When to Stop? A New Stop Criterion for Combinatorial Optimization Search Techniques

Yuval Cohen, Roy Gelbard and Marlene Amorim

Abstract This research draws on theories in discrete event simulation, order statistics and record-breaking statistics to develop a methodology for deciding when to stop a combinatorial optimization search. During the first optimization period, the objective function improves rapidly with the iterations, and the improvement slows gradually until it almost stalls. We adopt a popular method to detect the period of rapid improvements of the "warm-up period", and then we propose a special control chart technique to identify with a given certainty reaching a steady state. Then, we suggest using the theory of record-breaking to decide on a stopping criterion. In addition, the paper develops estimates for the optimum bounds and estimates for value and timing of the next expected improvement. The advantages of this approach are discussed.

Keywords Stop criterion · Convergence · Evolutionary · Swarm · Solution's quality · Bound

1 Introduction

The convergence towards the optimum of genetic algorithms and other related search techniques is well established (Holland 1992; Pandey et al. 2014). The objective value

Y. Cohen (✉)
Department of Industrial Engineering, Tel-Aviv Afeka College of Engineering, 38 Mivtsa Kadesh, 69988 Tel Aviv, Israel
e-mail: yuvalc@afeka.ac.il

R. Gelbard
Department of Business Administration, Bar-Ilan University, 5290002 Ramat-Gan, Israel
e-mail: r_gelbard@yahoo.com

M. Amorim
Department of Economics, Management and Industrial Engineering, University of Aveiro, 3810-193 Aveiro, Portugal
e-mail: mamorim@ua.pt

of all search techniques shows a typical initial rapid improvement period followed by continual decrease in improvement rate, as the search goes-on (Rawlins and Sushil 2014).

There are several popular stopping criteria such as: (1) number of generations (Liu et al. 2014), (2) number of generations without minimal improvement (Ha et al. 2014), (3) maximum optimization time (Glorieux 2015, p. 9), (4) optimization threshold value (Glorieux 2015, p. 9), (5) a threshold value for the difference between the objective best value and average value, over the generations (Toledo et al. 2014). However, setting the stop criterion in each case is more an art than a science. Martí et al. (2016) tried to integrate science into this decision, but ended up with complex models based on progress indicators with a Kalman filter which is used for data gathering. This paper suggests a more structured and reason-based strategy. It first identifies an initial "warmup period" (Rossetti et al. 2005), of significant improvement, then it tests when the objective function reaches a plateau (Robinson 2007), finally, it adopts the statistical inference of a record breaking process from a stationary random process (Glick 1978; Krug and Jain 2005).

2   Objectives

The objective of the research is to propose an effective and efficient stopping criterion (Hermadi et al. 2014; Kim 2013) that could be applicable to most search techniques (for example, Genetic Algorithms (GA), Simulated Annealing (SA), Ant Colony (AC), Particle Swarm Optimization (PSO), etc.). The suggested stopping technique should also be based on current scientific tools and on reason. Additionally, the proposed technique should be easy to apply and easily understood. Most importantly, the technique should offer ways to control its reliability.

To fulfil the objectives we set three different goals to be achieved:

1. The identification of the initial period of rapid improvement in the objective function.
2. The identification of the second period of the objective function convergence to a plateau.
3. The identification of a sufficiently good solution by assuming a stationary homoscedastic process.

3   Identifying the End of the Intensive Warm-up Period

The first step is to identify the end of the initial steep descent or "warmup period" (Rossetti et al. 2005). For that purpose, we adopt the following Definitions:

