Population Sizing Based on Landscape Feature

Jian Zhang, Xiaohui Yuan, and Bill P. Buckles

Department of Electrical Engineering and Computer Science
Tulane University, New Orleans, LA 70001

Abstract. Population size for Evolutionary Algorithms is usually an empirical parameter. We study the population size from aspects of fitness landscapes' ruggedness and Probably Approximately Correct (PAC) learning theory.

1 Introduction

Evolutionary Algorithms (EAs) have been applied to various kinds of problems [1,2]. One important aspect that affects the global convergence in EA is deciding population size, which has become a focus of research in recent years [3,4]. In order to estimate population size in real-coded EAs, we employ PAC learning theory and propose concepts on characterizing the fitness landscapes.

2 Population Size of Evolutionary Algorithms

The intuitive notion of *ruggedness* is related to the difficulty of optimizing over a given landscape [5]. Several distinctive approaches have been proposed to quantify ruggedness [6,7,8,9].

Intuitively, if the initial population of real-coded EAs is large enough to include points that are in the neighborhood of the global optimum, EAs have a great chance of locating the target point in subsequent generations. It is therefore necessary to define a metric inferring closeness to optima points.

Definition 1. *Granularity of Fitness Function*
A fitness function $f(x)$ can be represented by the linear combination of a set of orthonormal basis functions $\{\varphi_k(x)\} = \{\varphi(k_1 x), \varphi(k_2 x), \ldots, \varphi(k_n x)\}$, that is $f(x) = \sum_k a_k \varphi_k(x)$. The granularity of fitness function f is defined as $\tau = 1/max(k_i), i = 1, \ldots, n$, where k_i characterizes the frequency information of the basis function $\varphi(k_i x)$.

In order to infer the nearness of organisms in real value domain, we adopt the definition of ϵ-cover provided by Vidyasagar [10].

Theorem 1. *Given the fitness function $f(x)$ with granularity τ, where $f(x)$ is defined on $\mathcal{S} \subset R^n$, the PAC population size m is bounded by $\tilde{m}$, that is, $m \geq \tilde{m}$*

$$\tilde{m} = \lceil \frac{1}{\phi}(ln\lceil \frac{1}{\phi} \rceil + ln\frac{1}{\delta}) \rceil \tag{1}$$

where $\phi = g(\tau)/\mathcal{S}, g(\tau) = \tau$ when $n = 1$; $g(\tau) = \pi\tau^2/4$ when $n = 2$. $\lceil 1/\phi \rceil$ defines the size of hypothesis space, such that with confidence δ, $0 < \delta < 1$, the initial population forms an ϵ-cover of $\mathcal{S}$ with probability greater than $1 - \delta$ and $\epsilon = \tau$.

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 1624–1625, 2003.

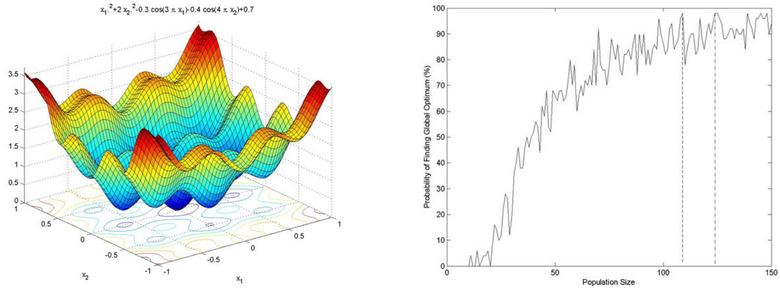

Fig. 1. Convergence graphs for Bohachevsky function: $f(x_1, x_2) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7, x_1, x_2 \in [-1, 1]$.

Assume the initial population is drawn according to a uniform distribution. If the initial population forms an ϵ-cover of the solution space, then no point in that space is more than ϵ-away from a member in ϵ-cover. Several approaches for measuring the ruggedness have been considered, such as the number of local minima and the correlation length. We adopted a granularity measure τ from the decomposition of fitness landscapes.

Figure 1 shows that enlarging the population size increases the probability of finding the global optimum. However, after it reaches a certain point, the improvement is no longer dramatic. The PAC population size is at the threshold that gives high convergence rate and minimizes the computational expense.

References

1. Dasgupta, D., Michalewicz, Z., eds.: Evolutionary Algorithms in Enigneering Applications. Springer, Berlin, Germany (1997)
2. Zhang, J., Yuan, X., Buckles, B.P.: An evolution strategies based approach to image registration. In: Proceeding of Genetic Evolutionary Computation Conference, New York (2002)
3. Goldberg, D.E., Rudnick, M.: Genetic algorithms and the variance of fitness. Technical report no. 91001, Illinois GAs laboratory, University of Illinois, Champaign-Urbana (1991)
4. Hernández-Aguirre, A., Buckles, B.P., Martinez-Alcantara, A.: The PAC population size of a genetic algorithm. In: the 12th Int'l Conf. on Tools with Artificial Intelligence. (2000) 199–202
5. Jones, T., Forrest, S.: Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In Eshelman, L., ed.: Proceedings of the Sixth International Conference on Genetic Algorithms, San Francisco, CA, Morgan Kaufmann (1995) 184–192
6. Sorkin, G.: Combinatorial optimization, simulated annealing and fractals. No. 61253, IBM Research Report RC13674 (1988)
7. Weinberger, E.D.: Correlated and uncorrelated fitness landscapes and how to tell the difference. Biological Cybernetics **63** (1990) 325–336
8. Kauffman, S.A., Levin, S.: Towards a general theory of adaptive walks on rugged landscapes. Journal of Theoretical Biology **128** (1987) 11–45
9. Palmer, R.: Optimization on rugged landscapes. Molecular Evolution on Rugged Landscapes: Proteins, RNA, and the Immune System (1991) 3–25
10. Vidyasagar, M.: A Theory of Learning and Generalization. Springer (1997)