Diploidy and Overdominance in Genetic Algorithms

by Garrett Bidwell
BBN

Alden Wright
University of Montana

Genetic Algorithms:

A family of randomized algorithms for heuristic search and optimization.

Based on analogy to natural genetics and evolution.

Uses a simulated evolution of a population of “individuals”.

Each individual is a potential solution to the problem.

Fitness

We assume that each individual has a *fitness*.

In an optimization problem, this will be derived from the objective function.

High fitness — good

Low fitness — bad

Objective: find individuals with high fitness.

The basic algorithm is:

1. Initialize population

2. Selection — survival of the fittest

3. Mating (crossover)

4. Mutation

5. Go to 2.

<u>Elaborated:</u>

1. Choose a random initial population.
2. Evaluate the fitness all individuals in the population.
3. Construct a new population as follows:
   - A. Select two individuals with high fitness
   - B. Mate these individuals—combine their genetic representations to produce a child individual
   - C. Possibly mutate the genetic representation of the child
   - D. Repeat steps A to C until a new population is built.
4. Go to step 2.

Genetic representation (haploid)

Each individual is represented as a *chromosome*.

A GA chromosome is a string of bits.

A biological chromosome can be viewed as string of DNA bases, or as a string of genes, where each gene codes for protein.

Haploid crossover (mating)

1-point crossover:

```
Parent 1:          0 1 0 1 1 1 0 1 0 1
Parent 2:          1 1 1 0 0 1 1 1 1 0
Crossover point:            |
Child 1:           1 1 1 0 0 1 0 1 0 1
Child 2:           0 1 0 1 1 1 1 1 1 0
```

Mutation: Each bit is mutated with a small probability.

Haploidy and diploidy.

Haploid organisms (like bacteria) have a single chromosome.

Diploid organisms (like vertebrates) have pairs of homologous chromosomes.

```
Locus:         0 1 2 3 4 5 6 7 8 9
Chromosome 1:  0 1 0 1 1 1 0 1 0 1
Chromosome 2:  0 1 1 0 1 0 1 0 0 1
```

A locus (position) may be *homozygous*, or have the same alleles.

A locus may be *heterozygous*, or have different alleles.

Dominance at one Locus:

Often, one allele is *dominant*.

Dominance map:

	0	1
0	0	1
1	1	1

If 1 is dominant over 0, then the phenotype of a 10 individual will be the same as the phenotype of a 11 individual.

Example:

If Rh positive is dominant over Rh negative, then an individual with one Rh positive gene and one Rh negative gene will have blood type Rh positive.

Overdominance at one locus:

Assume two alleles, 0 and 1.

Overdominance occurs when the heterozygote is more fit than either homozygote.

10 and 01 are more fit than either 00 or 11.

Fitness map:

	0	1
0	F_0	$F_1 + s$
F_1	$F_1 + s$	1

Example:

Homozygotes for the sickle-cell gene get sickle-cell anemia (low fitness).

Heterozygotes for the gene have better resistance to malaria (heterozygote advantage).

GA Problem: Maintain high population diversity.

For many problems, a few highly fit individuals and their descendents quickly dominate the population.

Population variance quickly decreases.

Alleles are lost.

The search is no longer global.

Diploidy is one method to attack this problem.

1 locus symmetric dominance:

Fitness map: Assume that phenotype 0 has fitness F_0, and phenotype 1 has fitness F_1.

Let $P_0 =$ proportion of 0 in the population.

Let $P_1 = 1 - P_0 =$ proportion of 1 in the population.

Fitness map:

	0	1
0	F_0	$P_0F_0 + P_1F_1$
1	$P_0F_0 + P_1F_1$	F_1

When population is mostly 0's, 0 is dominant.

1 locus symmetric overdominance

Fitness map:

	0	1
0	F_0	$P_0F_0 + P_1F_1 + s$
1	$P_0F_0 + P_1F_1 + s$	F_1

$s > 0$ is an overdominance parameter.

Theoretical properties:

We showed that the theoretical 1-locus symmetric overdominant model has a globally stable polymorphic equilibrium.

Multiple loci

Map diploid chromosomes to haploid chromosome.

For each locus, use the probabilistic dominance map:

	0	1
0	0	$0 : P_0$ $1 : P_1$
1	$0 : P_0$ $1 : P_1$	1

Use the haploid fitness to assign a fitnes to a diploid individual.

Add a “fitness bonus” s for each heterozygotic locus.