IGR Report for EPSRC Grant GR/R64742/01
A Rigorous Investigation into Estimation of Distribution Algorithms

Qingfu Zhang
Department of Computer Science, University of Essex, Colchester, CO4 3SQ.
email: *qzhang@essex.ac.uk*

1 Background

Estimation of Distribution Algorithms (EDAs) have been recognized as a major paradigm in evolutionary computation. There is no traditional crossover or mutation in EDAs. Instead, they explicitly extract global statistical information from the selected solutions (often called parents) and build a posterior probability distribution model of promising solutions, based on the extracted information. New solutions are sampled from the model thus built and fully or in part replace the old population. Since the dependence relationships in the distribution of the promising solutions are highly relevant to the variable interactions in the problem, EDAs are promising methods for capturing the structure of variable interactions, identifying and manipulating crucial building blocks, and hence efficiently solving hard optimization and search problems with interactions among the variables. Many EDA-like algorithms have been developed for various optimization and search problems in recent years. Instances of EDAs include Population-Based Incremental Learning (PBIL), Univariate Marginal Distribution Algorithm (UMDA), Mutual Information Maximization for Input Clustering (MIMIC), Combining Optimizers with Mutual Information Trees (COMIT), Factorized Distribution Algorithm (FDA), Bayesian Optimization Algorithm(BOA), Bayesian Evolutionary Algorithm (BEA), and Global Search Based on Reinforcement Learning Agents (GSBRL), to name a few [1].

Relatively little effort has been devoted to studying the working mechanisms of EDAs. Mühlenbein [3], González et al. [2] and Höhfeld & Rudolph [4] have studied the behaviours of UMDA and PBIL (the simplest versions of the EDA, which ignore all the variable interactions). Their results show that these algorithms can locate the optimum of a linear function but cannot solve problems with nonlinear variable interactions. In [5], Mühlenbein and Mahning discussed the convergence of FDA (Factorized Distribution Algorithm) for separable additively decomposable functions (ADFs). Since there are no overlaps in their objective functions, their FDA is equivalent to UMDA. Therefore, their work does not deal with the ability of FDA to solve problems with variable interactions. The theoretical study of the ability of EDAs for dealing with variable interactions is urgently needed in order to obtain a deeper understanding of EDAs.

Since it is impractical to calculate the actual posterior distribution of the promising solutions, most of the existing EDA-like algorithms model the distribution functions by probabilistic graph models or Bayesian networks. These algorithms can only take into account some selected dependence relationships that satisfy the triangulation constraints. This inherent shortcoming severely limits the ability of the algorithms in solving hard problems with other interaction structures. Besides, EDAs mainly explore the search space by random sampling from probability models. Therefore, EDAs in themselves are often very computationally expensive. Most, if not all, researchers currently use conventional local search techniques in EDAs for overcoming these shortcomings. Any other systematic methods for improving the performance of EDAs should benefit the applications of EDAs.

The first part of this project is on the theory of EDAs. We have theoretically studied the behaviours of two typical EDAs: UMDA and FDA, under widely-used selection schemes. These two algorithms can also be regarded as instances of ant colony optimization methods. We have shown that it is necessary and sufficient, in terms of convergence, to consider some selected crucial dependence relationships in EDAs for optimization of additively decomposable functions. These theoretical results provide a real insight into the working mechanisms of EDAs.

In the second part of this project, we have developed an efficient implementation of an EDA for global optimization, which is based on experimental techniques and a novel strategy for using two different local search techniques in EDAs. We have also proposed a new EDA operator, called guided mutation, for generating offspring through combination of global statistical information and the local information of solutions found so far. It usually provides better solutions than the random sampling approach used in the existing implementation of EDAs. We have developed an operator, called guided jump. Guided jump is based on EDA ideas. It collects statistical information about features during the search and then uses this information to help the search move away from the area it has exploited. We have also proposed using the guided local search in EDAs. These proposed approaches have been used for solving the quadratic assignment problem and the maximum clique problem.

