# A MODIFIED MULTI-POPULATION GENETIC ALGORITHM FOR PARAMETER IDENTIFICATION OF CULTIVATION PROCESS MODELS

Olympia N. Roeva, Kalin Kosev
*Centre of Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., Sofia 1113, Bulgaria*

Tanya V. Trenkova
*Institute of Water Problems, Bulgarian Academy of Sciences, 1 Acad. G. Bonchev Str., Sofia 1113, Bulgaria*

Keywords: Multi-population Genetic Algorithm, Modification, Mutation, Identification, Cultivation Process.

Abstract: In this work a modified multi-population genetic algorithm (MPGA) without the performance of the mutation operator is proposed. The idea is to reduce the convergence time and therefore to increase the identification procedure effectiveness for on-line application of the algorithm. Modified MPGA, classical multipopulation GA and two other modifications are tested for parameter identification problem of an E. coli non-linear fed-batch cultivation model. The contribution of each modification measure to the performance improvement is demonstrated. The obtained results show that the highest accuracy for parameter identification of the considered model is achieved with the multipopulation GA with Modification 1. The best calculation time is shown by the multipopulation GA without mutation.

## 1 INTRODUCTION

The most popular stochastic optimization method is the evolutionary computation. This is a class of methods based on the ideas of biological evolution, which is driven by the mechanisms of reproduction, mutation, and the principle of survival of the fittest. Several different types of evolutionary search methods have been developed independently. One of them are genetic algorithms (GAs) (Goldberg, 1989), which focuses on optimizing general combinatorial problems. The GAs are highly relevant for industrial applications, because they are capable of handling problems with non-linear constraints, multiple objectives, and dynamic components – properties that frequently appear in real-world problems.

The GAs are widespread optimization techniques and finding applications in a large scope of problems. The application of GAs in bioprocess optimization had been reported in early 1996. Since then GAs are used widely in the field of bioprocesses engineering as an alternative optimization tool to conventional methods (Na, 2002, Roeva, 2009).

Many variations of the standard genetic algorithm, as presented by Goldberg (Goldberg, 1989), can be found in the literature. Modifications and hybridizations have been motivated by a desire to improve the performance of the GA, and to adapt them to particular problem domains (Alsumait, 2010, Kim, 2007, Roeva, 2006). The purpose of this work is to propose a modification of GA, improving the convergence time of the algorithm for a specific problem – parameter identification of non-linear fed-batch cultivation process of *E. coli*.

Better results can be obtained by introducing many populations, called subpopulations compared to the standard GA. Each subpopulation evolves for a few generations isolated before one or more individuals are exchanged between the subpopulations. Thus the evolution of a species resulting in multi-population genetic algorithm (MPGA), is more similar to nature than the single population GA.

In this work a modified MPGA (MMPGA) is proposed and only the operator *crossover* is


N. Roeva O., Kosev K. and V. Trenkova T..
A MODIFIED MULTI-POPULATION GENETIC ALGORITHM FOR PARAMETER IDENTIFICATION OF CULTIVATION PROCESS MODELS.
DOI: 10.5220/0003080603480351
In *Proceedings of the International Conference on Evolutionary Computation* (ICEC-2010), pages 348-351
ISBN: 978-989-8425-31-7

performed. The direct replacement is used where parents are replaced by their offspring (Rowe, 1995). When two chromosomes crossover, they are both replaced by the resulting offspring, then all the original alleles are preserved. This ensures no loss of alleles in the subpopulations. Mutation operator is not performed. The examined GA are tested with a problem for parameter identification of non-linear model of fed-batch cultivation process of *E. coli*.

## 2 MULTI-POPULATION GENETIC ALGORITHM WITHOUT MUTATION

Generally the last operator in the GA is the mutation algorithm. The importance of its role is still a matter of debate. Most authors however, consider that mutation plays a secondary role in the genetic algorithm. Other authors define mutation as an opportunity to prevent the solution from entering in a local maximum.