F_{max} Largest value of the fitness function
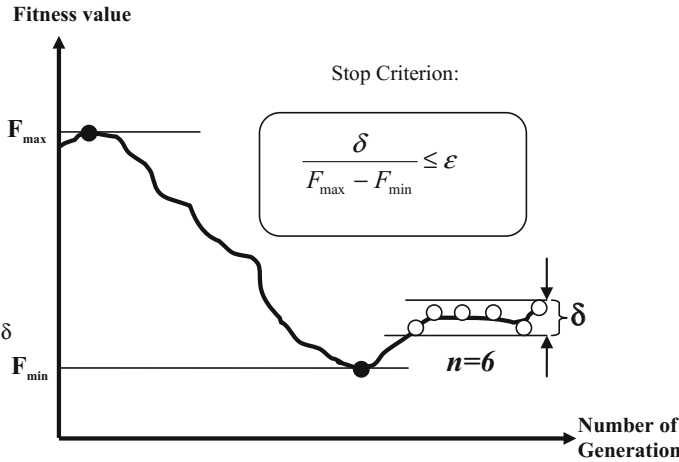F_{min} Smallest value of the fitness function

Fitness value

Stop Criterion:

$$\frac{\delta}{F_{max} - F_{min}} \leq \varepsilon$$

F_{max}

δ

F_{min}

$n=6$

δ

Number of Generation

Fig. 1 Example of identifying the warm-up period in evolutionary optimization

Total improvement $= F_{max}$-F_{min}

n   The number of last observations of the fitness function for consideration.
δ The range of fitness function value of the last n observations.
e   Threshold value for the ratio between δ and the total improvement

First we need to decide on the value of n (the number of last observations we consider). Once n is know (say $n = 6$), δ could be computed continuously as follows:

$$\delta = \text{Max}(X_t, X_{t-1}, \ldots X_{t-n}) - \text{Min}(X_t, X_{t-1}, \ldots X_{t-n}) \tag{1}$$

Note that δ is analogous to the total improvement, but is only related to the last n observations. Thus, the ratio between δ and the total-improvement is the ratio between the local improvement (of last n points) and the global improvement. This ratio is expected to decline through time and using an epsilon as a threshold is utilized as a stopping rule as described in Fig. 1.

Reaching the threshold described in Fig. 1, is telling us that the process finished its steep progress towards the optimum. However, it does not mean that the process reached a stationary condition or hit a plateau. It still remains up to a second procedure to ascertain the steady state of the process. This step is described in Sect. 4.

4   Identifying the Steady State

Once the initial period of intensive improvement is curtailed, a quest for steadiness begins. A second procedure is applied to make sure that the process is stationary with unchanging variance. A popular quality control based technique was adopted for this
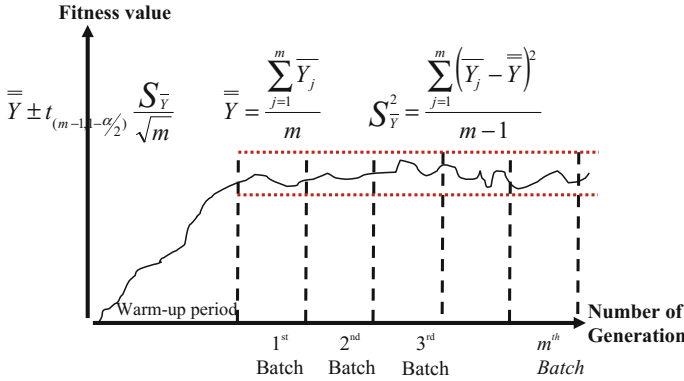
Fig. 2 An example for the method of batch means

purpose. The technique is widely used by the discrete event simulation community. The method is named the method of "Batch Means" (Robinson 2007). This method is sometimes used outside the world of discrete event simulations: for example, Rossetti et al. (2005).

In this scheme we define:

n the number of observations in each batch.
m the number of required batches.
$\overline{Y}_j$ the mean of batch j.
$\overline{\overline{Y}}$ the total mean of all observations.
$S_{\overline{Y}}^2$ the variance of the batch means.

