Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

A GENETIC ALGORITHM AND AN INDIFFERENCE-ZONE RANKING AND SELECTION FRAMEWORK FOR SIMULATION OPTIMIZATION

Henrik E. Hedlund
Mansooreh Mollaghasemi

Industrial Engineering and Management Systems
University of Central Florida
Orlando, FL 32816 U.S.A

ABSTRACT

A methodology for optimization of simulation models is presented. The methodology is based on a genetic algorithm in conjunction with an indifference-zone ranking and selection procedure under common random numbers. An application of this optimization algorithm to a stochastic mathematical model is provided in this paper.

1 INTRODUCTION

Many optimization problems in industrial engineering involve searching for the best choice among a large set of alternative solutions. Frequently these problems are trying to find answers to questions such as what is the shortest path to visit a set of cities? what is the best routing system for bus operation in a city to minimize customer-waiting time? what is the best facility layout to minimize material flow? what is the best levels for reorder point and reorder quantities to minimize inventory cost? These types of problems are often referred to as combinatorial optimization problems.

A combinatorial optimization problem can be defined by a search space S and a set of constraints $C = \{c_1, c_2, c_3 \ldots\}$, where the search space consists of a finite, or possibly countable infinite, set of candidate solutions (s), and an objective function f: $S \rightarrow \Re$ (Papadimitrion and Steiglitz, 1992). The objective of a combinatorial optimization problem $P_c(S,C,f)$ is to find an optimal solution among the feasible solutions $s' \in S$ that satisfy the constraints C in S.

Even though these types of problems are theoretically easy to solve, they can be computationally intractable. Consider a combinatorial optimization problem that has only four decision variables and each variable can only equal integer values from 1 to 10, then the solution space consists of 10^4 or 10,000 alternatives. Performing an exhaustive search to find the optimum can in general not be done within a timely fashion. Therefore, more intelligent optimization algorithms are needed to solve these types of problems.

It is critical that these intelligent optimization algorithms are able to balance the tradeoff between exploration and exploitation. The algorithms need to explore the solution space for new, improved solutions while exploiting already visited solutions. Classical algorithms, such as hill climbing, emphasize exploration of the search space and neglect exploitation, whereas random search algorithms emphasize exploitation and neglect exploration. However, optimization algorithms based on learning, such as genetic algorithms (GA), have proven to balance the tradeoff between the two.

Genetic algorithm (GA) is a population based search algorithm inspired by Darwinian evolutionary theory: survival of the fittest (Holland, 1975). The basic idea behind GAs is that a population of individuals with certain behaviors is exposed to a new environment, where each individual represents an alternative solution. Some of the individual's behaviors are better suited to meet the demand of the new environment. Through selecting and mating the individuals appropriately, desired behaviors are passed to their offsprings. These offsprings form the new generation of the population.

Unfortunately, this reproduction is never faultless, nor can individual genotypes remain free of random mutation. Introduction of random genetic variation in turn leads to formation of unique behavior characteristics. Frequently this random event creates individuals with less desirable characteristics, however, in some cases more desirable individuals are constructed.

The major drawback with GA is that it was originally developed to be used on deterministic problems. Therefore, the use on stochastic problems can be difficult. This is due to the stochastic nature of their responses. Neglecting this fact can cause the GA to be mislead during the search, hence, causing it not to perform better than a random search. This research proposes a new framework that combines a GA with an indifference-zone ranking and selection procedure. This hybrid approach would make sure that a correct decision is made about which direction the search should proceed in.

2 BACKGROUND

GA has successfully been used for optimization of simulation models, see Faccenda and Tenga (1992), Tompkins and Azadivar (1995), Tautou and Pierreval (1995), Azadivar and Wang (2000) among others. Its success is mainly due to the fact that it is generally more applicable than classical optimization algorithms since only the evaluation of the individual's behavior is needed during its search for finding new alternative solutions. Furthermore, it does not require information about the gradient's convexity or continuity of the response. It can also easily be hybridized to improve its performance and it is relatively robust under low levels of variance. Finally, GA can handle both quantitative (i.e. number of machines to employ) and qualitative (i.e. queue disciplines) decision variables. This feature may be the most attractive since optimization of simulation models often involves mixes of the two types of variables.

