Measuring the Evolvability Landscape to study Neutrality

Sébastien Verel, Philippe Collard, Manuel Clergue

HAL Id: hal-00164917

https://hal.science/hal-00164917

Submitted on 25 Sep 2007

1. FITNESS LANDSCAPES

We will use the definition of fitness landscapes from [4].

A *fitness landscape* is a triplet $(\mathcal{S}, \mathcal{V}, f)$ such as: $\mathcal{S}$ is the set of potential solutions, $\mathcal{V} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$ is the neighborhood function which associated to each solution $s \in \mathcal{S}$ a set of neighbor solutions $\mathcal{V}(s) \subset S$, and $f : \mathcal{S} \rightarrow \mathbb{R}$ is the fitness function which associates a real number to each solution.

1.1 Rugged Fitness landscapes

Weinberger [6] introduced the autocorrelation function and the correlation length of random walks to measure the ruggedness of fitness landscapes.

Given a random walk $(s_t, s_{t+1}, \ldots)$, the autocorrelation function ρ of a fitness function f is the autocorrelation function of time series $(f(s_t), f(s_{t+1}), \ldots)$:

$$\rho(k) = \frac{E[f(s_t)f(s_{t+k})] - E[f(s_t)]E[f(s_{t+k})]}{var(f(s_t))}$$

where $E[f(s_t)]$ and $var(f(s_t))$ are the expected value and the variance of $f(s_t)$. The correlation length $\tau = -\frac{1}{log(\rho(1))}$, measures how the autocorrelation function decreases and it summarizes the ruggedness of the landscape: the larger the correlation length, the smoother is the landscape.

1.2 Neutral fitness landscapes

The geometry of neutral fitness landscapes are based on the concept of neutral neighborhood and neutral networks.

For every $s \in S$, the *neutral neighborhood* of s is the set $\mathcal{V}_{neut}(s) = \{s' \in \mathcal{V}(s) \mid f(s) = f(s')\}$ and the *neutral degree* of s is the number of neutral neighbors of s. There is no exact definition of neutral fitness landscape but we can define a fitness landscape as neutral if there are "many" solutions with "high" neutral degree. In this case, we can imagine fitness landscapes with some plateaus called *neutral networks*. There is no significant difference of fitness between solutions on neutral networks and the population drifts around on them.

The Neutral Networks (NN) of a fitness landscape are connected graphs which are the connected components of graph $(\mathcal{S}, \mathcal{V}_{neut})$. Solutions in a NN have the same fitness value and there is a path of neutral neighbors between two solutions of a NN. The presence of NN takes place in the search dynamic. For example, the time of drift on NN during the search process depends on the properties of the networks.

ABSTRACT

This theoretical work defines the measure of *autocorrelation of evolvability* in the context of neutral fitness landscape. This measure has been studied on the classical MAX-SAT problem. This work highlight a new characteristic of neutral fitness landscapes which allows to design new adapted metaheuristic.

Categories and Subject Descriptors

I.2.8 [**Problem Solving, Control Methods, and Search**]: Heuristic methods—*fitness landscapes, neutrality, measures*

General Terms

Theory

Introduction

Fitness landscape, introduce by Wright [4] in evolutionary biology, is one of the powerful metaphor to model evolutionary process. The dominant view in this metaphor is an adaptive evolution where an uphill walk of a population on a mountainous fitness landscape in which it can get stuck on suboptimal peaks. In combinatorial optimization, this view also influence the design of metaheuristic: the geometry of multimodality or ruggedness describes the fitness landscape and the metaheuristics try to escape from local optima by using a probability to explore the landscape in simulated annealing, by using a memory in tabu search, or by preserving the diversity of population in evolutionary algorithm.

Another geometry of fitness landscapes, enlightened in molecular evolution by Kimura [2], takes an important place in optimization: the neutral fitness landscape. In the theory of neutral evolution, the overwhelming majority of mutations are either effectively neutral or lethal and in the latter case purged by negative selection. According to this theory, fitness landscapes are dominated by plateaus, called *neutral networks*. The landscapes from genetic programming or from applicative problems such as minimal linear arrangement are known to be neutral. It is difficult to decide whether neutrality is useful for optimization. This work propose to deeper describe neutral fitness landscapes in order to get some characteristics of theses problems and obtain a more complete view on neutrality which allows to design more useful metaheuristic.

2. THE AUTOCORRELATION OF EVOLVABILITY

Evolvability is defined by Altenberg [1] as "the ability of random variations to sometimes produce improvement". As enlighten by Turney [5] the concept of evolvability is difficult to define. As he puts it: "if s and s' are equally fit, s is more evolvable than s' if the fittest offspring of s is more likely to be fitter than the fittest offspring of s'". Following this idea the evolvability of a solution is defined by a function ef that assigns to every $s \in \mathcal{S}$ a real number which measure the evolvability. For examples, the evolvability function could be the maximum fitness from the neighborhood $ef(s) = max\{f(s') \mid s' \in \mathcal{V}(s)\}$.

We define the *autocorrelation of evolvability* for a neutral network N as the autocorrelation function of evolvability on neutral networks.