2 Key Advances and Supporting Methodology

2.1 Theory of Estimation of Distribution Algorithms

Like other evolutionary algorithms, EDAs with finite populations can be regarded as stochastic processes. Since EDAs maintain a probability model and a population of solutions at each generation, their transition probability is very difficult to formulate and analyze. Therefore, it is very hard to directly apply stochastic process theory to the study of EDAs. In this project, we assume that the population size is infinite. Then an EDA can be modelled as a deterministic sequence. The study of this sequence will not only be a vehicle for understanding the behaviour of the EDA with a large population, but also a prerequisite for studying the EDA stochastic process.

2.1.1 Models of EDAs with infinite populations

Let $Pop(t)$ and $Pop^s(t)$ be the population and the parent set, respectively, at generation t in an EDA. Let the underlying probability distribution functions for the points in $Pop(t)$ and $Pop^s(t)$ be $p(x,t)$ and $p^s(x,t)$, respectively. By the famous Glivenko-Canteli theorem, the empirical probability density functions induced by points in $Pop(t)$ and $Pop^s(t)$ will converge to $p(x,t)$ and $p^s(x,t)$ respectively, as the sizes of $Pop(t)$ and $Pop^s(t)$ tend to infinity. Therefore, $p(x,t)$ and $p^s(x,t)$ can be thought of as the population and the parent set at time t in EDA with infinite populations, respectively. Correspondingly, EDA with infinite population can be modelled as:
Step 1: Selection $p(x,t) \to p^s(x,t)$
Step 2: Variation $p^s(x,t) \to p(x,t+1)$
The above iteration of probability functions is called the limit model of EDA. An EDA limit model defines a deterministic sequence of probability functions. We have focused on the dynamic properties of this sequence.

2.1.2 Fixed Points and Stability

In the discrete search space, any probability model can be expressed by finite real parameters. Therefore, we can define a distance between any two probability models. As a result, we obtain a metric space containing all the probability models under consideration. Therefore, an EDA limit model can be considered as a nonlinear mapping in such a metric space. We can locate the fixed points of this nonlinear mapping and study the stability of these fixed points. We can also study the relationship between these fixed points and the optimal solutions for the optimization problem.

Intuitively, if a point in a sequence generated by iterating a nonlinear mapping is sufficiently close to a stable fixed point, then all the points thereafter remain close to this stable fixed point. It is stable to all small perturbations. If a fixed point is not stable, no matter how close to it the initial point is, some points thereafter may become far from it. For an asymptotically stable fixed point, if a point in the sequence is sufficiently close to it then the sequence will converge to it. The convergence is stable to small perturbations.

2.1.3 UMDA and Hyper-Cube ACO

In UMDA, all the variables in the probability model $p(x,t)$ are independent. Originally, UMDA was proposed as an instance of EDAs. We noticed that UMDA can also be regarded as an instance of Ant colony optimization (ACO) meta-heuristics. In fact, UMDA has the same limit model as the so-called Hyper-Cube ACO [6] when the search space is $\{0,1\}^n$. We have proved that the UMDA limit model with 2-tournament selection could become stuck at any local optimal point [8]. That means that UMDA or Hyper-Cube ACO in itself has little chance of finding the global optimum if there are many local optima. Therefore, we have to modify these algorithms or hybridize them with other techniques to solve complex problems in applications.

2.1.4 FDA

Most existing EDAs select some dependence relationships (i.e., multi-variable marginal probabilities of $p^s(x,t)$) to construct $p(x,t+1)$. To the best of our knowledge, no work on the convergence of these algorithms has been carried out so far. The factorized distribution algorithm (FDA) chooses a graphical model for building $p(x,t+1)$ based on the prior knowledge of the structure of the objective function.

- FDA with Proportional Selection [9]

We have proposed a generalized proportional selection (GPS) and proved that FDA with GPS converges globally for the optimization of continuous additively decomposable functions with overlap. Since the conventional proportional selection is a special case of GPS, FDA converges under conventional proportional selection. This proof is based on the theory of statistical graphical models. This work was started when the PI was a postdoctoral research fellow in GMD. It has been generalized and completed under this grant.

- FDA with Truncation Selection [7]

Truncation selection ranks all the points in the current population according to their objective function values and selects the best ones as parents. This selection is very popular in implementations of EDAs. We have proved that FDA for optimization of continuous additively decomposable functions with overlap converges to the global optimum.

