The Single Chromosome's Guide To Dating

Michael Ratford, Andrew Tuson, and Henry Thompson
Department of Artificial Intelligence, University of Edinburgh
80 South Bridge, Edinburgh EH1 1HN, U.K.
{miker,andrewt,ht}@dai.ed.ac.uk

Abstract

In nature, sexually reproducing organisms do not mate indiscriminately — the choice of mate has an impact upon their offspring's fitness. The investigation described here shows that, for a wide range of problems in the literature, using sexual selection proved to be a robust method for enhancing genetic algorithm performance. In addition, this investigation provides evidence for which parameters are important for a successful implementation.

1 Introduction

In genetic algorithms (GAs) [6], the crossover operator plays a central role. Informally, the rationale behind this is that the recombinative process combines portions of the two parent solutions, in the hope of combining sections of each solution associated with high fitness together, so to produce solutions of higher fitness. Standard implementations of the GA select the second parent either at random, or on the basis of fitness. However using different criteria to select the second parent may make crossover more effective by encouraging recombination between strings that have useful information to exchange, thus making GAs more useful optimisers.

In nature, sexually reproducing organisms do not mate indiscriminately. The reason for this is simply that half of the offspring's genes will come from the other parent, and the choice of mate will therefore have a large impact upon their children's fitness, and hence the survival of the parent's own genes. Therefore organisms effect some form of sexual selection such as choosing a mate so to control the balance between out-breeding and inbreeding. Rephrasing this in GA terms, crossover between solutions that are too similar should be discouraged as no useful search takes place; but there should be some similarity as information exchange is more likely to be meaningful.

Work has been performed on using mate choice in GAs: [10] described how the use of a 'seduction function' based upon an aesthetic measure to select the second parent can improve GA performance; [7] discussed the various ways that sexual selection can be used in evolutionary computation; and [4] showed that 'incest prevention', the prohibition of crossover between identical or very similar strings, can prevent premature convergence and give improved results.

An investigation of the potential of using sexual selection, and how it should be implemented in GAs is described in this paper.

2 Implementation

The publicly available GA implementation PGA was used as a testbed for this investigation. Due to lack of space, full details of the implementation are given in [9]. The first parent was selected in the usual fashion; with modifications being made in the selection of the second parent. A seduction function [10], based upon the similarity of each of the prospective second parents to the first parent, was then used to select the second parent. The process is summarised by the pseudo-code below (adapted from [10]):

```
until (termination condition met) do {
     parent1=select(population);
     parent2=seduce(parent1,population);
     operate(parent1,parent2,child1,child2);
     merge(child1,child2,population);
     }
```

2.1 Similarity Metrics

Three choices of a similarity metric were investigated: *Hamming distance* which measures the similarity in genotype space; *Euclidean distance* the distance between chromosomes' in phenotype space; and finally the number of building blocks common to both chromosomes. In each case the distance metric is normalised to a value between 0 and 1 by dividing the raw value by the maximum possible.

2.1.1 Seduction Functions

The similarity measures mentioned above need to be somehow processed to provide a measure of seduction, *sed*. The higher a chromosome's seduction is in relation to another, then the better it's potential as a suitor will be.

Two forms of the seduction function were investigated, with the parameters: *centre*, *width* and *depth*. A bell curve was defined as:

$$sed = \left(\frac{width}{width + (distance - centre)^4}\right)^{depth}$$

and a simpler, and less computationally expensive function was defined as:

$$sed = \begin{cases} 1 - depth + \frac{distance \times depth}{centre} & distance \leq centre \\ 1 - depth \times \frac{distance - centre}{1 - centre} & distance > centre \end{cases}$$

Incest prevention [4] was also implemented: a *threshold* for was provided, so that the final seduction measure was defined as:

$$thresholded\ seduction = \begin{cases} 0 & seduction \leq threshold \\ seduction & seduction > threshold \end{cases}$$

It is possible that using purely distance to select a second parent could be inefficient, allowing unfit parents to contaminate the reproductive process. Thus three options for a final *attraction* measure were used: seduction only, the average between seduction and normalised fitness, and seduction multiplied by *normalised fitness* defined as:

$$normalised\ fitness = \frac{\#\ solutions - fitness\ rank + 1}{\#\ solutions}$$

where the *fitness rank* is 1 for the fittest chromosome, 2 for the second fittest, and so on.

2.1.2 Selecting The Mate

As calculating the "overture" from every other chromosome in the population each time an individual was selected for breeding would clearly be unmanageable, tournament [1], marriage [11], and courtship selection [9] were used.

3 The Test Problems

This investigation tested the performance of sexual selection over a wide range of problems in the literature, so to provide a strong test of this technique.

