CHEMICAL ENGINEERING TRANSACTIONS, Volume 11, 2007
Edited by Sauro Pierucci
 ISBN 978-88-95608-00-6

A systematic procedure to set up the genetic algorithm parameters for large scale systems: application to a three-phase catalytic reactor

Rezende, M. C. A. F., Costa, A. C., Maciel Filho, R., Bártolo, P. J. S., Rezende, R. A.
Chemical Engineering School, State University of Campinas (UNICAMP)
CP 6066, CEP 13087-970, Campinas-SP, Brazil

The Genetic Algorithms require setting up parameters in the search for the best value of the objective function in an optimization process. The present work proposes the use of a factorial design technique, which makes use of the most meaningful effects on a response, in order to set up the genetic algorithms parameters applied to a three phase catalytic reactor. The genetic algorithm used is a real-coded genetic algorithm that operates directly on real values. The results have shown the potential of factorial design technique proposed to find out the most significant set of parameters without using the conventional procedure of trial-and-error which is not suitable for large scale systems.

1. Introduction

Genetic Algorithms (GAs) have been the most popular form of evolutionary algorithms and they have proved to be a versatile and effective approach for solving optimization problems (Baskar et al., 2003). GAs have the advantages of not exploiting the mathematical structure of the objective function and/or constraint and not requiring an initial feasible point. This makes the optimization procedure robust and attractive for the solution of large scale nonlinear systems. A real-coded GA is used in this work. During the procedure of optimization by GA, a set of parameters must be chosen. There are in the literature many works that suggest values for GA parameters, but there are few works that present how the suitable GA parameters choice is made.
The present work proposes a systematic procedure to set up the GA parameters for large scale systems applied to a three-phase catalytic reactor that produces 2-methyl-cyclohexanol. This procedure is a factorial design that determines the statistically significant GA parameters that exert influence on the search for the optimal solution at the optimization problem. The selection of the statistically significant GA parameters is important to decrease the computational effort, saving time, since the optimization trials can be done varying only these statistically significant GA parameters, guarantying high performance of the search and the convergence to global optimal solution. The factorial design requires less time than conventional procedure of trial-and-error and is more suitable for large scale systems.

2. Genetic Algorithms

GAs are global optimization algorithms inspired by Darwin's theory of the survival of the fittest. The GAs start with a random population of chromosomes that are a set of

solutions to the optimization problem. Each solution is evaluated by the fitness function that associates a value to the solution, determining the best ones. A new population is created using genetic operators. This procedure is repeated along the generations, until a termination criterion is satisfied. In this paper, the real-coded GA is employed.

2.1 Real coded GA

Two ways of encoding GA are binary or real. The real-coded GA eliminates the difficulties of achieving arbitrary precision in decision variables. The GA code used in this paper is the Fortran GA driver based on real code developed by Yedder (2007), with modifications. In the code the niching (sharing) technique is activated emphasizing distant solutions in the variable space to remain in the population. The elitism strategy is also included guarantying that the best individual is replicated into next generation. The number of generations is a termination criterion chosen, since the solutions are getting better along the generations. The genetic operators of selection, crossover and mutation are applied. The selection operator used is roulette wheel where each individual in the population is assigned a space on the roulette wheel, which is proportional to the individual relative fitness (Arumugam et al., 2005). The routine can apply two types of crossover and two types of mutation. Crossover operator can be multi-point crossover (expressed by multip) with four-points crossing sites or SBX crossover (simulated binary crossover) where the spread of children solutions around parent solutions can be controlled using a distribution index set in 2. With SBX crossover operator any arbitrary contiguous region can be searched, provided there is enough diversity maintained among the feasible parent solutions (Deb, 2000). Mutation operator can be non-uniform or gaussian mutation. Non-uniform mutation (denoted by nonuni) is a special dynamic mutation operator that improves single-element tuning and reduces the disadvantage of random mutation in the real encoding (Michalewicz, 1992). Gaussian mutation (expressed by varnor) consists in adding a random value from a Gaussian distribution to each element of an individual's vector to create a new offspring (Hussain, 1998).
Crossover and mutation operators are not performed on every individual, its frequency being controlled by a crossover probability (Pc) and mutation probability (Pm).

