1

Introduction

Martin Pelikan[1], Kumara Sastry[2], and Erick Cantú-Paz[3]

Summary. This chapter provides motivation for estimation of distribution algorithms and discusses the scope of this book. Additionally, the chapter provides a road map to the book and pointers to additional information.

1.1 Motivation for EDAs

Estimation of distribution algorithms (EDAs) [1, 5, 8, 11] address broad classes of optimization problems by learning explicit probabilistic models of promising solutions found so far and sampling the built models to generate new candidate solutions. By incorporating advanced machine learning techniques into genetic and evolutionary algorithms, EDAs can scalably solve many challenging problems, significantly outperforming standard genetic and evolutionary algorithms and other optimization techniques. In the recent decade, many impressive results have been produced in the design, theoretical investigation, and applications of EDAs.

An EDA evolves a population of candidate solutions to the given problem. Each iteration starts by evaluating the candidate solutions and selecting promising solutions so that solutions of higher quality are given more copies than solutions of lower quality. EDAs can use any standard selection method of genetic and evolutionary algorithms, such as binary tournament selection. Next, a probabilistic model is build for the selected solutions and new solutions are generated by sampling the built probabilistic model. New solutions are then incorporated into the original population using some replacement strategy, and the next iteration is executed unless the termination criteria have been met. EDAs usually differ in the representation of candidate solutions, the considered class of probabilistic models, or the procedures for learning and sampling probabilistic models. The pseudocode of an EDA follows:

M. Pelikan et al.: *Introduction*, Studies in Computational Intelligence (SCI) **33**, 1–10 (2006)
www.springerlink.com

```
Estimation of Distribution Algorithm (EDA)
  t := 0;
  generate initial population P(0);
  while (not done) {
    select population of promising solutions S(t);
    build probabilistic model P(t) for S(t);
    sample P(t) to generate O(t);
    incorporate O(t) into P(t);
    t := t+1;
  }
```

EDAs derive inspiration from two areas: genetic and evolutionary computation and machine learning. The remainder of this section discusses these two sources of inspiration.

1.1.1 Motivation from Genetic and Evolutionary Computation

EDAs borrow two important concepts from genetic and evolutionary computation:

1. Population-based search
2. Exploration by combining and perturbing promising solutions

Using a population of solutions as opposed to a single solution has several advantages; for example, it enables simultaneous exploration of multiple regions in the search space, it can help to alleviate the effects of noise in evaluation, and it allows the use of statistical and learning techniques to automatically identify problem regularities.

Exploration of the search space by combining and perturbing promising solutions can be effective because in most real-world problems, high quality solutions are expected to share features. By effective identification of important features and their juxtaposition, the global optimum can be identified even in problems where local operators fail because of exponentially many local optima and strong large-order interactions between problem variables.

1.1.2 Motivation from Machine Learning

EDAs use probabilistic models to guide exploration of the search space. Using probabilistic models enables the use of rigorous statistical modeling and sampling techniques to automatically discover and exploit problem regularities for effective exploration.

In most EDAs, probabilistic models are represented by graphical models [2, 4, 9], which combine graph theory, modularity and statistics to provide a flexible tool for learning and sampling probability distributions, and probabilistic inference. Graphical models provide EDAs with a powerful tool for identifying and exploiting problem decomposition, whereas evolutionary algorithms provide EDAs with robust operators for maintaining diverse populations of promising candidate solutions. Since most complex real-world systems are nearly decomposable and hierarchical [16], the combination of machine learning techniques and evolutionary algorithms should enable EDAs to solve broad classes of difficult real-world problems in a robust and scalable manner. This hypothesis was supported with a number of theoretical and empirical results [3, 6, 7, 10, 12–15].

1.2 Scope and Road Map

This book provides a selection of some of the important contributions to research and application of EDAs. There are three main areas covered in this book:

1. *Design of robust and scalable EDAs.* (Chaps. 2–6, 10 and 11)
2. *Efficiency enhancement of EDAs.* (Chaps. 7–9, 11, 13, and 15)
3. *Applications of EDAs.* (Chaps. 12–15)

The content of each chapter is discussed next.

