New Operators for Faster Convergence and Better Solution Quality in Modified Genetic Algorithm

Pei-Chann Chang, Yen-Wen Wang, and Chen-Hao Liu

Dept. of Industrial Engineering and Management, Yuan Ze University,
Chung-Li, Taiwan 32026
iepchang@saturn.yzu.edu.tw

Abstract. The aim of this paper is to study two new forms of genetic operators: duplication and fabrication. Duplication is a reproduce procedure that will reproduce the best fit chromosome from the elite base. The introduction of duplication operator into the modified GA will speed up the convergence rate of the algorithm however the trap into local optimality can be avoided. Fabrication is an artificial procedure used to produce one or several chromosomes by mining gene structures from the elite chromosome base. Statistical inference by job assignment procedure will be applied to produce artificial chromosomes and these artificial chromosomes provides new search directions and new solution spaces for the modified GA to explore. As a result, better solution quality can be achieved when applying this modified GA. Different set of problems will be tested using modified GA by including these two new operators in the procedure. Experimental results show that the new operators are very informative in searching the state space for higher quality of solutions.

1 Introduction

Since Holland proposed the genetic algorithm back in 1975, this adaptive system, which is biologically motivated, has been successfully applied to solve different application problems. In the standard GA, the population diversity is obtained and maintained using the genetic operators of crossover and mutation, which allow the GA to find more promising solutions and avoid premature convergence to a local maximum (Goldberg 1989). However, the use of the genetic operators has been the object of study of many researchers. Some important work related with crossover and mutation can be found in (Davis 1989; De Jong et al. 1992; Schaffer et al. 1991).

In addition to the traditional genetic operators, many researchers have presented new genetic domain-dependent operators, for instance, (D'Haeseleer 1993; Mathias et al. 1992). Nevertheless, no new biologically inspired genetic operators have been widely adopted since the advent of GAs. Mitchell et al. (1994) point out the importance of studying new genetic operators. Mitchell et al. (1994) and Mitchell (1996) state that it would be interesting to analyze if any of these biological mechanisms, incorporated in a GA, could lead to any significant advantages. Banzhaf et al. (1998) share the same opinion and they highlight the significance of imple-

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3611, pp. 983 – 991, 2005.

menting evolutionary approaches using mechanisms such as conjugation, transduction or transposition.

This paper tries to develop two new operators, which can be embedded in the original GA procedure. They are duplication operator and fabrication operator. Duplication operator just like a general biological cloning technology, it clones the chromosome from the parent to breed the offspring. Fabrication operator attempts to extract the superior gene structure from the chromosome, and will generate the new offspring. These two new operators will either duplicate or fabricate new chromosomes from the chromosome base. With these new duplicated or artificial chromosomes, the convergence rate and solution quality of the GA searching procedure will be improved greatly.

The rest of the paper is divided in five sections. Section 2 introduces the modified GA. The following section describes two new operators: duplication and fabrication and gives detailed procedure of these two operators. Section 4 is the experimental tests conducted to test the quality of solution generated by using these two new operators. Finally, conclusion is made and future direction of the research is provided.

2 The Modified GA

The evolution procedure of modified GA as shown in Figure 1 is pretty similar to the procedure in general GA except that two new operators are included in the procedure, i.e., duplication and fabrication operators. The modified GA starts with a randomly initialized population of candidate solutions and then assigns fitness value to each individual in the population. Individuals with highest fitness value will be extracted into the elite chromosome base for storage. Then roulette wheel selection procedure will be applied to select N pairs of parents for reproduction. Later on, the modified GA associates each individual candidate in the population with a fitness, which measures the quality of a solution. Selection chooses individuals probabilistically, according to their fitness. The higher the fitness, the more likely it is for an individual to be selected. Next, duplication operator will duplicate d% of best fit individual into the next generation population. Duplication operator is similar to Elitism, but duplication operator clones the elitist repeatedly and Elitism does not.

While fabrication operator will produce f% of individuals from the elite chromosome base by assigning job into the position using statistical inference. These individuals are called artificial chromosomes. After that, regular crossover and mutation operator will be applied again to generate new individual for the next generation population. Crossover and mutation produce new individuals: the first operator exchanges genetic information between two selected parents; mutation randomly changes one gene value to the generated offspring. The modified GA searches through an iterative process: the process of one generation involving selection, duplication, fabrication, crossover and mutation is called one cycle of iteration and is repeated until convergence is reached or the number of generations achieves the established limit.