3. Case Study and Formulation of the Optimization Problem

In this paper the o-cresol hydrogenation to obtain 2-methyl-cyclohexanol, which is carried out in a three phase catalyst slurry reactor.
The mathematical model of this reactor was developed by Vasco de Toledo et al. (2001) and respective equations can be found in Rezende et al. (2006). In order to perform the optimization of the reactor and to find the GA significant parameters, the GA real-coded is coupled with the non-linear mathematical model.
The optimizing variables in this model are linear velocity of gas (ug), linear velocity of liquid (ul), linear velocity of coolant (ur), hydrogen concentration in the gas phase in the reactor feed (Agf), hydrogen concentration in the liquid phase in the reactor feed (Alf), o-cresol concentration in the liquid phase in the reactor feed (Blf), feed reactor temperature (Tf) and the feed coolant temperature (Trf).
The objective function is given by the maximization of the productivity of 2-methyl-cyclohexanol, as calculated by Equation 1:

$$f(x) = \text{Productivity} = \frac{(Blf - Bl) * ul}{L} \quad (1)$$

where L is the reactor length that is equal to 2 meters.
Since, the productivity of the process is deeply dependent on the o-cresol conversion, the o-cresol conversion is defined as a constraint of the process by Equation 2:

$$g_1(x) = \text{Conversion} = \frac{Blf - Bl}{Blf} > 0.90 \quad (2)$$

The optimization problem can be written as in Equation 3:

$$\begin{aligned} &\max_x f = \max \text{ productivity} \\ &\text{subject to: Model equations} \\ &\qquad \text{Conversion} > 90\% \end{aligned} \quad (3)$$

where x is vector composed by the input variables (ug, ul, ur, Agf, Alf, Blf, Tf and Trf).

In order to manipulate the constraint of the optimization problem, the constraint handling method proposed by Deb (2000) and given by Equation 4is used.

$$F(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } g_j(\mathbf{x}) \geq 0 \ \ \forall j=1,2,...,nc \\ f_{max} + \sum_{j=1}^{m} \langle g_j(\mathbf{x}) \rangle & \text{otherwise} \end{cases} \quad (4)$$

The parameter f_{max} is the objective function value of the worst feasible solution in the population. The fitness of an infeasible solution not only depends on the amount of constraint violation, but also on the population of solutions at hand. However, the fitness of a feasible solution is always fixed and is equal to its objective function value.

4. Factorial Design

The factorial design method is a statistical technique that evaluates the process variables (factors) and determines which ones really exert significant influence on the final response (Costa et al., 2005).
The factors to be considered in the factorial design proposed in this work are the GA parameters, to know, population size (popsize), number of generations (genemax), crossover probability (Pc), mutation probability (Pm) and two qualitative factors: type of crossover (typec) and type of mutation (typem). The response is the variable of interest, which in this work is the productivity of 2-methyl-ciclohexanol (fitness function). In order to analyze the GA parameters influence on the fitness function, a factorial design was constructed using the Statistica software (StatSoft v.7.0).
Table 1 show the levels of the quantitative factors used in the factorial design. The values of zero level (central point) are based on the suggestions of the literature.

Table 1 - Levels of the quantitative factors used in the GA code.

GA Parameters (quantitative factors)	(-) level	Central point	(+) level
(1) popsize	64	80	96
(2) genemax	40	50	60
(3) Pc	0.64	0.80	0.96
(4) Pm	0.10	0.125	0.15

The variables typec and typem are considered qualitative factors in the factorial design. Qualitative factors are those where the factor settings are categorical in nature, and which, therefore, cannot be continuously adjusted. Center points cannot be added for qualitative factors; hence, to balance the design, when you request center points, Statistica software will construct full factorial designs for all qualitative factors at each center point, for all continuous factors (Statistica Eletronic Manual, 2004). The meanings of types of crossover (sbx and multip) and of the types of mutation (nonuni and varnor) were explained in item 2.1.
Table 2 shows the levels of the qualitative factors used in the factorial design.

Table 2 - Levels of the qualitative factors used in the GA code.

