UCD CASL
Complex & Adaptive Systems Laboratory

Genetic Programming: Syntax & Semantics

Michael O'Neill

Computational Toolbox Winterschool
Engelberg 8-12 Feb 2014

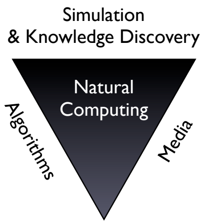
Overview

Genetic Programming: Syntax & Semantics

1. Setting the Stage
   - What is Natural Computing?
   - What is Evolutionary Computation?
   - An Introduction to Genetic Programming (GP)
2. Grammar-based GP
3. Semantic methods & Open Issues in GP

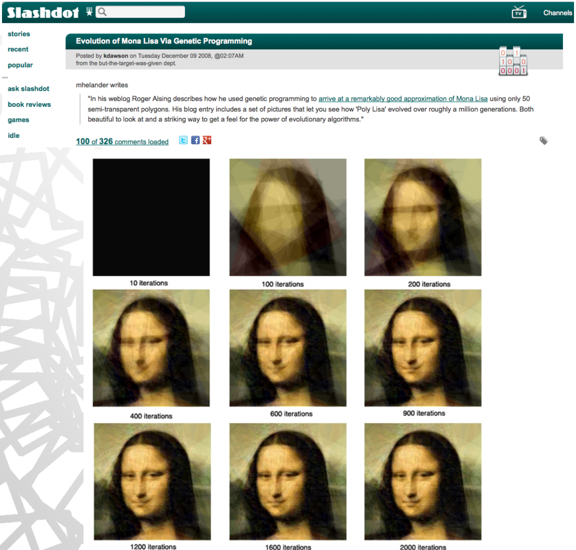
Setting the Stage

What is Natural Computing?

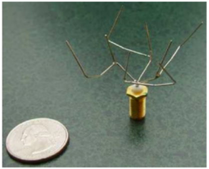


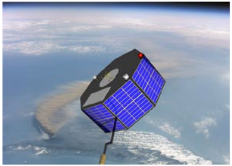
New Book: Brabazon, O'Neill, McGarraghy (2014). Natural Computing Algorithms. Springer.

Natural Computing Algorithms



Natural Computing Algorithms





Simulation of Natural Systems

SCIENTIFIC AMERICAN™

Subscribe: Digital / Print | Sign In / Register

Search ScientificAmerican.com

Subscribe | News & Features | Topics | Blogs | Multimedia | Education | Citizen Science | SA Magazine | SA Mind | Products

Home » Web Exclusives » November 2011

Web Exclusives | Technology

Tweet 71 | Like 1.1k

IBM Simulates 4.5 percent of the Human Brain, and All of the Cat Brain

A special online-only addition to November 2011's Graphic Science

By Mark Fischetti | October 25, 2011 | 25

Share Email

Supercomputers can store more information than the human brain and can calculate a single equation faster, but even the biggest, fastest supercomputers in the world cannot match the overall processing power of the brain. And they are nowhere near **as compact or energy efficient**.

Nevertheless, IBM is trying to simulate the human brain with its own cutting-edge supercomputer, called Blue Gene. For the simulation, it used 147,456 processors working in parallel with one another. IBM researchers say each processor is roughly equivalent to the one found in a personal computer, with one gigabyte of working memory.

So configured, Blue Gene simulated 4.5 percent of the brain's neurons and the connections among them called synapses—that's about one billion neurons and 10 trillion synapses. In total, the brain has roughly 20 billion neurons and 200 trillion synapses.

IBM describes the work in an intriguing paper (pdf) that compares various animal simulations done by its cognitive computing **research group** in Almaden, Calif. The group has managed to completely simulate the brain of a mouse (512 processors), rat (2,048) and cat (24,576). To rival the cortex inside your head, IBM predicts it will need to hook up 880,000 processors, which it hopes to achieve by 2019.