When the variance of the responses are high the performance of a GA can be inadequate. In the worst case it might not perform better than a random search. This occurs when bad decisions are made on which individuals to use for constructing the new population. In order to improve the performance of a GA in high variance environment statistical comparison procedures need to be used.

The two most commonly used statistical comparison procedures in simulation modeling are multiple-comparison and indifference-zone ranking and selection procedures. The primary difference between them is that the later results in a decision rather than an estimate of the difference. When optimizing a system the analyst is trying to make a decision about which system is the best. Therefore, indifference-zone procedures are preferred when used in conjunction with GA for optimization of stochastic processes.

Boesel (1999) proposed to combine both, a multiple-comparison procedure and a ranking and selection procedure with a GA. He used a multiple-comparison procedure to group together individuals during the search. The individuals in a group get the same selective probability. After termination of the search, Boesel proposed to use Nelson *et al*. (1998) screening and selection procedure to select the best individual from the visited solutions during the search.

Allen *et al*. (1999) proposed an optimization procedure based on GA in conjunction with both a ranking and selection and an indifference-zone ranking and selection procedure. Both of these statistical comparison procedures require known and equal variance, which can be an unrealistic assumption in simulation modeling.

3 METHODOLOGY

This section presents a framework that uses a GA in conjunction with an indifference-zone ranking and selection procedure for optimization of stochastic processes. The framework consists of three major components. The components are the initialization phase, the search phase, and the solution phase. See Figure 1 for a pseudo code of the proposed optimization algorithm.

```
t=0                              Initialization
initialize P^t                            phase
evaluate P^t
loop                               Search phase
P^{t+1} = select P^t
  recombine P^{t+1}
  mutate P^{t+1}
  evaluate P^{t+1}
  t=t+1
while not terminate P^{t+1}
present solution                 Solution phase
```

Figure 1: Pseudo Code of a Simple Genetic Algorithm, where P^t Denotes the Population at Generation t

A brief description of the three phases is presented next.

3.1 Initialization Phase

The first step of the proposed methodology is the initialization phase in which the parameters used in the algorithm are specified. Theses include, the indifference-zone (d^*), the initial number of replications (n_0), and the probability of overall correct selection (P). In addition, the parameters of the genetic algorithm including the probability of premature convergence (ψ), the probability of crossover (p_c), the probability of mutation (p_m), the probability of selecting the winner in a tournament (p_{tour}), and the subset size of immigration mutation (m_i) are specified.

After initializing the parameters an estimate of the necessary population size can be obtained. The population size depends on the indifference-zone (d^*) and the variance of the alternative solutions as well as the variance of the population. The variance of the alternatives and the population is estimated by performing an initial study. This study is also used for checking the efficiency of the CRN.

After estimating the necessary population size, the initial population is generated. The initial population consists of a mixture of randomly generated alternatives, user defined solution alternatives, and extreme points. This allows for good population diversity, hence, minimizing the risk of premature convergence of the search.

The next step in initialization phase is to estimate the alternatives fitness. This is done by using an indifference-zone ranking and selection procedure. By using an indifference-zone procedure an accurate estimate of the alternatives fitness is obtained. Furthermore, using an indifference-zone procedure gives a statistical guarantee that the fitness of the best alternative in the population is monotone increasing.

Usually the indifference-zone procedures are two stage procedures. That is, by performing an initial stage of n_0 replications an estimate of the required number of replications (N) can be obtained. By performing these N replications the alternatives can be ranked and a selection can be made under some probability of correct selection (P^*).

