On Estimation Distribution Algorithms

Roberto Santana, Alberto Ochoa
Center of Mathematics and Theoretical Physics, ICIMAF,
Calle 15 e/ C y D, Vedado, Havana, Cuba
Rsantana@cidet.icmf.inf.cu, Ochoa@cidet.icmf.inf.cu

Estimation Distribution Algorithms(EDA)[3] are Evolutionary Algorithms that do not use crossover or mutation operators, instead, they estimate the probability distribution of points in the selected set, and use this information to generate new points. Recent developments in EDA [4][6][7] have opened a new promising way for the creation of more efficient evolutionary algorithms.

The main goal of the dissertation research is the design of efficient Estimation Distribution Algorithms based on non simply connected probabilistic networks. This work is part of a more ambitious ongoing research on Low Cost Evolutionary Algorithms (LCEA)[5]. Other important areas to be covered with the dissertation are: 1) The integration of different theories about GA in order to explain successful results of EDA in the context of evolutionary computation. 2) Define ways of incorporating knowledge about the problem domain, prior to the beginning and during the function optimization . 3) To study the recognized sources of hardness for GA's in the framework of EDA. 4) Define measures to evaluate the quality of the different search strategies used. We now describe some aspect of the current research in these topics.

The construction of desired solutions via the beneficial mixing or exchange of building blocks has been strongly related to the question of linkage learning. In the theory of EDA the linkage learning issue is related with the factorization problem. A correct factorization would avoid the disruption of building blocks. Useful factorizations can be extracted from prior information about the problem[4]. Factorizations can be also identified along the evolution, in our experiments we have already used an algorithm able to do that[6]. Nevertheless the questions of what makes a certain factorization suitable, the relation between the selection operator and the factorizations, and the ways of making factorizations less costly in computational terms have not been clarified yet. We analyze these questions in our dissertation.

Although our approach is primarily based on the theory of graphical models, we also analyze other approaches to the dynamics of GA's. Particularly we analyze the possible use of majorization and replicators[2] associated to discrete probability distributions. We point out the links existing between the question of incorporating information about the problem domain with research done in the field of bayesian inference. In our consideration Graphical models support a powerful framework for analyzing the sources of hardness[1] for GA's, noise within and between schema partitions are being considered using marginal distributions. Furthermore, we have defined measures based on Population Genetics to evaluate the quality of the search strategy and to compare the adequacy of different factorizations to explore and exploit the space of solutions[8].

Bibliography

[1] Goldberg, D. E., Deb, K. , & Clark, J. H. (1992). Genetic Algorithms, noise and the sizing of populations. Complex Systems, 6, 333-362.
[2] Menon A., Mehrotra, K., Mohan C., and Ranka, S. (1997) Replicators, majorization and genetic algorithm: New models, connections and analytical tools. In Foundation of Genetic Algorithms - 4, pp. 155-180. Morgan Kaufman.
[3] Mühlenbein, H. (1998). The equation for response to selection and its use for prediction, Evolutionary Computation 5, pp. 303--346.
[4] Mühlenbein, H.; Mahnig, T., Ochoa A. (1998). Schemata, Distributions and Graphical Models in Evolutionary Optimization, to appear in Journal of Heuristics Vol 5, No. 2. Also at http://set.gmd.de/AS/ga.publineu.html#index.
[5] Ochoa A. (1997), How to deal with costly fitness functions in evolutionary computation. In Proceedings of the 13th ISPE/IEE International Conference on CAD/CAM, Robotics & Factories of the Future. pp. 788-793. Pereira.
[6] Ochoa A., Soto M., Santana R., Madera J. C., Jorge N. (1999). The Factorized Distribution Algorithm and The Junction Tree: A Learning Perspective. Proceedings of the II Second Symposium on Artificial Intelligence. ISAS' 99 pp.
[7] Pelikan, M.; Goldberg D. E., Cantú-Paz E.(1998). Linkage problem, Distribution estimation and Bayesian Estimation Networks. Illegal Report 98013. November 1998.

[8] Santana, R. (1999) Towards an intelligent genetic search: Defining measures of convergence. Submitted to ECCAI 99.