Simulation of Natural Systems

BBC | News | Sport | Weather | Travel | TV | Radio | More... | Search BBC News

NEWS

Watch ONE-MINUTE WORLD NEWS

News Front Page
Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Science & Environment
Technology
Entertainment
Also in the news

Video and Audio

Programmes
Have Your Say
In Pictures
Country Profiles
Special Reports

Related BBC sites
Sport
Weather
On This Day
Editors' Blog
BBC World Service

Page last updated at 12:46 GMT, Wednesday, 22 April 2009 13:46 UK

E-mail this to a friend | Printable version

Simulated brain closer to thought

By Jason Palmer
Science and technology reporter, BBC News, Prague

A detailed simulation of a small region of a brain built molecule by molecule has been constructed and has recreated experimental results from real brains.

This result completes the first phase of the brain simulation project

The "Blue Brain" has been put in a virtual body, and observing it gives the first indications of the molecular and neural basis of thought and memory.

Scaling the simulation to the human brain is only a matter of money, says the project's head.

The work was presented at the European Future Technologies meeting in Prague.

The Blue Brain project launched in 2005 as the most ambitious brain simulation effort ever undertaken.

While many computer simulations have attempted to code in "brain-like" computation or to mimic parts of the nervous systems and brains of a variety of animals, the Blue Brain project was conceived to reverse-engineer mammal brains from real laboratory data and to build up a computer model down to the level of the molecules that make them up.

The first phase of the project is now complete; researchers have modeled the neocortical column - a unit of the mammalian brain known as the neocortex which is responsible for higher brain functions and thought.

SEE ALSO
- Supercomputer to build 3D brain
  07 Jun 05 | Science & Environment
- Mouse brain simulated on computer
  27 Apr 07 | Technology
- IBM plans 'brain-like' computers
  21 Nov 08 | Science & Environment

RELATED INTERNET LINKS
- Science beyond fiction
- Blue Brain Project
- German Research Centre for Artificial Intelligence

The BBC is not responsible for the content of external internet sites

TOP SCIENCE & ENVIRONMENT STORIES
- Night-sky image is biggest ever
- Phantom Eye 'spy plane' unveiled
- Higgs discovery rumour is denied

| News feeds

MOST POPULAR STORIES NOW

SHARED | READ | WATCHED/LISTENED

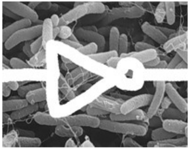
1. Rover's first find in Sharp focus
2. France 'opens Arafat murder case'
3. Isaac reaches hurricane strength
4. Virginity cream sparks Indian sex debate
5. World of Warcraft cut off in Iran

Synthesised Computing

FILED UNDER *Science*, *Alt*

Scientists build logic gates out of gut bacteria, then hopefully wash their hands

By Sharif Sakr posted Oct 24th 2011 1:42AM



Ever thought about upgrading your PC by breeding more cores? Or planting a few GBs of extra storage out in the yard? Us neither, until we heard that scientists at Imperial College in London have succeeded in building "some of the basic components of digital devices" out of genetically modified E.Coli. We've seen these germs exploited in a similar way before, but Imperial's researchers claim they're the first to make bacterial logic gates that can be fitted together to form more complex gates and potentially whole biological processors. Aside from our strange upgrade fantasies, such processors could one day be implanted into living bodies -- to weed out cancer cells, clean arteries and deliver medication exactly where it's needed. So much for Activia.

VIA *PhysOrg*
SOURCE *Imperial College London*

DISCUSS

Like 205 people like this. Be the first of your friends.

203 19 13 +1 14

TAGS *AND, bacteria, biological, biological computing, BiologicalComputing, biology, computing, E.Coli, germ, germs, gut, Imperial college, ImperialCollege, logic, logic gates, LogicGates, NAND, NOT, organic, organic computing, organic processor, OrganicComputing, OrganicProcessor, processor, stomach*

Synthesised Computing

The Economist

World politics | Business & finance | Economics | Science & technology | Culture | Blogs | Debate | The World in 2012 | Multimedia | Print edition

Technology Quarterly: Q1 2012

DNA computing

Computing with soup

Molecular computing: DNA is sometimes called the software of life. Now it is being used to build computers that can run inside cells

Mar 3rd 2012 | from the print edition

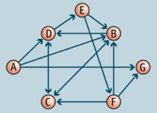
EVER since the advent of the integrated circuit in the 1960s, computing has been synonymous with chips of solid silicon. But some researchers have been taking an alternative approach: building liquid computers using DNA and its cousin RNA, the naturally occurring nucleic-acid molecules that encode genetic information inside cells. Rather than encoding ones and zeroes into high and low voltages that switch transistors on and off, the idea is to use high and low concentrations of these molecules to propagate signals through a kind of computational soup.

