Estimation of Distribution Algorithms

Petr Pošík

Czech Technical University, Faculty of Electrical Engineering, Dept. of Cybernetics
Technická 2, 166 27 Prague 6, Czech Republic
E-mail: posik@labe.felk.cvut.cz, phone: +420-2-24357228

1 Introduction

Evolutionary algorithms (EAs) are known in many areas as a powerful and robust optimization and searching tool. Classical EAs rely on the well-known two phases: selection and variation. Variation is usually carried out by means of perturbation of promissing individuals (searching local neigbourhoods), or by means of combining two promising individuals together (creating offsprings which embody some characteristics of both parents). However, classical EAs suffer from several problems. The linkage problem belongs among the most severe ones. It arises in situations when the individual components of chromosomes are not statistically independent of each other with respect to the fitness function. There exists no general way of EA modification that would enable the modified EA to account for the dependencies at hand. Usually, this problem is solved by constructing special crossover and mutation types of operators and by incorporating some problem-specific knowledge in them. The classical EA then looses its flavor of general problem solver and quickly becomes an algorithm highly specialized to the given problem.

2 Estimation of Distribution Algorithms

Recently, a new type of EAs emerged — *Estimation of Distribution Algorithms* (EDAs) [1]. Some researchers use names as *Probabilistic Model Building Genetic Algorithms* (PMBGAs), or *Iterated Density Estimation Algorithms* (IDEAs), but all these names describe basically the same concept. These algorithms don't rely on the 'genetic' principles anymore; instead, in each generation, they build an explicit probabilistic model of distribution of 'good' individuals in the search space. New individuals are created by sampling from this distribution. The *model-sample* step of EDA can be thought of as a generalized type of multiparent crossover operator. The strengths and weaknesses of a particular EDA are mainly determined by the used probabilistic model.

2.1 Probabilistic Models for Discrete Variables

The probabilistic models differ for EDAs in discrete and continuous spaces. The first EDAs were developed for the discrete spaces. They range from simple *Univariate Marginal Density Algorithm* (UMDA), which is comparable to simple genetic algorithm, to *Bayesian Optimization Algorithm* (BOA) [2] which uses Bayesian net as the underlying probabilistic model. Bayesian nets are able to encode general type of discrete probabilistic distribution, however, their learning from data involves either sophisticated methods for statistical dependency detection, or they are learnt by searching the space of possible Bayesian nets (usually by a greedy algorithm).

2.2 Probabilistic Models for Continuous Variables

In continuous spaces, the situation is even more complicated. The simplest continuous EDAs (continuous UMDAs) use models in which the joint probability density function (PDF) is factorized into a product of marginal univariate PDFs which take various forms: empirical histograms, normal (or any other well-known) distribution, finite mixtures of univariate Gaussians, etc. To take into account the dependencies between variables, we have to employ more complex models like Gaussian nets (GN), which results in *Estimation of Gaussian Networks Algorithm* (EGNA) [1]. GN has the power to encode general multi-dimensional Gaussian distribution, however, very often this type of probabilistic model is not sufficient. Then we should use even more flexible models which are empowered by (hard- or soft-) clustering, e.g. finite mixture of multidimensional Gaussians. To be objective, one must say that these models are capable in covering various types of interactions, however, learning them is not a trivial task. It is usually very time consuming and it must be performed using a kind of iterative learning scheme (usually by a variant of the expectation-maximization algorithm).

3 Original Contribution

My own research is aimed at the EDAs in continuous spaces. I have successfully implemented the UMDA in continuous domain. I compared the suitability of four different marginal probability models, namely the equi-width histogram (HEW), equi-height histogram (HEH), max-diff histogram (HMD), and univariate mixture of Gaussians (MOG), on a suite of test functions [3]. From the experiments the following conclusion can be made: the HEW model is the least flexible one and the behaviour of EDAs with this model is unsatisfactory in comparison with the other models. The performance of HEH and HMD histograms was comparable. The MOG model showed a bit worse performance, however, it used considerably less components than the histogram models and offers other advantages over the histogram models (easy extension to mixture of multidimensional Gaussians). Typical tracks of evolution of bin boundaries for histogram models and component centers of MOG for one of the test functions shows figure 1.

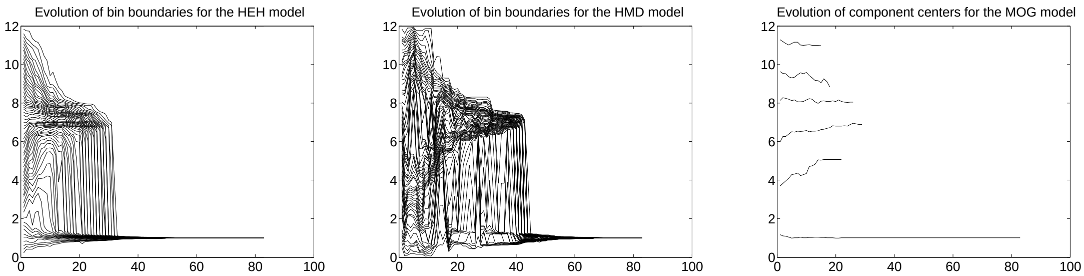

Figure 1: Two Peaks function — Evolution of bin boundaries for equi-height and max-diff histogram models and evolution of component centers for mixture of Gaussians model.

The above described algorithm was succesfully applied to vestibulo-ocular reflex (VOR) signal processing. By analyzing the VOR signal, physicians can recognize some pathologies of the balance system of a pacient in a non-invasive way. The principle is simple: the patient is situated in a chair which is then rotated in a defined way (following some reference signal – sine wave or sum of sine waves). The patient is said to visually track some points on surrounding walls and the movements of his eyes are monitored. The resulting eye signal must be first processed (it is distorted by the fast eye movements) to get 'eye-filtered' response to the reference signal. The differences in amplitudes and phases of the sine waves are the indicators of the balance system pathologies. EDA was applied in the signal processing phase in a co-evolutionary manner. It reaches much more accurate results (in terms of mean squared error) than conventionally used methods based on some form of interpolation.

In the near future, I would like to implement an EDA using mixture of principal component analysers (MPCA) and test it on several artificial and practical problems (e.g. on Hough's transformation used in image processing, or for hidden Markov models training). The aim of these comparative studies is to find out if it is worth to use such complex models (e.g. MPCA), in other words, if the time spent on learning the model each generation is lower than the time the simple EA needs to find a solution of comparable quality.

Acknowledgements

The project was supported by the Grant agency of the Czech Rep. with the grant No. GACR 102/02/0132 entitled "Use of Genetic Principles in Evolutionary Algorithms", and was supervised by Jiří Lažanský.

References

[1] Larrañaga, P., Lozano, J.A. *Estimation of Distribution Algorithms.* Kluwer Academic Publishers, 2001.

[2] Pelikan, M. *Bayesian Optimization Algorithm: From Single Level to Hierarchy.* IlliGAL Report No. 2002023. University of Illinois at Urbana-Champaign. 2002.

[3] Pošík, P. *Optimization of Evolutionary Algorithms.* Technical Report GL 164-03, Department of Cybernetics, Czech Technical University in Prague, 2003.