- De Jong's test functions (DJ1, 2, 3, and 5) [2].
- Modified binary F6 (BF6) [12].
- Deb's test functions (DEB1, 2, 3, and 4) [3].
- Himmelblau's function (HIMM) [5].
- Maximising 1s (MAX).
- Maximum contiguous block (MCBn).
- The Royal Road (RR) [8].

4 An Initial Study

An initial attempt was made to find trends for each parameter value so to provide a general feel for what was and was not important. The test functions used were DJ3, DJ5 and BF6. Due to space constraints a summary of results is given in this paper. Detailed results are available in [9].

The GA was initially run with and without mate selection, covering as wide a range of the parameter space as practicable. Each run had a population of 50 and a genotype size of 32. Each set of parameters was run 20 times. Each run was terminated when the maximum possible fitness was reached, the population had converged, or 100,000 function evaluations had been made. These results were then examined to detect trends for each parameter choice. A *t-test* was used to determine significance.

4.1 Does Sexual Selection Work?

The use of seduction only, rather than a combination of seduction and fitness, produced significantly better quality solutions for DJ3 and BF6.

4.2 The Seduction Function

There was no difference between either of the two functions (bell or simple) used.

4.3 The Similarity Metric

The results of using phenotypic or genotypic similarity measures varied. The phenotype was slightly better for DJ3, no different for DJ5, and the genotype measure did better for BF6; the choice is problem specific. The building block measure did extremely badly.

4.4 The Selection Method

Courtship selection outperformed marriage selection for DJ3 and BF6 with no difference for DJ5.

4.5 Incest Prevention

Using sexual selection in conjunction with incest prevention using a seduction threshold gave better results each time.

5 Expanding the Study

The study was then expanded to include all of the test problems, to give a convincing demonstration of sexual selection's abilities, and to check the earlier conclusions.

To see how good the current model was, each function was run 50 times with an educated guess at a good, standard genetic algorithm: a steady-state GA, with mutation rate 0.001, string length 30, two-point crossover, and rank-based selection. The functions were then rerun with the same algorithm together with an educated guess at a good sexual selection technique: a simple curve, seduction as an attraction measure, Hamming distance, incest prevention, and marriage selection. The results obtained (Tables 1 and 2) were heartening, with 11 out of the 13 functions having a significant (in **bold**) improvement in solution quality.

		Fitness		Evals	
Test	Opt	Mean	σ	Mean	σ
MAX	×25	29.34	0.7982	714.00	110.21
DJ1	×0	99.9637	0.0685	938.00	305.49
DJ2	×1	999.948	0.0626	846.00	244.08
DJ3	×20	54.12	0.9398	15790.00	28218.95
DJ5	×1	496.528	3.0968	982.00	263.76
BF6	×2	0.761324	0.2664	850.00	245.78
MCB5	×0	19.54	3.8184	33786.00	16758.13
RR	×11	0.9804	0.8182	79118.00	40414.82
DEB1	×15	0.999925	0.0003	1002.00	298.46
DEB2	×8	0.986123	0.0445	1022.00	291.40
DEB3	×7	0.999649	0.0006	998.00	315.09
DEB4	×8	0.993056	0.0157	974.00	267.69
HIMM	×1	199.386	0.9354	926.00	285.40

Table 1: Results for a Standard GA

		Fitness		Evals	
Test	Opt	Mean	σ	Mean	σ
MAX	×49	**29.98**	0.1414	718.00	95.70
DJ1	×1	**99.9926**	0.0178	1406.00	1217.75
DJ2	×1	**999.951**	0.0746	1266.00	290.92
DJ3	×46	**54.92**	0.2740	4074.00	11308.56
DJ5	×8	498.08	1.7311	1202.00	281.57
BF6	×8	**0.928968**	0.1749	3258.00	13974.38
MCB5	×2	**26.36**	2.8978	59586.00	20139.21
RR	×15	1.1332	0.8639	73506.00	43758.20
DEB1	×34	**0.999979**	9.3e-5	1230.00	256.35
DEB2	×11	**0.991279**	0.0417	1270.00	373.07
DEB3	×15	**0.999957**	0.0002	1182.00	230.74
DEB4	×6	**0.994211**	0.0151	1186.00	394.74
HIMM	×1	**199.924**	0.2239	1278.00	361.43

Table 2: Results for a GA with Sexual Selection

The fact that this technique was not tuned indicates that it is fairly robust to its parameter settings. Also, it was interesting to note that, unlike in Section 4.4, marriage selection proved to be the better choice. This is a demonstration of the dangers of making general statements from only a few problems.

