Optimum system identification incorporating mating technique in genetic algorithm

F.A. Zainuddin[1], M. F. Abd Samad[1, 2, *]

[1]Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
[2]Center for Advanced Computing Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.

*Corresponding author's email: mdfahmi@utem.edu.my

ABSTRACT: System identification is concerned with the derivation of a mathematical model from input and output data to explain the dynamical behavior of a system. As a popular search method, a genetic algorithm (GA) is used for selecting a model structure in this paper. It incorporates a mating technique within GA such that exploitation of the information carried by only one parent is sufficient in achieving mating that demonstrates high exploration capability. Two discrete-time systems of linear and nonlinear types are simulated and identified. The results show that a specific mating percentage using the proposed technique exists in enabling the search for optimal solutions sooner than without mating.

Keywords: *Genetic algorithm, Mating, System identification*

1. INTRODUCTION

System identification (SI) is known as a field of study where a mathematical relation between variables and terms of a process is determined. This is done based on observed input-output data with the aim to enable better control of a system [1]. The four main steps of SI involve data acquisition, selection of the model structure, parameter estimation and model validation. The selection of model structure, in particular, is a step where the variables and terms of a generally specified model, are scrutinized so that only significant variables and terms are kept to build the final model that satisfies a given objective function [2].

GA has proven its strength and durability in many areas and is therefore considered as an optimization tool for many researchers [3]. This paper aims to use GA for system identification yet incorporates a new mating technique that is aimed to maintain a suitable balance between exploration and exploitation of search space in the search for optimal models.

2. METHODOLOGY

GA begins with random initialization of a population of chromosomes of a specific number. The chromosomes that represent specific SI models are evaluated based on a specific optimality measurement, also called objective function (OF) of the models in identifying the system at hand. The objective function evaluates the model for both accuracy and parsimony (simplicity).

GA is well-known to undergo several processes – selection, crossover and mutation in its search for optimal results. It has become somewhat a stigma that crossover must be done by selecting more than 1 parent. However, the 2 selected parents may be of the same characteristics, causing the process to produce similar offspring to parents, thus suppressing the evolution.

By introducing a suitable mating technique after the selection process, it is possible to maintain high diversity in the outcome of crossover.

In this paper, the mating is done by first, taking parents from the mating pool (constructed after the selection process). Using binary representation, these parents are then copied, and then inverted to form a new set of parents. The mating is then achieved by pairing the original parent with its inverted self. In this way, a bigger search space may be explored and higher variability of offspring may be accomplished in the next generation.

The proposal was made by testing 2 discrete-time simulated models. These models are known as linear or nonlinear autoregressive with exogenous input (ARX or NARX) models. The models are as follows with both numbers of possible models = 2^{20} - 1 = 1048575:

a. Simulated Model 1 (Sim 1):
$$y(t) = 0.5y(t-1) - 0.2y(t-4) + 0.5y(t-8) + 0.6u(t-2) - 0.2u(t-9) + e(t) \tag{1}$$
Search space specification: nonlinearity, $l = 1$, lag order of output $n_y = 10$, lag order of input $n_u = 10$, time delay, $n_k = 1$
Number of correct regressor = 5 out of 20

b. Simulated Model 2 (Sim 2):
$$y(t) = 0.4y(t-1) + 0.4u(t-1) - 0.6u(t-3) - 0.7y(t-1)u(t-1) - 0.2y^2(t-2) + 0.2u^2(t-3) + e(t) \tag{2}$$
Search space specification: $l = 2$, $n_y = 2$, $n_u = 3$, $n_k = 1$
Number of correct regressor = 6 out of 20

The input $u(t)$ is generated randomly from a uniform distribution to represent a white signal. The values are selected from the interval [-1, 1]. The noise

$e(t)$ is also generated randomly from a uniform distribution to represent white noise and the values are from the interval [-0.01, 0.01]. Using the models, 500 data are generated to be used in identification.