Both Clark and Yang (1986) and Nelson and Matejcik (1995) proposed indifference-zone ranking and selection procedures for selecting the best of the k alternatives. In this research, a new indifference-zone procedure was developed that selects the m best of k alternatives. This selection is used during the search for improving the performance of the GA. Furthermore, it is used in the solution phase to select the m best alternatives found during the search.

3.2 Search Phase

The second phase of the proposed optimization algorithm is the search phase. By using the stochastic processes of selection, recombination, and mutation the population is able to find new and improved solutions.

Selection is the process of selecting which alternatives are going to be used for generating the new population. This process is based on the selective probabilities, assigned to each alternative based on its rank within the population, often referred to as ordinal based selection. This type of selection is often preferred over proportional selection that is based on alternatives relative fitness to the other (Miller, 1997).

The proposed indifference-zone procedure is also used in the selection process. This ensure that the m best alternatives are retained from one generation to the next, thus, minimizing the risk of losing good alternatives if they are not selected for reproduction or destroyed by recombination or mutation. Furthermore, this allows for a statistical guarantee that the best alternative in succeeding generations is monotone increasing.

The recombination is the process of mixing two alternatives to create two offsprings. This is the search primary mechanism to find new and improved solutions. Mutation is a background operation that ensures the search recovers from lost behavior information of the alternatives.

The search phase continuous until the stopping criterion is meet. In order to ensure that a broad search has been performed termination occurs only after convergence of the population.

3.3 Solution Phase

The final phase of the proposed optimization procedure is solution presentation. This phase presents the alternatives that can be considered for implementation. The solutions are selected using the same indifference-zone ranking and selection procedure used for evaluating the alternatives. That is the m best of the k alternatives is selected with some probability of correct selection. This selection allows for a guarantee that the presented solutions are the best with some probability of overall correct selection.

4 EXPERIMENT

The initial case study was performed on a stochastic model initial developed by Keys *et al.* (1995). The function is defined as

$$f_1(x_1,x_2) = -0.704376\left(1-x_1^2\right)\left(1-x_2^2\right)\left(4-x_1\right)\left(0.05^{x_1^2}-0.05^{x_2^2}\right)^2+\varepsilon_1$$

where x_1 and x_2 are limited in the interval [-1.000,1.000] and ε_1 is the random noise This problem was selected because it has a known optimum at $f_1(0.699,0.000)=-0.9999$. Furthermore, it has previously been used for comparison of several simulation optimization methodologies.

The model is constructed so that due to the random noise at least one of the local optimums can return a better fitness than the global optimum. Hence, causing an optimization algorithm to converge to a local optima.

The indifference-zone has the following setup during the experiment: $n_0 = 8$, $m = 78$, $k = 8$, $d^* = 0.05$, the selection is done at an overall confidence level of 0.90. The setup of the GA follows general setting presented in literature. The initial population of alternatives consists of the extreme points and randomly generated solutions.

A total of 42 iterations of the proposed optimization algorithm were performed before the termination criterion was met. During the search, the optimization algorithm visited a total of 1,180 alternatives, which is only 0.03% of the entire solution space (i.e., 1,180/4,004,001). The three best alternatives found during the search were (x_1, x_2) = (0.703, -0.003), (0.701, -0.005), and (0.699, -0.010) with an estimated average response of -1.02316, -1.02315, and –1.02253, respectively. These are selected with an overall probability of correct selection of 0.90. See the evolution of the population in Figure 2.

