The Relationship Between Evolvability and Bloat

Jeffrey K. Bassett
George Mason University
4400 University Dr., MS2A1
Fairfax, VA
jbassett@cs.gmu.edu

Mark Coletti
George Mason University
4400 University Dr., MS2A1
Fairfax, VA
mcoletti@gmu.edu

Kenneth A. De Jong
George Mason University
4400 University Dr., MS2A1
Fairfax, VA
kdejong@gmu.edu

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*Concept learning*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

General Terms

Algorithms, Performance

Keywords

Learning Classifier Systems, Pittsburgh Approach, evolvability, bloat, homologous crossover

ABSTRACT

Bloat is a common problem with Evolutionary Algorithms (EAs) that use variable length representation. By creating unnecessarily large individuals it results in longer EA runtimes and solutions that are difficult to interpret. The causes of bloat are still uncertain, but one theory suggests that it occurs when the phenotype (e.g. behaviors) of the parents are not successfully inherited by their offspring [5]. Noting the similarity to evolvability theory [1], which measures heritability of *fitness*, we hypothesize that reproductive operators with high evolvability will be less likely to cause bloat.

We set out to design a new crossover operator for Pittsburgh approach classifier systems that has high phenotypic heritability. We saw an opportunity using the nearest neighbor representation to perform crossover cuts in phenotype space rather than on the genomes. We demonstrate that our operator tends to be less susceptible to bloat and has higher evolvability than a standard Pittsburgh approach crossover operator. Our hope is that this will lead to a general approach to reducing bloat for any representation.

1. INTRODUCTION

Many Evolutionary Algorithms that use variable length representations suffer from bloat. Bloat occurs when the average genome size tends to grow as evolution progresses. The main side effect is that progress toward a solution slows, which is primarily caused by larger individuals taking longer to evaluate.

GECCO'09, July 8–12, 2009, Montréal Québec, Canada.
ACM 978-1-60558-325-9/09/07.

Several theories have been posed for the causes of bloat. McPhee and Miller demonstrate that in some cases bloat occurs when the phenotypic semantics of a parent are not accurately replicated in the offspring [5]. An operator that is very destructive, such as removing large parts of the genome, is likely to radically change the semantic behavior of the individual. Alternatively, adding a large number of genes can often be done without significantly changing the semantic behavior. When this is true there will be a bias towards larger genomes.

The relationship between parents and offspring is a critical component of evolvability theory, although it is concerned with the fitness of the individuals [1]. This theory states that the ability of parents to produce offspring that are more fit is often a good indicator of ultimate EA performance. The measure most often used to determine evolvability is either the covariance or correlation between parent and offspring fitness.

Two individuals that are semantically different are also likely to have different fitnesses while individuals that are semantically similar are likely to have similar fitnesses. Given this, we hypothesize that a reproductive operator that has a higher evolvability will also be less likely to cause bloat. This assumes that the fitness landscape is somewhat smooth, which is a necessary condition for any EA to make progress.

Our focus is finding methods for reducing bloat in Pittsburgh approach rule systems [6]. The standard Pittsburgh approach recombination operators can be quite destructive often deleting large numbers of genes. Since genes are position independent (they can hold any position on the genome with no change in meaning), the recombination operators typically take random numbers of genes from both parents. Therefore it is possible for a child to inherit most of its genes from one parent with almost all of the other genes going to its sibling.

2. METHODOLOGY

We design a crossover operator that limits the number of deletions, which should increase evolvability and thereby decrease bloat. The Nearest-Neighbor rule representation [3] provides an opportunity for achieving this since the conditions of each rule can be treated as a independent points in feature space. This allows us to define a recombination operator that performs cuts in feature space instead of along the genome. The advantage of this approach is that two very similar parents should produce two very similar offspring, and so we can expect a high fitness correlation.

We compare this operator's behavior with that of a stan-

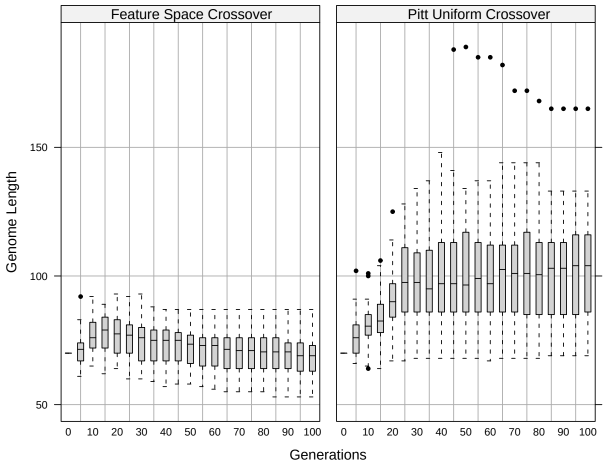

Figure 1: Plots of the genome length for the 2-spiral problem.