- FDA with Tournament Selection [8]

In K-tournament selection, K points are randomly chosen from the current population and the best point is selected to be a parent. $K = 2$ is a typical value used in many applications. For this reason, we consider 2-tournament selection in our study. We have investigated the stability of the limit model of FDA and proved that FDA will not be trapped on some local optima, but that UMDA may be trapped on all the local optima. Therefore, FDA has a better chance of locating the global optimum than UMDA for general objective functions. For a class of additively decomposable functions, we have proved that FDA can locate the global optimum if the dependence structure used in FDA matches the variable interaction structure in the problem.

2.2 Implementation of Estimation of Distribution Algorithms

EDAs in themselves have difficulty in solving complex optimization and search problems. In this project, we have proposed several systematic ways to improve the performance of EDAs.

2.2.1 Combination of EDA, Experimental Design Technique and Local Search [10]

We have developed a hybrid EDAs for the global continuous optimization problem. This algorithm maintains and improves a population of solutions in the feasible region. Initial candidate solutions are generated by uniform design techniques. These solutions scatter more evenly over the feasible solution region than points generated randomly. To generate a new population, a marginal histogram model is built, based on the global statistical information extracted from the current population, and then new solutions are sampled from the model thus built. An incomplete simplex method is applied to every new solution generated by uniform design or sampled from the histogram model. UOBDQA (unconstrained optimisation by diagonal quadratic approximation) is applied to several selected solutions at each generation. The incomplete simplex method is only allowed to run for $O(n)$ steps and its role is to help to locate promising solutions more accurately while UOBDQA is used to find good local minima and hopefully the global minima. We have studied the effectiveness of the main components of this algorithm on many well-known test problems. The experimental results show that our algorithm is better than four other recent evolutionary algorithms in terms of the solution quality and the computational cost.

2.2.2 Guided Mutation

One of the key issues in the design of evolutionary algorithms is how to generate offspring. Crossover combines parts of parent solutions for producing offspring while mutation changes part of a parent solution randomly. These operators mainly utilize information about the locations of the parent solutions involved. EDAs generate offspring based on the global statistical information collected from the search so far. However, the local information of solutions found so far has not directly been used in EDAs. We have proposed a new operator, called guided mutation, which combines the global statistical information and local information of solutions found so far to generate offspring. In guided mutation, one part of the offspring is directly copied from its parent, while the other part is sampled from a probability distribution model of promising solutions. It can be expected that the resultant offspring should fall in a promising area. Sampling also provides diversity for the search. Meanwhile, the offspring will not be very different from its parent, which usually is among the best solutions found so far. Guided mutation can have various implementations, depending on the topology of the search space and the probability model it uses. We have developed implementations of guided mutation for binary and permutation search spaces.

2.2.3 Guided Jump: Restart Strategy in Heuristics

In heuristics for search and optimization problems, the search may be trapped in a small area (which may contain several local optima) after a number of generations. Inspired by ideas from EDAs and guided local search, we have developed a new strategy called guided jump for helping a heuristic algorithm to jump out of the area in which it has been trapped. In guided jump, a probability model is learned by an EDA approach for characterizing the area the search has visited. In a sense, this model gives the probabilities that some features should be avoided in new solutions. The model is used to generate new starting points for the further search when the search has been trapped.

2.2.4 Hybrid EDA algorithm for the Quadratic Assignment Problem [13]

The Quadratic Assignment Problem (QAP) is a combinatorial optimization problem which arises in many applications such as facility location, scheduling, manufacturing and statistical data analysis. In the QAP, n departments have to be assigned to n locations such that the cost of the assignment, depending on the distances between the locations and the flows of materials between the departments, is minimized. In the proposed hybrid EDA, the guided local search (GLS) is applied to every new solution generated by guided mutation. The guided jump is employed to help the search move out of the area in which it has been trapped. The experiment results show that the proposed EDA outperforms guided local search.