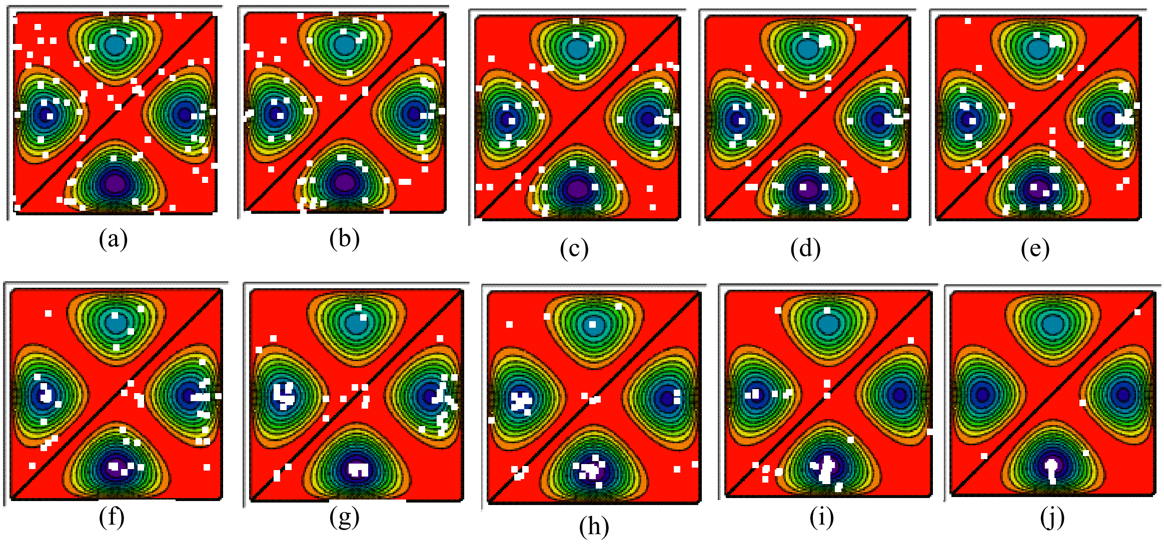

Figure 2: Evaluation of Population During Iteration 1 Through 5 ((a) Through (e)), Iteration 10 (f), Iteration 15 (g), Iteration 20 (h), Iteration 30 (i), and Iteration 42 (j) (White Squares Symbolize the Location of a Alternative Solution on the Contour Plot)

4.1 Comparison with other procedures

Hall and Bowden (1997) used this stochastic function for a comparison of the performance of Tabu Search (TS), Evolution Strategies (ES), and Simplex Algorithm (SA). In their study, each alternative was allowed to perform 20 iterations. Assuming that the proposed algorithm terminates after 20 iterations the following was found. During this partial search 606 alternatives were visited, that is 606/4,004,001 or 0.015% of the solution space. The three best solutions found during this partial search are (x_1, x_2) = (0.690, -0.010), (0.680, -0.010), and (0.722, -0.010), with an estimated response of -1.02165, -1.01866, and -1.01679, respectively.

As a performance measure Hall and Bowden used the percentage of successful iterations, where an iteration is considered to be successful if the Euclidean distance between the best alternative in the population and the true optimum is less than 1% of the range specified by x_1 and x_2. This performance measure gives a good indication of the optimization algorithms ability to converge to the global optimum under search restrictions. Table 1 presents the percentage of successful replications of the first 20 iterations of the proposed algorithm as compared with tabu search, evolutionary strategies, and Nelder-Mead simplex algorithm. The results indicate that the proposed algorithm performs better than tabu search and Nelder-Mead simplex algorithm, and at least as good as the evolutionary strategies.

Table 1: Percentage of Successful Iterations (Values for Tabu Search, Evolution Strategies, and Simplex Algorithm Taken from Hall and Bowden 1997)

	Proposed	Tabu Search	Evolution Strategies	Simplex Algorithm
f_1	50%	5%	45%	0%

In Hall and Bowden comparison study, the alternatives were replicated 2000 times at each of the 20 iterations. The proposed algorithm performed only eight replications on each of the alternatives. During the time it took for their study to evaluate one alternative, the proposed algorithm can search 2,000/8 = 250 alternatives. That is, the proposed algorithm can perform (250/78 ≈ 3.2) 3 complete iterations. Hence a significant reduction in the simulation effort is achieved by using proposed procedure.