In connection with the algorithm convergence, some authors propose increase of mutation rates (Louis, 1993). A GA converges when most of the population is identical, i.e. the diversity is minimal. Therefore the increasing of mutation rates is the usual way of maintaining the diversity. Although the high mutation rates may increase the diversity, its random nature raises problems. Mutation is as likely to destroy good schemes as bad ones and therefore elitist selection is needed to preserve the best individuals in a population (Louis, 1993).

The role of every operator in GA as well as the role of the *mutation* operator depends mainly on the specific problem. For the considered problem here – model parameter identification, the operator *mutation* can be eliminated. The experiments show that in this way the better convergence time is achieved without loss of the solution accuracy.

The proposed MMPGA works in a similar way compared to the SMPGA. The subpopulations evolve independently from each other for a certain number of generations (isolation time), like the single population GA. After the isolation time a number of individuals is distributed between the subpopulations (migration). The migration rate, the selection method of the individuals for migration and the scheme of migration determines how much genetic diversity can occur in the subpopulations and the exchange of information between subpopulations. The selection of the individuals for migration can be uniform at random (pick individuals for migration in a random manner) and fitness-based (select the best individuals for migration). There are many variants of the migration structure of the individuals between subpopulations. The most general migration strategy is that of unrestricted migration (complete net topology). Here, individuals may migrate from any subpopulation to another. For each subpopulation, a pool of potential immigrants is constructed from the other subpopulations. The individual migrants are then uniformly at random determined from this pool.

## 3 FED-BATCH CULTIVATION PROCESS OF *E. COLI*

As a test problem a fed-batch cultivation process of *E. coli* is considered. Cultivation of recombinant micro-organisms e.g. *E. coli*, in many cases is the most economical way to produce pharmaceutical biochemicals such as interleukins, insulin, interferons, enzymes and growth factors.

For the parameter identification real experimental data are used. Detailed description of the fed-batch cultivation process of *E. coli* strain *MC4110* is presented in (Arndt, 2004).

The mathematical model of the considered process has the form (Crueger, 1984):

$$\frac{dX}{dt} = \mu_{max}\frac{S}{k_S + S}X - \frac{F}{V}X \tag{1}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\mu_{max}\frac{S}{k_S + S}X + \frac{F}{V}(S_{in} - S) \tag{2}$$

$$\frac{dA}{dt} = \frac{1}{Y_{A/X}}\mu_{max}\frac{S}{k_S + S}A - \frac{F}{V}X \tag{3}$$

$$\frac{dV}{dt} = F \tag{4}$$

where: $X$ is the concentration of biomass, [g/l]; $S$ – concentration of substrate (glucose), [g/l]; $A$ – concentration of acetate, [g/l]; $F$ – feeding rate, [l/h]; $V$ – bioreactor volume, [l]; $S_{in}$ – substrate concentration of the feeding solution, [g/l]; $\mu_{max}$ – maximum growth rate, [$h^{-1}$]; $k_S$ – saturation constant, [g/l]; $Y_{S/X}$, $Y_{A/X}$ – yield coefficient, [-].

The optimization criterion is presented as a minimization of a distance measure $J$ between experimental and model predicted values of state variables as follows:

$$J = J_X + J_S + J_A \rightarrow min \tag{5}$$

$$J_X = \sum_{i=1}^{n} \left( X_{exp}(i) - X_{mod}(i) \right)^2 \quad (6)$$

$$J_S = \sum_{i=1}^{n} \left( S_{exp}(i) - S_{mod}(i) \right)^2 \quad (7)$$

$$J_A = \sum_{i=1}^{n} \left( A_{exp}(i) - A_{mod}(i) \right)^2 \quad (8)$$

where $X_{exp}$, $S_{exp}$, $A_{exp}$ are the vectors of experimental data for biomass, substrate and acetate, $X_{mod}$, $S_{mod}$, $A_{mod}$ – the vectors of simulated data, $n$ – is the number of data for each variable.

# 4 RESULTS AND DISCUSSION

For the problem of parameter identification of model (1) – (4), with an optimization criterion (5) four genetic algorithms are compared:

- *SMPGA*: Standard MPGA (Goldberg, 1989);
- *MPGA Modification 1*: modified MPGA based on modification proposed in (Roeva, 2006);
- *MPGA Modification 2*: MPGA without mutation – here proposed modification;
- *MPGA Modification 3*: MPGA realized using both *Modifications 1* and *2*.