GA Parameters (qualitative factors)	**(-) level**	**(+) level**
(5) typec	sbx	multip
(6) typem	nonuni	varnor

Since the factors are six, a 2^{6-1} fractional factorial design was chosen requiring thirty two optimization trials plus four optimization trials of central point. The spreadsheet containing the thirty six optimization trials is presented in Table 3. The last column brings the best fitness function (productivity) in the last generation for each run.

Table 3 - Fractional factorial design 2^{6-1} study results for the three-phase reactor.

Runs	popsize	genemax	Pc	Pm	typec	typem	Productivity (x10^{-4} kmol/m^3 s)
1	64	40	0.64	0.1	-1	-1	1.424923
2	96	40	0.64	0.1	-1	1	1.227468
3	64	60	0.64	0.1	-1	1	1.233618
4	96	60	0.64	0.1	-1	-1	1.424923
5	64	40	0.96	0.1	-1	1	0.974902
6	96	40	0.96	0.1	-1	-1	0.871991
7	64	60	0.96	0.1	-1	-1	1.121153
8	96	60	0.96	0.1	-1	1	0.974902
9	64	40	0.64	0.15	-1	1	1.198361
10	96	40	0.64	0.15	-1	-1	0.865833
11	64	60	0.64	0.15	-1	-1	0.865833
12	96	60	0.64	0.15	-1	1	1.335519
13	64	40	0.96	0.15	-1	-1	0.854199
14	96	40	0.96	0.15	-1	1	0.843818
15	64	60	0.96	0.15	-1	1	0.843818
16	96	60	0.96	0.15	-1	-1	0.923729
17	64	40	0.64	0.1	1	1	1.210146
18	96	40	0.64	0.1	1	-1	1.244694
19	64	60	0.64	0.1	1	-1	1.245954
20	96	60	0.64	0.1	1	1	1.210146
21	64	40	0.96	0.1	1	-1	1.380034
22	96	40	0.96	0.1	1	1	1.282908
23	64	60	0.96	0.1	1	1	1.282908
24	96	60	0.96	0.1	1	-1	1.380035
25	64	40	0.64	0.15	1	-1	1.407136
26	96	40	0.64	0.15	1	1	1.142728
27	64	60	0.64	0.15	1	1	1.143843
28	96	60	0.64	0.15	1	-1	1.407136
29	64	40	0.96	0.15	1	1	1.284729
30	96	40	0.96	0.15	1	-1	1.372017
31	64	60	0.96	0.15	1	-1	1.372017
32	96	60	0.96	0.15	1	1	1.285247
33 (C)	80	50	0.8	0.125	-1	-1	0.850719
34 (C)	80	50	0.8	0.125	1	-1	1.401577
35 (C)	80	50	0.8	0.125	-1	1	0.868171
36 (C)	80	50	0.8	0.125	1	1	1.143311

Table 4 presents the effect estimates of the parameters and interactions effects between the parameters. In bold are the parameters statistically significant.

Table 4 - Effect estimates on productivity for the fractional factorial design with two factor interactions calculated with 95% of confidence.

Factor	Effect	p	Factor	Effect	p
Mean/Interc.	**1.163901**	**0.000000**	**1 by 2**	**0.107217**	**0.032969**
			1 by 3	-0.019234	0.677743
(1)popsize	-0.003155	0.945488	1 by 4	0.028916	0.533773
			1 by 5	0.002923	0.949491
(2)genemax	0.029056	0.531829	1 by 6	0.019456	0.674255
			2 by 3	0.010846	0.814349
(3)Pc	-0.096241	0.052026	2 by 4	-0.003016	0.947890
			2 by 5	-0.028694	0.536878
(4)Pm	-0.084046	0.084872	2 by 6	-0.010938	0.812793
			3 by 4	0.022889	0.621407
(5)typec	**0.249594**	**0.000043**	**3 by 5**	**0.174755**	**0.001748**
			3 by 6	-0.019834	0.668350
(6)typem	-0.051520	0.247929	**4 by 5**	**0.106300**	**0.034266**
			4 by 6	0.044180	0.346225
			5 by 6	-0.084550	0.067876

Table 4 shows the effects and p-values. Although we use p-value analysis to evaluate the significance of the effects (effects with p-level ≤ 0.05 are considered significant), in a simulation work the p-value has no physical meaning, as there are no experimental random errors involved.

