Genetic Programming: Optimal Population Sizes for Varying Complexity Problems

Alan Piszcz
Department of Computer Science
University of Idaho
Moscow, ID 83844, USA
apiszcz@acm.org

Terence Soule
Department of Computer Science
University of Idaho
Moscow, ID 83844, USA
tsoule@cs.uidaho.edu

ABSTRACT

The population size in evolutionary computation is a significant parameter affecting computational effort and the ability to successfully evolve solutions. We find that population size sensitivity how much a genetic program's efficiency varies with population size is correlated with problem complexity. An analysis of population sizes was conducted using a unimodal, bimodal and a multi-modal problem with varying levels of difficulty. Specifically we show that a unimodal and bimodal and multimodal problems exhibit an increased sensitivity to population size with increasing levels of difficulty. We demonstrate that as problem complexity increases, determination of the optimal population size becomes more difficult. Conversely, the less complex a problem is the more sensitive the genetic program's efficiency is to population size.

Categories and Subject Descriptors

I.2.2 [**Artificial Intelligence**]: Automatic Programming, program synthesis.

General Terms

Algorithms, Performance, Experimentation.

Keywords

Population sizing, MAX binary tree problem, optimal, binomial-3, problem difficulty.

1. INTRODUCTION

Population size is recognized as a primary parameter influencing algorithm efficiency in evolutionary computation. If one chooses too small of a population one may not achieve a successful solution, whereas if one chooses too large of a population one may unable to obtain a result in a justifiable amount of time. Determining optimal population size remains an important open research question. The evolutionary computation community's interest in improving population sizing for genetic algorithms and genetic programming (GP) is evident in a wide range of past and current research. As GP techniques are applied to real-world problems, classified as 'human-competitive' by attaining one or more of the criteria established by Koza, population size becomes a significant concern [2].

Most previous research on population size in evolutionary computation has focused on finding the optimal population size, either empirically or through theoretical analysis. In this paper we address a related, but significantly different question: *what is the relationship between problem complexity and the 'window' of reasonable population sizes?* Specifically we hypothesize that: *the optimal population size window decreases as problem complexity increases.* This is an important question because if a problem does not exhibit a significant sensitivity towards population size (i.e. the window of reasonable population sizes is broad) then it may not be worth the effort required to find the optimal population size.

In this paper *complexity* refers to the number of modes or maxima in the fitness landscape that mathematically represents the problem. The term *difficulty* refers to the level of computational effort required to achieve a successful solution for a problem of a given complexity.

We attempt to address the following questions related to population sizing: Does the problem complexity influence the optimal population size? Does problem complexity and optimal population sizing have a significant impact on solution effort? Do large populations always provide and advantage regardless of problem complexity?

2. PROBLEM DESCRIPTION

To test our hypothesis that the optimal population size window decreases as problem complexity increases we implement three problem types. The MAX binary tree problem is used as the unimodal and bimodal test problem. For the multimodal case, the Binomial-3 problem allows us to test a problem by scaling the level of difficulty. Each of these problems allows one to control the level of difficulty through parameter setting.

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.
ACM 1-59593-186-4/06/0007.

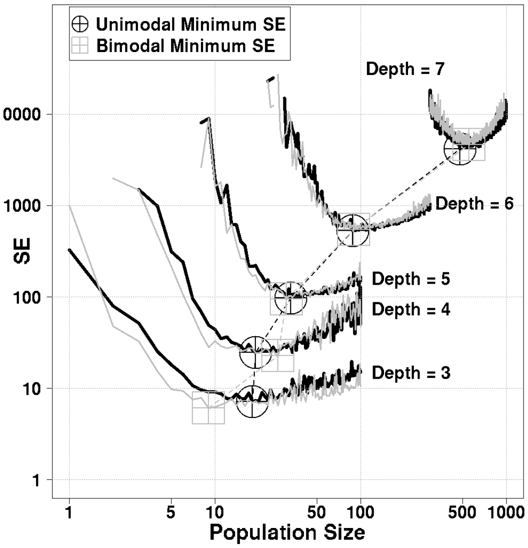

Figure 1: Experiment 1: Unimodal and Bimodal MAX Binary Tree Fitness Models Solution Effort Results