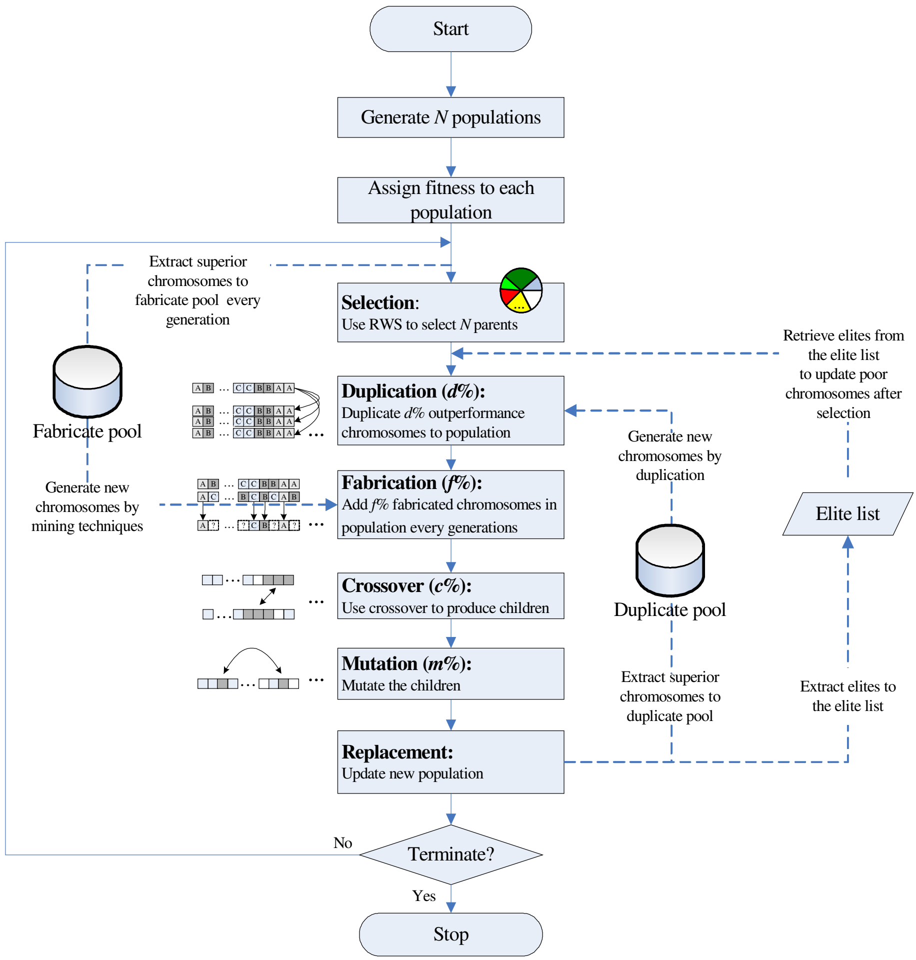

Fig. 1. The Modified Genetic Algorithm

3 New Operators

The cloning of "dolly" in 1997 is the first example in real life that duplication can be applied to copy an existing chromosome to retain the quality of solution in the next generation. Duplication is a very easy and fast computational procedure. However, the duplication operator can be applied to intelligently evolve population and fine-tune the solution quality that is the merits of this research try to achieve. Fabrication is also a new operator to be dealt with. Duplication is a reproducing procedure of a chromosome but fabrication is to make up a new chromosome with the help of statistical information from the elite chromosome base. With the advance of the data mining technology, more sophisticated chromosome can be produced through the fabrication procedure. Detailed explanations of these two new operators are given in the following two sections.

3.1 Duplication

Duplication is a very simple procedure in the modified GA and it makes a single copy of the best fit chromosome, similarity to Elitism. The difference of duplication and Elitism is that, Elitism only retains the best fit chromosome in the next generation, but duplication involves multiple copies of the best fit chromosome in the next generation. However, questions such as the particular individual to be copied? Or how many copies to be made during the evolution process? How does duplication affect the convergence rate of the GA searching procedure? They are interesting questions to be answered later on in this research.

Actually, duplication may increase convergence rate by introducing more elite chromosome into the population, however, there is a chance that the searching procedure might be trapped into local optimality. The percentage of duplication should be controlled within certain limit in order to keep diversity of the searching procedure in the population.