Chapter 2. *The Factorized Distribution Algorithm and the Minimum Relative Entropy Principle* by Heinz Mühlenbein and Robin Höns.
In this chapter, Mühlenbein and Höns discuss major design issues of EDAs using an interdisciplinary framework: The *minimum relative entropy* (*MinRel*) approximation. They demonstrate the relation between the Factorized Distribution Algorithm (FDA) and the *MinRel* approximation. Mühlenbein and Höns propose an EDA derived from the *Bethe–Kikuchi* approach developed in statistical physics and present details of a concave–convex minimization algorithm to solve optimization problems. The two algorithms are compared using popular benchmark problems – 2-d grid problems, 2-d Ising spin glasses, and Kaufman's $n - k$ function – with instances of up to 900 variables.

Chapter 3. *Linkage Learning via Probabilistic Modeling in the Extended Compact Genetic Algorithm (ECGA)* by Georges R. Harik, Fernando G. Lobo, and Kumara Sastry.
The first-generation genetic algorithms (GAs) are not very successful in automatically identifying and exchanging structures consisting of several correlated genes. This problem, referred in the literature as the *linkage-learning* problem, has been the subject of extensive research over the last few decades. Harik et al. explore the relationship between the linkage-learning problem and that of learning probability distributions over multivariate spaces. They argue that the linkage-learning problem and learning probability distributions are equivalent. Using a simple yet effective approach to learning distributions, and by implication linkage,

Harik et al. propose a GA-like algorithm – the extended compact GA – that is potentially orders of magnitude faster and more accurate than simple GAs.

Chapter 4. *Hierarchical Bayesian Optimization Algorithm* by Martin Pelikan and David E. Goldberg.
Pelikan and Goldberg describe the hierarchical Bayesian optimization algorithm (hBOA) and its predecessor, the Bayesian optimization algorithm (BOA), and outline some of the important theoretical and empirical results in this line of research. The hierarchical Bayesian optimization algorithm (hBOA) solves nearly decomposable and hierarchical optimization problems scalably by combining concepts from evolutionary computation, machine learning and statistics. Since many complex real-world systems are nearly decomposable and hierarchical, hBOA is expected to provide scalable solutions for many complex real-world problems.

Chapter 5. *Numerical Optimization with Real–Valued Estimation-of-Distribution Algorithms* by Peter A.N. Bosman and Dirk Thierens.
In this chapter, Bosman and Thierens focus on the design of real-valued EDAs for the task of numerical optimization. Here, both the problem variables as well as their encoding are real values, and concordantly, the type of probability distribution to be used for estimation and sampling in the EDA is continuous. Bosman and Thierens indicate the main challenges in real-valued EDAs and review the existing literature to indicate the current EDA practice for real-valued numerical optimization. They draw conclusions about the feasibility of existing EDA approaches and provide an explanation for some observed deficiencies of continuous EDAs as well as possible improvements and future directions of research in this branch of EDAs.

Chapter 6. *A Survey of Probabilistic Model Building Genetic Programming* by Yin Shan, Robert I. McKay, Daryl Essam, and Hussein A. Abbass.
While the previous chapters address EDAs that mainly operate on variables encoded into fixed-length chromosomes, there has been growing interest in extending EDAs to operate on variable-length representations, especially for evolving computer programs. In this chapter, Shan et al. provide a critical and comprehensive review of EDAs for automated programming. They discuss important lessons learned from genetic programming (GP) for better design of probabilistic models for GP. Shan et al. also present key strengths and limitations of existing EDAs for GP.

Chapter 7. *Efficiency Enhancement of Estimation of Distribution Algorithms* by Kumara Sastry, Martin Pelikan, and David E. Goldberg.
Estimation of distributions have taken problems that were *intractable* with first generation GAs and rendered them *tractable*, whereas *efficiency-enhancement* take EDAs from *tractability* to *practicality*. That is, efficiency-enhancement techniques speedup the search process of estimation of distribution algorithms (EDAs) and thereby enable EDAs to solve hard problems in *practical* time. Sastry et al. provide a decomposition and

a review of different efficiency-enhancement techniques for EDAs. They illustrate a principled approach for designing efficiency enhancement techniques by developing an evaluation-relaxation scheme in the Bayesian optimization algorithm, and a time-continuation method in the extended compact genetic algorithm.