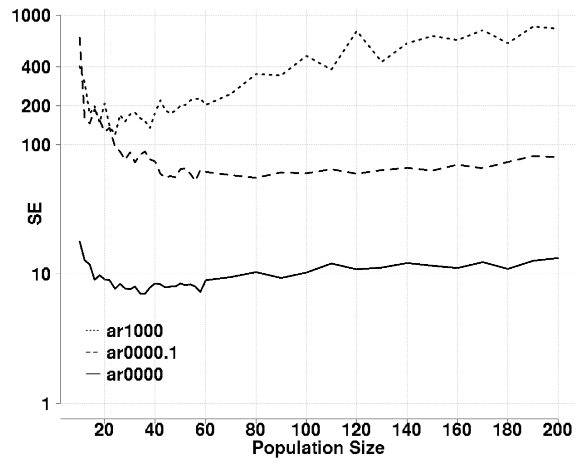

Figure 2: Experiment 2: Binomial 3: Solution effort.

3. RESULTS

Results for experiment 1 are shown in Figure 1 represent the MAX BTP SE curves around the optimum population range. The heavy black lines plot the results of unimodal MAX BTP solution effort over the log population ranges; the gray lines plot the solution effort for the bimodal MAX BTP. The minimum solution effort population size point is shown by the cross-hair circle for unimodal and square for the bimodal case. The label '**Depth =**' near each set of plot lines indicate the depth value for each MAX BTP sub experiment. The optimal population size in the unimodal problem is 9, 27, 31, 90 and 555 for depths 3, 4, 5, 6 and 7 demonstrating a consistent correlation of optimal size to problem complexity.

To extend the test of our hypothesis relating population size to problem complexity we choose to use the tunable difficultly symbolic regression problem introduced by Daida. We use a similar configuration and ERCs as Daida and perform the experiment with a parametric sweep of the population size parameter [1]. The sweeps of the population parameter range include a similar sequence to experiment 1.

Figure 2 shows the solution effort for three of the eight levels of difficulty tested. The solution effort required increases with problem difficulty and in the case of $\mathsf{R} = 1000$, where the difficulty is maximum we note a reduction in the effort required between population size 20 and 60. Thus, picking a population size in that range will significantly increase efficiency.

4. CONCLUSION

GP research continues to investigate population sizes in order to generate successful solutions with the least computational effort and to improve the probability of achieving solutions in more complex applications. As problems increase in complexity (i.e. unimodal to multimodal) it becomes more difficult to identify the optimal population size. This research provides a preliminary investigation at the population sizing sensitivity for a few problems of varying complexity with scalable difficulty.

We summarize the research results with answers to the questions raised in the introduction: *Does the problem complexity influence the optimal population size?* We confirm problem complexity is a factor of the optimal problem size for less complex problems. In more complex problems such as multimodal the precise optimal population size becomes more difficult to determine. Determining the value of locating optimal population sizes may outweigh the effort required to find it.

Do optimal population sizes have a significant effect on solution effort? The population size is a key factor in producing solutions and has a significant effect on all test problems. Determination of the optimal population size appears is more challenging as the problem complexity increases. In low complexity problems with high levels of difficulty (MAX BTP) we find the optimal population size has a significant payoff for reducing computational effort.

Do large populations always provide and advantage regardless of problem complexity? If we consider computational effort as a primary concern, the answer is no. Excluding computational effort and reviewing the quality of solutions for other attributes may open new questions regarding alternate values of population size.

5. REFERENCES

[1] J. M. Daida, J. A. Polito, S. A. Stanhope, R. R. Bertram, J. C. Khoo, and S. A. Chaudhary. What makes a problem GP-hard? analysis of a tunably difficult problem in genetic programming. In W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, editors, *Proceedings of the Genetic and Evolutionary Computation Conference*, volume 2, pages 982–989, Orlando, Florida, USA, 13-17 July 1999. Morgan Kaufmann.

[2] J. R. Koza. Human-competitive machine intelligence by means of genetic programming. *IEEE Intelligent Systems*, 15(3):76–78, May-June 2000.