The modified GA by including duplicate operator is described as follows:

1. Following the general GA's procedure, first generate initial populations, and calculate their fitness. Then, using the selection operator (i.e., the RWS) to produce candidate chromosomes for next generation.
2. To apply the "duplicate" operator, we will extract the best chromosome from the Duplicate Pool and then reproduce it N times (N is determined by the rate of duplicate, i.e., d %).
3. After that, we substitute the worst N chromosomes in the original populations by the duplicated chromosomes.
4. And then, just like the general GA's procedure, crossover, mutation, replacement.
5. During the replacement, we will update the "Duplicate Pool", i.e., the best chromosome will be retracted and recorded in Duplicate Pool.
6. Terminate or not? If not, go to 2. Else, 7.
7. Stop the GA and output the best solution.

3.2 Fabrication

Fabrication is the procedure to make up a set of new chromosome based on the elite chromosome base. There is a lot of gene information left in the elite chromosome base; however general GA searching procedure just reuses only 20% of the chromosomes generated. Fabrication will follow the sequence structure of each chromosome in the elite chromosome base, and according to the votes from the elite chromosome base; a job-position matrix M_{ij} can be formed, i.e., a dominance matrix describing the number of times job shown up in each position will be recorded in this matrix. For each chromosome, the gene represents the job and the sequence the position each job is assigned. We will count the number of times job showing up in each different positions and recorded in the matrix. Thus a dominance matrix generated from the chromosome base is formed. Next we will mine this matrix according to the votes from each elite chromosome, and fabricate an artificial chromosome.

The algorithm for fabricate operator is described as follows:

Let A : the set of cities have to be assigned

A' : the set of assigned-cities

B' : the set of weed-out cities

$|V_l|$: the highest number of vote of the $l-th$ sequence = $MAX(V_{kl}), \quad \forall k$

$|C_l|$: the city with the highest number of vote of the $l-th$ sequence

Processes of generate fabricated chromosomes:

Step 1. For all l in A, find l, where $|V_l|$ is the maximum

If there is any other l' where $|V_{l'}| = |V_l|$, go to step 3; otherwise, go to step 2.

Step 2. Remove city $|C_l|$ from A to A', and let it to hold the $l-th$ sequence in fabricated chromosomes. Then, go to step 4.

Step 3. Remove cities $|C_l|$ and $|C_{l'}|$ from A to B'. Then, go to step 4.

Step 4. If $A = \phi$, then, go to step 5; otherwise, go back to step 1.

Step 5. Random assign the cities in B' to the unassigned sequences in fabricated chromosomes.

In the following is a simple example for our fabricate operator:

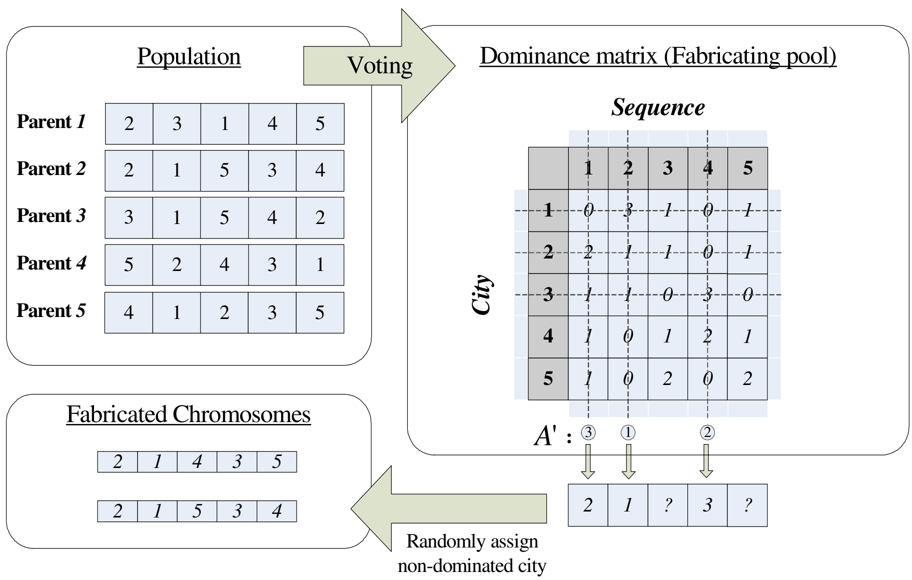

Fig. 2. The Diagram of Fabrication

4 The Test Functions and Results

First, we will analyze the duplication and fabrication results individually, explaining how the duplicating and fabricating operators can influence the performance of the GA. Empirical results show how we can choose the appropriate size for the duplicating sequences, depending on the size of population.