Table 4 shows that the most significant effect is the type of crossover (typec). The interactions popsize and genemax, pc and typec, pm and typec have shown to be of great significance. These results show a great influence of the type of crossover, the most important effect, but also show that the population size, number of generations, crossover and mutation probabilities must be considered in further optimizations.

Figure 1 shows the Pareto chart that reinforces the results presented in Table 4.

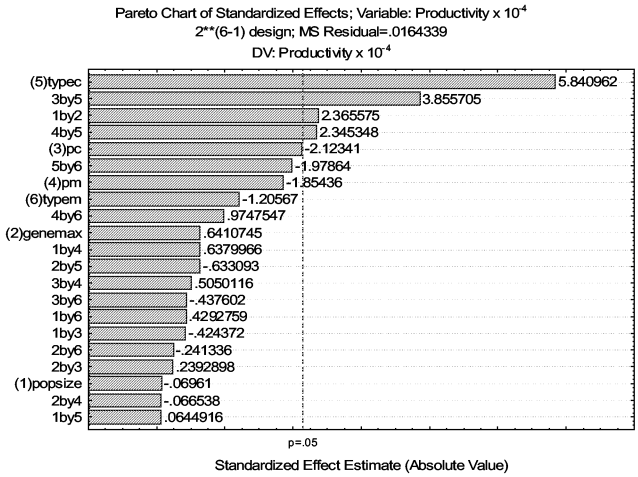

Figure 1 – Pareto Chart for the factorial design study.

In Pareto chart, the effect estimates of all parameters in the factorial design are plotted and statistically significant parameters are the ones that cross the vertical line of p=0.05.

5. Conclusions

A systematic procedure to set up the GA parameters applied to a three-phase catalytic reactor was developed. The factorial design technique was used in order to analyze six GA parameters and to determine the most meaningful effects on the productivity of 2-methyl-cyclohexanol. The results showed that the most statistically significant effect is the type of crossover and that the interaction between this parameter and the crossover and mutation probabilities was also of great significance. The interaction between the population size and the maximum number of generations was also important. In this way, these parameters must be considered in further optimizations.

6. Acknowledgements

The authors thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) of Brazil and The Centre for Rapid and Sustainable Product, Leiria, Portugal.

7. References

Arumugan, M.S., M.V.C. Rao and R. Palaniappan, 2005, Appl. Soft Comput., 38-52, 6.

Baskar, S., P. Subbaraj and M.V.C. Rao, 2003, Hybrid real coded genetic algorithm solution to economic dispatch problem. Comp. and Electrical Eng., 29, 407-419.

Ben Haj Yedder, A., 2006, FORTRAN Genetic Algorithm Driver. In http://cermics.enpc.fr/~benhaj/corps.html accessed in February 2th, 2007.

Costa, C.B.B., M.R. Wolf Maciel and R. Maciel Filho, 2005, Factorial Design Technique Applied to Genetic Algorithm Parameters in a Batch Cooling Crystallization Optimisation. Comp. and Chem. Eng., 29, 2229-2241.

Deb, K., 2000, An Efficient Constraint Handling Method for Genetic Algorithms. Computer Methods in Applied Mechanics and Engineering, 186, 311-338.

Hussain, T.S., 1998, In http://neo.lcc.uma.es/cEA-web/documents/hussain98.pdf accessed in February 22th, 2007.

Michalewicz, Z., 1992, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag, New York.

Rezende, M.C.A.F., R. Maciel Filho, A.C. Costa and R.A. Rezende, 2006, Algorithms for real-time integrated optimization and control: one layer approach. International Symposium on Advanced Control of Chemical Processes, Gramado, RS, 2, 833-838.

Statistica Eletronic Manual, 2004, In Statistica software (StatSoft v.7.0).

Vasco de Toledo, E.C., P.L. Santana, Wolf Maciel, M.R. and R. Maciel Filho, 2001, Dynamic Modelling of a three-phase catalytic slurry reactor. Chem. Eng. Sc., 56, 6055-6061.