The *autocorrelation of evolvability* on the neutral network N is the autocorrelation of series $(ef(s_0), ef(s_1), \ldots)$ where $(s_0, s_1, \ldots)$ is a neutral random walk on N.

A neutral random walk is a series of solutions $(s_0, s_1, \ldots)$ such as $s_{i+1} \in \mathcal{V}(s_i)$ and $f(s_{i+1}) = f(s_i)$. To extend this measure to the set of all neutral networks, the average of autocorrelation coefficient is computed. Several choices could be made to define evolvability function; in particular we call *autocorrelation of maximal evolvability* the autocorrelation when the evolvability is the maximum fitness from the neighborhood of a solution.

The evolvability gives the fitnesses of neutral network in the neighborhood. For example, the maximum evolvability of a solution is the fitness of the higher NN in the neighborhood. The autocorrelation of evolvability allows to describe the distribution of neutral networks around. If the correlation is large, the fitness in the neighborhood of a NN is distributed regularly, whereas if the correlation is low, the NNs around a NN is randomly distributed.

3. FIRST RESULTS ON THE MAX-K-SAT PROBLEM

In the following we present the first measures of autocorrelation of evolvability. The MAX-k-SAT is defined from the SAT optimization problem. SAT is a decision problem that asks whether a binary tuple can be found that satisfies all clauses in normal conjunctive form. Many studies deal with the solution space of the SAT problem [3], such phase transistion arround the threshold α_c defined as the ratio between the number of clauses and the number of variables. For $k = 3$, α_c is equal to 4.3.

The experiments are led in the same way as for preview landscapes with random instances of MAX-3-SAT. The number of variables is set to $N = 16$ and $N = 64$, the number of literals by clause is $k = 3$ and the number of clauses m describe respectively the sets $\{39, 59, 64, 69, 74, 79, 99\}$ and $\{200, 250, 265, 275, 285, 300, 350\}$. The average neutral degree decreases when m increases.

The figure 1 shows the correlation length of maximal evolvability. For all the parameters, the correlation is significant. The correlation length is around 2 for $N = 16$ and around 5 for $N = 64$. For all the value of the number of clauses, the autocorrelation functions are nearby and the variations of correlation length are weak according to the parameter α. The correlation decreases slowly according to the number of clauses, which is linked to the neutrality. The autocorrela-

tion of maximal evolvability do not shows a phase transition around the threshold $\alpha_c = 4.3$. The neutral networks are not randomly distributed in the fitness landscapes. Those first experiments show that the autocorrelation of evolvability is one of characteristic of applicative neutral fitness landscapes.
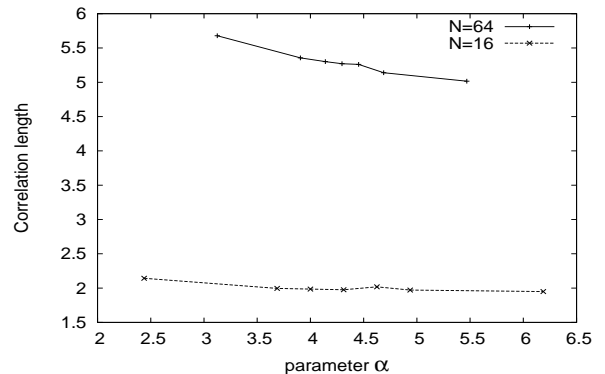

Figure 1: Correlation length of maximal evolvability for $N = 16$ and 64 and different number of clauses.

4. CONCLUSION

From the metaphor of neutral fitness landscapes, the neutral networks (NN) are the plateaus. For each NN, we have defined the autocorrelation function of evolvability. The first studies on applicative MAX-SAT problem has showed a new characteristic of those problem which can be compared to the preview additive landscapes.

The autocorrelation of evolvability is an useful measure which highlight a new characteristic of neutral fitness landscapes which could be study in real optimization problems. In spite of a lack of differential between fitness in a NN, evolvability could be exploit to guide the search process on a network. Futures works could take into account this information to design new metaheuristics adapted to neutral fitness landscapes.

5. REFERENCES

[1] G. Wagner. Complexes adaptations and the evolution of evolvability. In *Evolution*, pages 967–976, 1996.

[2] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK, 1983.

[3] M. Mezard and R. Zecchina. The random k-satisfiability problem: from an analytic solution to an efficient algorithm. *Phys. Rev. E*, 66(056126), 2002.

[4] P. F. Stadler. Fitness landscapes. In M. Lässig and Valleriani, editors, *Biological Evolution and Statistical Physics*, volume 585 of *Lecture Notes Physics*, pages 187–207, Heidelberg, 2002. Springer-Verlag.

[5] P. D. Turney. Increasing evolvability considered as a large scale trend in evolution. In P. Marrow and al, editors, *GECCO'99 , Workshop Program on evolvability*, pages 43–46, 1999.

[6] E. D. Weinberger. Correlated and uncorrelatated fitness landscapes and how to tell the difference. In *Biological Cybernetics*, pages 63:325–336, 1990.