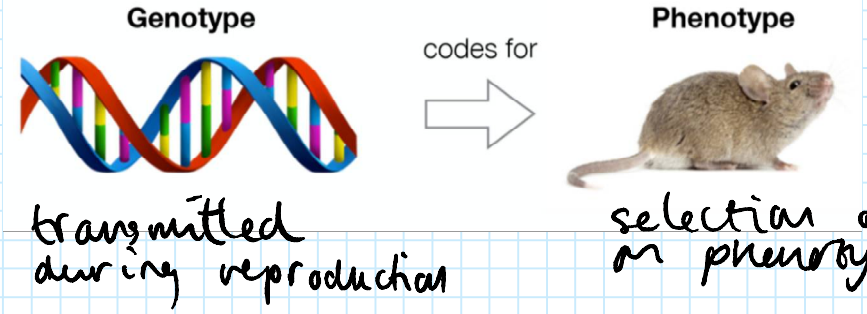
Genetic Algorithms

22 May 2017 15:34

Four Pillars of Evolution

- Population
- Diversity
- Heredity
- Selection

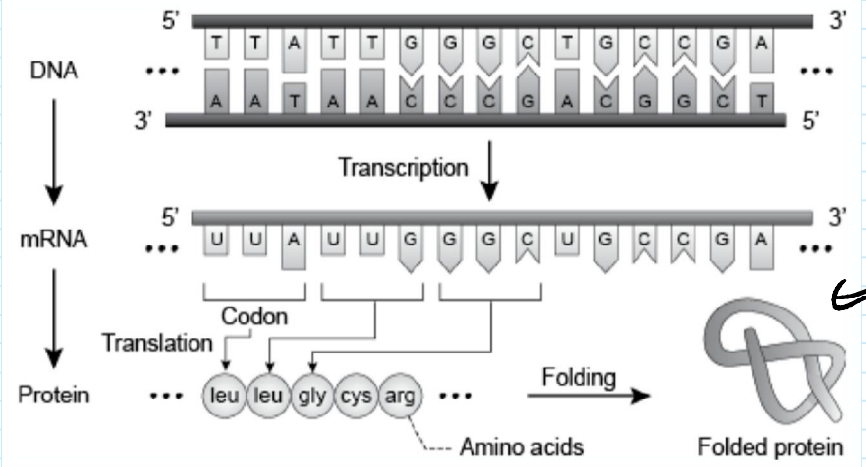
Genotypes & Phenotypes



transmitted during reproduction

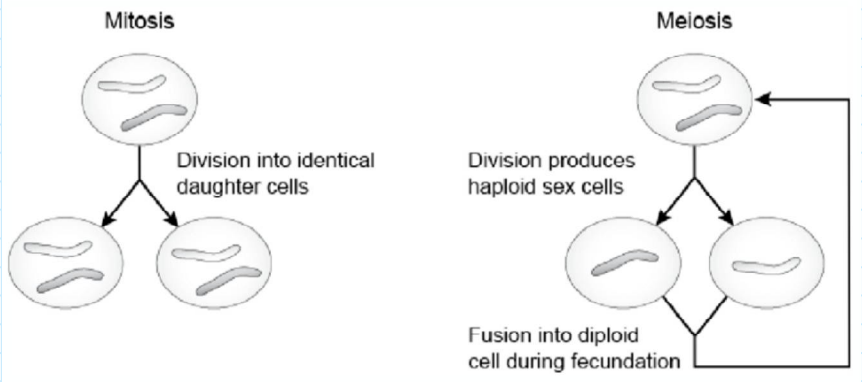
selection operates on phenotype

Genes to proteins



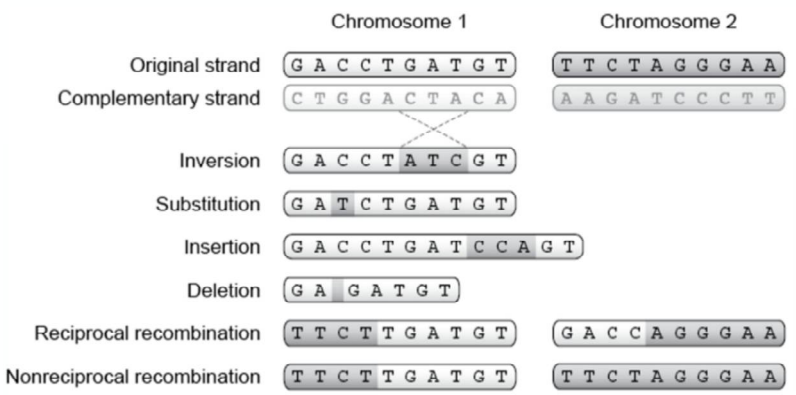
folded structure has function

Replication:



