A Bit-Wise Epistasis Measure for Binary Search Spaces

Cyril Fonlupt, Denis Robilliard, Philippe Preux

Laboratoire d'Informatique du Littoral
BP 719
62228 Calais Cedex, France
e-mail: {fonlupt,robillia}@lil.univ-littoral.fr

Abstract. The epistatic variance has been introduced by Davidor as a tool for the evaluation of interdependences between genes, thus possibly giving clues about the difficulty of optimizing functions with genetic algorithms (GAs). Despite its theoretical grounding in Walsh function analysis, several studies have shown its weakness as a predictor of GAs results. In this paper, we focus on binary search spaces and propose to measure epistatic effect on the level of individual genes, an approach that we call *bit-wise epistasis*. We give examples of this measure on several well-known test problems, then we take into account this supplementary information to improve the performances of evolutionary algorithms. We conclude by pointing towards possible extensions of this concept to real size problems.

1 Introduction

An important issue tackled by epistasis studies, in genetic algorithms (GAs), is understanding and characterizing the difficulty to optimize functions. In other words, how well are epistatic measures able to predict the difficulty of problems? Can they help in finding better solutions? There have been a lot of works dealing with this problem, both theoretical and empirical studies, initiated by a seminal paper by Davidor [1]. His epistasis variance have been shown to have a strong mathematical foundation, based on Walsh functions analysis (see [2–4] and also [5,6]), but still the correlation between problem hardness and epistasis is not straightforward. As it was argued in [3], even if it may provide some guidance for this matter, no definitive conclusion can be drawn. This is also confirmed by recent works from Rochet and Venturini [7], showing that one can change the representation of a problem, and thus achieve a lower epistasis, *without* changing the problem hardness. We think that this lack of accuracy may come from the global nature of this measure, which computation involves two levels of averaging. In this paper, we do not hope to provide a definitive answer to this question. Rather we propose a different measure of epistatic effects in binary search spaces, that we call *bit-wise epistasis*. We think our measure provides an increased accuracy over Davidor's proposition and we show that this may give another point of vue on the matter of problems hardness. This

accuracy is especially clear when dealing with and explaining the unsuccessful remappings of search space proposed in [7]. We present the definition and the principles of our epistasis measure in Sect. 2. We give examples of computations, applied to NK-landscapes, in Sect. 3. Then we study a set of common functions in Sect. 4, and explain why some proposed remapping have failed in Sect. 5. In Sect. 5, we use the information provided by our measures in order to enhance the quality of solutions found by evolutionary algorithms. This improvement is obtained through adding a simple stochastic mutation operator which rate is based on bit-wise epistasis values. This work is too preliminary to allow us to report improvements on real world problems, but we give hints at how one could use the concept of bit-wise epistasis in GAs.

2 A Definition of Bit-Wise Epistasis

Our approach is based on the fact that a strong epistatic relation may exist only on some genes within a genotype, the other genes being much more independent. We think that it may be interesting to have a detailed view of such interactions rather than mixing and merging them like it is done in the epistatic variance defined by Davidor. In the following, we assume the reader is familiar with the notion of *schema* (see also [8, 9]).

Let f be the fitness function from a binary search space $B = \{0,1\}^l$ in the set of reals $\mathbb{R}$ with l the length of genotypes.

Let $B' = \{0,1,\#\}^l$ the set of schemata associated with B. Let Σ_i the set of schemata of order $l-1$ such that their unique undefined loci is at the i^{th} position in the schema, i.e. $\Sigma_i = \{\sigma_0\sigma_1 \ldots \sigma_i \ldots \sigma_{l-1} \in B' \mid \sigma_{j\neq i} \in \{0,1\} \text{ and } \sigma_i = \#\}$.

Let $\alpha = \alpha_0\alpha_1 \ldots \alpha_{i-1}\#\alpha_{i+1} \ldots \alpha_{l+1}$ a schema in Σ_i. Let $X_\alpha, \bar{X}_\alpha$ be genotypes in B, members of α, with $X_\alpha = \alpha_0\alpha_1 \ldots \alpha_{i-1}\mathbf{0}\alpha_{i+1} \ldots \alpha_{l-1}$ and $\bar{X}_\alpha = \alpha_0\alpha_1 \ldots \alpha_{i-1}\mathbf{1}\alpha_{i+1} \ldots \alpha_{l-1}$. We call $d_i(\alpha)$ the *fitness difference*[1] *at gene i*:

$$d_i(\alpha) = f(X_\alpha) - f(\bar{X}_\alpha)$$

We define the *mean fitness difference at gene i* as the mean $d_i(\alpha)$ for all schemata $\alpha \in \Sigma_i$:

$$M_i = \frac{1}{2^{l-1}} \sum_{\alpha \in \Sigma_i} d_i(\alpha)$$

We call *bit-wise epistasis at gene i* the variance of the fitness difference at gene i:

$$\sigma_i^2 = \frac{1}{2^{l-1}} \sum_{\alpha \in \Sigma_i} [M_i - d_i(\alpha)]^2$$

[1] An alternative definition could be the absolute value $|d_i(\alpha)|$. We intend to discuss the reasons underlying our choice in a forthcoming paper.