Optimisation and Fitness Modelling of Bio-control in Mushroom Farming using a Markov Network EDA

Alexander Brownlee[1], Yanghui Wu[2], John McCall[1], Paul Godley[3], David Cairns[3], Julie Cowie[3]

[1]School of Computing, The Robert Gordon University Aberdeen, UK
{sb,jm}@comp.rgu.ac.uk

[2]Northwest A & F University Yangling, China
sxxawyh@hotmail.com

[3]Department of Computing Science, The University of Stirling Stirling, UK
{pgo,dec,jco}@cs.stir.ac.uk

ABSTRACT

We explore the application of an Estimation of Distribution Algorithm which uses a Markov Network to the problem of bio-control in mushroom farming. This falls into the category of "bang-bang control" problems and was previously used as an application for genetic algorithms with modified crossover operators. The EDA yields a small improvement in the solutions that are evolved. Moreover, the probabilistic models constructed closely match identifiable features in the underlying dynamics of the problem. We conclude that this is a useful by-product of the probabilistic modelling which can be further exploited.

Categories and Subject Descriptors

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

G.3 **[Probability and statistics]**: Probabilistic algorithms, Stochastic processes;

F.2.2. **[Analysis of Algorithms and Problem Complexity]**: Nonnumerical Algorithms and Problems

General Terms

Algorithms, Performance, Theory

Keywords

Estimation of Distribution Algorithms, Real-world applications, Modelling.

1. INTRODUCTION

Estimation of Distribution Algorithms (EDA) use probabilistic models to replace the selection and variation operators in traditional evolutionary algorithms. While the construction and sampling of a model can be more computationally expensive than genetic operators, the resulting search can use fewer function evaluations. A further potential advantage is that models produced using raw data can also reveal underlying information about the problem.

In this paper, we will be exploring the application of the EDA Distribution Estimation using Markov Networks (DEUM) to bio-control in mushroom farming. There is a small but growing literature on Markov Network EDAs (e.g. [5], [6], [7], [8]). In comparison to Bayesian networks, however, they have not been extensively studied in the EDA literature. Our aim is to explore the relationship between optimization and the nature of the probabilistic model learned by DEUM. We will describe how a clear relationship can be drawn between probabilistic model and the underlying shape of the problem.

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-130-9/08/07.

2. BIO-CONTROL IN MUSHROOM FARMING

When mushrooms are produced in commercial quantities, the quality and yield of the mushroom crop can be seriously damaged through infestation by sciarid flies. An important weapon in combatting sciarid fly is the use of the nematode worm, *Steinernema feltiae* which feeds on sciarid larvae.

In [1] a dynamic mathematical model is developed that expresses the life cycle of Sciarid larvae in the presence of periodic dosing with nematode worms. The model consists of a series of coupled differential equations and can be implemented numerically using standard techniques. The problem admits a *bang-bang control* strategy, that is a fixed dose of nematodes is either applied or not applied at each of a series of discrete potential intervention points during the mushroom cultivation period. Solutions to the problem will seek to minimise interventions whilst maximising the effect of the control on the sciarid larva population.

3. EVOLUTIONARY ALGORITHMS FOR BIO-CONTROL

3.1 Genetic Algorithms

In previous work, Godley et al. developed genetic algorithms for the bio-control problem using directed intervention crossover operators [2], [3], [4]. The operators, CalEB (Calculated Expanding Bin) and TInSSel (Targeted Intervention with Stochastic Selection), favour the reduction of intervention points and attempt to focus the location of interventions at time points where they seem to have most effect. Both use the number of interventions present in the parents to calculate the number required in the children. While these operators help a GA, with probabilistic modelling it could introduce a bias within the population that limits the accuracy of the model. The GA operators described above treated variables as independent – the inclusion in the model of interactions between variables will help the algorithm in detecting the number and distribution of interventions required. In this paper we make comparisons with TinSSel as it was shown in [2] to be the more efficient of the two approaches.

3.2 Estimation of Distribution Algorithms

Our motivation in this paper is to explore the relationship between optimisation and the probabilistic model built by an EDA. The probabilistic model used by an EDA is sampled to produce good solutions and so should reflect the objective. We apply Chain-DEUM, a bivariate extension of the EDA Distribution Estimation Using Markov networks $DEUM_d$ [10]. This algorithm operates on bitstring solutions and thus we encode the problem so each bit represents a time step with a value of '1' representing an intervention and '0' representing no intervention. As the bit string is a time series it is conceivable that neighbouring bits – which represent consecutive interventions – will have an interdependency. This leads us to use a chain model; we incorporate these interactions in the model by adding β terms for each neighbouring pair of bits. Thus, we develop the Markov network to model the energy distribution over the fitness function according to this relation:

$$-\ln F(x) = \begin{pmatrix} \alpha_0 V(x_0) + \beta_0 V(x_0) V(x_1) + ... \\ + \beta_{T-1} V(x_{T-1}) V(x_T) + a_T V(x_T) \end{pmatrix} / T \quad (3)$$