Computing with nucleic acids is much slower than using transistors. Unlike silicon chips, however, DNA-based computers could be made small enough to operate inside cells and control their activity. "If you can programme events at a molecular level in cells, you can cure or kill cells which are sick or in trouble and leave the other ones intact. You cannot do this with electronics," says Luca Cardelli of Microsoft's research centre in Cambridge, England, where the software giant is developing tools for designing molecular circuits.

At the heart of such circuits is Watson-Crick base pairing, the chemical Velcro that binds together the two strands of DNA's double helix. The four chemical "bases" (the letters of the genetic alphabet) that form the rungs of the helix stick together in complementary pairs: A (adenine) with T (thymine), and C

Showing the way

The experiment that launched DNA computing



1. Given a network of one-way roads linking seven towns, the aim is to determine if there is a route that starts in town A, finishes in town G and visits each town exactly once. First, short strands of DNA are made to represent the links in the road network. Because a road runs from town A to town B, AB strands are created. There are no roads between A and C, so no AC or CA strands are created. For each link in the network, about 100 trillion DNA strands

Natural Computing Algorithms

Sources of inspiration

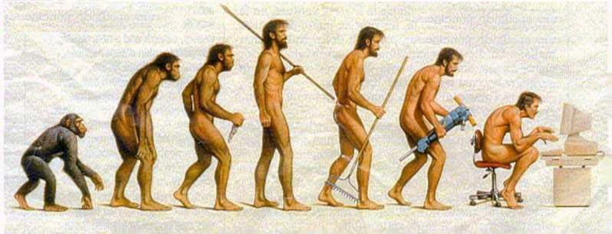
- Central Nervous System (*Neurocomputing*);
- Evolution (*Evolutionary Computation*);
- Molecular Dynamics (*Physical and Chemical Computing*);
- Immune Systems (*Immunocomputing*);
- Social Interaction amongst organisms (*Social Computing*);
- Language and Developmental Biology (*Developmental and Grammatical Computing*).

Not perfect imitation - exploit salient computational features

What is Evolutionary Computation?

Origin of the Species

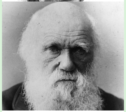
Million Years Ago	Event
?	Origin of Life
3500	Bacteria
1500	Eukaryotic Cells
600	Multicellular Organisms
1	Human Language



Origin of the Species

Milestones

∼ 200y.a. **Jean-Baptiste Lamarck**:
Lamarckism or soft-inheritance:
- Passing of lifetime acquired characteristics.



∼ 150y.a. **Charles Darwin**:
Theory of Natural Selection:
- Natural vs. Artificial Selection (a.k.a. breeding).

∼ 150y.a. **Gregor Johann Mendel**:
Mendelian Inheritance:
- Basis of Modern Genetics.

∼ 80y.a. **Fisher, Haldane & Wright**:
Population Genetics:
- Combined evolution, genetics, and statistical probabilities.

Origin of the Species

Milestones

$\sim$ 60y.a. **James D. Watson**:
Helix structure of DNA:
- Watson-Crick base paring of nucleotides.

$\sim$ 60y.a. **Francis Crick**:
Helix structure of DNA:
- Watson-Crick base paring of nucleotides.

$\sim$ 40y.a. **Motoo Kimura**:
Neutral Theory of Molecular Evolution:
- Variation at molecular level likely result of genetic drift.

$\sim$ 40y.a. **Richard Lewontin**:
Molecular Diversity:
- Evolution at molecular level.

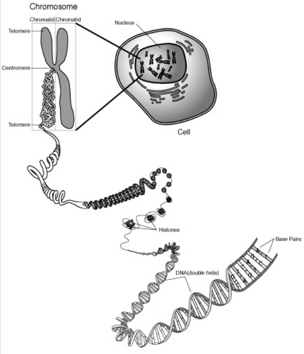
Origin of the Species

But...

Epigenetics: back to Lamarckism!