Passing Genes:

Genetic mutations occur during cell replication (4^{-10} per nucleotide per year)
Those that occur in sex cells can affect evolution
Recombination is a mutation that affects two homologous chromosomes



Introduce noise

Genome size within species stays constant

Genome size cross species varies wildly.

Bigger genome ≠ more complexity

Evolving Artificial systems:

Aspects we take from biology	Aspects we leave behind
• Genotype • Phenotype • Populations • Diversity • Selection • Inheritance	• Molecular dynamics • DNA transcription to mRNA to construct folded proteins • Coupled information strands (DNA being helical)

Designing an Evolutionary Algorithm:

- Devise genetic representation
- Define fitness function,
  $fit: Genome \to \mathbb{R}$
- Define crossover function,
  $cross: Genome \times Genome \to Genome$
- Define mutation function,
  $mut: Genome \to Genome$
- Define evolution process,
  $evolve: [Genome] \to [Genome]$

evolve: [Genome] → [Genome]

This is used to evolve the population by selection and generation.

select: [Genome] → [Genome], gen: [Genome] → [Genome]

∴ evolve pop = (gen ∘ select) pop

where select pop = take (sortBy fit pop) n

for constant n, then for gen,

$$\text{gen pop} = \begin{cases} \text{gen}(\text{pop}+[g]) & \text{if len pop} < N \\ \text{pop} & \text{otherwise} \end{cases}$$

where g = mut ∘ cross ∘ takeTwoRandom pop

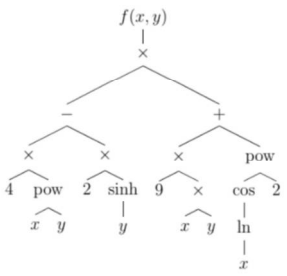
Thereby on each evolutionary cycle, the bottom (N-n) least fit genomes are culled, then the population is restored by randomly pairing individuals to reproduce.

Tree-based representations:

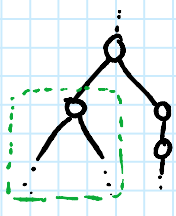
Genotype describes branching structure from a root to terminals

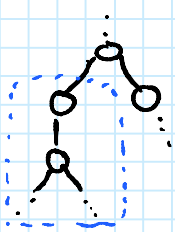
Used to evolve programs by encoding ASTs as genomes.

For example, $f(x,y) = (4x^y - 2\sinh y)(9xy + \cos^2 \ln x)$



Therefore crossover can be done by aligning trees as much a possible and mixing/matching.





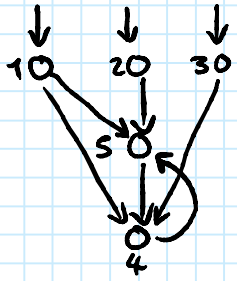
Mutation is done by randomly changing nodes or connective structure.

This feed forward tree structure is a

This feed forward tree structure is a less general form of evolving neural networks.

The NeuroEvolution of Augmenting Topologies (NEAT) algorithm, which defines genes for nodes and connections

Phenotype:

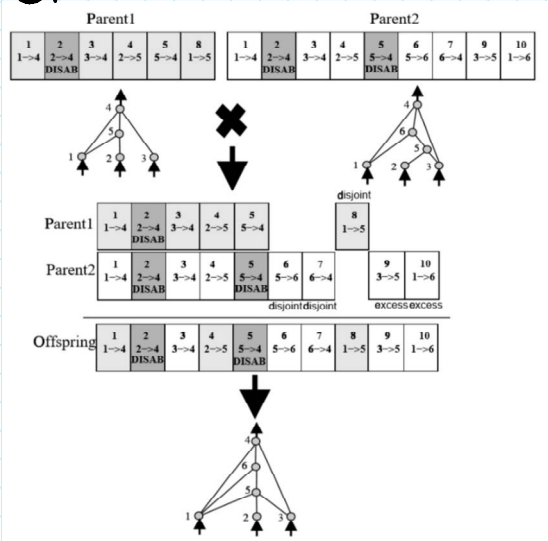


Genotype:

