More Concise and Robust Linkage Learning by Filtering and Combining Linkage Hierarchies[1]

Peter A. N. Bosman[a] D. Thierens[b]

[a] *Centrum Wiskunde & Informatica, Amsterdam, The Netherlands, Peter.Bosman@cwi.nl*
[b] *Utrecht University, Utrecht, The Netherlands, Dirk.Thierens@cs.uu.nl*

1 Introduction

Exploiting a problem's structure to arrive at the most efficient optimization algorithm is key in many optimization disciplines. In evolutionary computation, especially for solving discrete optimization problems from a black-box optimization (BBO) perspective, linkage learning is an important research line because if important linkages are disrupted during variation, optimization will not proceed efficiently [4].

Estimation-of-distribution algorithms (EDAs) are well-known for building and using models to exploit problem structure [2, 3]. Models in EDAs represent probability distributions and linkage information is processed via probabilistic dependency relations within these distributions. Although EDAs can be very powerful, estimating complete distributions might be more than what is required to respect important linkage relations. Here, we therefore consider the class of Genepool Optimal Mixing Evolutionary Algorithms (GOMEAs) as they exploit linkage information by integrating greedy local search, genetic recombination and fitness-based selection [1] based on linkage models, which can typically be learned more efficiently.

Recent results indicate that the use of hierarchical linkage models in GOMEAs leads to the best performance [1]. There are, arguably, however still potential inefficiencies. In this paper, we consider ways to filter these out. We further consider a way to combine the strengths of different linkage models.

2 GOMEA

GOMEA uses a population $\mathcal{P}$ of n solutions. Every generation, a selection $\mathcal{S}$ of n solutions is created by tournament selection (with tournament size 2). Linkage learning is then performed on $\mathcal{S}$, after which $\mathcal{S}$ is discarded. Through variation a set of offspring $\mathcal{O}$ of n solutions is generated that ultimately replaces $\mathcal{P}$.

The general linkage model in GOMEA is the Family Of Subsets (FOS). A FOS $\mathcal{F}$ contains subsets of a set L, i.e. it is a subset of the powerset of L: $\mathcal{F} \subseteq \mathcal{P}(L)$. In GOMEA, L contains all indices of variables, i.e. $L = \{0, 1, \dots, l-1\}$. A FOS $\mathcal{F}$ can be written as $\mathcal{F} = \{\boldsymbol{F}^0, \boldsymbol{F}^1, \dots, \boldsymbol{F}^{|\mathcal{F}|-1}\}$ where $\boldsymbol{F}^i \subseteq \{0, 1, \dots, l-1\}, i \in \{0, 1, \dots, |\mathcal{F}|-1\}$.

Variation in GOMEA is called Genepool Optimal Mixing (GOM). GOM is applied to each solution in the population. A solution is first cloned. Then, each linkage set $\boldsymbol{F}^i \in \mathcal{F}$ is iteratively considered. For each $\boldsymbol{F}^i$, a parent solution is randomly picked from the population and if the parents' values for the variables in $\boldsymbol{F}^i$ differ from those in the current solution, these values in the current solution are overwritten. The change is accepted in case this leads to an improvement or equal quality and is undone otherwise.

3 Filtering and Combining Linkage Hierarchies

The use of hierarchical linkage models in GOMEA has so far been found to be the most robust and to be highly efficient on a variety of problems [1]. There are, arguably, however still potential inefficiencies. Hierarchical models contain linkage hierarchies, i.e. linkage sets $\boldsymbol{F}^i$ and $\boldsymbol{F}^j$ exist such that $\boldsymbol{F}^i \subset \boldsymbol{F}^j$. We focus on such relations here in order to remove linkage sets that may be superfluous.