Origin of the Species

Genetics



Chromosomes:

- composed of Deoxyribonucleic acid: **Genetic fingerprint of individuals**;
- Located in nucleus (eukaryotes) or cytoplasm (prokaryotes);
- Double helix of base pairs: Adenine, Thymine, Guanine and Cytosine;
- Sequence of genes;
- Exons and Introns;
- Genome.

Sequence Space

- Individual
  - Chromosome .....AGGCACCGTAGTTTAATAAGGGCTA....
  - Gene .....AGGCACCGTAGTTTAATAAGGGCTA....
  - Exon .....AGGCACCGTAGTTTAATAAGGGCTA....
  - Intron .....AGGCACCGTAGTTTAATAAGGGCTA....
  - Genome
- Genome lives in Sequence Space

Organism	Length
Small Virus	10000
Bacterium	4 Million
Humans	3.5 Billion

Evolutionary Computation

Brief History

- Evolution with computers can be traced back to 1948 (Turing);
- First PhD in Computer Science (John Holland, 1959) popularised Genetic Algorithms;
- 1960s:

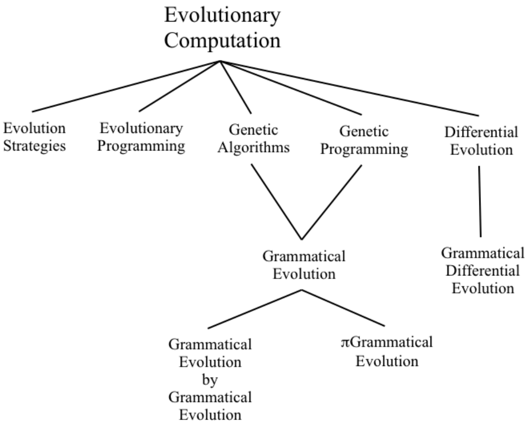
Genetic Algorithms
(Evolutionary Programming)

vs.

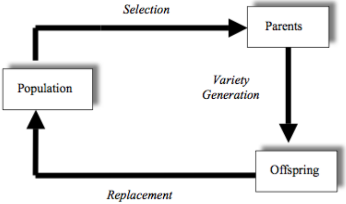
Evolution Strategies

- 1985: First Conference;
- 1992: Genetic Programming (1^{st} instance 1958!);
- 1990s: Unified under EC.

Evolutionary Computation



Evolutionary Computation

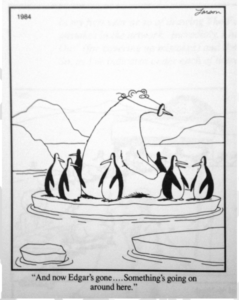


$$x[t+1] = r(v(s(x[t])))$$

Evolutionary Algorithm

```
Initialise Population;
While (termination condition FALSE):
   select Parents;
   create Offspring;
   Update Population;
EndWhile
```

Let's Evolve a Smart Polar Bear



- World's largest carnivore;
- Descendent of Brown Bear - Separate evolution for last 4-5 million years;
- Clear/White Fur;
- 4 Legs;
- Furred Soles;
- Broad Forepaws;
- Large and Stocky:
  - 1.8-2.5m length (tip of nose to tail);
  - 150-800kg.

Colour	Legs	Soles	Forepaws	Length	Weight
White	4	Furred	30.4cm	2.2m	785.4kg
Category	Integer	Boolean	Float	Float	Float

Polar Bear Example

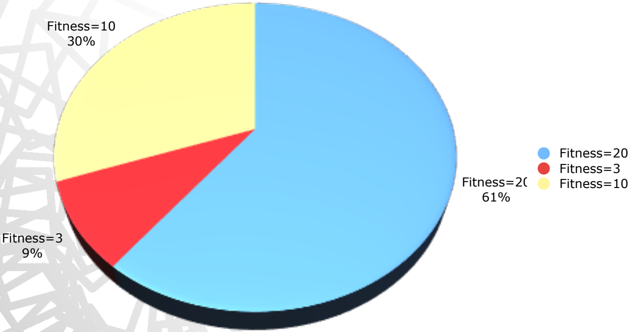
Colour	Legs	Soles	Forepaws	Length	Weight	**Fitness**
White	4	Furred	30.4cm	2.2m	785.4kg	**20 years**