Setting values to n, and m (typically greater than 6) generates m groups of n observations each. The central limit theorem (CLT) ensures that the batch means are distributed approximately as Normal IIDs, and thus by the procedure tests for outliers from the following range:

$$\overline{\overline{Y}} \pm t_{(m-1,1-\alpha/2)} \frac{S_{\overline{Y}}}{\sqrt{m}} \tag{2}$$

If there are no outliers, we adopt the end-point of last batch (*mth* batch) to the beginning of the third step of the overall procedure. Otherwise, a repeated procedure of *m*-means will be applied where the first batch will start right after the last outlier (Fig. 2).

Table 1 Record number *r* and its simulated statistical parameters (Glick 1978)

r	2	3	4	5	6	7	8
E(n)	4	14	40	109	291	781	2097
Median = M_r	4	10	26	69	183	490	1316
M_r/M_{r-1}		2.50	2.60	2.65	2.65	2.68	2.69

5 Identifying Sufficiently Satisfying Record

Once the second step of batch means is done, a steady state process could be assumed, for the observations that follow the last batch. In that region, a third step is taken using record-breaking theory for deciding when to stop the search. For presenting the record-breaking theory and our proposed stopping rule, the following definition are given.

Definitions

r	Record number (r = 2: second record, r = 3: third record, etc.)
n	Number of generations/observations
E(n)	mean number of trials/observations until a certain record number.
R_n	number of records in *n* generations/observations.
M_r	Median number of trials/observations related to the *rth* record
M_r/M_{r-1}	The ratio of current median to median of last record

Table 1 shows results of an extensive simulation trial for a process of drawing *n* iid variables, until the 8th record is achieved. Table 1, shows for each *rth* record: the mean average observations *(E(n))*. However *E(n)* approaches infinity as *n* grows, while median is stable. Therefore, we prefer dealing with medians. Median is denoted (M_r), and the ratio of current median to last one (M_r/M_{r-1}) is based on (Glick 1978). This ratio approaches $e = 2.718$ as *r* grows.

Thus:

$$\frac{Median\{n_{r+1}\}}{Median\{n_r\}} = e = 2.718 \tag{3}$$

This means that the median number of observations between the current record to next record is 2.718 times the total number of observations so far (until current record). Setting a threshold for number of generations without improvement is easy. For example, limiting the search for achieving next record by maximum of additional million generations has a median of last record: at, or after, a total of: $1,000,000/2.718 = 367,918$ generations.

Additional way to set a threshold is to use the number of records directly. In that case, the threshold could be determined using the Chebyshev's bound for ensuring that the solution is in a certain percentile of the total population of solutions. Table 2

Table 2 Expectation, variance, and standard deviation for n i.i.d. observations (Glick 1978)

Number observations n	$E(R_n)$	$V(R_n)$	$\sigma(R_n)$	$E(R_n)+3\sigma(R_n)$
100	5.19	3.55	1.88	10.84
200	5.88	4.24	2.06	12.06
300	6.28	4.64	2.15	12.74
400	6.57	4.93	2.22	13.23
500	6.79	5.15	2.27	13.60
600	6.97	5.33	2.31	13.90
700	7.13	5.49	2.34	14.16
800	7.26	5.62	2.37	14.37
900	7.38	5.74	2.40	14.57
1000	7.49	5.8564	2.42	14.75
1,000,000	14.39	12.75	3.57	25.10

shows results for several thousand repetitions of generating n iid observations and keeping R_n at each repetition.

For each n, the mean $E(R_n)$ and variance $V(R_n)$ number of records was computed. This enables to find a number of records that ensures with high probability that the best of n observations was found. The rightmost column of Table 2, uses 3 standard deviations above the mean. Using Chebyshev's one-sided bound ensures that the best solution among n is found with a probability of 94.4%:

$$\Pr\{\text{no better records exist for } n|k = 3\} = 1 - (1/(2(k^2)))$$
$$= 1 - ((1/2)(1/3^2)) = 17/18 = 0.944 \tag{4}$$

Thus, a process of waiting for the next 15 records is equivalent to waiting at the minimum 1200 generations, and waiting for 25 records is equivalent to observing 1,000,000 generations.

