# An Application of a Multivariate Estimation of Distribution Algorithm to Cancer Chemotherapy

Alexander Brownlee
School of Computing,
The Robert Gordon University
Aberdeen, UK
sb@comp.rgu.ac.uk

Martin Pelikan
Dept. of Mathematics and Computer Science,
University of Missouri
St. Louis, USA
pelikan@cs.umsl.edu

John McCall
School of Computing,
The Robert Gordon University
Aberdeen, UK
jm@comp.rgu.ac.uk

Andrei Petrovski
School of Computing,
The Robert Gordon University
Aberdeen, UK
ap@comp.rgu.ac.uk

## ABSTRACT
Chemotherapy treatment for cancer is a complex optimisation problem with a large number of interacting variables and constraints. A number of different heuristics have been applied to it with varying success. In this paper we expand on this by applying two estimation of distribution algorithms to the problem. One is UMDA and the other is hBOA, the first EDA using a multivariate probabilistic model to be applied to the chemotherapy problem. While instinct would lead us to predict that the more sophisticated algorithm would yield better performance on a complex problem like this, we show that it is outperformed by the algorithms using the simpler univariate model. We hypothesise that this is caused by the more sophisticated algorithm being impeded by the large number of interactions in the problem which though present, do not complicate the search for optima.

## Categories and Subject Descriptors
I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search;

G.3 [**Probability and Statistics**]: Probabilistic Algorithms, stochastic processes;

J.3 [**Life and Medical Sciences**]: Health

## General Terms
Algorithms, Performance, Theory.

## Keywords
Estimation of Distribution Algorithms, Complexity, Medical Applications.

## 1. CANCER CHEMOTHERAPY
Cancer chemotherapy treatment is a complex optimisation problem with a large number of variables and constraints. A delicate balance of drug doses must be struck which controls tumour size while minimising toxic side effects. Finding a treatment regime from the numerous possibilities can be viewed as a classic search problem suited to evolutionary algorithms.


*GECCO'08*, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-131-6/08/07.

## 2. APPLICATION OF EAs
Previous research [5, 9] has shown that a large range of probabilistic algorithms can be applied to the chemotherapy problem with varying degrees of success – these have included genetic algorithms (GAs), particle swarm optimization and estimation of distribution algorithms (EDAs). Related work applying evolutionary algorithms to chemotherapy can also be found in [1, 4, 7, 10].

EDAs are a development of GAs, replacing the traditional crossover and mutation operators with the building and sampling of a probabilistic model. Previous applications of EDAs to chemotherapy have used algorithms employing a univariate model [9]. Here we show the application of the multivariate EDA, the Hierarchical Bayesian Optimisation Algorithm (hBOA) [8] which uses a Bayesian network to factorise the joint probability distribution of good solutions. For comparison we also applied a further univarate EDA – the Univariate Marginal Distribution Algorithm (UMDA) [6]. We report the performance of each and then compare this to previously reported results.

As in most of our previous experiments here we use a binary representation where clusters of four bits represent the level of one drug in one treatment interval. There are 10 drugs and 10 treatment intervals overall giving a 400 bit chromosome length.

### 2.1 Problem Complexity
Like many real-world problems, the chemotherapy problem has a large number of interactions between variables. We ran the Linkage Detection Algorithm [3] 30 times on the problem to find the interactions (linkage) present between the 400 variables. Of the 79800 possible bivariate interactions, the mean number of interactions discovered was 50491 with a standard deviation of 109.5. This means that to perfectly learn the fitness function would require a highly complex probabilistic model which would be computationally expensive to construct. While it must be noted that this does not mean that a multivariate model is needed for optimisation of this problem, this does provide motivation to consider the application of a multivariate EDA to the problem.

## 3. EXPERIMENTAL PROCEDURE
For each experiment, the algorithms were run systematically with a number of different parameter configurations to find their optimal performance. We ran two experiments on each algorithm.

The first investigated algorithm efficiency. Often the fitness function is the most computationally expensive part of an EA – especially with complex real-world fitness functions. Thus it is

important to reduce the number of fitness evaluations the algorithm requires to find a solution. For this problem, we measure the number of evaluations required to find a feasible solution (one in which all constraints are satisfied and the patient survives). The second experiment recorded the quality of solution found. In this case, this is the best fitness value found by the algorithm in a fixed number of evaluations (to allow comparison with our previous results this was set to 200 000).

