

On the Locality of Representations

Franz Rothlauf

Department of Information Systems 1, University of Mannheim
68131 Mannheim/Germany
franz@rothlauf.com

Abstract. It is well known that using high-locality representations is important for efficient evolutionary search. This paper discusses how the locality of a representation influences the difficulty of a problem when using mutation-based search approaches. The results show that high-locality representations do not change problem difficulty. In contrast, low-locality representations randomize the search process and make problems that are phenotypically easy for mutation-based search more difficult and phenotypically difficult problems more easy.

1 Metrics and Locality

When considering representations it must be distinguished between phenotypes \mathbf{x}_p and genotypes \mathbf{x}_g . Thus, an optimization problem can be decomposed into two parts. The first maps the genotypic space Φ_g to the phenotypic space Φ_p , and the second maps Φ_p to the fitness space \mathbb{R} :

$$f_g(\mathbf{x}_g) : \Phi_g \rightarrow \Phi_p,$$

$$f_p(\mathbf{x}_p) : \Phi_p \rightarrow \mathbb{R},$$

where the overall optimization problem is defined as $f = f_p \circ f_g = f_p(f_g(\mathbf{x}_g))$. The genotype-phenotype mapping f_g is the used representation and f_p is the used fitness function.

When using search algorithms, a metric has to be defined on the search space Φ . Based on the metric the distance $d_{\mathbf{x}_a, \mathbf{x}_b}$ between two individuals $\mathbf{x}_a \in \Phi$ and $\mathbf{x}_b \in \Phi$ describes how similar the two individuals are. The larger the distance, the more different two individuals are. Two individuals are neighbors if the distance between two individuals is minimal. For example, when using the Hamming metric for binary strings the minimal distance between two individuals is $d = 1$.

If we use a representation f_g there are two different search spaces, Φ_g and Φ_p . Therefore, different metrics can be used for Φ_g and Φ_p . In general, the metric used on the phenotypic search space Φ_p is determined by the specific problem that should be solved and describes which problem solutions are similar to each other. In contrast, the metric defined on Φ_g is not given a priori but depends on the used genotypes. As different genotypes can be used to represent the same phenotypes, different metrics can be defined on Φ_g . Therefore, in general, different metrics are used for Φ_p and Φ_g which imply a different neighborhood structure in Φ_g and Φ_p .

The locality of a representation describes how well neighboring genotypes correspond to neighboring phenotypes. The locality of a representation is high if all neighboring genotypes correspond to neighboring phenotypes. In contrast, the locality of a representation is low if some neighboring genotypes do not correspond to neighboring phenotypes.

2 Influence on Problem Difficulty

The phenotypic difficulty of an optimization problem depends on the metric that is defined on the phenotypes and the function f_p which assigns a fitness value to every phenotype. Based on the phenotypic metric a local search operator can be defined (for the phenotypes). By the use of a representation which assigns a genotype to every phenotype a new genotypic metric is introduced that depends on the used search operator and that can differ from the phenotypic metric. Therefore, the character of the search operator can be different for genotypes and phenotypes. If the locality of a representation is high, then the mutation operator has the same effect on the phenotypes as on the genotypes. As a result, genotypic and phenotypic problem difficulty is the same and the difficulty of a problem remains unchanged by the use of an additional representation f_g . Phenotypically easy problems remain genotypically easy and phenotypically difficult problems remain genotypically difficult. Figure 1 (left) illustrates the effect of mutation for high-locality representations. The search operator (mutation) has the same effect on the phenotypes as on the genotypes.

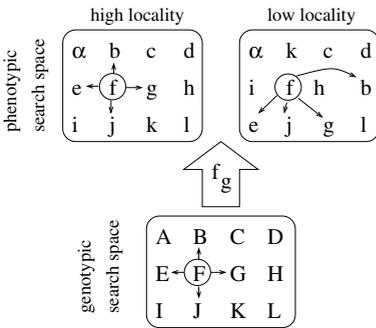


Fig. 1. The effect of mutation for high versus low-locality representations

Finally, we have to consider deceptive and misleading problems. On average, the use of low-locality representations transform such problems into easier problems as the problems become more randomized. Therefore, mutation-based search is less misled by the fitness landscape and the problem difficulty for mutation-based search is reduced. On average, low-locality representations “destroy” the deceptiveness of the problem and make it easier.

3 Conclusions

This work discusses that representations can easily change the difficulty of problems. Only when using high-locality representations problem difficulty does not change. Phenotypically easy problems remain overall easy and phenotypically difficult problems remain overall difficult when using mutation-based search approaches. In contrast, low-locality representations randomize the search process. Therefore, problem difficulty changes and phenotypically easy problems which guide mutation-based search to good solutions become more difficult and phenotypically difficult problems that mislead mutation become more easy.

The situation is different when focusing on low-locality representations. Then the influence of the representation on the difficulty of a problem depends on the considered optimization problem. If the considered problem f_p is easy and the structure of the search space guides the mutation-based search method to the optimal solution, a low-locality representation f_g randomizes the problem and makes the overall problem f more difficult. When using low-locality representations a small change of a genotype does not correspond to a small change of the phenotype but larger changes of the phenotype are possible (compare Fig. 1 (right)).