Colour	Legs	Soles	Forepaws	Length	Weight	**Fitness**
Brown	4	Furred	29.9cm	1.1m	203.7kg	**3 years**

Colour	Legs	Soles	Forepaws	Length	Weight	**Fitness**
White	4	No Fur	15.4cm	1.8m	771.6kg	**10 years**

Average Fitness of Population= 11 years
Best Individual Fitness = 20 years

Polar Bear Example (Selection)



Polar Bear Example (Variation)

Parents:

Colour	Legs	Soles	Forepaws	Length	Weight	**Fitness**
White	4	Furred	30.4cm	2.2m	785.4kg	**20 years**
Brown	4	Furred	29.9cm	1.1m	203.7kg	**3 years**
White	4	No Fur	15.4cm	1.8m	771.6kg	**10 years**

Offspring:

Colour	Legs	Soles	Forepaws	Length	Weight	**Fitness**
White	4	Furred	31.2cm	2.2m	798.1kg	**23 years**
White	4	Furred	29.5cm	1.9m	778.1kg	**15 years**
White	4	No Fur	15.4cm	1.7m	741.6kg	**7 years**

Polar Bear Example (Replacement)

- Several approaches possible;
- Generational population (offspring replace parents).

New Population:

Colour	Legs	Soles	Forepaws	Length	Weight	**Fitness**
White	4	Furred	31.2cm	2.2m	798.1kg	**23 years**
White	4	Furred	29.5cm	1.9m	778.1kg	**15 years**
White	4	No Fur	15.4cm	1.7m	741.6kg	**7 years**

Average Fitness of Population= 15 years
Best Individual Fitness = 23 years

Evolutionary Computation

"Black Art" of EC

- Population-based search;
- Stochastic;
- **Design representation**;
- **Design fitness measure**;
- **Design algorithm (e.g., balanced variety generation operators and selection pressure)**.

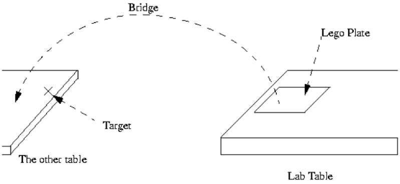
Evolutionary Computation

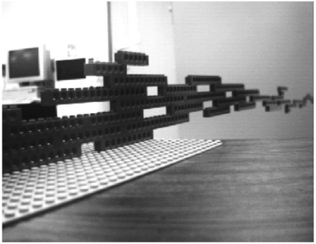
Applications

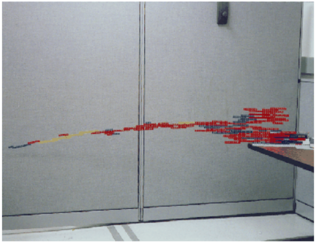
- Too many to list!;
- Engineering;
- Design;
- Sound Synthesis;
- Circuit Design;
- Games;
- Financial Modelling;
- Bioinformatics;
- Human-competitive results.

Evolutionary Computation

Funes & Pollack (1997)



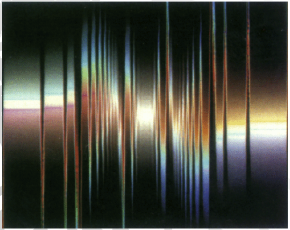




Evolutionary Computation

Karl Sims (1991)






Evolutionary Computation

Al Biles (1993)
GenJam

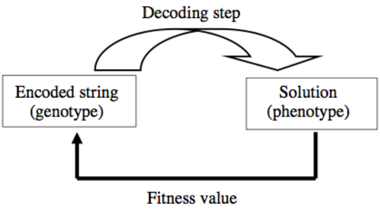
Evolutionary Computation

[video]

Genetic Algorithm

Overview

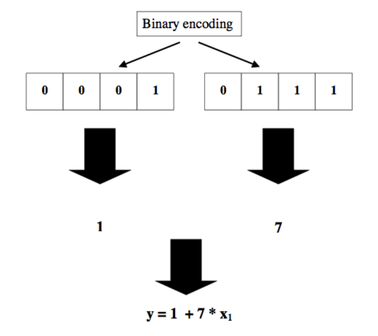
- Holland (1975), Goldberg (1989);
- Binary String individuals;
- Evolutionary search operates on encoding of solution (genotype);
- Decoding: genotype-phenotype map.



