
Variable Length Genomes for Evolutionary Algorithms

C.-Y. Lee

Department of Mechanical Engineering
California Institute of Technology
Pasadena, CA 91125
cinyoung@caltech.edu

E.K. Antonsson

Department of Mechanical Engineering
California Institute of Technology
Pasadena, CA 91125

Abstract

A variable length genome has been developed to search variable dimensionality/complexity spaces efficiently. A simple proof of concept experiment shows the utility of such a representation.

1 INTRODUCTION

Most evolutionary algorithms (with the notable exception of genetic programming) have difficulties searching spaces of variable dimensionality/complexity. This is primarily a result of the fixed length representations of canonical genomes. We introduce a variable length genome to counteract this problem.

While canonical genomes are single strings of numbers, they have an implied second string that identifies each gene (*i.e.*, each gene's position is known). We extend canonical genomes to a variable length representation by making this second identifying string explicit. Furthermore, the identifying genes are allowed to take on any real value, rather than integer values as in canonical genomes. Mutation and crossover remain unchanged from the standard implementations. For example, in two point crossover, a range of identifying genes is chosen over which genes are swapped. Below are two genomes before crossover.

1st genome:
1 0 0 1 0 1 1 1 1
.1 |.2 .3 .4 .5 .6 .7| .8 .9
2nd genome:
0 1 0 0 1 1
|.11 .15| .81 .85 .89 .9

After crossover with a range of 0.11 to 0.70, the

genomes become

1st genome:
1 0 1 1 1
.1 |.11 .15| .8 .9
2nd genome:
0 0 1 0 1 1 0 0 1 1
|.2 .3 .4 .5 .6 .7| .81 .85 .89 .9

It should be apparent that if genomes have the same identifying strings, this representation is equivalent to the canonical one. Henceforth, we call the above genome exG, as in extended genome.

2 RESULTS

A simple proof of concept experiment is conducted to determine the efficiency of the exG in variable dimensionality searches. The problem is to find a target string of length 36. The fitness function used is

$$\text{if } l_i \neq l_t : F = |l_i - l_t|$$

$$\text{if } l_i = l_t : F = \|g_i - g_t\|$$

where l_i is test genome length, l_t is target length, and $\|\cdot\|$ denotes the euclidean metric of of the test and target genomes. A (10+60) evolution strategy selection scheme is used. Results of the algorithm's convergence to the correct length are shown below. Initial range indicates the range of string lengths that the algorithm was initialized over.

Initial Range	Generations	σ
3-6	12	2.04
3-28	5.45	1.16
3-100	2.82	0.94
3-35	4.55	1.50
7-31	5.73	1.35
15-23	6.09	0.67