Chapter 8. *Design of Parallel Estimation of Distribution Algorithms* by Jiri Ocenasek, Erick Cantú-Paz, Martin Pelikan, and Josef Schwarz.

In this chapter, Ocenasek et al. focus on the parallelization of Estimation of Distribution Algorithms (EDAs) and present guidelines for designing efficient parallel EDAs that employ parallel fitness evaluation and parallel model building. They employ scalability analysis techniques to identify and parallelize the main performance bottlenecks to ensure that the achieved speedup grows almost linearly with the number of utilized processors. Ocenasek et al. demonstrate their proposed approach on the parallel Mixed Bayesian Optimization Algorithm (MBOA) and verify it with experiments on the problem of finding ground states of 2-d Ising spin glasses.

Chapter 9. *Incorporating a priori Knowledge in Probabilistic-Model Based Optimization* by Shumeet Baluja.

Complex dependency networks that can account for the interactions between parameters are often used in advanced EDAs; however, they may necessitate enormous amounts of sampling. In this chapter, Baluja demonstrates how a priori knowledge of parameter dependencies, even incomplete knowledge, can be incorporated to efficiently obtain accurate models that account for parameter interdependencies. This is achieved by effectively putting priors on the network structures that are created. These more accurate models yield improved results when used to guide the sample generation process for search. Baluja demonstrates the results on a variety of graph coloring problems, and examines the benefits of a priori knowledge as problem difficulty increases.

Chapter 10. *Multiobjective Estimation of Distribution Algorithms* by Martin Pelikan, Kumara Sastry, and David E. Goldberg.

Many real-world optimization problems contain multiple competing objectives and that is why the design of optimization techniques that can scalably discover an optimal tradeoff between given objectives (Pareto-optimal solutions) represents an important challenge. Pelikan et al. discuss EDAs that address this challenge. The primary focus is on scalability on discrete multiobjective decomposable problems and the multiobjective hierarchical BOA (mohBOA), but other approaches to multiobjective EDAs are also discussed.

Chapter 11. *Effective and Reliable Online Classification Combining XCS with EDA Mechanisms* by Martin Butz, Martin Pelikan, Xavier Llorà, and David E. Goldberg.

Learning Classifier Systems (LCSs), such as XCS and other accuracy-based classifier systems, evolve a distributed problem solution online.

During the learning process, rule quality is assessed iteratively using techniques based on gradient-descent, while the rule structure is evolved using selection and variation operators of evolutionary algorithms. While using standard variation operators suffices for solving some problems, it does not assure an effective evolutionary search in many difficult problems that contain strong interactions between features. Butz et al. describe how advanced EDAs can be integrated into XCS in order to ensure effective exploration even for problems in which features strongly interact and standard variation operators lead to poor XCS performance. In particular, they incorporate the model building and sampling techniques from BOA and ECGA into XCS and show that the two proposed algorithms ensure that the solution is found efficiently and reliably. The results thus suggest that the research on combining standard LCSs with advanced EDAs holds a big promise and represents an important area for future research on LCSs and EDAs.

Chapter 12. *Military Antenna Design Using a Simple Genetic Algorithm and hBOA* by Tian-Li Yu, Scott Santarelli, and David E. Goldberg.
In this chapter, Yu et al. describe the optimization of an antenna design problem via a simple genetic algorithm (SGA) and the hierarchical Bayesian optimization algorithm (hBOA). Three objective functions are designed in an effort to find a solution that meets the system requirements/specifications. Yu et al. show empirical results that indicate that the SGA and hBOA perform comparably when the objective function is "easy" (that is, traditional mask). When the objective function more accurately reflects the true objective of the problem (that is, "difficult"), however, hBOA consistently outperforms the SGA both computationally and the optimal antenna design obtained via hBOA also outperforms that obtained via the SGA.