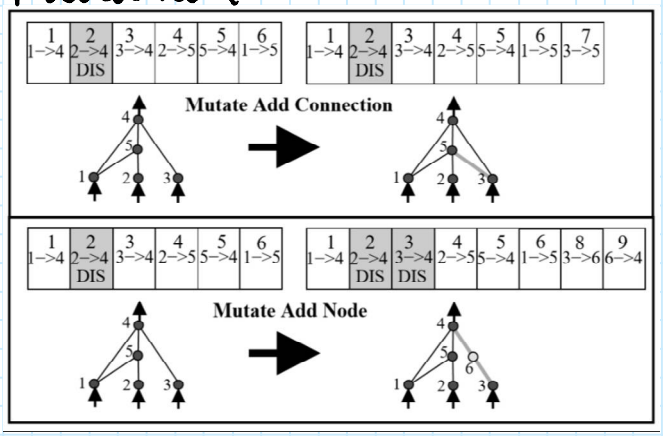
Node Genes	Node 1 Sensor	Node 2 Sensor	Node 3 Sensor	Node 4 Output	Node 5 Hidden		
Connect. Genes	In 1 Out 4 Weight 0.7 Enabled Innov 1	In 2 Out 4 Weight-0.5 **DISABLED** Innov 2	In 3 Out 4 Weight 0.5 Enabled Innov 3	In 2 Out 5 Weight 0.2 Enabled Innov 4	In 5 Out 4 Weight 0.4 Enabled Innov 5	In 1 Out 5 Weight 0.6 Enabled Innov 6	In 4 Out 5 Weight 0.6 Enabled Innov 11

So we can see the genome encodes all the information to express the phenotypes, however there is also information about the evolutionary history. This is done through "innovation" markers and gene disabling. Therefore the mutation and crossover functions are defined,

Crossover:



Mutation:



More Advanced Selection

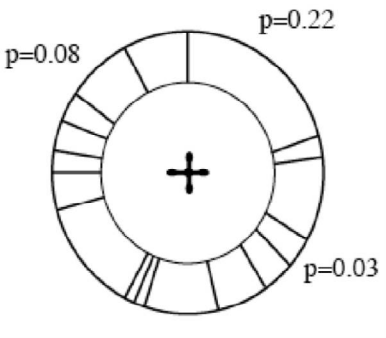
Key idea: Reduce population size and regenerate.

Proportional selection:

Roulette Wheel selection assigns probabilities

Proportional selection:

Roulette Wheel selection assigns probabilities to genomes in accordance with their fitnesses.

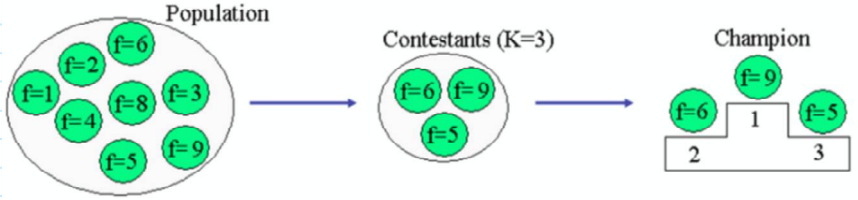


prob: Genome × [Genome] → [0,1]

prob g pop = (fit g) / sum (map fit pop)

This method is effective for encouraging diversity, but is expensive.

Tournament selection



- Pick k individuals with uniform random choice.
- Choose individual with highest fitness and make a copy, which is put in a separate pool.
- When the pool reaches a critical mass, crossover is performed to regenerate the population.
- This can be made probabilistic by assigning P[contestant i is selected] = $p(1-p)^i$
- Larger tournament size k increases selection pressure.

Regeneration:

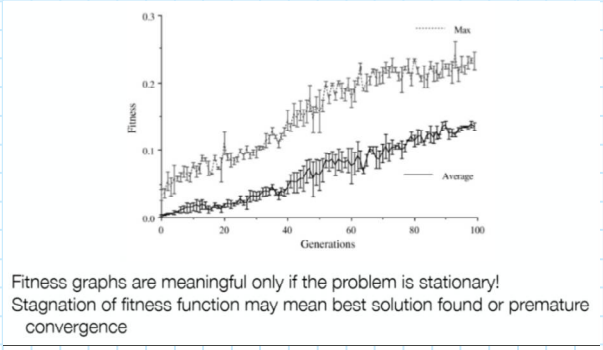
Generational replacement: all individuals replaced.
Elitism: maintain n best from previous generation.
Generational rollover: insert offspring at the place of worst individuals.