5.1 Incest Prevention Revisited

Is it just incest prevention that is causing the improvement? When the same tests were performed with a seduction threshold of 0.1 and selection based on ranked fitness only, the results were actually worse than for those tests using a standard GA *without* sexual selection. Thus either the initial conclusions (in Section 4.5) about its worth were wrong or incest prevention, as used here, is doing its work as an intrinsic part of sexual selection — follow-up results using sexual selection without a seduction threshold were inconclusive, providing better results than those from using sexual selection with incest prevention for seven test functions and worse results for five. The small size of the differences suggest that the shape of the seduction curve renders explicit incest prevention unnecessary.

5.2 Varying the Aesthetic

The runs were repeated using a phenotype measure of similarity, where the algorithm did generally slightly worse than when using a genotype measure.

This may, at first, appear slightly surprising, as in genotype space the fitness landscape will be more disjointed and ragged. However crossover will often have more chance of success between chromosomes that are genotypically similar, as phenotypically similar chromosomes may be either side of a *Hamming cliff*, where a small difference in natural numbers is equivalent to a large difference in the binary representation. In these cases crossover will be a poor tool for local search.

The results when the building block aesthetic were used were poor, worse than not using mate selection in most cases.

5.3 The Effect of Crossover Rate

Varying the crossover rate showed that the higher the crossover rate, the better the results achieved, although in more evaluations. This result applied when using a standard GA, but the rate of increase was much higher with sexual selection (Figure 1); ie. sexual selection makes crossover more effective.

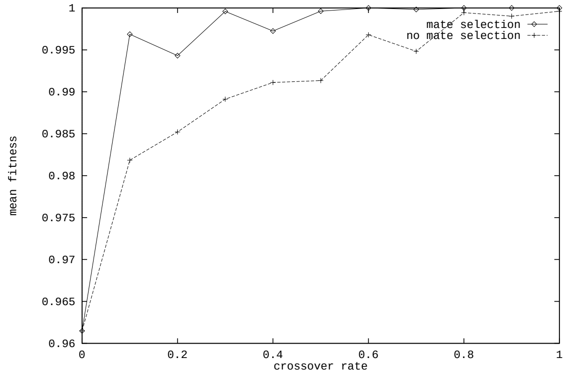

Figure 1: Mean Fitness Against $p(Xover)$, DEB4

6 Conclusion

On the basis of the results obtained here, sexual selection appears to have made crossover a more effective search operator. Significant increases in solution quality were obtained for most of the problems. Although of course some care has to taken on how it is implemented, it appears that the approach is fairly robust to its settings in many cases, and some useful guidelines to the application of this technique have arisen from this investigation. This bodes well for its application to problems in the real-world.

7 Acknowledgements

Thanks to the EPSRC for their support of Michael Ratford and Andrew Tuson via studentships with references 95411068 and 95306458.

References

[1] A. Brindle. *Genetic Algorithms for Function Optimization.* PhD thesis, University of Alberta, 1981.

[2] K. A. De Jong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, University of Michigan, 1975.

[3] K. Deb. Genetic algorithms in multimodal function optimization. Master's thesis, University of Alabama, Department of Engineering Mechanics, 1989. TCGA Report No. 89002.

[4] L. J. Eshelman and J. D. Schaffer. Preventing premature convergence in genetic algorithms by preventing incest. In R. K. Belew and L. B. Booker, editors, *Proceedings of the Fourth International Conference on Genetic Algorithms.* Morgan Kaufmann, 1991.

[5] D. M. Himmelblau. *Applied Nonlinear Programming.* McGraw-Hill, 1972.

[6] J. H. Holland. *Adaptation in Natural and Artificial Systems.* Ann Arbor: The University of Michigan Press, 1975.

[7] G. F. Miller. Exploiting mate choice in evolutionary computation: Sexual selection as a process of search, optimization, and diversification. In T. C. Fogarty, editor, *Proceedings of the 1994 AISB Workshop on Evolutionary Computing.* Springer-Verlag, 1994.

[8] M. Mitchell, S. Forrest, and J. H. Holland. The Royal Road for Genetic Algorithms: Fitness Landscapes and GA Performance. In *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, Cambridge, MA, 1991. MIT Press.

[9] M. Ratford. The Single Chromosome's Guide To Dating. Master's thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.

[10] E. Ronald. When Selection Meets Seduction. In Larry J. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, San Francisco, CA, 1995. Morgan Kaufmann.

[11] P. Ross. Personal communication, 1996.

[12] J. D. Schaffer, R. A. Caruna, L. J. Eshelman, and R. Das. A study of control parameters affecting online performance of genetic algorithms for function optimization. In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms.* Morgan Kaufmann, 1989.