

COGS2010

Evolutionary Computation I Introduction

Janet Wiles

Adapted and expanded from Lectures by Jennifer Hallinan

What is Evolution?

- Theory of the origins of diversity of living organisms based on variation and natural selection
- Its basis is a change in the frequencies of genes in a population

Example: change in gene frequencies

Genome: ataggcggccaatttaaacgcgcga
Gene: gggccaat
Alternative alleles: gggcgaat

Genes code for proteins that have useful structural and regulatory functions.

Suppose gggccaat occurs in 70% of the population and gggcgaat occurs in 30%

Over time, if one allele is more advantageous for one environment and the other for a different environment, their frequencies are likely to change in populations that are isolated in those environments.

Biological Evolution

- Genotypes are the instructions for building bodies
- Phenotypes are the 3-D agents that are built.
- From Genotype to phenotype is a multi-step, multi-stage process, involving many levels of emergent complexity, self-organisation, and development.
- The basis of selection is phenotypic variability

Phenotypic Variability

- Adaptation: Selection acts on the phenotype
- Sources of differences between phenotypes include:
 - Different genotypes (DNA)
 - Genetic variability arises through crossover (sex) and mutation
 - Different environments
 - Interactions between the developing organism and its environment

Computer Models of Evolution

- How to represent an agent?
 - Genotype, eg a sequence of symbols
 - Phenotype, eg a computer program
 - Environment
 - Developmental process
- How to represent the evolutionary process?
 - Mutation
 - crossover

Types of genetic mistake

Correct sequence: (known as the wild type)
THE CAT AND THE RAT ARE FAT

Transposition:
THE ACT AND THE RAT ARE FAT

Missing Letter:
THE CAA NDT HER ATA REF AT

Repeated Letter:
THE CCC ATA NDT HER ATA REF

Substituted Letter:
THE COT AND THE RAT ARE FAT

(From the Genetic Interest Group www.gig.org.uk)

Crossover between two members of a population

Parents

THE CAT AND T*HE RAT ARE FAT
THE DOG AND T*HE BUG ARE MAD

Offspring

THE CAT AND **THE BUG ARE MAD**
THE DOG AND THE RAT ARE FAT

Evolutionary Computation

- According to the Hitch-hiker's Guide to Evolutionary Computation:
 - "Evolutionary Algorithm is an umbrella term used to describe computer-based problem solving systems which use computational models of some of the known mechanisms of evolution as key elements in their design and implementation".
- This includes any way of solving problems which involves adaptive change in a population of computational agents.
 - EAs are typically characterized by population of candidate solutions encoded as character strings ("chromosomes") which are modified by genetic operators and have fitness proportional reproduction

Types of Evolutionary Algorithm (EA)

- Genetic algorithms (GAs):
 - Typically use fixed length character strings to represent genetic information. Agents undergo crossover and mutation to search the solution space.
- Evolutionary Strategies (ES)
 - Uses mutation as the search operator and a population size of 1
- Genetic Programming (GP)
 - Agents typically use tree-based representations rather than fixed length chromosomes

Computational models (non-Biological Evolution)

- Genotypes are the instructions for solving problems
- Phenotypes are the solutions.
- From genotype to phenotype is typically a simple process (often the genotype and phenotype are identical, and only the genotype is used)
- Genetic variability is typically derived from mutation and/or crossover

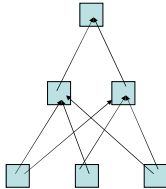
A simple GA

1. Initialize the population
2. For as many generations as required:
 1. Assess fitness
 2. Generate the children
 - Select 2 parents
 - Is crossover used?
 - Yes: identify the position and perform crossover
 - Mutate

Initialize the population

Decisions:

- How to represent the population?
 - Most often binary strings are used: 01000101
 - Sometimes real values: 0.3 0.5 0.6
- Each value is a "gene", defined by its location:
 - 01000101 has 8 genes
 - 0.3 0.5 0.6 has 3 genes
- Each alternative value of a gene is an "allele"
 - {0, 1}
 - infinite [1,0]
 - possibly a finite set of values {0, 0.1, 0.2, 0.3, ...}
- What do bits represent?
 - Nodes in a graph
 - Weights in a neural network
 - Features
- How many genes would be required to represent the modifiably weights in a 3-2-1 neural network?



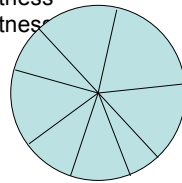
Assess Fitness

Decision:

- What's the objective function?
 - Depends on the representation
 - Must yield a single number
- A GA aims to maximize the fitness

Select the Parents

- Parent Selection is based on fitness
- Often fitness proportional selection (also called "Roulette wheel" selection)
 - Individuals are assigned an fitness proportional to their relative fitness
- Rank selection
- Tournament selection



Example: fitness functions

- Consider a population of individuals, each one has a genome of 12 bits, each bit is 0 or 1.
- The fitness function is the number of 1 bits in the chromosome
- What is the fitness of the following 3 individuals?
 - 1011 0000 1000
 - 0111 0001 0001
 - 0011 1011 1111

Example: fitness proportional selection

- Consider a population of 10 individuals which each have fitness values as follows
 - 5, 8, 6, 9, 8, 5, 3, 2, 3, 1
- What would be their expected number of offspring using fitness proportional selection?
 - Total pop fitness is $5+8+6+9+8+5+3+2+3+1=50$
 - Expected % of offspring is 5/50, 8/50 etc
 - 10%, 16%, ...
 - Expected number of times a parent is 2, 3.2, ...

