A Critical and Empirical Study of Epistasis Measures for Predicting GA Performances: A Summary

S. Rochet[1] G. Venturini[2] M. Slimane[2] E.M. El Kharoubi[2]

[1] IUT,
Dép. Génie Electrique,
150, Avenue du Maréchal Leclerc,
13300 Salon de Provence, France.

[2] Laboratoire d'informatique,
Université de Tours,
64, Avenue Jean Portalis,
37200 Tours, France.
venturini,slimane@univ-tours.fr
Tel: (+33)-2-47-36-14-14,
Fax: (+33)-2-47-36-14-22

Abstract. Epistasis measures have been developed for measuring statistical information about the difficulty of a function to be optimized by a genetic algorithm (GA). We give first a review of the work on these measures such as the epistasis correlation. Then we try to relate the epistasis correlation to the overall performances of a binary GA on a set of 14 functions. The only relation that seems to hold strongly is that a high epistasis correlation implies GA-easy, as indicated by the GA theory of deceptiveness. Then, we show that changing the representation of the search space with transformations that improve epistasis measures does not imply the same increasing in the genetic algorithm performances. These both empirical studies seem to indicate that the generality of epistasis measures is limited.

1 Introduction

Characterizing the difficulty of functions is a key point of the genetic algorithms (GAs) theory (see a survey in (Venturini et al. 1997)). The theory of deceptiveness was probably the first one to be studied (Goldberg 1987) (Whitley 1991). It has, however, several limitations (Grefenstette 1992) like for instance dealing only with static fitness computed on the whole search space rather than dynamic fitness computed on the population. Recently, several statistical measures have been developed for characterizing the difficulty of functions for GAs, and thus for predicting the performances of the GA. The fitness/distance correlation evaluates the correlation that may exist between the distance of individuals to the closest global optimum of the fitness function f and their fitness (Jones and Forrest 1995). This measure however has the drawback of requiring the knowledge of f global optima, which is of course not realizable in practice, or not

realistic because it this case one does not need any search method to find the optimum. Other statistical measures are for instance the operator correlation (Manderick et al. 1991) which can be used in virtual GAs (Grefenstette 1995). This measure evaluates for instance the correlation that may exist between the fitness of parents and their offspring with various operators. Such operators can thus be selected according to this correlation. Finally, epistasis measures (Davidor 1991) approximate the fitness function with another function computed by considering the fitness of schemata independently from each others. This bit by bit approximation measures the dependence that may exist between bits in the representation.

In this paper, we describe with more details epistasis measures and we concentrate on the use of such measures for predicting the performances of the GA. Epistasis measures have the advantage of being easily computable from a set of points generated randomly in the search space. For this reason, they seem to be more interesting than other measures such as fitness/distance correlation. We aim here at answering one main question about epistasis measures, a question which as far as we know has not been answered yet: are these measures useful for predicting the performances of the GA in practice? We provide at least two empirical answers to this question. The first one consists in computing the epistasis of several standard functions and in running the GA on these functions. One can then try to relate the epistasis with the observed performances of the GA. As will be seen, this relation is not strong for the tested functions. Then, the second answer is provided by using epistasis measures in order to guide the search for a better binary representation that could improve the GA performances. Our tests suggest that the epistasis measure of a sample of points can be improved but that the GA performances remain at the same level.

The paper is organized as follows: section 2 provides an introduction to epistasis measures. Section 3 describes our first tests where epistasis measures are compared to GA performances on a set of 14 binary functions. Section 4 introduces the basic techniques for changing the representation in a binary GA and provides experimental results. Section 5 concludes on the usefulness of epistasis measures.

2 Overview of epistasis measures

¿From initial work on epistasis measures for GAs (Davidor 1991), one may define the variance of epistasis as follows: let X denote the fitness function that goes from $S = \{0,1\}^l$ to R. To compute the epistasis variance, one must generate a set S_a of binary strings $s_1, ..., s_n$ and evaluate the fitness of those strings $X(s_1), ..., X(s_n)$. Then one must compute a function W that approximates X by using the information contained in the sample of n strings. A binary string of length l can be denoted by $s = b_1, b_2, ..., b_l$ where $b_i \in \{0,1\}$. The value of W for this string equals:

$$W(s) = [X(b_1 * ...*) - X(*...*)] + [X(*b_2 * ...*) - X(*...*)] + ...$$