GA-Selection Revisited from an ES-Driven Point of View

Michael Affenzeller, Stefan Wagner, Stephan Winkler

Institute for Formal Models and Verification
Johannes Kepler University
Altenbergerstrasse 69
A-4040 Linz - Austria
{michael,stefan,stephan}@heuristiclab.com

Abstract. Whereas the selection concept of Genetic Algorithms (GAs) and Genetic Programming (GP) is basically realized by the selection of above-average parents for reproduction, Evolution Strategies (ES) use the fitness of newly evolved offspring as the basis for selection (survival of the fittest due to birth surplus). This contribution proposes a generic and enhanced selection model for GAs considering selection aspects of population genetics and ES. Some selected aspects of these enhanced techniques are discussed exemplarily on the basis of standardized benchmark problems.

1 Introduction

In contrast to other heuristic optimization techniques Genetic Algorithms and certainly also Genetic Programming (GP) take a fundamentally different approach by considering recombination (crossover) as their main operator. The essential difference to neighborhood-based techniques is given by the fact that recombination is a sexual operator, i.e. properties of individuals from different regions of the search space are combined in new individuals. Therefore, the advantage of applying GAs to hard optimization problems lies in their ability to scan broader regions of the solution space than heuristic methods based upon neighborhood search do. Nevertheless, also GAs are frequently faced with a problem which, at least in its impact, is quite similar to the problem of stagnating in a local but not global optimum. This drawback, called premature convergence in the terminology of GAs, occurs if the population of a GA reaches such a suboptimal state that the genetic operators are no longer able to produce offspring that are able to outperform their parents (e.g. [5], [1]). This happens if the genetic information stored in the individuals of a population does not contain that genetic information which would be necessary to further improve the solution quality. Therefore, in contrast to the present contribution, the topic of premature convergence is considered to be closely related to the loss of genetic variation in the entire population in GA-research [11], [15]. In this contribution we do not identify the reasons for premature convergence in the loss of genetic variation in general but more specifically in the loss of what we call essential genetic information, i.e. in the loss of alleles which are part of a global optimal solution. Therefore, we will denote the genetic information of the global optimal (which may be unknown a priori) solution as essential genetic information in the following. If parts of this essential genetic information are missing, premature convergence is already predetermined in some way.

A very essential question about the general performance of a GA is, whether or not good parents are able to produce children of comparable or even better fitness (the building block hypothesis implicitly relies on this). In natural evolution, this is almost always true. For GAs this property is not so easy to guarantee. The disillusioning fact is that the user has to take care of an appropriate coding in order to make this fundamental property hold.

In order to somehow overcome this strong requirement we try to get to the bottom of reasons for premature convergence from a technical as well as from a population genetics inspired point of view and draw some essential interconnections.

The basic idea of the new selection model is to consider not only the fitness of the parents in order to produce a child for the ongoing evolutionary process. Additionally, the fitness value of the produced child is compared with the fitness values of its own parents. The child is accepted as a candidate for the further evolutionary process if and only if the reproduction operator was able to produce a child that could outperform the fitness of its own parents. This strategy guarantees that evolution is presumed mainly with crossover results that were able to mix the properties of their parents in an advantageous way. I.e. **survival of the fittest alleles is rather supported than survival of the fittest individuals** which is a very essential aspect for the preservation of essential genetic information stored in many individuals (which may not be the fittest in the sense of individual fitness).

2 Some Basic Considerations about GA-Selection

In terms of goal orientedness, selection is the driving force of GAs. In contrast to crossover and mutation, selection is completely generic, i.e. independent of the actually employed problem and its representation. A fitness function assigns a score to each individual in a population that indicates the 'quality' of the solution the individual represents. The fitness function is often given as part of the problem description or based upon the objective function. In the Standard GA the probability that a chromosome in the current population is selected for reproduction is proportional to its fitness. However, there are also many other ways of accomplishing selection. These include linear-rank selection or tournament selection (cf. e.g. [7], [10]).