2.2.5 Hybrid EDA algorithm for the Maximum Clique Problem [14]

The maximum clique problem (MCP) is to find a maximum clique for a given graph. The MCP is one of the best known problems of combinatorial optimization. In many applications such as coding theory, computer vision and mobile networks, the underlying problem can be formulated as an instance of the MCP. The MCP is NP-complete. Moreover, there is no $n^{1-\epsilon}$ approximation algorithm for the MCP unless $P = NP$. These facts indicate that the MCP is very hard to solve. To further study the performance of our new hybrid EDA method, we have applied it for the MCP where we have employed a very simple repair operator, instead of guided local search, to repair every solution generated by the guided mutation. Experimental results show that our algorithm outperforms the best evolutionary algorithm reported so far on 37 DIMACS benchmark graphs

2.2.6 Combination of EDA and Differential Evolution [15]

Differential Evolution (DE) is a novel evolutionary algorithm which has achieved great success in solving the global continuous optimization problem. It mainly uses the distance and direction information from the current population to guide the further search. We have observed that DE does not make the best use of the global information in the search. We have proposed a combination of DE and EDA (DE/EDA) for the global continuous optimisation problem. DE/EDA combines global information extracted by EDA with local information obtained by DE to create promising solutions. We have compared DE/EDA with DE and EDA on several famous test functions. Experimental results demonstrate that DE/EDA outperforms DE and EDA.

2.2.7 Heuristic for Maximin Design Problems [11] [16]

In experimental design, if the experimental region is a bounded rectangle, Latin hypercube sampling and orthogonal array sampling can generate space-filling designs. However these samplings are of very limited use when, as is often the case in EDAs and other applications, the experimental region is not a rectangle. Maximin designs explicitly spread the design points (sampling points) as much as possible in a region. We have designed a population-based heuristic for maximin design in a non-rectangular area. This heuristic uses a problem-specific combination operator to generate new designs and uses a local search to improve them. The experimental results shows that this heuristic is very promising.

3 Research Plan Review

We have achieved all the original objectives stated in the proposal. We have also been able to apply our hybrid EDAs for solving the MCP, which was not planned in the original proposal. The reason for doing so is that we wished to verify the usefulness of our methodology for different search and optimization problems. In the very early stage of this project, we realized that UMDA and hycube-ACO have the same limit model (in fact, there is no difference between ACO and EDAs in principle). Then we studied the behaviour of the limit model of UMDA and Hycube-ACO for accomplishing objective 2 set in the original plan. During our study on EDA+Experimental Design+Local Search for global optimization problems, it became quite clear that an efficient implementation of such an algorithm may need to locate n points on an area which is not a rectangle, this problem turns out to be an experimental design problem which has just received the attention from statistician. We have developed a heuristic for solving the maximin distance design in non-rectangular spaces, and a combination of DE and EDA for the global optimisation problem, which are out of the original plan.

4 Research Impact and Benefits to Society

We have made some contributions to the theory of estimation of distribution algorithms. Such results will be helpful to researchers and practitioners in understanding the capabilities of EDAs and thus enable them to apply or modify them for solving real-world problems. There are still many unsolved questions in EDAs, but our research should pave the way for further study in the theory.

Combinations of different techniques are often needed for solving hard real-world problems. We have proposed a very efficient combination of EDAs, experimental design and local search for the global continuous optimization problem. We have also proposed a hybrid algorithm which uses guided mutation, guided jump and guided local search/conventional local search for the QAP and MCP. These algorithms are easily modifiable for solving other optimization and search problems. We believe that these algorithms will benefit practitioners in designing practical algorithms.

5 Explanation of Expenditure

Due to visa and passport problems, the RA, Mr. Jianyong Sun arrived later than planned. The duration of his contract was shortened to 17 months and the resultant saving on the RA cost was used for hiring Mr. Hui Li as a part-time Research Assistant.

6 Dissemination Activities or Further Research

Two journal papers on the theory of EDAs have been accepted for publication in *IEEE Transactions on Evolutionary Computation*, one journal paper on the theory of EDAs has been accepted for publication in *Complexity*, one journal paper on the hybrid EDA for global optimisation has been accepted for publication in *Engineering Computations*, 2 journal papers on implementations have been submitted to *Evolutionary Computation* and *Journal of Information Science*. 4 conference papers on implementations have been published in 2002 UK Workshop on Computational Intelligence, 2003 Genetic and Evolutionary Computation Conference, 2003 International Conference of Metaheuristic and XVI Conference of the European Chapter on Combinatorial Optimisation.