Speciation:

Section population into disjoint sets, whereby only members of the same set (species) can reproduce. Interspecies selection can be done with any of the aforementioned methods

reproduce. Interspecies selection can be done with any of the aforementioned methods. The idea is this encourages higher diversity by allowing experimentation that is currently yielding poor fitness, but may develop to an optimal solution.

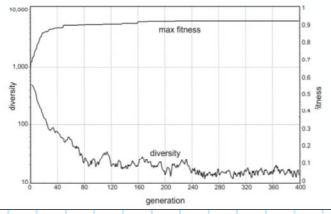
Monitoring Evolutary Process



Fitness graphs are meaningful only if the problem is stationary!
Stagnation of fitness function may mean best solution found or premature convergence

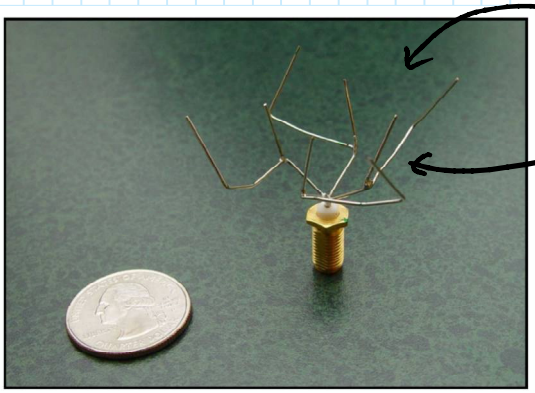
Diversity tells whether the population has potential for further evolution
Measures of diversity depend on genetic representation
E.g., for binary and real valued, use sum of Euclidean or Hamming distances

$$D_a(P) = \sum_{i,j \in P} d(g_i, g_j)$$



Example Applications:

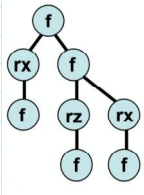
- Hardware:

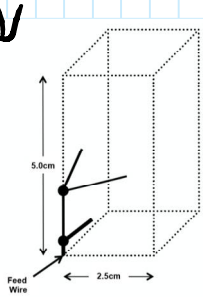


Evolved antennae

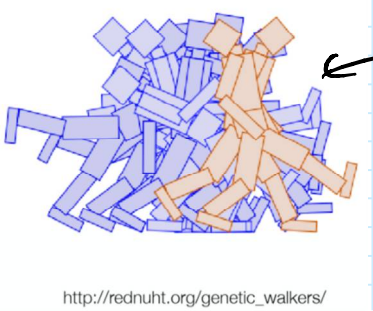
Tree structure to evolve Fitness based on simulated physical performance

Naïve human approach





- Robotic Controllers



simulated body with constrained movements in environment physically modelled.

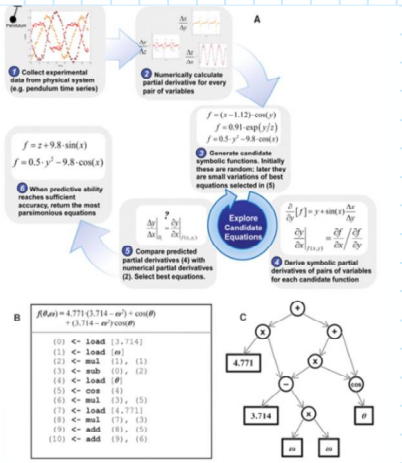
- Physical Models

  Fitness based on ability to predict collected data.

  Genomes encode equations as trees.

  Evolving from single node means the simplest models are given an advantage - therefore overfitting is avoiding.

Evolution of laws of physics



Distilling Free-Form Natural Laws from Experimental Data
Schmidt M., Lipson, H. (2009) Science, Vol. 324, no. 5923, pp. 81 - 85.

- Bioinformatics (RNA structure prediction).
- Modelling climate
- Data clustering
- Automated design.

Technological Concerns

- Validating black boxes
- Local optima
- Artificially evolving schema for biological genomes and physically implementing them.