Theoretical Aspects of Evolutionary Algorithms

Ingo Wegener*

FB Informatik, LS2, Univ. Dortmund, 44221 Dortmund, Germany
wegener@ls2.cs.uni-dortmund.de

Abstract. Randomized search heuristics like simulated annealing and evolutionary algorithms are applied successfully in many different situations. However, the theory on these algorithms is still in its infancy. Here it is discussed how and why such a theory should be developed. Afterwards, some fundamental results on evolutionary algorithms are presented in order to show how theoretical results on randomized search heuristics can be proved and how they contribute to the understanding of evolutionary algorithms.

1 Introduction

Research on the design and analysis of efficient algorithms was quite successful during the last decades. The very first successful algorithms (Dantzig's simplex algorithm for linear programming and Ford and Fulkerson's network flow algorithm) have no good performance guarantee. Later, research was focused on polynomial-time algorithms (see Cormen, Leiserson, and Rivest (1990)) and this type of research has been extended to approximation algorithms (see Hochbaum (1997)) and randomized algorithms (see Motwani and Raghavan (1995)). Indeed, designing and implementing an efficient algorithm with a proven performance guarantee is the best we can hope for when considering an algorithmic problem. This research has led to a long list of efficient problem-specific algorithms. Moreover, several paradigms of algorithms have been developed, among them divide-and-conquer, dynamic programming, and branch-and-bound. There are general techniques to design and analyze algorithms. However, these paradigms are successful only if they are realized with problem-specific modules. Besides these algorithms also paradigms for the design of heuristic algorithms have been developed like randomized local search, simulated annealing, and all types of evolutionary algorithms, among them genetic algorithms and evolution strategies. These are general classes of search heuristics with many free modules and parameters. We should distinguish problem-specific applications where we are able to choose the modules and parameters knowing properties of the considered problem and problem-independent realizations where we design a search heuristic to solve all problems of a large class of problems. We have to argue why one should investigate such a general scenario. One main point is that we obtain

* This work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center "Computational Intelligence" (531).

F. Orejas, P.G. Spirakis, and J. van Leeuwen (Eds.): ICALP 2001, LNCS 2076, pp. 64–78, 2001.

the frame of a general search heuristic where some details may be changed in problem-specific applications. Moreover, there are at least two situations where problem-independent algorithms are of particular interest. First, in many applications, one has not enough resources (time, money, specialists,...) to design a problem-specific algorithm or problem-specific modules. Second, often we have to deal with "unknown" functions which have to be maximized. This scenario is called black box optimization. It is appropriate for technical systems with free parameters where the behavior of the system cannot be described analytically. Then we obtain knowledge about the unknown function only by "sampling". The t-th search point can be chosen according to some probability distribution which may depend on the first $t-1$ search points $x_1, \dots, x_{t-1}$ and their function values $f(x_1), \dots, f(x_{t-1})$. One main idea of all randomized search heuristics is to "forget" much of the known information and to make the choice of the probability distribution only dependent on the "non-forgotten" search points and their f-values.

Our focus is the maximization of pseudo-boolean functions $f : \{0,1\}^n \rightarrow \mathbb{R}$ which covers the problems from combinatorial optimization. We investigate and analyze randomized search heuristics which are designed to behave well on "many" of the "important and interesting" pseudo-boolean functions. Obviously, they cannot beat problem-specific algorithms and, also obviously, each randomized search heuristic is inefficient for most of the functions. The problem is to identify for a given randomized search heuristic classes of functions which are optimized efficiently and to identify typical functions where the heuristic fails. Such theoretical results will support the selection of an appropriate search heuristic in applications. One may also assume (or hope) that the search heuristic behaves well on a function which is "similar" to a function from a class where it is proved that the heuristic is efficient. Moreover, the proposed results lead to a better understanding of search heuristics. This again leads to the design of improved search heuristics and gives hints for a better choice of the parameters of the search heuristic. Finally, analytical results support the teaching of randomized search heuristics.

In black box optimization the black box (or oracle) answers queries x with $f(x)$ where $f : \{0,1\}^n \rightarrow \mathbb{R}$ is the function to be maximized. Since queries are expensive, the search cost is defined as the number of queries. For a fixed search heuristic let X_f be the random number of queries until "some good event" happens. The good event in this paper is that a query point is f-maximal. Then we are interested in the expected optimization time $E(X_f)$ and the success probability function $s(t) := \text{Prob}(X_f \leq t)$. This is an abstraction from the real problem, since obtaining the f-value of some optimal x does not imply that we know that x is optimal. In applications, we additionally need good stopping rules.

Our focus is on evolutionary algorithms which have been developed in the sixties of the last century and which have found many applications during the last ten years. Evolutionary algorithms are described in many monographs (Fogel (1995), Goldberg (1989), Holland (1975), Schwefel (1995)) and in a more recent