5 FUTURE RESEARCH

Future areas of research include applying this procedure to simulation models. Compare the optimization procedures as well as the indifference-zone ranking and selection procedures performance with other techniques.

Another area of future research includes the development of a similar framework to be used for optimization of multi-response stochastic systems.

REFERENCES

Allen, T., M. Bernshteyn, and L. Yu. (1999). "A Heuristic Algorithm for Discrete Stochastic Optimization." *Working paper from Ohio State University*

Azadivar, F. and J. Wang. (2000). "Facility Layout Optimization Using Simulation and Genetic Algorithms." *International Journal of Production Research*. Vol. 38, No. 17. pp. 4369-4383.

Boesel, J. (1999). "Search and Selection for Large-Scale Stochastic Optimization." *Unpublished Ph.D. Dissertation form Northwestern University*.

Clark, G.M. and W.M. Yang. (1986). "A Bonferroni Selection Procedure when Using Common Random Numbers with Unknown Variance." In J. Wilson, J. Henriksen, and S. Roberts (Eds.) *Proceedings of the 1986 Winter Simulation Conference*. pp. 313-315.

Faccenda, J.F. and R.F. Tenga. (1992). "A Combined Simulation/Optimization Approach to Process Plant Design." In J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson (Eds.) *Proceedings of the 1992 Winter Simulation Conference*. pp. 1256-1261.

Hall, J.D. and R.O. Bowden. (1997). "Simulation Optimization by Direct Search: A Comparative Study." *Proceedings of the 6th Industrial Engineering Research Conference*. pp. 298-303.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan.

Keys, A., L. Rees, and A. Greenwood. (1995). "A Performance-Based Study of Nonparametric Metamodeling Techniques for the Classification of Response Surfaces in Simulation Optimization." *Working Paper from Virginia Tech*.

Miller, B.L. (1997). "Noise, Sampling, and Efficient Genetic Algorithms." *IlliGAL Report No. 97001*. University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, IL.

Nelson, B.L. and F.J. Matejcik. (1995). "Using Common Random Numbers for Indifference-zone Selection and Multiple Comparisons in Simulation." *Management Science*. Vol. 41, No. 12. pp. 1935-1945.

Nelson, B.L., J. Swann, D. Goldsman, and W. Song. (1998). "Simple Procedures for Selecting the Best Simulated System when the Number of Alternatives is Large." *Technical Report*, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

Papdimitriou, C.H. and K. Steiglitz. (1992) *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Englewood Cliffs, NJ.

Tautou, L. and H. Pierreval. (1995). "An Evolutionary Algorithm for the Optimization of a Simulated Manufacturing System Producing Plastic Yogurt Pots." In T.I. Ören and L.G. Birta (Eds.) *Proceedings of the SCSC*. pp. 1096-1100.

Tompkins, G., and F. Azadivar. (1995). "Genetic Algorithms in Optimizing Simulated Systems." In C. Alexopoulos, K. Kang, W.E. Lilegdon, and D. Goldsman (Eds.), *Proceedings of the 1995 Winter Simulation Conference*. pp. 757-762.

AUTHOR BIOGRAPHIES

HENRIK E. HEDLUND is a Ph.D. candidate in the Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, Florida. He received his M.S. in Industrial Engineering from the University of Central Florida. His research interests are statistical aspects of simulation modeling, simulation optimization, and modeling and analysis of complex systems. He is a member of INFORMS. His email address is <heh76410@pegasus.cc.ucf.edu>.

MANSOOREH MOLLAGHASEMI is an associate professor of Industrial Engineering and Management Systems at the University of Central Florida. She received her Ph.D. from the University of Louisville in 1991. She holds a B.S. and a M.S. in Chemical Engineering from the same University. Her research interests involve simulation modeling and analysis, multiple response simulation optimization, and neural networks. She is a member of IIE and INFORMS. Her email and web addresses are <mollagha@mail.ucf.edu> and <ie.engr.ucf.edu/ver40/faculty/mollaghasemi.htm>.