On the Usefulness of Linkage Processing for solving MAX-SAT¹

Krzysztof L. Sadowski^a Peter A.N. Bosman^b Dirk Thierens^a

^a Utrecht University, Utrecht, The Netherlands ^b Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

1 Introduction

Mixing of partial solutions is a key mechanism used for creating new solutions in many Genetic Algorithms (GAs). However, this mixing can be disruptive and generate improved solutions inefficiently. Exploring a problem's structure can help in establishing less disruptive operators, leading to more efficient mixing. One way of using a problem's structure is to consider variable linkage information.

This paper focuses on exploring different methods of building family of subsets (FOS) linkage models, which are then used with the Gene-pool Optimal Mixing Evolutionary Algorithm (GOMEA) to solve MAX-SAT problems. The GOMEA framework provides an efficient mechanism for mixing partial solutions and generating new candidate solutions, given a FOS linkage model [1]. We wish to examine if learning variable linkage information, and representing this linkage within a FOS model can be beneficial in creating highly fit solutions more efficiently.

2 FOS Linkage Model and GOMEA Instances

A linkage model has the capacity to represent groups of variables for a given problem, which make an important contribution to the fitness of solutions. A family of subsets (FOS) linkage model is a set of subsets of a main set, which contains all problem variables. Each subset of the FOS contains between one and l-1 items, where l is the number of problem variables. Individual algorithms from the GOMEA family are distinguished by how the linkage models are constructed.

GOMEA is a family of genetic algorithms which generate new solutions through mixing. Initially, a random population of solutions is created and evaluated. During solution mixing, each subset of the FOS model is used as the variation operator, or a mask, which is applied to copy values from a randomly picked donor solution from the population to the current solution and the effect is evaluated immediately. Any improvement to the fitness of the solution from applying a FOS subset mask is kept [1]. This process is repeated for every solution in the population. In the Black Box setting we consider three GOMEA Instances which generate the FOS linkage model differently.

The Linkage Tree Genetic Algorithm (LTGA) is a GOMEA instance that learns the linkage between the problem variables, by generating a linkage tree at each generation. Building this tree is accomplished by a hierarchical clustering algorithm which uses mutual information to determine dependencies between variables [2]. Each variable is first assigned to a single cluster. Based on the mutual information new clusters are created by joining together already existing clusters. This process continues until only one cluster (containing all variables) is left. FOS model subsets generated by LTGA consist of the variable clusters present in the linkage tree.

Random-Tree GOMEA has a linkage structure identical to LTGA. While the FOS linkage structure is still based on a linkage tree, the actual linkage information is not based on learned variable dependen-

¹The full paper has been published in the *Proceedings of the Genetic and Evolutionary Computation Conference - GECCO13*, pages 853–860, 2013.

cies. Mutual information used to determine variable dependencies and cluster proximities is completely random. No real linkage information is discovered.

Univariate GOMEA is a very simple instance of the GOMEA family. All the problem variables are considered to be independent, resulting in no linkage between the variables being modeled. In other words, the FOS linkage structure only consists of singleton subsets.

3 Results

Figure below shows how different algorithms from the GOMEA family perform on representative benchmarks (averaged over 30 independent runs). *Unif* is a random and uniformly distributed benchmark, while *am55* and *c880* are structured and not uniform. The only algorithm which attempts to learn linkage information is LTGA. The results show that Univariate-GOMEA is only efficient for small



population sizes of uniformly distributed random problems. LTGA becomes more efficient for larger populations. Problems with a better defined underlying structure favor linkage learning in all cases. LTGA outperforms the Random-Tree and Univariate GOMEA instances. This indicates that learning variable linkage can be a powerful tool in the MAX-SAT domain.

4 Summarized Additional Results and Conclusions

In the full version of this paper White Box setting results are presented, where problem specific information can be explored. We also provide comparisons with other SAT solving algorithms such as hBOA, GSAT and Walksat, showing how linkage-learning GOMEA instances (such as LTGA enhanced with a local search) often outperform them.

Given experimental data in the full version of this paper, we conclude that exploring variable linkage of MAX-SAT problems helps in generating better solutions more efficiently. The results hold in terms of evaluations performed in both black and white-box settings, and in terms of the total run-time the algorithms need to generate near-optimal solutions.

References

- Peter A N Bosman and Dirk Thierens. Linkage neighbors, optimal mixing and forced improvements in genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '12, pages 585–592, New York, NY, USA, 2012. ACM.
- [2] D. Thierens. The linkage tree genetic algorithm. In Parallel Problem Solving from Nature PPSN XI, pages 264–273, Berlin, 2010. Springer–Verlag.