Effectiveness of Mutation Operators in the Genetic Search for Optimal Model in a Sorting-Out GMDH Algorithm

Olha Moroz

International Research and Training Center for Information Technologies and Systems of the NAS and MES of Ukraine

Address: 40, Glushkov Ave, Kyiv, 03680 GSP, Ukraine

olga_moroz@irtc.org.ua

Abstract. *The paper presents a research of different mutation operator effectiveness in searching-out COMBI-GA algorithm used to find model of optimal structure quickly because of avoiding exhaustive search. The obtained experimental results demonstrate that this algorithm performs well when solving artificial inductive modelling task.*

Keywords.

GMDH, Combinatorial algorithm, Genetic Algorithm.

1 Introduction

GMDH algorithms [[1]] belong to the most effective means for solving inductive modelling problems like structur and parametric identification, forecasting, object and processes modeling from statistical or experimental data samples under uncertainty conditions. These algorithms include two main groups, sorting-out and iterative ones. The main example of the first group of algorithms is combinatorial GMDH algorithm COMBI [[2]], the second is multilayer iterative algorithm MIA GMDH of neural network type.

In the COMBI algorithm is free of main MIA GMDH drawbacks such as high probability of loss of relevant arguments and retention of irrelevant ones. A hybrid structure of COMBI and Genetic Algorithms (GA) for over ten years are developing by foreign scientists [[3]]. However combinatorial algorithm has a significant drawback: for searching the best model by a given external criterion it uses exhaustive search of all possible model structures in a given basic set of functions (e.g. polynomial) and accordingly has exponential complexity. As a result, its usage is unreasonable for problems with more than 30 input variables. So studying the efficiency of hybridization between COMBI and GA is relevant and promising.

Hybrid algorithm COMBI-GA [[4]] is able to largely eliminate these deficiencies using genetic algorithms to search optimal model, which is a powerful tool of direct global search in optimization problems, have a clear structure and easy implementation. The effectiveness of GA significantly depends on the method of coding individuals in a population, the initial population size, genetic operators, fitness functions and stop criteria of the algorithm. Therefore, the detailed investigation of the impact of each factor on the GA performance is an important task.

The most significant impact on the GA performance make a crossover operator, which mostly affects on a variety of individuals in the population, and thus guarantees the solution of the problem. But often GA gets stuck in local optima. This phenomenon usually occurs when crossover operator is poorly chosen under given specific tasks and also when the fitness function is close to the optimum value or all simultaneously.The main mechanism of GA to solve this problem is the mutation operator with successful selection of the intensity and probability.

The aim of this paper is to analyze the effectiveness of two types of mutations, genetic and chromosomal ones, and their influence on speed of COMBI-GA algorithm, depending on their level of intensity and crossover operator probability. In the numerical experiments all other parameters of GA are fixed.

2 Algorithm COMBI-GA

A general definition of the inductive modelling problem may be done as follows. Let us given: a data set of n observations after m input $x_1, x_2, \ldots, x_m$ and one output y variables. The GMDH task is to find a model $y=f(x_1,x_2,\ldots x_m, \theta)$ with *minimum variance of prediction error*, where θ is unknown vector of model parameters. The optimal model is $f^*= arg\ min_{\Phi}\ C(f)$, where $C(f)$ is a given model quality criterion, Φ is a set of models, $f \in \Phi$. This problem consists of a discrete optimization task for finding the best model and continuous optimization task for estimation of parameters.

The two tasks are generally solved by the combinatorial algorithm COMBI GMDH in the following way:

1) transformation of the initial data according to the chosen system of basic functions;

2) generation of the complete set of all possible structures of partial models in the selected basis;

3) use LSM to estimate coefficients of each partial model;

4) calculation the selection criterion for each model;

5) successive selection best models by this criterion and choosing the model of optimal complexity with the minimum criterion value.