However, all evenly mentioned GA-selection principles have one thing in common:
They all just consider the aspect of sexual selection, i.e. mechanisms of selection only come into play for the selection of parents for reproduction. The enhanced

selection model which will be described in the following section defies this limitation by considering selection in a more general sense.

Selection and Selection Pressure: In the terminology of population genetics the classical GA selection concept is known as sexual selection. In the population genetics view, sexual selection covers only a rather small aspect of selection which appears when individuals have to compete to attract mates for reproduction. The population genetics basic selection model considers the selection process in the following way:

random mating $\rightarrow$ selection $\rightarrow$ random mating $\rightarrow$ selection $\rightarrow$ ......

I.e. selection is considered to depend mainly on the probability of surviving of newborn individuals until they reach pubescence which is called viability in the terminology of population genetics. The essential aspect of offspring selection in the interpretation of selection is rarely considered in conventional GA selection. The classical (μ, λ) Evolution Strategy in contrast does this very well: Reconsidering the basic functioning of a (μ, λ) ES in terms of selection, μ parents produce λ $(\lambda \geq \mu)$ offspring from which the best μ are selected as members of the next generation. In contrast to GAs where selection pressure is predetermined by the choice of the mating scheme and the replacement strategy, ES allow an easy steering of selection pressure by the ratio between μ and λ. The selection pressure steering model introduced in Section 3 picks up this basic idea of ES and transforms these concepts for GAs in order to be have an adjustable selection pressure (independent of the mating scheme and replacement strategy) at one's disposal.

Our advanced selection scheme allowing self-adaptive steering of selection pressure aims to transform the basic ideas for improving the performance of GAs. In doing so the survival probability is determined by a comparison of the fitness of the newly generated individual with the fitness values of its parents.

A very important consequence of selection in population genetics as well as in evolutionary computation is its influence on certain alleles. As a matter of principle there are four possibilities for each allele in the population:

- The allele is fixed in the population.
- The allele is lost in the population.
- The allele frequency converges to an equilibrium state.
- The allele frequency remains unchanged.

The basic approaches for retarding premature convergence discussed in GA literature aim at the maintenance of genetic diversity. The most common techniques for this purpose are based upon preselection [3], crowding [4], or fitness-sharing [6]. The main idea of these techniques is to maintain genetic diversity by the preferred replacement of similar individuals [3], [4] or by the fitness-sharing of individuals which are located in densely populated regions [6]. While methods based upon [4] or [6] require some kind of neighborhood measure depending on the problem representation, [6] is additionally quite restricted to proportional selection. Moreover, these techniques have the common goal to maintain genetic diversity which is very important in natural evolution where a rich gene pool is the guarantor in terms of adaptiveness w.r.t. changing environmental conditions.

In case of artificial genetic search as performed by a GA the optimization goal does not change during the run of a GA and the fixing of alleles of high quality solutions is desirable in the same manner as the erasement of alleles which are definitely not part of a good solution in order to reduce the search space and make genetic search more goal-oriented. I.e. we claim that pure diversity maintenance mechanisms as suggested in [3], [4], or [6] do not support goal-oriented genetic search w.r.t the locating of global optimal solutions.

3 An Enhanced Selection Model for Self-Adaptive Steering of Selection Pressure

The basic idea to create and evaluate a certain amount (greater or equal population size) of offspring, to be considered for future members of the next generation, is adapted from Evolution Strategies. Self-adaption comes into play when considering the question which amount of offspring is necessary to be created at each round, and which of these candidates are to be selected as members of the next generation, i.e. for the ongoing evolutionary process. In order to keep the concepts generic, no problem specific information about the solution space is allowed to be used for stating the self-adaptive model. Thus, it is desirable to systematically utilize just the fitness information of the individuals of the actual generation for building up the next generation of individuals, in order to keep the new concepts and methods generic. In principle, the new selection strategy acts in the following way:

The first selection step chooses the parents for crossover either randomly or in the well-known way of GAs by roulette-wheel, linear-rank, or some kind of tournament selection strategy. After having performed crossover and mutation with the selected parents we introduce a further selection mechanism that considers the success of the apparently applied reproduction in order to assure the proceeding of genetic search mainly with successful offspring in that way that the used crossover and mutation operators were able to create a child that surpasses its parents' fitness. Therefore, a new parameter, called success ratio ($SuccRatio \in [0, 1]$), is introduced. The success ratio gives the quotient of the next population members that have to be generated by successful mating in relation to the total population size. Our adaptation of Rechenberg's success rule [8] for GAs says that a child is successful if its fitness is better than the fitness of its parents, whereby the meaning of 'better' has to be explained in more detail: is a child better than its parents, if it surpasses the fitness of the weaker, the better, or is it in fact some kind of mean value of both?

For this problem we have decided to introduce a cooling strategy similar to Simulated Annealing. Following the basic principle of Simulated Annealing we claim that an offspring only has to surpass the fitness value of the worse parent in order to be considered as 'successful' at the beginning and while evolution proceeds the child has to be better than a fitness value continuously increasing between the fitness of the weaker and the better parent. As in the case of Simulated Annealing, this strategy effects a broader search at the beginning whereas at the end of the search process this operator acts in a more and more directed way. Having filled up the claimed ratio ($SuccRatio$) of the next generation with successful individuals in the above meaning, the rest of the next generation ($(1 - SuccRatio) \cdot |POP|$) is simply filled up with individuals randomly chosen from the pool of individuals that were also created by crossover but did not reach the success criterion. The actual selection pressure $ActSelPress$ at the end of a single generation is defined by the quotient of individuals that had to be considered until the success ratio was reached and the number of individuals in the population in the following way: $ActSelPress = \frac{|POP| SuccRatio + |POOL|}{|POP|}$.

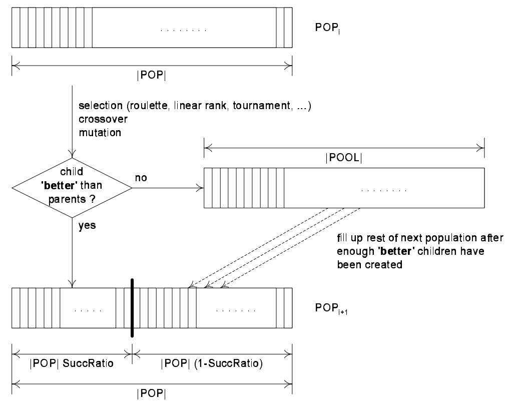

Fig. 1. Flowchart for embedding the new selection principle into a Genetic Algorithm.

Figure 1 shows the operating sequence of the above described concepts. With an upper limit of selection pressure ($MaxSelPress$) defining the maximum number of children considered for the next generation (as a multiple of the actual population size) that may be produced in order to fulfill the success ratio, this new model also functions as a precise detector of premature convergence:

If it is no longer possible to find a sufficient number of (SuccRatio $\cdot |POP|$*) offspring outperforming their own parents even if (*$MaxSelPress \cdot |POP|$*) candidates have been generated, premature convergence has occurred.*

As a basic principle of this selection model a higher success ratio causes higher selection pressure. Nevertheless, higher settings of success ratio and therefore of selection pressure do not necessarily cause premature convergence as the preservation of fitter alleles is additionally supported and not only the preservation of fitter individuals.

Also it is possible within this model to state selection pressure in a very intuitive way that is quite similar to the notation of selection pressure in Evolution Strategies. Concretely, we define the actual selection pressure as the ratio of individuals that had to be generated in order to fulfill the success ratio to the population size. For example, if we work with a population size of say 100 and it would be necessary to generate 1000 individuals in order to fulfill the success ratio, the actual selection pressure would have a value of 10. Via these means we are in a position to attack several reasons for premature convergence as illustrated in the following sections. Furthermore, this strategy has proven to act as a precise mechanism for self-adaptive selection pressure steering, which is of major importance in the migration phases of parallel evolutionary algorithms. The aspects of offspring selection w.r.t. parallel GAs are combined in the parallel SASEGASA-algorithm [1].