4.1 Experimental Test for Parameter Setup

We setup the parameters using three different levels for duplication and fabrication and they are 10%, 30% and 50%.

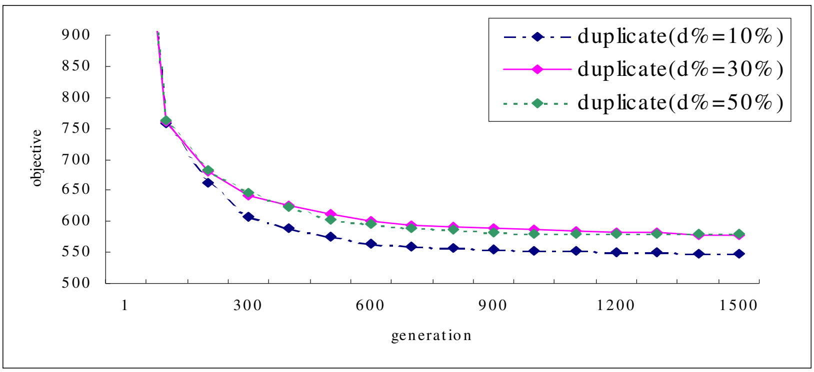

Fig. 3. Test for duplication percentage: d%

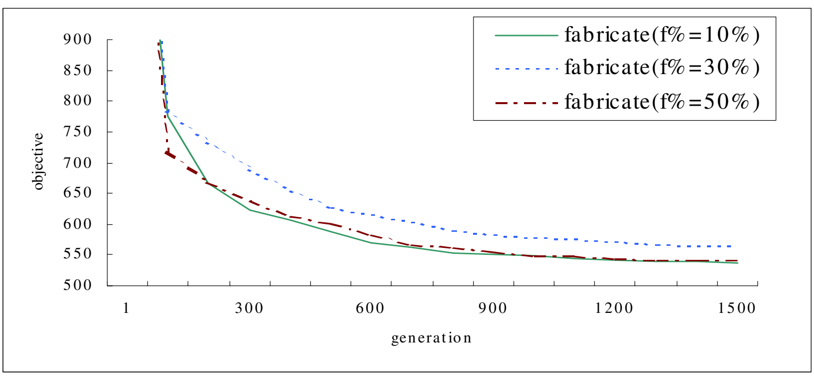

Fig. 4. Test for fabrication percentage: f%

The final result of this testing can be shown in the figures above. The duplication and fabrication operator can further improve the convergence of the algorithm however the duplicating rate and fabricating rate should be controlled at 10%. That is when increasing duplicating rate and fabricate rate over 10% there is a chance for the algorithm to be trapped into local optimality.

4.2 Testing for TSP Problem

The benchmark problem of 51-cities Traveling Salesman Problem from TSPLIB is applied to examine the performance of duplicate and fabricate operators..

Table 1. The mean and std. of final results

	$MEAN(f_{\min})$	$STD(f_{\min})$
Pure GA	515.3	17.47
GA-Duplicate	484.08	18.16
GA-Fabricate	479.78	9.55

From the results above, the modified GA obtains an improvement in terms of efficiency and solution quality when compared to general GA. Both mean objective values from GA-duplication and GA-fabrication are much less than those of pure GA. This implies that duplication and fabrication operators can further improve the quality of solution and rate of convergence. In addition, the standard deviation of GA-fabrication is much less than that of GA-duplication. This indicates that the fabrication operator seems posses a better robust performance during the searching procedure.

4.3 Testing for Continuous Function

The evolutionary parameters for our modified GA are designed as follows:

Table 2. Evolutionary parameters used for testing

Number	Items	Values
1	The scale of the population	50
2	The maximum number of iteration times	200
3	Crossover rate	0.85
4	Mutation rate	0.1
5	Number of calculation times	10
6	Length of binary code of each variance	8
7	Percentage of duplicate (d%)	10%
8	Percentage of artificial (f%)	10%

The five benchmark functions are chosen and tested in our numeric experiments and they are listed in the followings:

Example 1: GP (Goldstein-Price function)

$$f(x_1,x_2)=[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2+6x_1x_2+3x_2^2)] \times[30+(2x_1-3x_2)^2\times(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)] \quad (1)$$