6 Results

We compared the results of the proposed GA stopping technique with optimal values of the benchmark RCPSP solutions on PSPLIB (Kolisch and Sprecher 1997; Kolisch et al. 1999). We used 30 solutions for 4 different network sizes: 120, 90, 60 and 30 activities. The stopping parameters we used were: (1) For the intensive improvement stage: $n=8$, $\varepsilon=0.005$; (2) for the batch means $n=10$, $m=5$ and $\alpha=0.05$; (3) setting the start point at the end of step-2 and waiting for 15 more records. The results are summarized in Table 3.

Thus, preliminary results (summarized in Table 3) show that the suggested scheme ensures high quality solutions, albeit in some cases adding some computational time.

Table 3 Results of the GA stopping rules vs. benchmark optimum (Kolisch et al. 1999)

No. of activities	Benchmark cases	# reach optimum	% reach optimum (%)	Within 1% of the optimal value (%)
30	30	28	93	100
60	30	24	80	100
90	30	18	60	98
120	30	14	47	93

7 Conclusion

The paper suggests a new approach for stopping criterion for search techniques. First, the length of the warmup period with intensive improvement is identified. Then, batch means method is applied for reaching a flat part of the fitness function—to be approximated as a static distribution process. This approximation enables using record breaking statistics for stopping inference rules. Preliminary results show high quality solutions could be obtained using this scheme. It is left for future research to examine the impact of most model parameters, and the trade-off between solutions quality and its time-performance. For that purpose far more experimentation is needed, as well as sensitivity analysis for each parameter and their combinations.

References

Glick N (1978) Breaking records and breaking boards. Am Math Monthly 2–26

Glorieux E (2015) Constructive cooperative coevolution for optimising interacting production stations. Doctoral dissertation, University West, Trollhättan: Sweden

Ha MP, Kumar L, Ananthapadmanabha T (2014) A novel approach for optimal allocation of a distributed generator in a radial distribution feeder for loss minimization and tail end node voltage improvement during peak load. Int Trans Electr Comput Eng Syst 2(2):67–72

Hermadi I, Lokan C, Sarker R (2014) Dynamic stopping criteria for search-based test data generation for path testing. Inf Softw Technol 56(4):395–407

Holland J (1992) Adaptation in natural and artificial systems. The MIT Press; Reprint edition (originally published in 1975)

Kim JL (2013) Genetic algorithm stopping criteria for optimization of construction resource scheduling problems. Constr Manag Econ 31(1):3–19

Kolisch R, Sprecher A (1997) PSPLIB-a project scheduling problem library: OR software-ORSEP operations research software exchange program. Eur J Oper Res 96(1):205–216

Kolisch R, Schwindt C, Sprecher A (1999) Benchmark instances for project scheduling problems. Int Ser Oper Res Manag Sci 197–212

Krug J, Jain K (2005) Breaking records in the evolutionary race. Physica A 358(1):1–9

Liu HL, Gu F, Zhang Q (2014) Decomposition of a multi-objective optimization problem into a number of simple multi-objective sub-problems. IEEE Trans Evol Comput 18(3):450–455

Martí L, García J, Berlanga A, Molina JM (2016) A stopping criterion for multi-objective optimization evolutionary algorithms. Inf Sci 367:700–718

Pandey HM, Chaudhary A, Mehrotra D (2014) A comparative review of approaches to prevent premature convergence in GA. Appl Soft Comput 24:1047–1077

Robinson S (2007) A statistical process control approach to selecting a warm-up period for a discrete-event simulation. Eur J Oper Res 176(1):332–346

Rossetti MD, Li Z, Qu P (2005) Exploring exponentially weighted moving average control charts to determine the warmup period, In: Proceedings of the 2005 winter simulation conference, 771–780

Rawlins GJE, Sushil JL (2014) Syntactic analysis of convergence in genetic algorithms. Found Genet Algorithms (FOGA 2), 2:141

Toledo CFM, Oliveira L, França PM (2014) Global optimization using a genetic algorithm with hierarchically structured population. J Comput Appl Math 261:341–351