4 Empirical Discussion

The empirical section is subdivided into two parts:
The first subsection aims to highlight the main message of the paper (preservation of essential alleles). As the scope of the present work does not allow a deeper and more sophisticated analysis of different problem situations, the second part of the experimental discussion gives some references to related contributions which include a more detailed and statistically more relevant experimental discussion on the basis of several benchmark but also practical problems on which we have applied the new selection model recently. All empirical work shown and referred in this section have been implemented and performed using the HeuristicLab environment [12].[1]

4.1 Conservation of Essential Genetic Information

This subsection aims to point out the importance of mutation for the recovery of essential genetic information in the case of conventional GAs in order to oppose these results with the results being achieved with the enhanced selection model discussed in this paper. By reasons of compactness, the results are mainly shown on the basis of diagrams and give only a brief description of introduced operators, parameter settings, and test environments. Furthermore, the chosen benchmark instance is of rather small dimension in order to allow the observation of essential alleles during the run of the algorithm.

The results displayed in Figure 2 (left diagram) show the effect of mutation for reintroducing already lost genetic information. The horizontal line of the diagram shows the number of iterations and the vertical line stands for the solution quality. The bottom line indicates the global optimal solution which is known for this benchmark test case. The three curves of the diagram show the performance of a Genetic algorithm with no mutation, with a typical value of 5% mutation as well as a rather high mutation rate of 10%. For each of the three curves the lower line stands for the best solution of the actual population and the

[1] For more detailed information concerning HeuristicLab the interested reader is referred to http://www.heuristiclab.com/

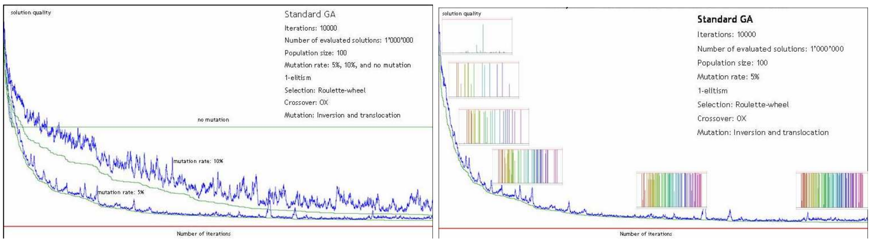

Fig. 2. The effect of mutation for certain mutation rates (left diagram) and the distribution of essential genetic information for a mutation rate of 5% (right diagram) both in case of a standard GA for the ch130 benchmark TSP.

upper line shows the average fitness value of the population members. The results with no mutation are extremely weak and the quality curve stagnates very soon and far away from the global optimum. The best and average solution quality are the same and no further evolutionary process is possible - premature convergence has occurred. As already stated before, mutation is a very essential feature of standard GAs in order to avoid premature convergence. But also a rather high mutation rate of 10% produces results which are not very satisfying and indeed the best results are achieved with a mutation rate which is very typical for GA applications - namely a mutation rate of 5%. Considering a standard benchmark problem like the ch130 (a 130 city TSP taken from the TSPLib [9]) with one single best solution allows to consider the edges of the shortest path as the essential alleles whose preservation during the run can be observed. The following figures indicate the spreading of essential alleles during the runs of the certain algorithms. This is visualized by inserting bar charts which have to be considered as snapshots after a certain number of iterations approximately corresponding to the position in the figure. The higher a certain bar (130 bars for a 130-city TSP) the higher the relative occurrence of the corresponding essential allele in the population.

The right diagram of Figure 2 shows the distribution of essential alleles over the iterations for a standard GA with a mutation rate of 5%. The interesting thing is that some minor ratio of essential alleles is rapidly fixed in the population and the majority of essential alleles which are still missing have disappeared in the entire population. During the further run of the algorithm it is only mutation which can reintroduce this essential genetic information. As it could be seen in Figure 2, without mutation premature convergence would already have occurred at this early state of evolutionary search. But with an appropriate mutation rate (5% in this example) more and more essential alleles are discovered ending up with quite a good solution. But there is still a gap to the global optimum caused by that alleles which could not be recovered due to mutation. The next figures will show how the new selection concept is able to close this gap and make the algorithm much more independent of mutation.