***MPGA Modification 1.*** The reproduction determines which chromosomes will be chosen as the basis of the next generation. Generating populations from only two parents may cause loss of the best chromosome from the last population. The obtained good solution may be destroyed by either the crossover or the mutation or both operations. Thereby, the best solution in GA pops up from the new population may be inferior to the old generations. The aim of the modification (Roeva, 2006) is to prevent this disadvantage.

The modified GA possesses a structure similar to the standard GA. However, the modified GA distinguishes itself from the standard GA in a way the reproduction is processed after both the crossover and mutation have been performed. Thus the deterioration problem never happens since the best solution from the current generation will be superior to or at least the same as the past.

Using the modification proposed in (Roeva, 2006), a modified MPGA is realized – (*MPGA Modification 1*).

***MPGA Modification 2.*** Modified multi-population genetic algorithm proposed in this paper is considered as *MPGA Modification 2*.

***MPGA Modification 3.*** The *MPGA Modification 3* is realized based on application of the two modifications described above. A MPGA without mutation operator and reproduction processed after the crossover operator is developed.

All numerical experiments are done on *Windows Vista* platform, with an Intel Core2Duo, 2.16 GHz, 3GB DDRIII RAM.

The considered GAs are realized in *Matlab 7.5* environment.

As a suitable genetic operators and parameters different authors propose different solutions, depending on the specific problem. The defined in this work GA operators and parameters are based on previous studies of the considered problem here – parameter identification of cultivation process model (see Roeva, 2007).

The parameter identification problem of the model (1) – (4) is solved on the basis of real experimental data for process variables – biomass, substrate and acetate (Arndt, 2004).

The obtained results (the values of the optimization criterion (*J*), as well as the values of the criteria $J_X$, $J_S$, $J_A$ and the convergence time (*T*)) from the four GA – *SMPGA*, *MPGA Modification 1, MPGA Modification 2, MPGA Modification 3* are presented in Tables 1. For each MPGA are presented estimates and criterion mean values of 30 runs (average). The results for minimal (min time) and maximum (max time) computing time are also shown. The results show that the algorithm produces the same estimations with more than 85% coincidence.

Considering the three indicators (average value, minimal time and maximum time) it is clearly noticeable that *MPGA Modification 1* finds the solution for less computing time compared to SMPGA. Using this modification, the best accuracy of the solution is also achieved. The obtained results confirm that the modification (Roeva, 2006) prevents the loss of "the best chromosome" and achieves an increase in solution accuracy.

The best computing time for the three indicators is shown by *MPGA Modification 2* (see Table 4). The elimination of the operator *mutation* decreases considerably the computing time. Minimal solution time of 195.36 s is achieved. The error in this case is slightly higher. The value of the optimization criterion for minimal time solution is 4.0749 compared to the one of *MPGA Modification 1* – 3.9090. The mutation operator changes the individual representation by introducing new genetic material to the gene pool. For this reason, mutation operator tends to preserve or increase the diversity

of the population. As the new material is completely untested, mutation operator often ends up decreasing the fitness of an individual and increasing the convergence time.

Table 1: Results from parameter identification – criteria values and convergence time.

| GA | Indicator | $J_X$ | $J_S$ | $J_A$ | $J$ | $T$, s |
|---|---|---|---|---|---|---|
| *SMPGA* | average | 2.3831 | 1.2583 | 0.0004 | 3.6418 | 316.0752 |
| | min time | 2.4876 | 1.2554 | 0.0004 | 3.7434 | 288.1806 |
| | max time | 2.4363 | 1.1442 | 0.0005 | 3.5810 | 359.8007 |
| *Modif. 1* | average | 2.4198 | 1.3892 | 0.0012 | 3.8102 | 277.8347 |
| | min time | 2.3131 | 1.5951 | 0.0008 | 3.9090 | 259.2737 |
| | max time | 2.2126 | 1.2794 | 0.0019 | 3.4939 | 297.6343 |
| *Modif. 2* | average | 2.4615 | 1.3631 | 0.0021 | 3.8267 | 203.0400 |
| | min time | 2.4631 | 1.6115 | 0.0003 | 4.0749 | 195.3601 |
| | max time | 2.3878 | 1.6648 | 0.0004 | 4.0530 | 214.7198 |
| *Modif. 3* | average | 2.4276 | 1.3830 | 0.0023 | 3.8130 | 240.6050 |
| | min time | 2.5629 | 1.5171 | 0.0043 | 4.0844 | 208.8073 |
| | max time | 2.4597 | 1.0401 | 0.0008 | 3.5005 | 264.4685 |

