Genetic Algorithms

CS 419/519

Evolutionary Computing: the Origins

Outline

- Historical perspective
- Biological inspiration:
  - Darwinian evolution theory (simplified!)
  - Genetics (simplified!)
- Motivation for EC

Historical perspective

- 1948, Turing:
  proposes "genetical or evolutionary search"
- 1962, Bremermann:
  optimization through evolution and recombination
- 1964, Rechenberg:
  introduces evolution strategies
- 1965, L. Fogel, Owens and Walsh:
  introduce evolutionary programming
- 1975, Holland:
  introduces genetic algorithms
- 1992, Koza:
  introduces genetic programming

Historical perspective

- 1985: first international conference (ICGA)
- 1990: first international conference in Europe (PPSN)
- 1993: first scientific EC journal (MIT Press)
- 1997: launch of European EC Research Network EvoNet

Historical perspective

EC in the early 21st Century:

- 3 or 4 major EC conferences, about 10 small related ones
- 4 scientific core EC journals
- 1000+ EC-related papers published each year(estimate)
- uncountable (meaning: many) applications
- uncountable (meaning: ?) consultancy and R&D firms
- part of some university curricula

Vocabulary

- Gene – a section of DNA that encodes a trait (e.g. eye color); the unit of heredity
- Alleles – different forms (values) of a gene (e.g. brown eyes and blue eyes result from different alleles for the eye color gene)

Vocabulary

- Genome – all of the genetic information for an individual
- Chromosome – a sequence of genes; a genome consists of 23 pairs of chromosomes

Vocabulary

- Genotype – the combination of alleles for an individual; may refer to entire genome or to the alleles for a specific locus in the genome
- Phenotype – an individual's observable characteristics; influenced by genotype and environment
- Heritable – a characteristic that can be passed from parent to offspring

Darwinian Evolution: Survival of the fittest

- All environments have finite resources
  (i.e., can only support a limited number of individuals)
- Life forms have basic instinct / lifecycles geared towards reproduction
- Therefore some kind of selection is inevitable
- Those individuals that compete for the resources most effectively have increased chance of reproduction
- Note: fitness in natural evolution is a derived, secondary measure, i.e., we (humans) assign a high fitness to individuals with many offspring

Darwinian Evolution: Diversity drives change

- Phenotypic traits:
  - Behavior / physical differences that affect response to environment
  - Partly determined by inheritance, partly by factors during development
  - Unique to each individual, partly as a result of random changes
- If a phenotypic trait:
  - Leads to higher chances of reproduction
  - Can be inherited

  then it will tend to increase in subsequent generations, leading to new combinations of traits …

Darwinian Evolution: Summary

- Population consists of set of diverse individuals
- Combinations of traits that are better adapted tend to increase representation in population

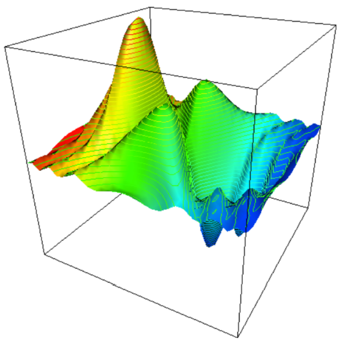
  Individuals are "units of selection"
- Variations occur through random changes yielding constant source of diversity, coupled with selection means that:

  Population is the "unit of evolution"
- Note the absence of "guiding force"

Adaptive landscape metaphor (Wright, 1932)

- Can envisage population with *n* traits as existing in a *n+1*-dimensional space (landscape) with height corresponding to fitness
- Each different individual (phenotype) represents a single point on the landscape
- Population is therefore a "cloud" of points, moving on the landscape over time as it evolves – adaptation

Adaptive landscape metaphor (Wright, 1932)



Adaptive landscape metaphor (cont'd)

- Selection "pushes" population up the landscape
- Genetic drift:
  - random variations in feature distribution as some members die or do not reproduce
  - can be positive or negative
  - can cause the population to "melt down" hills, thus crossing valleys and leaving local optima
  - no guarantee of population recovering from negative effects

Genetics: Natural

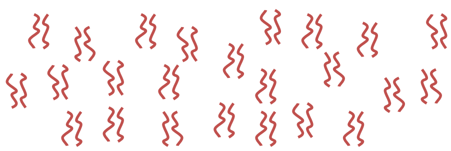
- The information required to build a living organism is coded in the DNA of that organism
- Genotype (DNA inside) determines phenotype
  - (environment also plays a role)
- Genes → phenotypic traits is a complex mapping
  - One gene may affect many traits (pleiotropy)
  - Many genes may affect one trait (polygeny)
  - (i.e. there is not a one-to-one mapping)
- Small changes in the genotype lead to small changes in the organism (e.g., height, hair colour)