Genetic Algorithm

Representation

- Fixed-length chromosome;
- Each locus 1 bit;
- Fixed-size genes;
- Encode reals, ints.



Genetic Algorithm

Encoding

- Integers (binary vs. gray):
  - n bits encode 2^{n-1} ints.
- Reals:

$$x = \frac{\text{decoded integer}}{2^{n-1}}$$

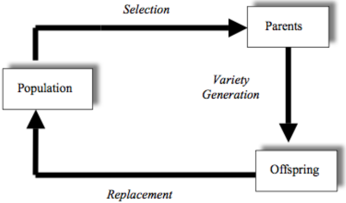
- Interval $a \longrightarrow b$:

$$z = a + x(b - a)$$

Integer Value	Binary Code	Gray Code
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100

- Beware of hamming cliffs!
- small phenotype change requires large genotype change
- 3 to 4 requires 3-bit changes
- Gray 1-bit change

Evolutionary Algorithm

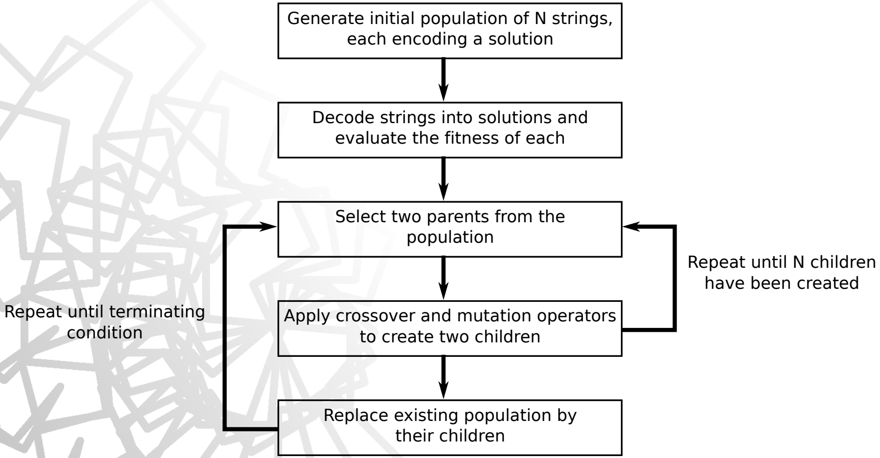


$$x[t+1] = r(v(s(x[t])))$$

Pseudo-code

```
Initialise Population;
While (termination condition FALSE):
   select Parents;
   create Offspring;
   Update Population;
EndWhile
```

Genetic Algorithm



Example Problem: ONEMAX

Definition

- **Maximise numbers of 1s in a bit string of length** n;
- Example: $n = 8$, popsize $= 4$, ...

1. Generate initial population:
   - Randomly assign 1 or 0 to each locus.
2. Calculate fitness:
   - Count number of 1s.

Candidate	String	Fitness
A	00000110	2
B	11101110	6
C	00100000	1
D	00110100	3

Example Problem: ONEMAX

Fitness

- Could normalise and standardise fitness:
  - Normalise between 0.0 and 1.0;
  - Standardise 1.0 is best, 0.0 is worst.

Candidate	String	Fitness	Normalised Fitness
A	00000110	2	0.25
B	11101110	6	0.75
C	00100000	1	0.125
D	00110100	3	0.375

Example Problem: ONEMAX

Selection

- Fitness proportionate selection:
  - Let f_i = fitness of individual i;
  - Average population fitness:

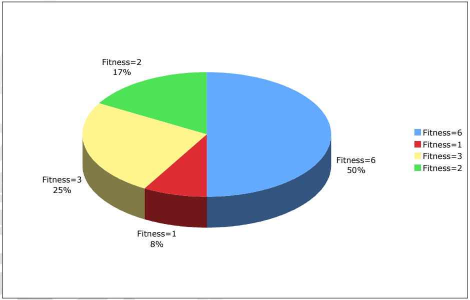
$$\bar{f} = \frac{1}{n}\sum_{i=1}^{N} f_i$$