The PI has given a seminar on EDAs at Xian Jiangtong University, Xidian Universiry and Jiangsu University in China. We are now collaborating with researchers from Nanyang Technical University in Singapore to apply our hybrid methods for solving hard routing problems in telecommunications. The preliminary results are very promising. We are also discussing with researchers from Leeds University about the possibility of combining our EDA algorithms with their heuristics for solving port scheduling problems.

EDAs are still in their infancy. Our theoretical results are on EDAs with infinite populations. It is very urgent to study EDAs with finite populations. The theory of stochastic stability should be useful for such a study, since EDAs can be regarded as members of stochastic adaptive algorithms. EDAs use probability models to express knowledge learned from the search. Exploring other machine learning and knowledge discovery techniques for designing search methods should be a very interesting research topic.

References

[1] P. Larranaga and J. A. Lozano, *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*, Kluwer Academic Publishers, Boston, 2001.

[2] C. González, J.A. Lozano and P. Larrañaga. *Analyzing the PBIL Algorithm by Means of Discrete Dynamical Systems,* Complex Systems. vol. 12, pp. 465-479, 2000.

[3] H. Mühlenbein, *The Equation for Response to Selection and Its Use for Prediction,* Evolutionary Computation, vol. 5, pp. 303-346, 1998.

[4] M. Höhfeld and G. Rudolph, *Towards a Theory of Population-Based Incremental Learning*, Proceedings of the 4th IEEE Conference on Evolutionary Computation, Indianapolis, IEEE Press, 1997, pp. 1-5.

[5] H. Mühlenbein and T. Mahning, *Convergence Theory and Application of the Factorized Distribution Algorithm*, Journal of Computing and Information Technology, vol. 7, pp. 19-32, 1999.

[6] M. Zlochin, M. Birattari, N. Meuleau and M. Dorigo, *Model-based search for combinatorial optimization.* Technical Report TR/IRIDIA, 2002.

[7] Q. Zhang, *On the Convergence of a Factorized Distribution Algorithm with Truncation Selection*, Complexity, Accepted for publication, 2003.

[8] Q. Zhang, *On Stability of Fixed Points of Limit Models of Univariate Marginal Distribution Algorithm and Factorized Distribution Algorithm*, IEEE Trans. on Evolutionary Computation, Accepted for publication, 2003.

[9] Q. Zhang, H. Muehlenbein, *On the Convergence of a Class of Estimation of Distribution Algorithms*, IEEE Trans. on Evolutionary Computation, Accepted for publication, 2003.

[10] Q. Zhang, J. Sun, E. Tsang and J. Ford, *Hybrid Estimation of Distribution Algorithm for Global Optimsation*, Engineering Computations, Accepted for publication, 2003.

[11] Q. Zhang, H. Li and E.P. Tsang, *Estimation of Distribution Algorithm with Local Search for Approximating the Discrepancy of a Finite Set of Points*, Abstract, XVI Conference of the European Chapter on Combinatorial Optimisation, Molde, Norway, 2003.

[12] Q. Zhang, J. Sun, E.P.K. Tsang and J. A. Ford, *Estimation of Distribution Algorithm Based on Mixture: Preliminary Experimental Results*, Proceedings of UKCI-02, pp. 251-257, 2002.

[13] Q. Zhang, J. Sun, E. P. K. Tsang and J. Ford, *Combination of Guided Local Search and Estimation of Distribution Algorithm for Solving Quadratic Assignment Problem,* Proceedings of the Bird of a Feather Workshops, Genetic and Evolutionary Computation Conference, pp.42 - 48, 2003.

[14] Q. Zhang, J. Sun and E. P. K. Tsang, *Applying Estimation of Distribution Algorithm to the Maximum Clique Problem*, Abstract, International Conference of Metaheuristic, 2003. The full paper has been submitted to Evolutionary Computation, 2003.

[15] J. Sun, Q. Zhang and E. Tsang, *DE/EDA, A New Evolutionary Algorithm for Global Optimisation,* submitted to Journal of Information Science, 2003.

[16] L. Duan and Q. Zhang, *A Population-Based Heuristic for Maxmin Problem*, 2003, in preparation.