Genetics: Genes and the Genome

- Genes are encoded in strands of DNA called chromosomes
- In most cells, there are two copies of each chromosome (diploid)
- The complete genetic material in an individual's genotype is called the Genome
- Within a species, most of the genetic material is the same

Genetics:
Example: Homo Sapiens

- Human DNA is organised into chromosomes
- Human body cells contain 23 pairs of chromosomes which together define the physical attributes of the individual:

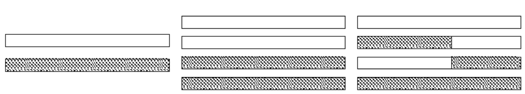


Genetics:
Reproductive Cells

- Gametes (sperm and egg cells) contain 23 individual chromosomes rather than 23 pairs
- Cells with only one copy of each chromosome are called haploid (as opposed to diploid)
- Gametes are formed by a special form of cell splitting called meiosis
- During meiosis the pairs of chromosomes undergo an operation called *crossing-over*

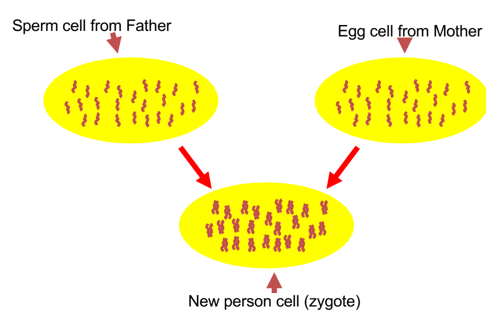
Genetics:
Crossing-over during meiosis

- Chromosome pairs align and duplicate
- Inner pairs link at a *centromere* and swap parts of themselves



- Outcome is one copy of maternal/paternal chromosome plus two entirely new combinations
- After crossing-over one of each pair goes into each gamete
- Because there are 23 chromosomes (in humans), and resulting gametes get one of each, it is highly likely that the gametes are distinct from the parent genome facilitating variation in offspring.

Genetics:
Fertilisation



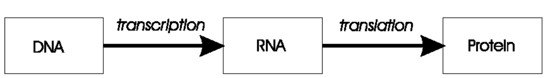
Genetics: After fertilisation

- New zygote rapidly divides creating many cells all with the same genetic contents
- Although all cells contain the same genes, depending on, for example where they are in the organism, they will behave differently
- This process of differential behaviour during development is called ontogenesis
- All of this uses, and is controlled by, the same mechanism for decoding the genes in DNA

Genetics: Genetic code

- All proteins in life on earth are composed of sequences built from 20 different amino acids
- DNA is built from four nucleotides in a double helix spiral: purines Adenine, Guanine; pyrimidines Thymine, Cytosine
- Triplets of these form *codons*, each of which codes for a specific amino acid
- Much redundancy:
  - purines complement pyrimidines (A with T; C with G)
  - $4^3 = 64$ possible codons which code for 20 amino acids
  - genetic code = the mapping from codons to amino acids
- For all natural life on earth, the genetic code is the same !

Genetics: Transcription, translation



A central claim in molecular genetics: only one way flow

Genotype → Phenotype

Genotype ↚ Phenotype

Lamarckism (saying that acquired features can be inherited) is thus wrong!

Genetics: Mutation

- Occasionally some of the genetic material changes very slightly during this process (replication error)
- This means that the child might have genetic material information not inherited from either parent
- This can be
  - catastrophic: offspring in not viable (most likely)
  - neutral: new feature does not influence fitness
  - advantageous: strong new feature occurs
- Redundancy in the genetic code forms a good way of error checking

Motivation for evolutionary computing

- Nature has always served as a source of inspiration for engineers and scientists
- The best problem solvers known in nature are:
  - the (human) brain that created "the wheel, New York, wars and so on" (Douglas Adams' Hitch-Hikers Guide to the Galaxy)
  - the evolution mechanism that created the human brain (Darwin's Origin of Species)

- Answer 1 → neurocomputing
- Answer 2 → evolutionary computing

Motivation for evolutionary computing

- Developing, analyzing, applying problem solving methods a.k.a. algorithms is a central theme in mathematics and computer science
- Time for thorough problem analysis and tailored algorithm design decreases
- Complexity of problems to be solved increases
- Consequence: ROBUST, GENERAL PROBLEM SOLVING technology is needed

GAs can "think" outside the (human) box

Space station boom design for vibration reduction
Note that there is no symmetry and no obvious design logic