Where $-2 < x_i < 2$, $i=1,2$, $\min f(x_1,x_2)=3$

Example 2: BR (Branin)

$$f(x_1,x_2)=a(x_2-bx_1^2+cx_1-d)^2+e(1-f)\cos(x_1)+e \quad (2)$$

Where $a=1$, $b=5.1/(4\pi^2)$, $c=5/\pi$, $d=6$, $e=10$, $f=1/(8\pi)$, $-5\le x_1\le 10$, $0\le x_2\le 15$, $\min f(x_1,x_2)=5/(4\pi)$

Example 3: RO (Rosenbrock function)

$$f(x_1,x_2)=(1-x_1)^2+105(x_2-x_1^2)^2 \quad (3)$$

Where $-2 < x_i < 2$, $i=1,2$, $\min f(x_1,x_2)=0$

Example 4: RA (Rastrigin function)

$$f(x_1,x_2)=x_1^2+x_2^2-\cos(18x_1)-\cos(18x_2) \quad (4)$$

Where $-1\le x_i\le 1$, $i=1,2$, $\min f(x_1,x_2)=-2$

Example 5: SH (Shubert function)

$$f(x_1,x_2)=\left\{\sum_{i=1}^{5} i\cos((i+1)x_1+i)\right\}\left\{\sum_{i=1}^{5} i\cos((i+1)x_2+i)\right\} \quad (5)$$

Where $-10\le x_i\le 10$, $i=1,2$, $\min f(x_1,x_2)=-186.7309$

From table 5, it is found that in most of continuous benchmark problems, our modified GA outperforms the general GA in solution quality and convergence rate except for the RO. problem. But when we observe the convergence charts, it is clearly to know that, the modified GA still has achieved a higher rate of convergence.

Table 3. Comparison of the final results

	The average generation times when the f_{max} is reached for the first time	
Continuous Problem	Pure GA	Duplicate & Fabricate GA
1. GP.	228	51
2. BR.	>300	178
3. RO.	158	256
4. RA.	>300	20
5. SH.	>300	271

5 Conclusions

This research develops two new operators: Duplication and Fabrication. Duplication is a very useful operator that can speed up the convergence rate after a series of experimental tests including TSP and Continuous test problem sets. The modified GA-duplication can have a faster convergence rate and produces near optimal solution within 100 generations when compared with traditional GA&Ellite. Fabrication is another new operator that deserves our attention. With the advanced in data mining technique, we can introduce more sophisticated chromosome into the GA procedure. One interesting phenomena we observe that Fabrication operator can further improve the quality of the solution by almost 10% if properly adjusted in the GA-fabrication procedure. Furthermore, duplication and fabrication performance depend essentially on two factors: the percentage of them and when to use them. We would like to explore more application of these two operators in the future.

References

1. Banzhaf, W., Nordin, P., Keller, R.E., Francone, F.D.: Genetic Programming – An Introduction – On the Automatic Evolution of Computer Programs and its Applications. CA: Morgan Kaufmann (1998).
2. Davis, L.: Adapting Operator Probabilities in Genetic Algorithms. In J. D. Schaffer (ed.), Proceedings of the Third International Conference on Genetic Algorithms, CA: Morgan Kaufmann (1989) 61-69.
3. Jong, K. A. D., Spears , W. M.: A Formal Analysis of the Role of Multi-Point Crossover in Genetic Algorithms. Annals of Mathematics and Artificial Intelligence, Vol. 5, No. 1, (1992) 1-26.
4. D'Haeseleer, P.: Context Preserving Crossover in Genetic Programming. In Proceedings of the 1994 IEEE World Congress on Computational Intelligence, Vol. 1, (1993) 256-261.
5. Goldberg, D. E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Publishing Company, Inc. (1989).
6. Mathias, K., Whitley, D.: Genetic Operators, the Fitness Landscape and the Traveling Salesman Problem. In R. Männer and B. Manderick (eds.), Proceedings of Parallel Problem Solving from Nature Vol. 2, (1992) 219-228.
7. Mitchell, M., Forrest, S.: Genetic Algorithms and Artificial Life. Artificial Life, Vol. 1, No. 3 (1994) 267-289.
8. Mitchell, M.: An Introduction to Genetic Algorithms. MIT Press (1996).
9. Schaffer, J. D., Eshelman, L. J.: On Crossover asan Evolutionarily Viable Strategy. In R. K. Belew, L. B. Booker (eds.), Proceedings of the Fourth International Conference on Genetic Algorithms, (1991) 61-68.