Comparing Genetic Operators with Gaussian Mutations in Simulated Evolutionary Processes Using Linear Systems

D. B. Fogel[1] and J. W. Atmar[2]

[1] ORINCON Corporation, 9363 Towne Centre Dr., San Diego, CA 92122, USA
[2] AICS Research, Inc., University Park, NM 88003, USA

Abstract. Evolutionary optimization has been proposed as a method to generate machine learning through automated discovery. Specific genetic operations (e.g. crossover and inversion) have been proposed to mutate the structure that encodes expressed behavior. The efficiency of these operations is evaluated in a series of experiments aimed at solving linear systems of equations. The results indicate that these genetic operators do not compare favorably with more simple random mutation.

1 Introduction

The common view of evolutionary processes is a successive selection of the best of a sequence of variants produced by random mutation. Holland (1975) has proposed this process to be equivalent to an enumeration of all possible coding structures. To be successful, adaptive plans require the use of specific sophisticated genetic operators, such as crossover and inversion. However, in several evolutionary simulations (Fogel 1964; Fogel et al. 1966; Atmar 1976: Fogel and Fogel 1986; Fogel 1988) rapid convergence toward optimal behavior was observed, even when using the slightest of mutations. Further, simulating crossover explicitly has resulted in poor overall performance in a variety of studies (Grefenstette et al. 1985; Fogel and Fogel 1986; Fogel 1988). It is of interest to quantify the benefits, if any, of using crossover and inversion as opposed to simple random mutation in a variety of environments.

2 Background

Some of the first simulations of evolutionary processes were conducted by Fogel (1962, 1964; and Fogel et al. 1966) where the task of predicting any stationary or nonstationary environment with respect to an arbitrary payoff function was chosen. Iterative mutation and selection was used to evolve a logic most suitable for resolving the problem at hand. The behavior of each "organism" was portrayed by a finite state machine, a general representation that does not constrain the stimulus/response transduction to be linear, passive, or without hysteresis.

This evolutionary programming was conducted as follows: original finite state machines were measured in their ability to predict each next event in their experience with respect to whatever payoff function had been prescribed. Progeny were then created through random mutation of these "parent" machines. Their predictive ability was scored in a similar manner to their parents. Those finite state machines judged to be superior survived to become the new parents. An actual prediction was made when the predictive fit score demonstrated that a sufficient level of credibility has been achieved. The surviving machines generated the prediction, indicated the logic of this prediction and became the progenitors for the next sequence of progeny. The sequence of predictor machines demonstrated phyletic learning, the inductive generation of hypothesis concerning the relevant regularities found within the experienced environment in light of the give payoff function.

Rather than place primary importance on behavior, others (e.g. Holland 1975; DeJong 1975; Fourman 1985) emphasize the mutational operations which are performed on the given coding structures, specifically mimicking the operations on natural chromosomes (e.g. crossover and inversion). Simple mutation, randomly altering a single component of a chosen coding structure, is also incorporated but only to assure that all possible values have some probability of being included. The effectiveness of these genetic operators as compared to simple random mutation in linear systems of varying degrees of interactivity is quantified here.

3 Methods and Materials

Several experiments were conducted to explore these hypotheses concerning the importance of specific genetic operators. These experiments required solving systems of linear equations by evolutionary optimization. Consider a system of linear equations of n dimensions:

$$b_i = \sum_{j=1}^{n} a_{ij}(x_j), \qquad i = 1, \ldots, n.$$

Let the vector x_j represent the coding structure of n "gene" products, the vector b_i represent n phenotypic behavioral responses, and the coefficients a_{ij} represent the respective contribution of each x_i component to each b_i response. Any such system will be pleiotropic (single genes expressing themselves through multiple effects) unless $a_{ij} = 0$ for all $i = j$.

A fitness function measures the quality of the evolved behavioral responses $a_{ij}x_j$, to the required response vector b_i. Unless the matrix a_{ij} is singular, there exists an ideal behavioral response ($a_{ij}x_{\text{ideal}}$). The fitness criterion is:

$$E = \sum_{i=1}^{n} E_i,$$

where

$$E_i = \left| \sum_{j=1}^{n} a_{ij} - b_i \right|, \qquad i = 1, \ldots, n.$$

E, the total amount of behavioral error will equal zero when $x_j = x_{\text{ideal}}$.

For the purposes of this comparison, the adaptive process is simulated with an arbitrary coefficient matrix a_{ij} of rank 10 and a randomly chosen desired response vector b_i. An initial population of 150 vectors x_j (genotypes) is taken at random with components distributed normally with zero mean and a standard deviation of 30. These vectors are randomly assigned to be one of three types. The first is subject to mutation only by random alteration of each of its components by a standard normally distributed random variable (i.e. zero mean and unit variance). The second is mutated by crossover (80% chance per offspring) and inversion (50% chance per offspring) operators, double mutations being allowed. The third is mutated by crossover, inversion and also given a 1% chance per offspring of random mutation by altering components by a standard normal random variable. This mutation rate is typical in the experiments conducted by DeJong (1975); Fourman (1985); and Davis (1985).