We call this the Markov Fitness Model (MFM). This model can be sampled according to the marginal probabilities:

$$p(x_i = 1) = \frac{1}{1 + e^{2(\alpha_i + \beta_i \beta_{i+1} x_{i+1})/T}} \quad (4)$$

T is a temperature coefficient which is varied according to a cooling scheme as the algorithm progresses through a series of generations. Combined, these give us the following workflow:

1. Generate random population
2. Repeat:
    1.1. Build model by substituting values from population in to (3) and fitting coefficient values of MFM
    1.2. Sample model with Gibbs sampler using marginal probabilities from (4)
3. :Until maximum number generations reached

No explicit selection operator is used, as selective pressure comes from the incorporation of fitness into the model. Full details of the $DEUM_d$ approach may be obtained from [10].

4. RESULTS AND CONCLUSIONS

The EDA finds slightly fitter solutions than the previous approach using a modified GA. The best solutions found by repeated runs are also more consistent in their placement of interventions. The EDA performs better without seeding the initial population with individuals having few interventions; with a heavily seeded population the modified GA outperforms the EDA. The underlying system uses several coupled differential equations and so is quite complex. This provides motivation for further research into how successfully this approach could be applied to more general bang-bang control problems.

The bio-control problem is to determine the number and distribution of intervention points. The analysis of the probabilistic model constructed after a single generation shows that it is possible to get good approximate information on this from a small number of evaluations. Depending on available data, this approach might be useful for learning bang-bang control systems from data as an alternative to mathematical modelling where little expert knowledge exists.

This work has also revealed that while initialisation control can be helpful or at worst neutral for a GA, it hinders the EDA due to biasing of the model. In this case this is because a reduced number of interventions in the population make it difficult for the EDA to easily find critical regions for intervention. Solutions with large numbers of initial interventions, whilst of poor fitness, can still provide useful information to the modelling process about where interventions should occur.

The most important observation from these experiments is the precision with which Chain-DEUM can identify critical regions for intervention and how it can be explicitly related to coefficients in the probabilistic model. This is significant because switching is key to bang-bang control and can be used directly to influence human decisions. Usually evolutionary algorithms produce results which are difficult to analyse with respect to theoretical understanding of the problem. Here there is a clear match which can be easily understood.

5. ACKNOWLEDGMENT

Yanghui Wu was supported in this work by the Chinese Government Scholarship Fund.

6. REFERENCES

[1] A.Fenton, R.L.Gwynn,A.Gupta, R.Norman, J.P.Fairbairn, and P.J.Hudson. Optimal application strategies for entomopathogenic nematodes: integrating theoretical and empirical approaches. Journal of Applied Ecology 2002 39,481-492.

[2] P. M. Godley, D. E. Cairns, and J. Cowie. Directed intervention crossover applied to bio-control scheduling. In *IEEE CEC 2007: Proceedings of the IEEE Congress On Evolutionary Computation*, 2007.

[3] P. M. Godley, D. E. Cairns, and J. Cowie. Maximising the efficiency of bio-control application utilising genetic algorithms. In EFITA / WCCA 2007: Proceedings of the 6th Biennial Conference of European Federation of IT in Agriculture, Glasgow, Scotland, UK, 2007. Glasgow Caledonian University.

[4] P. M. Godley, J. Cowie and D. E. Cairns, Novel Genetic Algorithm Approaches for Time-Series Problems, Doctoral Symposium on Engineering Stochastic Local Search Algorithms pp47-51, IRIDIA, 2007 ISSN: 1781-3794

[5] R. Santana. Probabilistic Modeling Based on Undirected Graphs in Estimation Distribution Algorithms. Ph.D. dissertation, Institute of Cybernetics, Mathematics and Physics, Havana, Cuba, 2003.

[6] R. Santana. Estimation of Distribution Algorithms with Kikuchi Approximation. Evolutonary Computation, vol. 13, no. 1, pp. 67-98, 2005.

[7] S. Shakya. DEUM: A Framework for an Estimation of Distribution Algorithm Based on Markov Random Fields. Ph.D. dissertation, The Robert Gordon University, Aberdeen, UK, April 2006.

[8] S. Shakya, John McCall. Optimization by Estimation of Distribution with DEUM Framework Based on Markov Random Fields. International Journal of Automation and Computing.04(3),July 2007,262-272