So let us take a closer look at the distribution of essential genetic information in the population when using the enhanced selection concepts. The left diagram of Figure 3 shows the quality curve and the distribution of essential alleles for a mutation rate of 5% (which was able to achieve the best results in case of a standard GA).

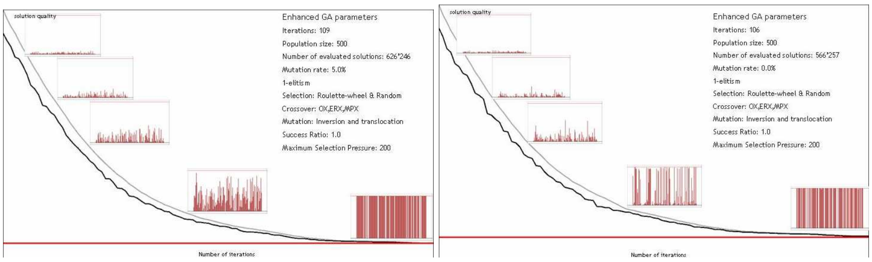

Fig. 3. The distribution of essential genetic information when using the enhanced selection concept considering the ch130 benchmark TSP with 5% mutation (left diagram) and with no mutation (right diagram).

When applying the GA with the new selection principle to the same benchmark test case one can see that the global optimal solution is detected in only about 100 iterations. Nevertheless, the computational effort is comparable to the standard GA as much more individuals have to be evaluated at each iteration step due to the higher selection pressure. Considering the distribution of essential alleles we see a totally different situation. Almost no essential alleles get lost and the ratio of essential alleles continuously increases in order to end up with a final population that contains almost all pieces of essential genetic information and therefore achieving a very good solution. This shows that the essential alleles are preserved much more effectively and indicates that the influence of mutation should be much less. But is this really the case? In order to answer this question, let us consider the same example with the same settings - only without mutation. And indeed the assumption holds and also without mutation the algorithm finds a solution which is very close to the global optimum (see right diagram of Figure 3). The essential alleles interfuse the population more and more and almost all of them are members of the final population. Reconsidering the standard GA without mutation the algorithm was prematurely converging very soon with a very bad total quality.

4.2 References to Recent Related Works

The basic concepts of the enhanced selection ideas as published in the present paper have already emerged more than one year ago. As the actual focus (like also stated in the present contribution) is to study the properties of the new selection concepts systematically, the potential w.r.t. achievable advancements in global solution quality were obvious immediately. Therefore, the main aim of the first works in this area was to check the generality of the new algorithmic concepts by applying them to various theoretically as well as practically relevant problems. And indeed this worked out very well and it was possible to demonstrate similar effects and achievements in global solution quality in various areas of application under very different problem codifications with exactly that enhanced generic selection techniques as being proposed in this paper.

While the last subsection considered only relatively small TSP instances in order to illustrate some selected aspects, journal article [1] includes a detailed and comprehensive empirical analysis also based on TSP instances of much higher dimension. Furthermore, [1] gives a comprehensive solution analysis based on several real valued n-dimensional test functions (like the n-dimensional Rosenbrock, Rastrigin, Griewangk, Ackley, or Schwefel's sine root function). Also here it is possible to locate the global optimal solution in dimensions up to $n = 2000$ with exactly the same generic extensions of the selection model as being stated here - only the crossover- and mutation-operators have been replaced with standard operators for real-valued encoding.

