Theoretical Computer Science 598 (2015) 64–86

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Comprehensive characterization of the behaviors of estimation of distribution algorithms

Carlos Echegoyen *, Roberto Santana, Alexander Mendiburu, Jose A. Lozano

Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, The University of the Basque Country (UPV/EHU), Paseo Manuel de Lardizábal 1, 20018 San Sebastian, Spain

A R T I C L E I N F O

Article history:
Received 16 March 2014
Received in revised form 6 February 2015
Accepted 14 April 2015
Available online 20 April 2015
Communicated by B. Doerr

Keywords:
Heuristic optimization
Machine learning
Estimation of distribution algorithms
Factorizations
Equivalence classes
Neighborhood systems

A B S T R A C T

Estimation of distribution algorithms (EDAs) are a successful example of how to use machine learning techniques for designing robust and efficient heuristic search algorithms. Understanding the relationship between EDAs and the space of optimization problems is a fundamental issue for the successful application of this type of algorithms. A step forward in this matter is to create a taxonomy of optimization problems according to the different behaviors that an EDA can exhibit. This paper substantially extends previous work in the proposal of a taxonomy of problems for univariate EDAs, mainly by generalizing those results to EDAs that are able to deal with multivariate dependences among the variables of the problem. Through the definition of an equivalence relation between functions, it is possible to partition the space of problems into equivalence classes in which the algorithm has the same behavior. We provide a sufficient and necessary condition to determine the equivalence between functions. This condition is based on a set of matrices which provides a novel encoding of the relationship between the function and the probabilistic model used by the algorithm. The description of the equivalent functions belonging to a class is studied in depth for EDAs whose probabilistic model is given by a chordal Markov network. Assuming this class of factorization, we unveil the intrinsic connection between the behaviors of EDAs and neighborhood systems defined over the search space. In addition, we carry out numerical simulations that effectively reveal the different behaviors of EDAs for the injective functions defined over the search space $\{0,1\}^3$. Finally, we provide a novel approach to extend the analysis of equivalence classes to non-injective functions.

1. Introduction

Recently, there has been an increasing interest in the application of machine learning techniques in combinatorial and stochastic optimization [1–4]. The use of machine learning in optimization provides a more efficient exploitation and representation of the information gathered about the search space. One consolidated example of the methods that incorporate machine learning to optimization are estimation of distribution algorithms (EDAs) [5–7]. Strong evidence of their popularity is the development of new and more complex EDAs [8–10], their application both to real and academic problems [11–13] and the studies of fundamental issues in order to better understand how these algorithms perform [14–16].

* Corresponding author.
E-mail addresses: carlos.echegoyen@ehu.es (C. Echegoyen), roberto.santana@ehu.es (R. Santana), alexander.mendiburu@ehu.es (A. Mendiburu), ja.lozano@ehu.es (J.A. Lozano).

http://dx.doi.org/10.1016/j.tcs.2015.04.015
0304-3975/

EDAs are a class of evolutionary algorithms (EAs) [17]. Commonly, EAs search for the best solutions of a problem by maintaining a population of individuals (solutions) that evolves from one generation to the next. The evolution is carried out by selecting a subset of promising individuals from the population and applying recombination operators that create new individuals (with the aim of obtaining higher quality solutions at each step). In particular, EDAs use machine learning methods to extract relevant features of the search space through the selected individuals of the population. The learning algorithm allows to estimate a new probability distribution over the search space at each step of the EDA. Thus, each of the candidate solutions has an associated probability of being sampled, which varies during the optimization process. Consequently, given a problem, the ideal objective of an EDA is to get higher probability values for the highest quality solutions throughout an iterative process. However, in practice, it is hard to know to what extent the EDA is working as desired.