  - Individual j selected with probability:

$$p_j = \frac{f_j}{\sum_{i=1}^{N} f_i}$$

Candidate	String	Fitness	Normalised Fitness	p_j
A	00000110	2	0.25	.17
B	11101110	6	0.75	.50
C	00100000	1	0.125	.08
D	00110100	3	0.375	.25

Example Problem: ONEMAX



$$r \in \left[0, \sum_{i=1}^{N} f_i\right) \qquad \sum_{i=1}^{j-1} f_i \leq r < \sum_{i=j}^{N} f_i$$

Example Problem: ONEMAX

Variation

3. Select two parents:

Candidate B	Candidate C
11101110	00100000

4. Crossover:

Candidate E	Candidate F
01101110	10100000

5. Mutation:

Candidate E	Candidate F
01001110	10100000

Example Problem: ONEMAX

Variation

3. Select two parents:

Candidate B	Candidate D
11101110	00110100

4. Crossover:

Candidate G	Candidate H
00111110	11100100

5. Mutation:

Candidate G	Candidate H
10111110	11000100

Example Problem: ONEMAX

Replacement

- Generational Replacement Strategy:

6. Replace parents with offspring:

Candidate	String	Fitness
A	00000110	2
B	11101110	6
C	00100000	1
D	00110100	3

Candidate	String	Fitness
E	01001110	4
F	10100000	2
G	10111110	6
H	11000100	3

7. Unless termination criteria met, go to step 3.

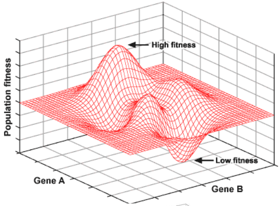
Exploration vs. Exploitation

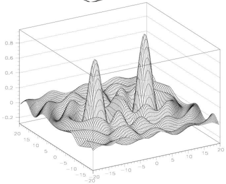
Black art of EC

- Careful choice of **Selection**, **Variation** and **Replacement** operators.
- Rate of convergence;
- Local Optima.

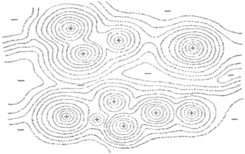
Exploration vs. Exploitation

Adaptive Landscape





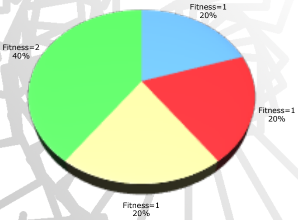
- Wright (1932)

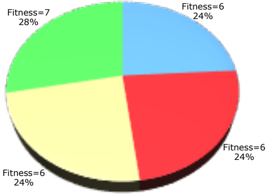


Selection

Roulette Wheel Selection (Fitness Proportionate Selection)

- High selection pressure earlier on;
- Premature convergence;
- Low selection pressure later:
  - Similar fitness values;
  - Uniform probability of selection;

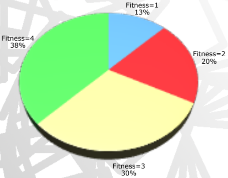




Selection

Rank Selection

- Rank from worst to best and calculate rescaled fitness;
- Linear ranking: $f_{\text{rank}} = 2 - P + 2 \times (P - 1) \times \frac{(\text{rank}-1)}{(n-1)}$
- Non-proportional selection.



Ranking	1	2	3	4
Fitness	1	2	3	4
f_{rank}	0.5	0.8	1.2	1.5
p_j	0.125	0.2	0.3	0.375

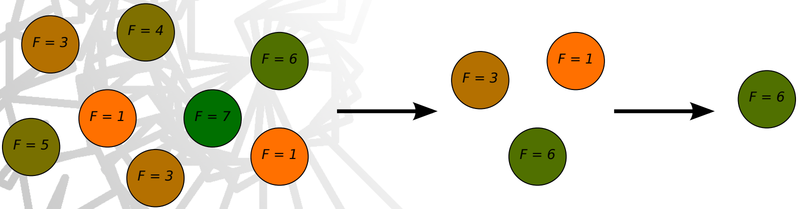
Ranking	1	2	3	4
Fitness	1	3	6	7
f_{rank}	0.5	0.8	1.2	1.5
p_j	0.125	0.2	0.3	0.375

P=1.5

Selection

Tournament Selection

- Select t individuals at random;
- Best of t individuals becomes parent.
- Selection pressure easily adjustable ($t \in [1..N]$);
- Can force fair tournament and unique parents.



