The Influence of Binary Representations of Integers on the Performance of Selectorecombinative Genetic Algorithms

Franz Rothlauf
University of Bayreuth, D-95440 Bayreuth/Germany, rothlauf@uni-bayreuth.de

1 Introduction

The discussion regarding the benefits of different binary representations of integers has a long tradition in Genetic and Evolutionary Algorithms (GEAs). However, besides the commonly used binary and gray encoding there are many other possible representations. We illustrate how the performance of selectorecombinative GAs depends on the used representation. We obtain three results: The choice of a proper binary representation is crucial for GEA success. When using selectorecombinative GAs the binary encoding can outperform the gray encoding. Encodings exist that are better than the gray or binary encoding.

2 Genotype-Phenotype and Phenotype-Fitness Mappings

When using a representation the fitness function f can be decomposed into

$$f_g(\boldsymbol{x}_g) : \Phi_g \to \Phi_p,$$
$$f_p(\boldsymbol{x}_p) : \Phi_p \to \mathbb{R},$$

where $f = f_p \circ f_g = f_p(f_g(\boldsymbol{x}_g))$, Φ_g is the genotypic and Φ_p is the phenotypic search space. The genotype-phenotype mapping f_g is the used representation. With a bitstring of length l we can represent 2^l different phenotypes. Therefore, the number of different representations is $2^l!$. For $l = 3$, there are $2^3! = 40\ 320$ different representations. To reduce the number of different f_g we limit ourself to $l = 3$ and assume that $\boldsymbol{x}_g = 000 \in \Phi_g$ is always assigned to $x_p = 0 \in \Phi_p$. Then, for $l = 3$ the number of different representations is reduced to $(2^l - 1)! = 7! = 5040$.

f_p represents the fitness function and assigns a fitness value $f_p(\boldsymbol{x}_p)$ to every phenotype $\boldsymbol{x}_p \in \Phi_p$. For our investigation we use the easy problem $f_p(x_p) = x_p$.

3 Experimental Results

We concatenate 20 sub-problems of size $l = 3$. The fitness of an individual is calculated as the sum over the fitness of the 20 sub-problems. We use a selectorecombinative GA only using uniform crossover, tournament selection without replacement of size 2, and a population size $n = 20$. We performed 250 runs for each of the 5040 different genotype-phenotype mappings. A sub-problem is correctly solved if the GA is able to find the best solution $x_p = 7$. Figure 1 shows the distribution of the number of correctly solved sub-problems at the end of a GA run for all 5040 different types of genotype-phenotype mappings.

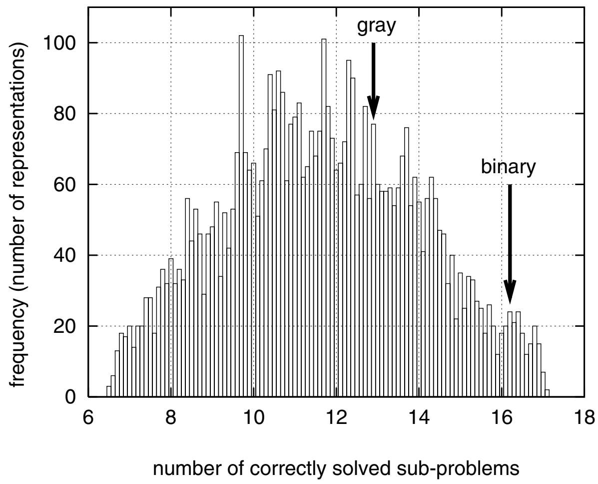

Figure 1: Experimental results

4 Conclusions

The results show that different genotype-phenotype mappings, that means assigning the genotypes $\boldsymbol{x}_g \in \{0,1\}^3$ in a different way to the phenotypes $x_p \in \{0,\dots 7\}$, strongly change the performance of GAs. Furthermore, for the considered easy problem the binary encoding (16.2 of the 20 sub-problems are correctly solved) performs better that the gray encoding (only 12.9 out of 20 sub-problems are correct). Finally, representations exist that outperform the binary encoding. If we can theoretically describe the properties of these encodings we can solve integer optimization problems more efficiently.