In general, although the application of EDAs can be relatively easy, predicting any aspect of their behavior or determining the quality of the search are non-trivial matters. This scenario motivates the study of the underlying mechanisms that govern EDAs in order to better understand their behavior when solving problems. For instance, there are works that analyze the convergence of specific EDAs to the global optimum and the local optima under ideal conditions [14,18,19], works that analyze the time complexity [20] of univariate EDAs in relation to the problem size or other works that conduct EDA runs on benchmark problems to study the structural models learned by the algorithm [21], or the probability values generated during the search [22]. The current work, which provides new knowledge in this research field, presents an exhaustive classification of the different behaviors that certain EDAs can exhibit, and studies how these behaviors are related with the local optima of the optimization problems.

The behavior of an EDA, when it is applied to a given objective function, can be accurately described by the sequence of probability distributions generated at each generation. We use this description of the EDA in order to identify and classify the different behaviors. Given the complexity of this task, a number of assumptions have to be made: i) we consider EDAs with infinite populations (although the taxonomy of problems developed throughout the paper is also valid for finite populations), ii) the selection scheme is based on the rank of the solutions and iii) the algorithm is applied in the space of injective functions. By considering infinite populations it is possible to see the EDA as a deterministic dynamical system. This type of EDA avoids the random errors caused by sampling finite sets of solutions, which is desirable for certain theoretical analysis. Section 3 explains in more detail this type of EDA.

A crucial element in this work is the definition of an equivalence relation between objective functions. The equivalence relation, which is based on the sequences of probability distributions generated during the search, partitions the space of functions into equivalence classes. We prove that the EDA exhibits the same performance for all the functions belonging to the same class and that the performance of the algorithm is different in each class. Thus, it is possible to group the optimization problems according to the behavior of the algorithm, creating a taxonomy.

This novel research line was initiated in [23], where the study is restricted to EDAs that assume independence among the variables of the problem. In the current work, the study is generalized to EDAs that are able to deal with multivariate dependences among the variables of the problem. In order to present a self-contained paper, we will need to include some definitions and explanations already included in [23]. Nevertheless, the current work provides significant advances which are summarized as follows. Firstly, in order to create partitions of the space of problems, the necessary and sufficient condition to identify equivalent functions that was presented in [23] is generalized here to EDAs that implement any factorization given by a Bayesian network [24]. Secondly, based on this condition, we carry out a novel detailed description of the functions belonging to a class and count the number of equivalent functions. Due to the wide casuistry that the use of general Bayesian networks entails, we have restricted the analysis to factorizations given by chordal Markov networks, which are a subclass of Bayesian networks. We argue that this type of models are widely used in EDAs. For instance, they are present in the earliest EDAs [25], in current research lines of EDAs [26] and also in theoretical works [27]. Thirdly, we study the connection between the equivalence classes generated by an EDA that implements a chordal Markov network and the local optima of the functions belonging to each class. In order to do that, we define a distance associated to the factorization implemented by the EDA. Then, according to this distance, we show that the functions in the same class have the same number of local optima and in the same ranking positions. In addition, we show that the algorithm cannot converge to a solution which is not a local optimum. These facts reveal the intrinsic connection between neighborhood systems and EDAs and open the path for the implementation of informed neighborhood search schemes within the EDA that take advantage of the factorization for the definition and exploitation of neighborhood systems. Finally, we conduct extensive numerical simulations of the EDAs that introduce a Bayesian network of three variables. These EDAs implement tournament selection [14] and they are applied to the injective functions defined over the search space $\{0,1\}^3$. Through the numerical analysis, we complement the theoretical study by analyzing the complexity of the problems belonging to each class. The difficulty of the problems is presented in relation to the local optima due to the relevant role that this problem descriptor plays in EDAs.

The rest of the paper is organized as follows. Section 2 formally introduces optimization problems and presents EDAs. Section 3 describes the EDAs with infinite populations considered in the current work. In Section 4, the concept of equivalence between functions is presented and discussed. Section 5 introduces the factorizations given by Bayesian networks and the equivalence condition. In Section 6 an in-depth description of the functions belonging to a class is carried out. Section 7 studies the relationship between EDAs and local optima. Section 8 presents numerical experiments to provide complemen-