Chapter 13. *Feature Subset Selection with Hybrids of Filters and Evolutionary Algorithms* by Erick Cantú-Paz.
The performance of classification algorithms is affected by the features used to describe the labeled examples presented to the inducers. Therefore, the problem of feature subset selection has received considerable attention. Approaches to this problem based on evolutionary algorithms typically use the wrapper method, treating the inducer as a black box that is used to evaluate candidate feature subsets. However, the evaluations might take a considerable time and the wrapper approach might be impractical for large data sets. Alternative filter methods use heuristics to select feature subsets from the data and are usually considered more scalable than wrappers to the dimensionality and volume of the data. In this chapter, Cantú-Paz describes hybrids of evolutionary algorithms and filter methods applied to the selection of feature subsets for classification problems. The proposed hybrids are compared against each of their components, two feature selection wrappers that are in wide use, and another filter-wrapper hybrid. Cantú-Paz investigates if the proposed evolutionary

hybrids present advantages over the other methods in terms of accuracy or speed. He uses decision tree and naive Bayes classifiers on public-domain and artificial data sets. The experimental results in this chapter suggest that the evolutionary hybrids usually find compact feature subsets that result in the most accurate classifiers, while beating the execution time of the other wrappers.

Chapter 14. *BOA for Nurse Scheduling* by Jingpeng Li and Uwe Aickelin.
Li and Aickelin have shown that schedules can be built mimicking a human scheduler by using a set of rules that involve domain knowledge. Li and Aickelin present a Bayesian Optimization Algorithm (BOA) for the nurse scheduling problem that chooses such suitable scheduling rules from a set for each nurse's assignment. Based on the idea of using probabilistic models, the BOA builds a Bayesian network for the set of promising solutions and samples these networks to generate new candidate solutions. Computational results from 52 real data instances demonstrate the success of this approach. The authors also suggest that the learning mechanism in the proposed algorithm may be suitable for other scheduling problems.

Chapter 15. *Searching for Ground States of Ising Spin Glasses with Hierarchical BOA and Cluster Exact Approximation* by Martin Pelikan and Alexander K. Hartmann.
In this chapter, Pelikan and Hartmann apply the hierarchical Bayesian optimization algorithm (hBOA) to the problem of finding ground states of Ising spin glasses with $\pm J$ and Gaussian couplings in two and three dimensions. The authors compare the performance of hBOA to that of the simple genetic algorithm (GA) and the univariate marginal distribution algorithm (UMDA). The performance of all tested algorithms is improved by incorporating a deterministic hill climber based on single-bit flips. The results in the chapter show that hBOA significantly outperforms GA and UMDA on a broad spectrum of spin glass instances. The authors also describe and incorporate the cluster exact approximation (CEA) into hBOA and GA to improve their efficiency. The results show that CEA enables all tested algorithms to solve larger spin glass instances and that hBOA significantly outperforms other compared algorithms even in this case.

1.3 Additional Information

1.3.1 Conferences

Most EDA researchers present their results at the following conferences:

- *Congress on Evolutionary Computation (CEC)*; IEEE
- *Genetic and Evolutionary Computation Conference (GECCO)*; SIGEVO, ACM Special Interest Group for Genetic and Evolutionary Computation
- *Parallel Problem Solving from Nature*

1.3.2 Journals

The following journals publish majority of EDA articles:

– *Evolutionary Computation*; MIT Press
– *Genetic Programming and Evolvable Machines*; Springer
– *IEEE Transactions on Evolutionary Computation*; IEEE Press

A number of EDA articles can also be found in the following journals:

– *Computational Optimization and Applications (COAP)*; Kluwer
– *Information Sciences*; Elsevier
– *International Journal of Approximate Reasoning*; Elsevier
– *New Generation Computing*; Springer

1.3.3 World Wide Web

The following search engines can be used to search for many EDA papers:

– CiteSeer, Scientific Literature Digital Library
  `http://citeseer.ist.psu.edu/`
– Google Scholar
  `http://scholar.google.com/`

More papers can be found on personal and institutional Web pages of the researchers that contributed to this book or were cited in the references therein.

