

Goal-Oriented Preservation of Essential Genetic Information by Offspring Selection

Michael Affenzeller
Inst. of Formal Models and
Verification
Johannes Kepler University
Altenbergerstrasse 69
A-4040 Linz, Austria
michael@heuristiclab.com

Stefan Wagner
Inst. of Formal Models and
Verification
Johannes Kepler University
Altenbergerstrasse 69
A-4040 Linz, Austria
stefan@heuristiclab.com

Stephan Winkler
Inst. for Design and Control of
Mechatronic Systems
Johannes Kepler University
Altenbergerstrasse 69
A-4040 Linz, Austria
stephan@heuristiclab.com

ABSTRACT

This contribution proposes an enhanced and generic selection model for Genetic Algorithms (GAs) and Genetic Programming (GP) which is able to preserve the alleles which are part of a high quality solution. Some selected aspects of these enhanced techniques are discussed exemplarily on the basis of standardized benchmark problems.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search—*heuristic methods*

General Terms

Algorithms theory

Keywords

Genetic Algorithms, selection, adaptation/self adaptation

1. INTRODUCTION

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. In natural evolution, this is almost always true. For Genetic Algorithms 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.

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 evenly 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

Copyright is held by the author/owner.
GECCO'05, June 25–29, 2005, Washington, DC, USA.
ACM 1-59593-010-8/05/0006.

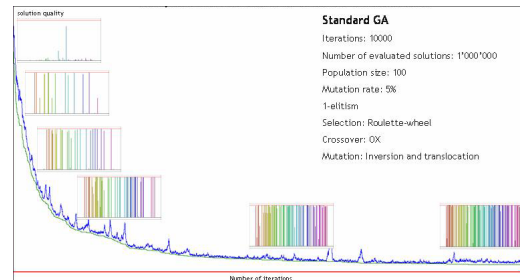


Figure 1: The distribution of essential genetic information in case of a standard GA for the ch130 benchmark TSP.

the preservation of essential genetic information stored in many individuals (which may not be the fittest in the sense of individual fitness).

The empirical Discussion is subdivided into two parts: The first subsection aims to highlight the main message of the paper (preservation of essential alleles) whereas the second part 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.

2. PRESERVATION OF ESSENTIAL GENETIC INFORMATION

Figure 1 shows the distribution of essential alleles (in the concrete case of a 130 city TSP an essential allele is one of the 130 edges of the optimal tour) 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. 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 those alleles which could

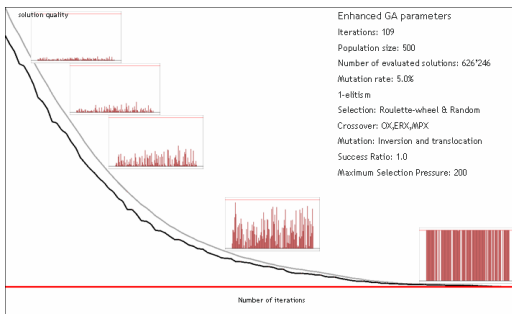


Figure 2: The distribution of essential genetic information when using the enhanced selection concept considering the ch130 benchmark TSP.

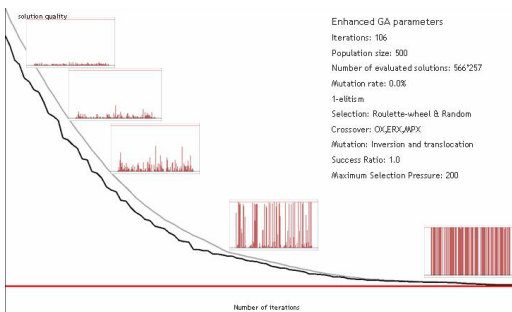


Figure 3: The distribution of essential genetic information when using the enhanced selection concept considering the ch130 benchmark TSP without mutation.

not be recovered due to 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 next curve (Figure 2) shows the quality curve and the distribution of essential alleles for 5% mutation which was able to achieve the best results in case of a standard GA.

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. 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.

3. REFERENCES TO RECENT RELATED WORKS AND OUTLOOK

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 [3] and [4] 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. ¹

4. 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] 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.
- [4] 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.

¹First results tables for the thyroid and Wisconsin breast cancer data-sets are shown on <http://www.heuristiclab.com/results/regression>.