## 4. ANALYSIS AND CONCLUSIONS

In both experiments we see that the GA is outperformed by all the EDAs, reinforcing the findings of previous comparisons between GA and EDA. As described in section 2.1 the multi-drug chemotherapy problem is highly complex with a large number of variable interactions. On many other problems with multivariate interactions hBOA has been shown to significantly outperform univariate EDAs [8]. Indeed, in many cases the difference between the algorithms is tractable vs. intractable. Given these facts, it is perhaps unexpected that the univariate algorithms do better than hBOA, both in the number of evaluations and the final fitness reached. We believe the reason for this is that a large number of the interactions within the problem are unnecessary for optimisation. Thus hBOA constructs a complex model of the problem which is very expensive to search and ultimately counterproductive – in essence being preoccupied by the additional complexity while the simpler algorithms are able to ignore it. The bigger model will also need more computation, both in building the model and in the number of solution evaluations. Intuitively we would expect a more precise model of the data to result in a more precise location of optima. However if the benefits gained are small, the computation required to do so may render the algorithm uncompetitive. It would be interesting to further explore the relationship between the costs and benefits of detecting multivariate interactions for other multivariate EDAs and on similar highly multivariate problems.

Further, though a complex problem, this instance of the chemotherapy problem uses a relatively simple tumour growth model. More complex models including factors like more antagonistic drugs and cancers are likely to include higher order interactions. Multivariate EDAs such as hBOA could exploit this to give a more pronounced performance improvement.

It is feasible that given an unlimited number of fitness evaluations the algorithm employing a more complex model will be more capable of finding the true global optimum, but with a huge computational expense. In comparison, the algorithms using a univariate model can get very close to the global optimum – in this case close enough that the reduced overhead is worthwhile. The addition of a local search operator such a deterministic hillclimber could improve the performance of UMDA and PBIL further and this would be another area for further investigation.

A technical report [2] with full results from the experiments is also available.

## 5. ACKNOWLEDGMENT

M.P. was sponsored by the National Science Foundation under CAREER grant ECS-0547013, by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant FA9550-06-1-0096, and by the University of Missouri in St. Louis through the High Performance Computing Collaboratory sponsored by Information Technology Services, and the Research Award and Research Board programs. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the Air Force Office of Scientific Research, or the U.S. Government. Some experiments were done using the hBOA software developed by Martin Pelikan and David E. Goldberg at the University of Illinois at Urbana-Champaign.

## 6. REFERENCES

[1] AGUR, Z., HASSIN, R. AND LEVY, S. 2006. Optimizing Chemotherapy Scheduling Using Local Search Heuristics. Operations research 54, pp. 829-846.

[2] BROWNLEE, A., PELIKAN, M., MCCALL, J. AND PETROVSKI, A. 2008. An Application of a Multivariate Estimation of Distribution Algorithm to Cancer Chemotherapy. MEDAL Technical Report No. 2008005. Available from: http://medal.cs.umsl.edu/files/2008005.pdf

[3] HECKENDORN, R.B. AND WRIGHT, A.H. 2004. Efficient Linkage Discovery by Limited Probing. Evolutionary computation 12, pp. 517-545.

[4] LIANG, Y., LUENG, K.S. AND MOK, T.S.K. 2006. Automating the drug scheduling with different toxicity clearance in cancer chemotherapy via evolutionary computation. In GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation, Seattle, Washington, USA, pp. 1705-1712.

[5] MCCALL, J., PETROVSKI, A. AND SHAKYA, S. 2008. Evolutionary Algorithms for Cancer Chemotherapy Optimization. In Computational Intelligence in Bioinformatics, G.B. FOGEL, D.W. CORNE AND Y. PAN, Eds. Wiley, pp. 265-296.

[6] MÜHLENBEIN, H. AND PAAß, G. 1996. From Recombination of Genes to the Estimation of Distributions I. Binary Parameters. In PPSN IV: Proceedings of the 4th International Conference on Parallel Problem Solving from Nature, Springer-Verlag, pp. 178-187.

[7] OCHOA, G., VILLASANA, M. AND BURKE, E.K. 2007. An evolutionary approach to cancer chemotherapy scheduling. Genetic Programming and Evolvable Machines 8, pp. 301-318.

[8] PELIKAN, M. 2005. Hierarchical Bayesian Optimization Algorithms. Springer Verlag.

[9] PETROVSKI, A., SHAKYA, S. AND MCCALL, J. 2006. Optimising cancer chemotherapy using an estimation of distribution algorithm and genetic algorithms. In GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation, Seattle, Washington, USA, pp. 413-418.

[10] TAN, K.C., KHOR, E.F., CAI, J., HENG, C.M. AND LEE, T.H. 2002. Automating the drug scheduling of cancer chemotherapy via evolutionary computation. Artificial Intelligence in Medicine 25, 169-18