1.3.4 Online Source Code

Source code of various EDAs can be downloaded from the following sources:

– Extended Compact Genetic Algorithm, C++; F. G. Lobo, G. R. Harik
  Bayesian Optimization Algorithm, C++; M. Pelikan
  Bayesian Optimization Algorithm with Decision Graphs; M. Pelikan
  `http://www-illigal.ge.uiuc.edu/`
– Learning Factorized Distribution Algorithm (LFDA); H. Mühlenbein, T. Mahnig; `http://www.ais.fraunhofer.de/~muehlen/`
– Adaptive mixed Bayesian optimization algorithm (AMBOA); J. Ocenasek
  `http://jiri.ocenasek.com/`
– Real-coded Bayesian Optimization Algorithm; C.-W. Ahn
  `http://www.evolution.re.kr/`
– Probabilistic Incremental Program Evolution (PIPE); R. P. Salustowicz
  `http://www.idsia.ch/~rafal/`
– Naive Multi-objective Mixture-based Iterated Density-Estimation Evolutionary Algorithm (MIDEA), Normal IDEA-Induced Chromosome Elements Exchanger (ICE), Normal Iterated Density-Estimation Evolutionary Algorithm (IDEA); P. A. N. Bosman
  `http://homepages.cwi.nl/~bosman/`

– Java applets for several real-valued and permutation EDAs; S. Tsutsui
`http://www.hannan-u.ac.jp/~tsutsui/index-e.html`

References

[1] Baluja, S. (1994). Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning. Tech. Rep. No. CMU-CS-94-163, Carnegie Mellon University, Pittsburgh, PA

[2] Buntine, W. L. (1991). Theory refinement of Bayesian networks. *Proceedings of the Uncertainty in Artificial Intelligence (UAI-91)*, pages 52–60

[3] Inza, I., Larrañaga, P., Etxeberria, R., and Sierra, B. (2000). Feature subset selection by Bayesian network-based optimization. *Artificial Intelligence*, 123(1–2):157–184

[4] Jordan, M. I., editor (1999). *Learning in Graphical Models.* MIT, Cambridge, MA

[5] Larrañaga, P. and Lozano, J. A., editors (2002). *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation.* Kluwer, Boston, MA

[6] Li, J. and Aickelin, U. (2003). A Bayesian optimization algorithm for the nurse scheduling problem. *Proceedings of the IEEE Congress on Evolutionary Computation 2003 (CEC-2003)*, pages 2149–2156

[7] Mühlenbein, H. and Mahnig, T. (1999). FDA – A scalable evolutionary algorithm for the optimization of additively decomposed functions. *Evolutionary Computation*, 7(4):353–376

[8] Mühlenbein, H. and Paaß, G. (1996). From recombination of genes to the estimation of distributions I. Binary parameters. *Parallel Problem Solving from Nature*, pages 178–187

[9] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann, San Mateo, CA

[10] Pelikan, M. (2005). *Hierarchical Bayesian optimization algorithm: Toward a new generation of evolutionary algorithms.* Springer, Berlin Heidelberg New York

[11] Pelikan, M., Goldberg, D. E., and Lobo, F. (1999). A survey of optimization by building and using probabilistic models. IlliGAL Report No. 99018, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, IL

[12] Pelikan, M., Sastry, K., and Goldberg, D. E. (2002). Scalability of the Bayesian optimization algorithm. *International Journal of Approximate Reasoning*, 31(3):221–258

[13] Rothlauf, F., Goldberg, D. E., and Heinzl, A. (2000). Bad codings and the utility of well-designed genetic algorithms. IlliGAL Report No. 200007, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, IL

[14] Santarelli, S., Goldberg, D. E., and Yu, T.-L. (2004). Optimization of a constrained feed network for an antenna array using simple and competent genetic algorithm techniques. *Proceedings of the Workshop Military and Security Application of Evolutionary Computation (MSAEC-2004)*

[15] Sastry, K. (2001). Efficient atomic cluster optimization using a hybrid extended compact genetic algorithm with seeded population. IlliGAL Report No. 2001018, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, IL

[16] Simon, H. A. (1968). *The Sciences of the Artificial*. The MIT, Cambridge, MA

Springer

http://www.springer.com/978-3-540-34953-2

Scalable Optimization via Probabilistic Modeling
From Algorithms to Applications
Pelikan, M.; Sastry, K.; Cantú-Paz, E. (Eds.)
2006, XX, 349 p., Hardcover
ISBN: 978-3-540-34953-2