General structure of the COMBI-GA algorithm. The developed algorithm [??] generates gradually an incomplete set of most promising partial model structures using genetic principles to search for the optimal complexity model. Generally it has structure like that:

1. Generation of a random set of partial model structures of a given size as an initial population of the COMBI-GA.
2. Computing coefficients of every partial model using least squares method (LSM).
3. Calculation of an external criterion values (fitness function of the GA) for each model, for example, the regularity criterion as typical for GMDH.
4. Current selection of the best partial models (elite selection in GA) or reduction–rejection of worst individuals from the parent and offspring populations. Formation of new population of the same size.
5. Check a stopping criterion for example achieving a given accuracy or number of iterations. Stop if it is fulfilled; otherwise go to the next step.
6. Use of genetic operators (crossover and mutation) with given probability to selected individuals in the population and forming a set of partial model structures for next generation. Go to step 2.

3 Ideas of experiments

In numerical experiments individuls (model structures) of GA initial population is encoded by binary vectors of structural elements 0 and 1, indicating absence or presence of a particular argument in the model respectively. As a fitness function of each individuals we use standard GMDH external regularity criterion based on dividing the data into two samples for training and testing: the first is used to generate the different structures of models and evaluating their parameters, and the second to calculate the accuracy of the generated models.

According to results of crossover comparison study [[5]], we consider here two the best crossover operators – one-point and universal [[6]] that are used to form the offspring population from selected elite individuals. The other GA operators and parameters are fixed. The stop criterion of the hybrid algorithm is achieving given accuracy for the difference of minimum values of the fitness function on two adjacent iterations.

To produce a new offspring by the one-point (OP) crossover, a chromosome point (area between neighboring bits in a row) is randomly selected, the fragment of binary string from beginning of the first parent chromosome to the point is copied, and the rest is copied from the second parent.

Universal crossover instead of crossover point uses an additional binary mask with length equal to that of the chromosome's set. Each offspring is formed by applying the mask to every parent with the some rule (0+0=0, 0+1=1, 1+1=0). The mask can be set randomly.

Gene mutation of an individual is the inversion of a given number of randomly selected bits of chromosome with probability 0.2. A given number of inverted bits forms the level of mutation intensity. Also the paper deals with next gene mutation variation: several genes are selected randomly in a chromosome, which number may be random; only zero values of the selected genes are reversed. The chromosomal mutation is turning whole chromosome or its fragment in 180^0. The level of intensity here is the length of the fragment.

The research was carried out on the test task using laptop with a dual core processor (2.7 GHz). There were artificially generated 50 input arguments by uniformly distributed random numbers generator, and only 25 of the inputs were used as relevant ones. The task of inductive modeling using COMBI-GA is searching the true dependence on the relevant arguments. The optimal model is searching in the class of linear functions.

3 The results of computing experiments

During the experiments it was found that mutation operator in fact make no effect on the COMBI-GA algorithm with universal crossover operator. The optimal model was obtained for 1.4 sec.

For the one-point crossover operator, using mutations can significantly speed up the process of finding the true model for 1.3–1.7 sec. (2.4–5.5 sec. without mutation) Choose the mutation operator and its intensity strongly depends on the probability of crossover operator. In particular, chromosomal mutation (intensity 3 to 15) gives better result (more stable and more accurate) with small crossover probability (0.6–0.8).

If the probability of crossover operator is high (0.8–1), it is better to use gene mutation with a few inverted positions (<10). This provides a better diversity and quick finding the optimal model.

Corresponding results are shown in Fig. 1 and Fig. 2. The pair of modified gene mutation and one-point crossover gives best COMBI-GA effectiveness. Fig. 3 shows mutations influence on the convergence of the COMBI-GA algorithm.

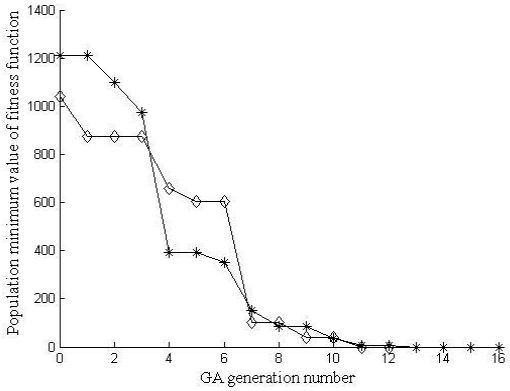

Fig. 1 COMBI-GA convergence without using (*) and with using (◊) mutation, crossover probability 0.9.