Variation

Crossover and Mutation

- Mutation introduces novelty:
  - Too little. . . stuck in local optima;
  - Too much. . . random search!
- Crossover should exploit (share) good subsolutions;
- Exploration/exploitation balance;
- Adaptive mutation.

Variation

Alternative Crossover

- 2-point Crossover:

Parent 1	Parent 2
11101110	00110100

Offspring 1	Offspring 2
11110110	00101100

- Uniform Crossover:

Parent 1	Parent 2
11101110	00110100

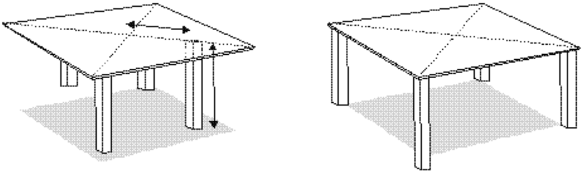
Offspring 1	Offspring 2
11100100	00111110

Replacement

Alternative Replacement

- Generational:
  - Children replace parents;
- Elitism:
  - Keep best fitness individual(s);
  - Generational for remainder.
- Steady-state:
  - Sort parents and offspring ($2N$), choose N best;
  - Can apply at variation operator level:
    - Two parents produce two children;
    - Best of four individuals make it to offspring population.

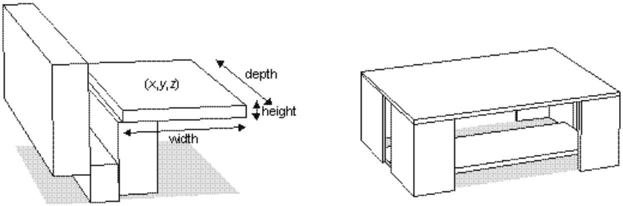
Example Encodings



Table

- Consisting of fixed top and four legs defined by:
  - Length of leg 1, distance of leg 1 from centre;
  - Length of leg 2, distance of leg 2 from centre;
  - Length of leg 3, distance of leg 3 from centre;
  - Length of leg 4, distance of leg 4 from centre.
- Genotype: 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000
  (l1, d1, l2, d2, l3, d3, l4, d4)

Example Encodings



Table

- Consisting of several 3-dimensional blocks:
  - x_1, y_1, z_1, width_1, height_1, depth_1;
  - x_2, y_2, z_2, width_2, height_2, depth_2;
  - ...
- Genotype: 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 (x_1 y_1 z_1 width_1 height_1 depth_1 ...)

Example Encodings

Polar Bear

Colour	Legs	Soles	Forepaws	Length	Weight
White	4	Furred	30.4cm	2.2m	785.4kg
Category	Integer	Boolean	Float	Float	Float
000	0100	1	0101010011001101	11001010	1101011001011100

$$x = \frac{\text{decoded integer}}{2^{n-1}} \qquad a \rightarrow b \qquad z = a + x(b-a)$$

Example Encodings

Tokaido-Sanyo Shinkansen N7000

Assessing Performance

Monitoring

- Never draw conclusions from a single run;
- Use sufficient number of runs ($R > 30$);
- Use statistical measures (averages, medians, std. dev. , etc);
- Record as much data from your population as possible:
  - Mean, Best, Worst Fitness at each generation;
  - Diversity (genotypes, phenotypes);
  - Graph progress of these.
- Use controls:
  - Compare to equivalent Random Search.

GA Literature

Sample of references...

- Holland (1975). Adaptation in Natural and Artificial Systems;
- Goldberg (1989). Genetic Algorithms in Search, Optimization and Machine Learning;
- Mitchell (1996). An Introduction to Genetic Algorithms;
- Fogel D (ed) (1998). Evolutionary Computation: The Fossil Record. IEEE Press;
- Fogel (2000). Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. IEEE Press;
- Goldberg (2002). The Design of Innovation: Lessons from and for Competent Genetic Algorithms. Kluwer;
- Rothlauf F (2002) Representations for Genetic and Evolutionary Algorithms. Physica-Verlag.