Theory

Chapter 11

Overview

- Motivations and problems
- Holland's Schema Theorem
  - Derivation, Implications, Refinements
- Dynamical Systems & Markov Chain Models
- Statistical Mechanics
- Reductionist Techniques
- Techniques for Continuous Spaces
- No Free Lunch ?

Why Bother with Theory?

- Might provide performance guarantees
  - Convergence to the global optimum can be guaranteed providing certain conditions hold
- Might aid better algorithm design
  - Increased understanding can be gained about operator interplay etc.
- Mathematical Models of EAs also inform theoretical biologists
- Because you never know ….

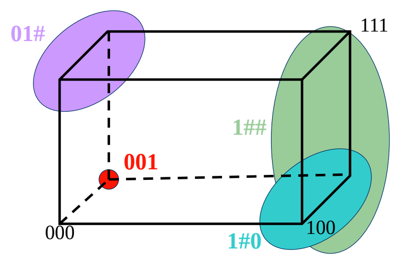
Problems with Theory ?

- EAs are vast, complex dynamical systems with many degrees of freedom
- The type of problems for which they do well, are precisely those it is hard to model
- The degree of randomness involved means
  - stochastic analysis techniques must be used
  - Results tend to describe average behaviour
- After 100 years of work in theoretical biology, they are still using fairly crude models of very simple systems ….

Holland's Schema Theorem

- A schema (*pl.* schemata) is a string in a ternary alphabet (0,1 **#** = "don't care") representing a hyperplane within the solution space.
  - E.g. 0001# #1# #0#, ##1##0## etc
- Two values can be used to describe schemata,
  - the ***Order*** (number of defined positions) = 6,2
  - the ***Defining Length*** - length of sub-string between outmost defined positions = 9, 3

Example Schemata



H	o(H)	d(H)
001	3	2
01#	2	1
1#0	2	2
1##	1	0

Schemata

- Aggregation
  - Simpler to analyse the effect of evolutionary operators of schemata then on individuals
  - Need to show how
    - operator increase fitness of the schemata
    - Or disrupts the fitness of schemata in population
- Implicit parallelism
  - a schema represents many possible individuals
  - Manipulating one schema affects many individuals
  - Computationally efficient
  - Often cited as on of the performance advantages of EA mechanism

Schema Fitnesses

- The true "fitness" of a schema *H* is taken by averaging over all possible values in the "don't care" positions, but this is effectively sampled by the population, giving an estimated fitness *f(H)*
- With Fitness Proportionate Selection
  $P_s(\text{instance of } H) = n(H,t) * f(H,t) / (\langle f \rangle * \mu)$
  therefore proportion in next parent pool is:
  $m'(H,t+1) = m(H,t) * f(H,t) / \langle f \rangle$

<f>: mean population fitness

Schema Disruption I

- One Point Crossover selects a crossover point at random from the *l-1* possible points
- For a schema with defining length *d* the random point will fall inside the schema with probability $= d(H) / (l-1)$.
- If recombination is applied with probability P_c the survival probability is $1.0 - P_c * d(H)/(l-1)$

Schema Disruption II

- The probability that bit-wise mutation with probability P_m will NOT disrupt the schemata is simply the probability that mutation does NOT occur in any of the defining positions,

$P_{survive}$ (mutation) $= (1 - Pm)^{o(H)}$

$= 1 - o(H) * P_m + \text{terms in } P_m^2 + \ldots$

- For low mutation rates, this survival probability under mutation approximates to $1 - o(h) * P_m$

o(h): Order of Schema h

The Schema Theorem

- Put together, the proportion of a schema *H* in successive generations varies as:

$$m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{<f>} \cdot \left[1 - \left(p_c \cdot \frac{d(H)}{l-1}\right)\right] \cdot [1 - p_m \cdot o(H)].$$

- Condition for schema to increase its representation is:

$$\frac{f(H)}{<f>} > \left[1 - \left(p_c \cdot \frac{d(H)}{l-1}\right)\right] \cdot [1 - p_m \cdot o(H)]$$

- Inequality is due to convergence affecting crossover disruption, exact versions have been developed

Implications 1: Operator Bias

- One Point Crossover
  - less likely to disrupt schemata which have **short** defining lengths relative to their order, as it will tend to keep together adjacent genes
  - this is an example of ***Positional Bias***
- Uniform Crossover
  - No positional bias since choices independent
  - BUT is far more likely to pick 50% of the bits from each parent, less likely to pick (say) 90% from one
  - this is called ***Distributional Bias***
- Mutation
  - also shows Distributional Bias, but not Positional

Operator Biases ctd

- Operator Bias has been extensively studied by Eschelman and Schaffer (empirically) and theoretically by Spears & DeJong.
- Results emphasise the importance of utilising all available problem specific knowledge when choosing a representation and operators for a new problem

Implications 2:The Building Block Hypothesis

- Closely related to the Schema Theorem is the “Building Block Hypothesis” (Goldberg 1989)
- This suggests that Genetic Algorithms work by discovering and exploiting “building blocks” - groups of closely interacting genes - and then successively combining these (via crossover) to produce successively larger building blocks until the problem is solved.
- Has motivated study of ***Deceptive*** problems
  - Based on the notion that the lower order schemata within a partition lead the search in the opposite direction to the global optimum
  - If interested look up Walsh functions

Criticisms of the Schema Theorem

- It presents an inequality that does not take into account the constructive effects of crossover and mutation
  - Exact versions have been derived
  - Have links to Price’s theorem in biology
- Because the mean population fitness, and the estimated fitness of a schema will vary from generation to generation, it says ***nothing*** about gen. t+2 etc.
- “Royal Road” problems constructed to be GA-easy based on schema theorem turned out to be better solved by random mutation hill-climbers
- BUT it remains a useful conceptual tool and has historical importance

Theories

- Many others exist
  - Landscape Metrics
  - Markow Chains
  - Vose’s Dynamical Systems Model
  - Statistical Mechanics Analysis
  - Reductionist Approaches
- Read the book…

Other Landscape Metrics

- As well as epistasis and deception, several other features of search landscapes have been proposed as providing explanations as to what sort of problems will prove hard for GAs
  - fitness-distance correlation
  - number of peaks present in the landscape
  - the existence of plateaus
  - all these imply a neighbourhood structure to the search space.
- It must be emphasised that these only hold for one operator

Markov Chain Analysis

- A system is called a Markov Chain if
  - It can exist only in one of a finite number of states
  - So can be described by a variable X^t
  - The probability of being in any state at time *t+1* depends only on the state at time *t*.
- Frequently these probabilities can be defined in a transition matrix, and the theory of stochastic processes allows us to reason using them.
- Has been used to provide convergence proofs
- Can be used with *F* and *M* to create exact probabilistic models for binary coded Gas, but these are huge

No Free Lunch Theorems

- **IN LAYMAN'S TERMS,**
  - **Averaged over all problems**
  - **For any performance metric related to number of distinct points seen**
  - **All non-revisiting black-box algorithms will display the same performance**
- **Implications**
  - **New black box algorithm is good for one problem => probably poor for another**
  - **Makes sense not to use "black-box algorithms"**
- **Lots of ongoing work showing counter-examples**

No Free Lunch

- Assumptions
  - Black box – uses no problem instance or specific knowledge
  - Nonrevisiting – means in a generate and test scenario the same point is not used twice
  - EAs revisit but avoid this by keeping a history and excluding children already seen
- Implications
  - If you make an algorithm that appears to be the best for solving some class of problems then it pays for it by not being very good for another class.
  - For a given problem can avoid the NFL problem by building in problem-specific knowledge