An Analysis of Recombination in Some Simple Landscapes

David A. Rosenblueth[1] and Christopher R. Stephens[2]

[1] C_3 - Centro de Ciencias de la Complejidad
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM
A. Postal 20-726, México D.F. 01000
[2] C_3 - Centro de Ciencias de la Complejidad
Instituto de Ciencias Nucleares, UNAM
Circuito Exterior, A. Postal 70-543, México D.F. 04510

Abstract. Recombination is an important operator in the evolution of biological organisms and has also played an important role in Evolutionary Computation. In neither field however, is there a clear understanding of why recombination exists and under what circumstances it is useful. In this paper we consider the utility of recombination in the context of a simple Genetic Algorithm (GA). We show how its utility depends on the particular landscape considered. We also show how the facility with which this question may be addressed depends intimately on the particular representation used for the population in the GA, i.e., a representation in terms of genotypes, Building Blocks or Walsh modes. We show how, for non-epistatic landscapes, a description in terms of Building Blocks manifestly shows that recombination is always beneficial, leading to a "royal road" towards the optimum, while the contrary is true for highly epistatic landscapes such as "needle-in-a-haystack".

1 Introduction

Recombination is an important operator in the evolution of biological organisms. It has been seen as an important element also in the context of Evolutionary Computation (EC), especially in Genetic Algorithms (GAs) [4, 5] and to a lesser extent in Genetic Programming (GP) [6]. However, there has been much debate as to its utility [2].

Importantly, when talking about its utility one has to distinguish two different questions — one in which we ask if recombination itself is useful or not; versus another associated with how we apply it. The former is naturally linked to the probability, p_c, of implementing recombination, versus, $p_c(m)$, the conditional probability that, given that recombination is implemented, what recombination mask is used, i.e., the recombination distribution.[1] Importantly, as we will discuss, neither of these questions can be answered in a way that is independent of

---

[1] Here we restrict attention to recombination that is implementable via a binary recombination mask only. This covers all the standard recombination operators but not variants such as multi-parent recombination.

A. Hernández Aguirre et al. (Eds.): MICAI 2009, LNAI 5845, pp. 716–727, 2009.

the fitness landscape on which the population moves. This takes us to a level of complexity which is much more than just asking whether recombination itself is useful.

In this paper using a simple diagnostic as to the utility of recombination in a given generation of a simple GA we will examine how the recombination distribution and fitness landscape affect in an inter-dependent way the efficacy of recombination. We will do this in the context of some simple toy landscapes, trying to indicate what lessons may be learned for more realistic problems. We will also consider the problem by exploiting different descriptions of the population — in terms of Building Blocks and Walsh modes — as well as the canonical representation —the genotype. We will see that in terms of these alternative representations the action of recombination and an understanding of when and under what circumstances it is useful are much more amenable to analysis.

2 $\varDelta$ - The Selection-Weighted Linkage Disequilibrium Coefficient

In this section we introduce the chief diagnostic we will use to examine the utility of recombination. As we are interested in the interaction of selection and recombination we will leave out mutation. The evolution of a population of length ℓ strings is then governed by the equation [8]

$$\langle P_I(t+1)\rangle = P_I'(t) - p_c \sum_m p_c(m)\varDelta_I(m,t) \tag{1}$$

where $P_I'(t)$ is the selection probability for the genotype I. For proportional selection, which is the selection mechanism we will consider here, $P_I'(t) = (f(I)/\bar{f}(t))P_I(t)$, where $f(I)$ is the fitness of string I, $\bar{f}(t)$ is the average population fitness in the tth generation and $P_I(t)$ is the proportion of genotype I in the population. Finally, $\varDelta_I(m,t)$, is the Selection-weighted linkage disequilibrium coefficient [1], for the string I and associated with the recombination mask $m = m_1m_2\dots m_\ell$, such that if $m_i = 0$ the ith bit of the offspring is taken from the ith bit of the first parent, while, if $m_i = 1$ it is taken from the ith bit of the second parent. Explicitly,

$$\varDelta_I(m,t) = (P_I'(t) - \sum_{JK} {\lambda_I}^{JK}(m)P_J'(t)P_K'(t)) \tag{2}$$

where ${\lambda_I}^{JK}(m) = 0,\ 1$ is an indicator function associated with whether or not the parental strings J and K can be recombined using the mask m into the offspring I. For example, for $\ell = 2$, ${\lambda_{11}}^{11,00}(01) = 0$, while ${\lambda_{11}}^{10,01}(01) = 1$. The contribution of a particular mask depends, as we can see, on all possible parental combinations. In this sense, $\varDelta_I(m,t)$ is an exceedingly complicated function.

From equation (1), we can see that if $\varDelta_I(m) > 0$ then recombination leads, on average, to a higher frequency of the string I than in its absence. In other words, in this circumstance, recombination is giving you more of I than you