Each vector is assigned a fitness score according to the above fitness function and is then mutated yielding offspring vectors. Rather than simply selecting the best 150 vectors to serve as new parents, each vector is given a probability of survival based on the ratio of its error score relative to other scores in the population of trial vectors. During the course of one generation, each genotype (trial vector) must directly compete with ten other genotypes. The probability of attaining a "win" is equal to the opponent's fitness score divided by the sum of both vector's fitness scores. For example, should a vector with a fitness score of 8.0 compete against a vector having a fitness score of 10.0, the probability of the first vector obtaining a victory is 10/18 or 0.556. Once competition has been completed, the 150 genotypic vectors with the most "wins" become the basis set of the next generation. If the crossover and inversion operators were to provide an advantage over simple random mutation alone, it would be expected that those vectors which undergo such operations would quickly dominate the population.

4 Experimental Findings

Linear systems of equations provide a convenient mechanism for examining the effects of these operators in domains of varying degrees of interactivity. Five sets of trials were conducted. Each varied the degree of interactivity of the a_{ij} matrix by setting the probability of an off-diagonal entry being non-zero to 0.0, 0.25, 0.5, 0.75, and 1.0, respectively. When this probability is zero, the domain is minimally pleiotropic; each independent "gene product" contributes only to the fitness of its respective component. When the probability is 1.0, the domain is fully pleiotropic. That is, each "gene product" contributes to the total behavioral error summed over all behavioral responses. Each experiment consisted of 100 randomly chosen systems. In each system, evolution was halted after 5000 offspring had been evaluated. The results are presented in Table 1.

Advantage quickly accrued to those genotypes which were altered by random mutation alone. The observed number of trials in which simple random mutation dominated the population is significantly greater than would be expected under a null hypothesis of the behavior of the evolutionary process being independent of the utilized mutation operation ($P < 0.0001$, using nonparametric chi-square test). Further, insufficient evidence exists to suggest that the degree of interactivity of the linear system of equations is associated with the degree to which those genotypes undergoing random mutation alone will overtake the population ($P > 0.12$).

In light of these results, a second experiment was conducted to investigate the difference in efficiency between systems using crossover and those not using crossover with varying rates of mutation. Again, a system of ten linear equations was used. The matrix a_{ij} was chosen by setting the entries to random integers between 0 and 9. The b_i vector was chosen so as to make all of the components of $x_{\text{ideal}} = 1$. Initially, the population was composed of 100 vectors, with each component normally distributed with zero mean and a

Table 1

Degree of interactivity		Percentage of trials when more than 50% of the population consisted of the given type
0	Random Mutation Alone	78% (70%)[a]
	Crossover/Inversion	9% (5%)
	Crossover/Inversion/Mutation	13% (4%)
25	Random Mutation Alone	75% (72%)
	Crossover/Inversion	10% (3%)
	Crossover/Inversion/Mutation	15% (3%)
50	Random Mutation Alone	83% (83%)
	Crossover/Inversion	9% (7%)
	Crossover/Inversion/Mutation	8% (4%)
75	Random Mutation Alone	88% (86%)
	Crossover/Inversion	2% (0%)
	Crossover/Inversion/Mutation	10% (6%)
100	Random Mutation Alone	76% (73%)
	Crossover/Inversion	10% (6%)
	Crossover/Inversion/Mutation	14% (6%)

[a] The values in parentheses indicate the percentage of trials in which the given type had completely taken over 100% of the population

Table 2. Using a sign-test on the differences between the best evolved solution to each of the ten trials, no significant evidence existed to suggest that Method #2 was superior to Method #3. In fact, Method #3 discovered a superior solution in six of the 10 trials. No solutions discovered by Method #1 were superior. The values in parentheses indicate the standard deviations

	METHOD #1	METHOD #2	METHOD #3
	P(Crossover) = 0.8 P(Mutation) = 0.01	P(Crossover) = 0.8 P(Mutation) = 1.0	P(Crossover) = 0.0 P(Mutation) = 1.0
AFTER 10,000 OFFSPRING			
Mean Population Score	54.19 ($\pm$24.404)	21.808 ($\pm$3.468)	28.697 ($\pm$12.118)
Mean Population Variance	4.582 ($\pm$6.125)	141.725 ($\pm$40.372)	170.694 ($\pm$203.548)
Mean Best Score in Population	52.365 ($\pm$22.690)	11.451 ($\pm$2.847)	15.176 ($\pm$7.084)

standard deviation of 30. After 10,000 offspring, the size of the population was limited to 75 vectors. All other conditions were retained.

In the first set of ten trials, each parent vector was mutated by crossover (80% chance/parent) and by a standard normal random variable to any component (1% chance/parent). The second set of ten trials retained all of the previous conditions, except the chance of a simple random mutation was increased to 100%. The third set of ten trials did not utilize crossover; each parent was mutated only by simple random mutation. Ten thousand offspring were generated in each trial. As seen in Table 2, no advantage was realized by optimization using crossover.