GA is used to represent ARX or NARX model structure using binary representation [2]. The specifications of GA are as follows: population size = 200, maximum generation = 100, mutation probability = 0.01 and crossover probability = 0.6. This paper uses roulette-wheel selection, bit-flip mutation and single-point crossover. The elitism strategy is also used.

The objective function (OF) used in evaluating the optimality of the model is parameter magnitude-based information criterion 2 [4], written as follows:

$$PMIC2 = \sum_n ((y(t) - \hat{y}(t))^2 + \sum_j \frac{1}{\theta}; \ \theta \neq 0 \qquad (3)$$

where $y(t)$ and $\hat{y}(t)$ are correct and predicted output, respectively, n is the number of data, θ is the value of parameter while j is the maximum number of parameters in the model.

3. RESULT AND DISCUSSION

Figure 1 shows the average results using 5 runs of 0% to 50% mating techniques for Sim 1 to Sim 2. The graphs are labelled single 0.0 to single 0.5 representing 0% mating to 50% mating using single-point crossover. Based on Figure 1, in the final generation, 30% mating (single 0.3) has the lowest OF value compared to the other percentages of mating.
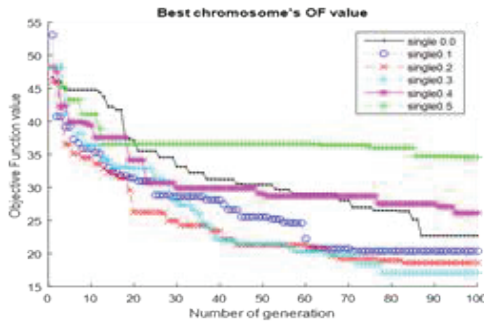

Figure 1 Best chromosome's OF value for Sim 1

Figure 2 shows that 10% mating (single 0.1) has the lowest average OF value in the final generation compared to the other percentages of mating.

The study thus far shows that based on a specific percentage, many optimal models were found. A further check reveals that the selected models (with the lowest OF) were very similar to the simulated ones. By incorporating a mating technique, in comparison to without mating (0% mating), the behavior of the search changes. New points of solution were able to be
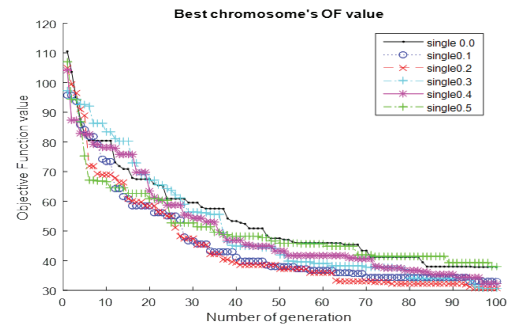
discovered.

Figure 2 Best chromosome's OF value for Sim 2

4. CONCLUSION

GA speeds up the identification of the discrete-time model without the need to evaluate all possible models. In some of the runs, the correct model is almost able to be identified using *PMIC2* as the OF. This is important in achieving a model which is both accurate and parsimonious. Incorporation of a mating technique in GA is shown to be able to find a lower OF model than without mating. At this stage, it is worth noting that there is high feasibility to warrant more studies be done on discrete-time system identification using the mating technique.

REFERENCES

[1] K. J. Keesman, "System identification: an introduction", Springer-Verlag London, 2011.

[2] M. F. Abd Samad, H. Jamaluddin, R. Ahmad & M. Shafiek, "Deterministic Mutation-Based Algorithm for Model Structure Selection in Discrete-Time System Identification", *International Journal of Intelligent Control and Systems*, vol. 16, no.3, pp. 182-190, 2011.

[3] A. B. Hassanat & E. A. Alkafaween, "On enhancing genetic algorithms using new crossovers", *International Journal of Computer Applications in Technology*, vol. 55, no.3, pp. 202-212, 2017.

[4] M. F. Abd Samad & A. R. Mohd Nasir, "Parameter magnitude-based information criterion in identification of discrete-time dynamic system", *Journal of Mechanical Engineering (JMechE)*, no.1, pp.119-128, 2017.