Example: rank selection

- Consider a population of 10 individuals, which each have fitness values as in the previous case
 - 5, 8, 6, 9, 8, 5, 3, 2, 3, 1
- What would be their expected number of offspring using *rank* selection?
 - Rank order is 10, 8.5, 8.5, 7, 5.5, 5.5, 3.5, 3.5, 2, 1
 - Values are 9, 8, 8, 6, 5, 5, 3, 3, 2, 1
 - Expected number of offspring are based on rank (not values) 1.9 1.6 1.6 1.3 1.0 1.0 0.6 0.6 0.3 0.1 (adjusted for equal scores)

Mutation

- Single bit mutation changes the value of a gene from one allele to another
- Eg flip a bit from 0 to 1
Original 0000 0011 0010
Mutated 0000 0011 0000
- What would mutation be in a neural network?

Next generation

- Total replacement
 - Child population replaces the parent population
 - Increases the variability
 - Maximum fitness may decrease
- Preservation of the fittest population (also called elitism)
 - Preserves the fittest individual(s) of the previous generation
 - Proportion of population preserved may vary

Example

- Problem: Evolve a string of 1s using elitism
- Initialize the population
0000 0010 0100
1011 1000 1100
0001 1110 1110
- Assess the fitness
2
6
7
Total fitness is 15; Average fitness 5
- Select parents
1011 1000 1100 & 0001 1110 1110
0000 0010 0100 & 0001 1110 1110

cont

- Crossover: choose a crossover point
1011 1*000 1100
0001 1*110 1110
Yields two new offspring
1011 1110 1110
0001 1000 1100
- Mutation
1011 1**0**10 1110
1001 1000 1100
- New population
0001 1110 1110 (elite member)
1011 1**0**10 1110 (new variable offspring)
1001 1000 1100
- Assess fitness
7
8
5
Average fitness 6.6

How do GAs work?

- GAs perform a parallel exploration of their search space.
- If the solution to a task can be represented as a set of N real-valued parameters, then the job of finding this solution can be thought of as a search through an N-dimensional space.
- In general, if the solution to a task can be represented computationally, then the search space is the set of all possible configurations that may be represented using that scheme.

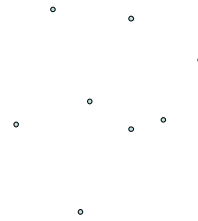
The exploration / exploitation dilemma

- How to effectively search a space, following promising leads, but not ignoring other potential pathways
- In EAs, the genetic operators mutation and crossover embody the modeller's theory about how to best achieve this balance

A classic optimisation problem The Travelling Salesman

- A salesman has N cities to visit, and must visit every city just once.
- His expenses are paid only if he chooses the path through the cities that minimizes his distance travelled.
- How should he find the shortest path?

A set of cities



- For small numbers of cities, the optimal tour is relatively easy to find.
- But the problem scales very badly
- The best known algorithms represent each city as a node on a graph, with a single path through each node.
- They start with random order of cities, then gradually refine the tour by swapping groups of cities on the tour
- This is a type of evolutionary strategy

What use are GAs?

- Problem solving
 - Optimization
 - Especially combinatorial problems
- Problem generation
 - Insights into evolution
 - Insights into the problem

Modelling with GAs

- Models simplify the real world
 - Including models of evolutionary processes
- What are the right simplifications?
 - Genome representation
 - Population of individuals
 - Genetic operators to maintain variability
 - Explicit or implicit fitness of individuals
 - Fitness proportional selection

Real biology

- but actually, it's not that simple...
- Most traits are controlled by many genes
- Each gene usually contributes to many traits
- Diploidy – each gene has two copies of each chromosome, one from each parent
- Alleles (alternative forms of a gene) may be dominant, co-dominant or recessive
- And DNA is double stranded, so each chromosome encodes the same information twice – once as the genes on one strand, and once as their reverse on the other strand
- Even inheritance is not simple
 - Some genetic material is inherited solely from the mother in cellular organelles in the cytoplasm, such as mitochondria and chloroplasts which have their own DNA.
 - Genes can be turned on or off by one or other parent, and the maternal and paternal switches oppose each other

Summary

- EAs are computational models inspired by biological evolution
- An EA may be broadly defined as a computational system that incorporates adaptive change in a population of individuals
- May be used in problem solving or problem generation (ie modeling the evolutionary process)
- Have been used for
 - Optimization
 - Automatic programming
 - Machine and robot learning
 - Economic models
 - Immune system models
 - Ecological models
 - Population genetics models
 - Interactions between evolution and learning
 - Models of social systems
- EAs are widely used in conjunction with neural nets.

Future Directions

- Evaluation of artificial life (Alife) systems
 - What is the difference between YACS and systems that teach something important and general?
- Representations: what works and why?
- Genotype to phenotype mappings
 - Is development a useful computational addition to an EA?
- Fitness evaluation
 - For Alife, implicit fitness is often of more interest than explicit calculations