dard Pitt recombination operator. We use a test suite consisting of two fairly simple man-made concept learning problems (a 3x3 checkerboard, and a 2 intertwined spirals) as well as a "real-life" problem predicting diabetes in the Pima Indian tribe [2]. We seek to find out if the evolvability of our operator is actually higher, if it is effective at reducing bloat, and if we must sacrifice the solution quality to achieve this. Assuming we are successful this may demonstrate a general approach to reducing bloat by customizing the reproductive operators.

3. RESULTS

For all problems, experiments were performed that compared the new feature space crossover with the more standard Pitt approach uniform crossover. The results showed that the new crossover did a better job of controlling bloat in all cases, reducing the number of rules by a statistically significant amount. Figure 1 shows just the results for the two-spiral problem with 70 initial rules. Similar plots for the other problems were all qualitatively similar. In no case was the fitness statistically different (within 95% confidence) when compared to runs using the standard crossover.

4. CONCLUSIONS

In our experiments the feature space crossover operator has been successful at reducing bloat, especially when compared to the Pitt uniform crossover operator. Other results indicate that the Pitt 1 and 2-point crossover operators cause even more bloat than the uniform, which is why they we not included in this study.

With this operator we have the opportunity to test our hypothesis that operators with high evolvability tend to cause less bloat. We have chosen to measure evolvability using the approach described by Manderick et. al. [4] called operator fitness correlation.

Figure 2 shows the operator correlations of both crossover operators averaged over the runs of the checkerboard problem with 9 initial rules. The Pitt uniform crossover gradually decreases to almost zero over the course of the run, while

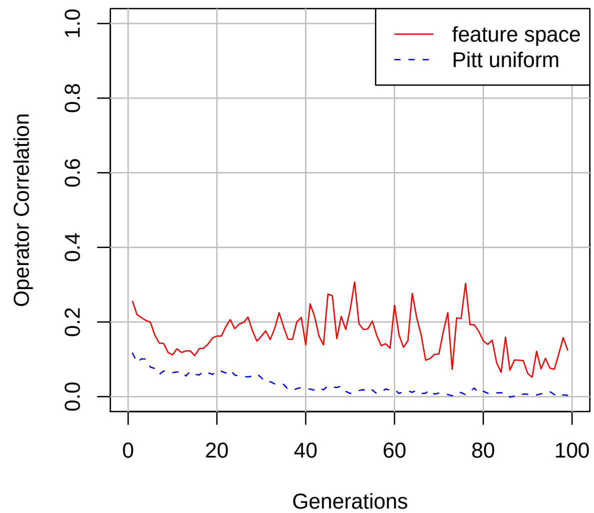

Figure 2: Correlation between parent and offspring fitness for the checkers problem with 9 initial rules.

the feature space crossover is able to maintain a steady level throughout the run. This pattern is consistent for every experiment we have discussed here.

5. REFERENCES

[1] L. Altenberg. The evolution of evolvability in genetic programming. In J. Kenneth E. Kinnear, editor, *Advances in Genetic Programming*, pages 47–74. MIT Press, 1994.

[2] A. Asuncion and D. Newman. UCI machine learning repository, 2007.

[3] M. Fuchs and A. Abecker. Optimized nearest-neighbor classifiers using generated instances. In *KI '96: Proceedings of the 20th Annual German Conference on Artificial Intelligence*, pages 71–83, London, UK, 1996. Springer-Verlag.

[4] B. Manderick, M. de Weger, and P. Spiessens. The genetic algorithm and the structure of the fitness landscape. In R. K. Belew and L. B. Booker, editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 143–150, San Diego, California, 1991. Morgan Kauffmann Publishers.

[5] N. F. McPhee and J. D. Miller. Accurate replication in genetic programming. In L. J. Eshelman, editor, *Proceedings of the Sixth Int. Conf. on Genetic Algorithms*, pages 303–309, San Francisco, CA, 1995. Morgan Kaufmann.

[6] S. F. Smith. *A Learning System Based on Genetic Adaptive Algorithms.* PhD thesis, The University of Pittsburgh, 1980.