[1] The full version of this paper is in the proceedings of the *Genetic and Evolutionary Computation Conference (GECCO-2013).*

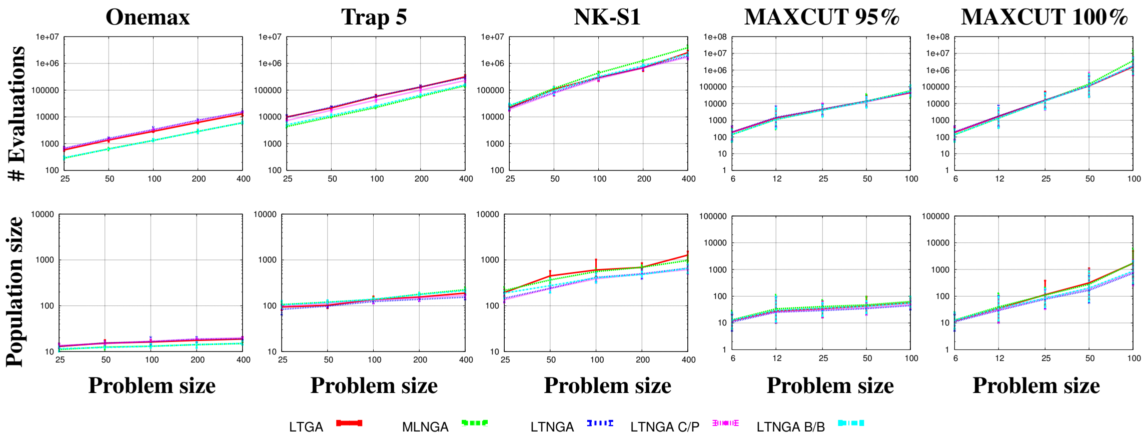

Figure 1: Scalability of LTGA, MLNGA and LTNGA variants on all problems.

To remove a parent linkage set $\boldsymbol{F}^j$, we use the well-known likelihood-ratio statistical hypothesis test. We filter out $\boldsymbol{F}^j$ if for every $X \in \boldsymbol{F}^i$ and every $Y \in (\boldsymbol{F}^j \setminus \boldsymbol{F}^i)$, X and Y are tested to be independent.

To remove a child linkage set $\boldsymbol{F}^i$, we consider the linkage strength (LS), which we define to be the average mutual information between all pairs of variables in that set. If $\text{LS}(\boldsymbol{F}^j)$ is larger than $\text{LS}(\boldsymbol{F}^i)$, this means that on average, the linkages between variables in $\boldsymbol{F}^j$ are stronger than the linkages between variables in $\boldsymbol{F}^i$. Thus, $\boldsymbol{F}^j$ could be said to be a bigger, more interesting, building block and $\boldsymbol{F}^i$ can be disregarded.

Different hierarchical linkage models (FOS instances) model have been used with GOMEA. Here, we consider the linkage tree (LT) and the multiscale linkage neighbors (MLN) model. A natural question is how the best of both worlds can be captured. The most straightforward approach is to combine the LT and MLN models, i.e. learn a (filtered) LT model, learn a (filtered) MLN model and take their union. A key question is whether such an approach is indeed as viable as the generality of the FOS model appears to allow it to be.

4 Some Results and Conclusions

Figure 1 shows the scalability of GOMEA variants on various optimization problems of varying complexity (details omitted in this abstract). Results show the use of either the LT model (LTGA), the MLN model (MLNGA), the combination model (LTNGA) and filtered versions of the latter model (LTNGA B/B uses all filters). All results are averaged over 100 runs and are for the minimally required population size to solve the problem at hand reliably (99/100 times). When all filters are in place, the combination model performs similar to the best of LTGA and MLNGA on all problems.

In the full version of this paper more in-depth comparisons are presented, also on other MAXCUT instances that lead to a clear performance difference between the LT model and the MLN model when used in GOMEA with the combination model LTN performing as good as the best of these two models. LTNGA thereby can be considered to be more robust. The filtering techniques we proposed work well on the considered problems and make the linkage models more concise without negatively affecting performance. The resulting algorithm, LTNGA, has state-of-the-art performance in terms of number of required evaluations and requires very little time to run, even for large problems, contrary to most state-of-the-art EDAs.

References

[1] P.A.N. Bosman and D. Thierens. Linkage neighbors, optimal mixing and forced improvements in genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference — GECCO–2012*, pages 585–592, New York, New York, 2012. ACM Press.

[2] J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea. *Towards a New Evolutionary Computation. Advances in Estimation of Distribution Algorithms.* Springer–Verlag, Berlin, 2006.

[3] M. Pelikan, K. Sastry, and E. Cantú-Paz. *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications.* Springer–Verlag, Berlin, 2006.

[4] D. Thierens and D. E. Goldberg. Mixing in genetic algorithms. In *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 38–47, San Francisco, CA, 1993. Morgan Kaufmann Publishers Inc.