

Lecture 10: Schema Theory

Schema Theory

- ▶ Schema theory is Holland's original theoretical explanation for how GAs work
- ▶ The theory is based on the concept of a *schema* (plural *schemata*)
 - ▶ Definition: For chromosomes of length l in an alphabet A , a schema is a subset of the space A^l in which all the chromosomes share a particular set of defined values
 - ▶ Schemata are represented as strings in the original alphabet extended with a wildcard symbol (e.g. $*$)
 - ▶ E.g. for $A = \{0, 1\}$, the schema $(1 + *)$ represents the chromosomes $\{(100), (101), (110), (111)\}$
 - ▶ Schemata have two characteristics, *order* and *length*
 - ▶ Definition: The order of a schema is the number of defined (non-wildcard) positions it has
 - ▶ Definition: The length of a schema is the distance between its first and last defined positions

Schema Theory

- ▶ Schemata can be interpreted in three obvious ways
 - ▶ *Set-theoretically*
 - ▶ Schemata define subsets of the set of all possible chromosomes
 - ▶ *Geometrically*
 - ▶ Schemata define hyperplanes on an l -dimensional hypercube
 - ▶ *Functionally*
 - ▶ Schemata define periodic functions of different frequency

Implicit Parallelism

- ▶ Schema theory is based on the idea of *implicit parallelism*
- ▶ Any chromosome is a member of 2^l schemata
 - ▶ For each locus the chromosome can be represented by a schema having either the chromosome's actual value at its corresponding locus, or the wildcard $*$
- ▶ A population of size N could contain up to $N2^l$ different schemata
 - ▶ However the usual figure will be much lower, due to similar chromosomes, particularly as the population converges
- ▶ By evaluating a chromosome's fitness, according to Holland's ideas, we are simultaneously evaluating many schemata
 - ▶ I.e. we are parallelising the search of the solution space

The Schema Theorem

- ▶ The *Schema Theorem* analyses what happens to schemata as the GA executes
- ▶ We shall present the theorem based on the Simple GA
 - ▶ I.e. fitness-proportional, generational selection, 1X and single-point mutation.
- ▶ First we must define the fitness of a schema S at time t as the mean fitness of the strings in the population that are members of it

$$f(S, t) = \frac{\sum_{x \in S \cap P_t} f(x)}{|S \cap P_t|}$$

- ▶ Next, we define the fitness ratio of the schema's fitness to the population mean fitness

$$r(S, t) = \frac{f(S, t)}{\bar{f}(t)}$$

The Schema Theorem

- ▶ First we analyse how the expected number of instances of a schema changes through selection alone
- ▶ From the definition of fitness proportional selection, this is given by

$$E(N_{S,t+1}) = r(S, t)N_{S,t}$$

- ▶ This ignores the constructive and destructive effects of genetic operators on schemata

The Schema Theorem

- ▶ Next we calculate the probability with which a schema may be *destroyed by crossover* alone as

$$1 - \chi \frac{l_S}{l - 1} P_d(S, t)$$

- ▶ where l_S is the length of schema S , and P_d is the probability the other parent is not a member of schema S
- ▶ N.B. we ignore the possibility of another instance of schema S being created through crossover
 - ▶ The Schema Theorem derives results for what happens to a schema, *independent* of other schemata in the population
 - ▶ Hence the probability above is a lower-bound on the probability that a schema is transmitted to the next generation
 - ▶ To make the probability fully independent of other schemata in the population, we can set $P_d(S, t) = 1$, so the independent lower bound is

$$1 - \chi \frac{l_S}{l - 1}$$

The Schema Theorem

- ▶ Now we calculate the probability with which a schema may be *destroyed by mutation* alone

$$1 - \mu k_S$$

- ▶ where k_S is the order of schema S
- ▶ Again, this probability is a lower bound, as we are not considering the possibility that an instance of schema S can be created through mutation
- ▶ Also, the actual probability is

$$(1 - \mu)^{k_S} \geq 1 - \mu k_S$$

- ▶ Our simplified version (the r.h.s. of the inequality) comes from assuming μ is small, so we can ignore terms μ^2 for $q > 1$

The Schema Theorem

- ▶ Finally, we put everything together to give the Schema Theorem

$$E(N_{S,t+1}) \geq (1 - \chi \frac{l_S}{l - 1} - \mu k_S) r(S, t) N_{S,t}$$

- ▶ Interpreting the above expression as involving a selection component and a loss component gives a nice parallel with the Breeder's Equation from quantitative genetics

$$R = sh^2$$

- ▶ where R is the response to selection, s is the selection coefficient, and h^2 is the heritability coefficient (N.B. *not* h raised to the 2nd power)

- ▶ The Schema Theorem is used to make several further arguments
 - ▶ The Building Block Hypothesis
 - ▶ The Two-Armed Bandit Analogy
 - ▶ The Principle of Minimum Alphabets

- ▶ From the Schema Theorem we know that the representation of a schema S in the population should increase when

$$r(S, t) \geq 1 + \chi \frac{f_S}{\bar{f}} + \mu k_S$$

- ▶ Short, low-order schemata will be unlikely to be disrupted by crossover and mutation, hence require a fitness ratio only slightly above 1 to spread
- ▶ Long, high-order schemata are much more vulnerable to disruption, so require correspondingly higher fitness ratios
- ▶ So, a GA should work well when it can combine short, low-order schemata (*building blocks*) to form better solutions
- ▶ The assumption that this is how a GA works is termed the *building block hypothesis*

- ▶ Holland drew an analogy between competition amongst schemata, and the two-armed bandit
- ▶ The two-armed (or n -armed) bandit is a problem from statistical decision theory
 - ▶ Given a 'bandit' with two (or n) arms each returning different, noisy rewards, what is the best trial-allocation strategy to maximise future expected reward?
- ▶ Holland argued that by allocating exponentially increasing numbers of trials to the superior schemata, the GA approximates the optimal strategy
 - ▶ For the GA there are actually several bandits, all with many more than two arms
 - ▶ As we are solving a non-linear problem, the order in which the bandits are solved is likely to be important
 - ▶ Despite this, Holland argued the theory could still be applied by placing a lower bound on the loss incurred by choosing an incorrect schema competition winner

The Principle of Minimal Alphabets

- ▶ Holland also argued for the optimality of binary encoding, using schema theory
- ▶ The *implicit parallelism* argument suggests we should try to maximise the number of schemata processed simultaneously
- ▶ The number of possible schemata for an alphabet A is $|A + 1|^l$
- ▶ This will be maximised when chromosome length l is maximised, which to store some fixed amount of information occurs when $|A|$ is minimised
 - ▶ The smallest possible value of $|A|$ is 2, i.e. binary encoding
 - ▶ So the maximum number of possible schemata is achieved by binary encoding

Summary

- ▶ The Schema Theory tracks changes in frequency of representation of subsets of the search space (schemata), independent of the other schemata in the population
 - ▶ Its formulation reveals the conflicting pressures of positive selection, and destructive genetic operators
- ▶ It has been used to make some proposals about how GAs (should) work
 - ▶ The Building Block Hypothesis
 - ▶ The Two-Armed Bandit Analogy
 - ▶ The Principle of Minimum Alphabets
- ▶ But is it a good explanation of GA behaviour?