5 Conclusions

The experimental evidence suggests the claim that sophisticated genetic operators are required to ensure successful adaptation is in error. In any positively entropic system, mutability is guaranteed. There will be errors in informational transcription. Competition is likewise guaranteed in any fi-

nite arena. Selection is the consequence of competition. Evolution toward optimality becomes inevitable, even when using only very small mutations. But more to the point, a sequence of variants produced by mutation and competition is not equivalent to an enumeration of structures. Only a small portion of the state space is searched under the search techniques described here.

Successful adaptive procedures must retain a sufficient link between parent and offspring to ensure that advances are maintained. Evolution succeeds not by random sampling, which is doomed to failure, but by incorporating through successive stages those behaviors which are appropriate. The crossover and inversion operators do not always maintain this behavioral link, no more than do other macromutations. A large shuffling of information takes place each generation. In pleiotropic systems with information structures of large size, the use of such mutation operations forces the evolution to approach a random search, where the behavior of each offspring becomes only minimally related to the behavior of the parent. The experimental results generated here are similar to those obtained by Reed et al. (1967) where "under conditions simulating polygenic control of quantitative characters, crossing [did] not enhance the speed of se-

lective adaptation."

Further, the crossover and inversion operations common to "genetic algorithms" provided no significant benefit, irrespective of the degree of interactivity of the "gene products" involved in the linear systems of equations. Rather, random mutation consistently generated more efficient searches. If a specific algorithm were to show promise over a general algorithm, it would be expected to do so under special circumstances. Intuitively, here that special circumstance would occur when the degree of interactivity is zero. While specific circumstances (other than linear equations) may well exist for which crossover and inversion operations are especially appropriate, those conditions cannot be the hallmark of a broadly useful algorithm.

"Genetic algorithms," which rely on crossover and inversion, have been almost universally described by their students (e.g., Holland 1975; Bethke 1981) as being fundamentally different than "evolutionary programming techniques, which rely on random mutation and hill-climbing" (Grefenstette 1986). That claim is in error. As long as selection occurs between individual values mapped into an error scoring algorithm, optimization will occur automatically as a hill climbing technique. Indeed, all optimization techniques either climb hills (maximizing appropriateness) or descend to the bottom of troughs (minimizing functional error), dependent upon the observer's choice of perspective. "Genetic algorithms" are based on specific mechanistic natural genetical systems. Rather than being fundamentally different from random mutation, as claimed, crossover and inversion are merely a subset of all random mutations. As in all subsets, their applicability will be strongly problem dependent, if advantageous at all.

References

Atmar JW (1976) Speculation on the evolution of intelligence and its possible realization in machine form. Ph.D. Dissertation, New Mexico State University

Bethke AD (1981) Genetic algorithms as function optimizers. Ph.D. Dissertation, University of Michigan

Davis L (1985) Applying adaptive algorithms to epistatic domains. Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Los Angeles, Calif., August, pp 162-164

DeJong KA (1975) Analysis of the behavior of a class of genetic adaptive systems. Ph.D. dissertation, University of Michigan

Fogel DB (1988) An evolutionary approach to the traveling salesman problem. Biol Cybern 60:139-144

Fogel LJ (1962) Autonomous automata, Industr Res 4:14 - 19

Fogel LJ (1964) On the organization of intellect, Ph.D. Dissertation, UCLA

Fogel LJ, Owens AJ, Walsh MJ (1966) Artificial intelligence through simulated evolution. Wiley, New York

Fogel LJ, Fogel DB (1986) Artificial intelligence through evolutionary programming. Final report under contract # PO-X56-9-1102C, Army Research Institute, October

Fourman MP (1985) Compaction of symbolic layout using genetic algorithms. In: Grefenstette JJ (ed) Proceedings of an International Conference on Genetic Algorithms and Their Applications. Carnegie-Mellon University, pp 141 - 153

Grefenstette JJ, Gopal R, Rosmaita B, Van Gucht D (1985) Genetic algorithms for the traveling salesman problem. In: Grefenstette JJ (ed) Proceedings of an International Conference on Genetic Algorithms and Their Applications. Carnegie-Mellon University, pp 160 - 168

Grefenstette JJ (1986) Optimization of control parameters for genetic algorithms. IEEE Trans Syst Man Cybern 16:122 - 128

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor

Reed J, Toombs R, Barricelli NA (1967) Simulation of biological evolution and machine learning. J Theor Biol 17:319 - 342

David B. Fogel
ORINCON Corporation,
9363 Towne Centre Drive
San Diego, CA 92122
USA

Errata added here:

(1) The third equation on page 112 has been corrected to properly place the absolute valuesymbols. The error in the original was pointed out by J.D. Schaffer (pers. comm. 1991) with thanks.

(2) Note that the notation for the matrix and its entries a_{ij} and the vector and its entries b_i is somewhat equivocal. For consistency to the original publication, no change has been made here. The use of the term should be clear in context.

Please contact:

David B. Fogel
Natural Selection, Inc.
3333 N. Torrey Pines Ct., Suite 200
La Jolla, CA 92037
dfogel@natural-selection.com
www.natural-selection.com

for more information.