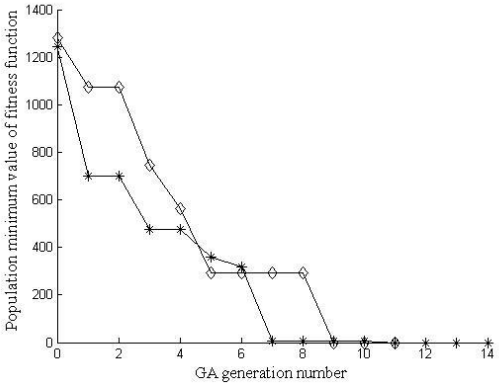

Fig. 2 COMBI-GA convergence without using (*) and with using (◊) mutation, crossover probability 0.7.

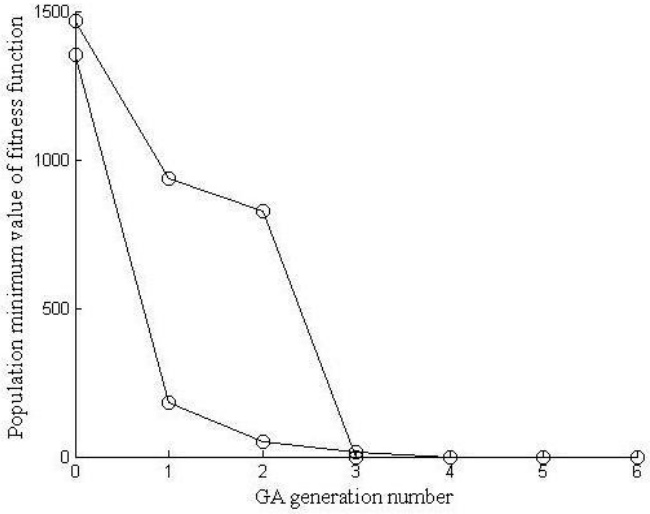

Fig. 3 COMBI-GA convergence using new variant of gene mutation for one-point crossover with different probability 0.7 (3 generations) and 0.9 (6 generations)

4 Conclusion

The use of GA is very effective to search true model in sorting-out GMDH algorithms for solving inductive modelling tasks with large number of input variables (much more than 30) during really small time.

A mutation operator is able to accelerate calculating speed of the algorithm depending on the kind of task, crossover operator and its probability. The best way to accelerate is to use the new bioinspired gene mutation operator, which is less dependent on crossover probability and has good influence on effectivenes of the COMBI-GA algorithm.

In the future, it would be reasonable theoretically substantiate the results obtained in practice when solving inductive modelling tasks using the COMBI-GA algorithm.

References

[1] Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наук. думка. (1985). 216 с.

[2] Степашко В.С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей. Автоматика. № 3 (1981). С. 31-36.

[3] Мороз О.Г., Степашко В.С. Огляд гібридних структур МГУА-подібних нейронних мереж та генетичних алгоритмів. Збірник наукових праць "Індуктивне моделювання складних систем. Випуск 7. (2015). С. 173-191.

[4] Мороз О.Г., Степашко В.С. Перебірний алгоритм МГУА з генетичним пошуком моделі оптимальної складності. Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту». Збірник матеріалів. (Залізний Порт, 26-30 травня 2016). Херсон: Вид-во ПП Вишемирський В.С. (2016). С. 297-299.

[5] Мороз О.Г Порівняння ефективності операторів кросинговеру при застосуванні генетичного пошуку кращих моделей в перебірному алгоритмі МГУА. VI Всеукраїнська школа-семінар молодих вчених і студентів «Сучасні комп'ютерні інформаційні технології» (АСІТ-2016). Збірник матеріалів. Тернопіль: Вид-во ТНЕУ. (2016). С. 77-79.

[6] Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. Москва.: ФИЗМАТЛИТ. (2006). 320с.