When the operator *mutation* is eliminated in the proposed in (Roeva, 2006) modification of GA – *MPGA Modification 3,* the convergence time is decreased compared to *MPGA Modification 1* – from 259.27 s to 208.80 s. The proposed *MPGA Modifications 2* and *3* considerably decrease the convergence time of GA, and in the same time the increase of the error is slightly smaller.

# 5 CONCLUSIONS

Based on performed numerical experiments the following conclusions for the performance of the examined MPGA could be generalized:

1. Applying *MPGA Modification 1*, the estimates of the considered model parameters with highest accuracy are obtained. The value of the optimization criterion $J$ is 3.4939 obtained for a time of 297.6343 s.

2. By the *MPGA Modification 2* the best convergence time is achieved. The average results are: $J$ = 3.8267 and $T$ = 203.04 s. The obtained minimal time for solution finding is 195.36 s with an optimization criterion value of 4.0749.

As a result from the conducted experiments and analysis of the received data the multi-population genetic algorithm without *mutation* (*MPGA Modification 2*) is defined as suitable for on-line application for optimization and control of bioprocesses. This is the algorithm with the best convergence time and in the same time the accuracy of the model is comparable with the higher accuracy achieved by *MPGA Modification 1*.

# ACKNOWLEDGEMENTS

This work is partially supported by the National Science Fund Grants DMU 02/4 and DID-02-29.

# REFERENCES

Alsumait, J. S., Sykulski, J. K., Al-Othman, A. K., 2010. A hybrid GA–PS–SQP method to solve power system valve-point economic dispatch problems, *Applied Energy*, 87, 1773-1781.

Arndt, M., Kleist, S., Miksch, G., Friehs, K., Flaschel, E., Trierweiler, J., Hitzmann, B., 2004. A feedforward-feedback substrate controller based on a kalman filter for a fed-batch cultivation of *Escherichia coli* producing phytase, *Computers and Chemical Engineering*, 29(5), 1113-1120.

Crueger, W., Grueger, D., 1984. *Biotechnology. A textbook of industrial microbiology*. Sci. Tech. Inc.

Goldberg, D., 1989. *Genetic algorithms in search, optimization and machine learning*. Addison-Weslcy Publishing Company, Reading, Massachusetts.

Kim, D. H., Abraham, A., Cho, J. H., 2007. A hybrid genetic algorithm and bacterial foraging approach for global optimization, *Information Sci.*, 177, 3918-3937.

Na, J.-G., Chang, Y. K., Chung, B. H., Lim, H. C., 2002. Adaptive Optimization of fed-batch culture of yeast by using genetic algorithms, *Bioprocess and Biosystems Engineering*, 24, 299-308.

Roeva, O., 2006. A modified genetic algorithm for a parameter identification of fermentation processes, *Biotech. & Biotechnol. Equipm.* 20(1), 202-209.

Roeva, O., 2007. Multipopulation genetic algorithm: A tool for parameter optimization of cultivation processes models, *LNCS*, Springer, 4310, 255-262.

Roeva, O., Tzonkov, St., 2009. A genetic algorithm for feeding trajectory optimization of fed-batch fermentation processes, *Int. J. Bioautomation*, 12, 1-12.

Rowe, J., East, I., 1995. Direct replacement: A genetic algorithm without mutation which avoids deception, *Lecture notes in Artificial Intel.*, Springer, 956, 41-48.

Louis, S. J., Rawlins, G. J. E., 1993. Predicting convergence time for genetic algorithms, Technical Report TR370, Department of Computer Science, Indiana University, https://www.cs.indiana.edu/cgi-bin/techreports/TRNNN.cgi?trnum=TR370.