENES3	41	21141	5765	22039	35637	89741	511797	2977996
BV3	7.99e-6	1.69e-5		3.2e-12			6.38e-5	1.3e-6
RT3	2	3.1	1.79	2	0.35	39.22	0.2595	2.19
ENES4	79	20898	13508	19125	6446	2230597	890683	2110889
BV4	1.31e-6	5.782e-5		6.1e-11			4.75e-5	1.76e-5
RT4	2	3.0	1.77	1.1	0.35	44.647	0.1699	2.29
ENES5	74	6318	76210	51845	13836	190134	451992	29449
BV5	-10.327	-10.403		-9.83			-9.98	-10.09
RT5	2	0.25	0.80	0.96	0.35	2.944	0.1115	1.39
ENES6	120	6075	744250	363685	259477	4440948	1.49e+7	879409
BV6	-10.101	-10.207		-9.75			-9.62	-9.59
RT6	2	0.42	0.66	0.53	0.35	3.419	0.0821	2.71
ENES7	120	6804	1877	10661	9925	1534.1	60219	33468
BV7	-4.6876	-4.687		-4.687			-4.6876	-4.69
RT7	2	1.28	1.11	3.9	0.35	36.90	0.1872	3.11
ENES8	501	14823	10083	41765	236348	26277	234698	20233341
BV8	-9.66	-9.66		-9.66			-9.66	-9.66
RT8	2	1.25	0.68	2.6	0.35	50.33	0.151	5.03
ENES9	176	4131	5308	11343	74720	232496	45783.2	6149
BV9	-1.499	-1.499		-1.49			-1.4976	-1.487
RT9	2	1.62	1.35	1.8	0.35	27.30	0.0956	4.50
ENES10	372	26973	44733	61729	1032627	15727653	443436	1071086
BV10	-1.499	-1.50		-1.49			-1.4976	-1.47
RT10	2	1.78	1.46	0.8	0.35	54.20	0.1029	6.75

TABLE 2: Results for the six symmetric TSP problems

	Fre-Mer	Pfa	Dor-Gam
eil51.tsp - b	426	485	426
eil51.tsp - a	426.7	502.5	432.63
eil51.tsp - B	426	426	426
eil51.tsp - A	426.0	427.2	428.06
kroA100.tsp - b	21282	28534	21296
kroA100.tsp - a	21291.7	30216.3	21780.45
kroA100.tsp - B	21282	21282	21282
kroA100.tsp - A	21282.0	21338.0	21420.0
d198.tsp - b	15780	20464	16278
d198.tsp - a	15787.1	20756.3	16617.38
d198.tsp - B	15780	15854	15888
d198.tsp - A	15780.0	15999.5	16054.00
att532.tsp - b	27753	42601	29978
att532.tsp - a	27802.1	43157.4	30859.00
att532.tsp - B	27686		28147
att532.tsp - A	27699.2		28522.80
rat783.tsp - b	8833	13800	9534
rat783.tsp - a	8846.7	14120.6	9940.00
rat783.tsp - B	8806		9015
rat783.tsp - A	8809.5		9066.80
fl1577.tsp - b	22350		25171
fl11577.tsp - a	22407.8		25533.50
fl1577.tsp - B	22286		22977
fl11577.tsp - A	22306.8		23163.17

TABLE 3: Results for the five asymmetric TSP problems

	Fre-Mer	Pfa	Dor-Gam
p43.atsp - b	2824	2815.5	2813
p43.atsp - a	2826.65	2821.3	2817.55
p43.atsp - B	2810	2810	2810
p43.atsp - A	2810.00	2810.6	2811.95
ry48p.atsp - b	14481	15382	14507
ry48.atsp - a	14645.8	15794.8	14878.90
ry48.atsp - B	14422	14422	14422
ry48.atsp - A	14440.0	14485.3	14565.45
ft70.atsp - b	38683	39905	39776
ft70.atsp - a	39020.3	40311.5	40422.65
ft70.atsp - B	38673	38673	38781
ft70.atsp - A	38683.8	38707.8	39099.05
krol124p.atsp - b	36351	41021	37035
krol124p.atsp - a	36705.2	42065.7	38971.26
krol124p.atsp - B	36230	36230	36241
krol124p.atsp - A	36235.3	36328.0	36857.00
ftvl70.atsp - b	2815	3211	2903
ftv170.atsp - a	2887.0	3431.8	2988.40
ftv170.atsp - B	2755	2762	2774
ftv170.atsp - A	2766.1	2817.3	2826.47