But also in practical applications like the Optimization of Production Planning in a Real-World Manufacturing Environment based on an extended formulation of the Job-shop Scheduling Problem [2] a significant increase in solution quality could be accomplished with the described methodology. Especially in combination with Genetic Programming self-adaptive selection pressure steering has already proven to be very powerful. In [13] and [14] we report first results achieved in the context of nonlinear structure identification based on time-series data of a diesel combustion engine. Concretely the aim of this project is the development of models for the NOx emission. Already until now it has become possible with a GP-based approach equipped with offspring selection to identify models which are superior to the models achieved with conventional GP-techniques and also superior to machine learning techniques which have also been considered in earlier stages of this project. Very recently we have adapted this GP-approach for the application on symbolic as well as logistic regression problems. First results achieved on benchmark classification problems (taken from the UCI machine learning repository) indicate a high potential also in these areas of application.[2]

5 Conclusion and Future Perspectives

This paper discusses a new generic selection concept and points out its ability to preserve essential genetic information more goal-oriented than standard concepts. Possibly the most important feature of the newly introduced concepts is that the achievable solution quality can be improved in a non-problem specific manner so that it can be applied to all areas of application for which the theory of GAs and GP provides suitable operators. Further aspects worth mentioning concern the robustness and self-adaptiveness of the population genetics and ES inspired measures: Selection pressure can be steered self-adaptively in a way

[2] First results tables for the thyroid and Wiskonsin breast cancer data-sets are shown on http://www.heuristiclab.com/results/regression.

that the amount of selection pressure actually applied is that high that further progress of evolutionary search can be achieved. Possible future research topics in that area are certainly to open new areas of application also and especially in GP related applications where the aspect of preservation of essential genetic information is especially important as the disruptive properties of GP-operators tend to add impurities into the genetic information for the ongoing search when using standard parent selection models.

References

1. M. Affenzeller and S. Wagner. SASEGASA: A new generic parallel evolutionary algorithm for achieving highest quality results. *Journal of Heuristics, Special Issue on New Advances on Parallel Meta-Heuristics for Complex Problems*, 10(3):239–263, 2004.
2. R. Braune, S. Wagner, and M. Affenzeller. Applying genetic algorithms to the optimization of production planning in a real-world manufacturing environment. *Proceedings of the European Meeting on Cybernetics and Systems Research - EMCSR 2004*, pages 41–46, 2004.
3. D. Cavicchio. *Adaptive Search using Simulated Evolution.* unpublished doctoral thesis, University of Michigan, Ann Arbor, 1970.
4. K. DeJong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, Department of Computer Science, University of Michigan, 1975.
5. D. Fogel. An introduction to simulated evolutionary optimization. *IEEE Trans. on Neural Network*, 5(1):3–14, 1994.
6. D. Goldberg. *Genetic Alogorithms in Search, Optimization and Machine Learning.* Addison Wesley Longman, 1989.
7. Z. Michaliwicz. *Genetic Algorithms + Data Structurs = Evolution Programs.* Springer-Verlag Berlin Heidelberg New York, 3. edition, 1996.
8. I. Rechenberg. *Evolutionsstrategie.* Friedrich Frommann Verlag, 1973.
9. G. Reinelt. TSPLIB - A traveling salesman problem library. *ORSA Journal on Computing 3*, 3:376–384, 1991.
10. E. Schoeneburg, F. Heinzmann, and S. Feddersen. *Genetische Algorithmen und Evolutionsstrategien.* Addison-Wesley, 1994.
11. R. Smith, S. Forrest, and A. Perelson. Population diversity in an immune system model: Implications for genetic search. *Foundations of Genetic Algorithms*, 2:153–166, 1993.
12. S. Wagner and M. Affenzeller. HeuristicLab: A generic and extensible optimization environment. *Accepted to be published in: Proceedings of the International Conference on Adaptive and Natural Computing Algorithms (ICANNGA) 2005*, 2005.
13. S. Winkler, M. Affenzeller, and S. Wagner. Identifying nonlinear model structures using genetic programming techniques. *Proceedings of the European Meeting on Cybernetics and Systems Research - EMCSR 2004*, pages 689–694, 2004.
14. S. Winkler, M. Affenzeller, and S. Wagner. New methods for the identification of nonlinear model structures based upon genetic programming techniques. *Proceedings of the 15th International Conference on Systems Science*, pages 386–393, 2004.
15. Y. Yoshida and N. Adachi. A diploid genetic algorithm for preserving population diversity - pseudo-meiosis GA. *Lecture Notes in Computer Science*, 866:36–45, 1994.