

## Holland’s GA Schema Theorem

- ❖ Objective – provide a formal model for the effectiveness of the GA search process.
- ❖ In the following we will first approach the problem through the framework formalized by Holland [1] and popularized by Goldberg [2].
- ❖ This concentrates on providing a model for the expectation of schema survival, where this naturally represents a limitation in itself.
- ❖ We will then move on to consider further drawbacks of the original ‘Schema Theorem’ and some of the recent attempts to provide more representative Schema Theorems.
- ❖ Consider the case of a canonical GA,
  - Binary alphabet;
  - Fixed length individuals of equal length,  $l$ ;
  - Fitness Proportional Selection;
  - Single Point Crossover;
  - Gene wise mutation.
- ❖ **Definition 1** – Schema,  $H$ .
  - A schema is a subset of the space of all possible individuals for which all the genes match the template for schema  $H$ .
- ❖ If  $A$  denotes the alphabet of gene alleles then  $A \square *$  is the schema alphabet, where  $*$  is the ‘wild card’ symbol matching any allele value.
  - E.g. for the binary alphabet  $A \square \{0, 1, *\}$  where  $*$   $\square \{0, 1\}$

### Example

- ❖ For a binary individual with the gene sequence  $\{0\ 1\ 1\ 1\ 0\ 0\ 0\}$ , then it follows that one (of many) matching schema might have the form,  $H = [*\ 1\ 1\ * \ 0\ * \ *]$
- ❖ The schema  $H = [0\ 1\ * \ 1\ *]$  identifies the chromosome set,

0	1	0	1	0
0	1	0	1	1
0	1	1	1	0
0	1	1	1	1

- ❖ Needless to say, not all schema are created equal, thus,  $1\ * \ * \ * \ * \ * \ *$  does not tell us as much as  $0\ 1\ * \ * \ 1\ 1\ 0$ .

- ❖ Moreover, schema  $1^{* * * * *} 0$  spans the entire length of an individual whereas  $1^{*} 1^{* * * * *}$  does not.
- ❖ **Definition 2** – Schema Order,  $o(H)$ .
  - Schema order,  $o(\cdot)$ , is the number of non ‘\*’ genes in schema  $H$ .
  - **Example**,
    - $o(1^{* * * *} 0^{* * * *}) = 1$
- ❖ **Definition 3** – Schema Defining Length,  $\Delta(H)$ .
  - Schema Defining Length,  $\Delta(H)$ , is the distance between first and last non ‘\*’ gene in schema  $H$ .
  - **Example**,
    - $\Delta(1^{* * * *} 0^{* * * *}) = 4 - 1 = 3$
- ❖ **Note**
  - For cardinality,  $k$ , there are  $(k + 1)^l$  schema in string of length ‘ $l$ ’.
  - We are now in a position to incrementally introduce the effect of the various selection and search operators associated with the above definition for the special case of a canonical GA with binary genes.

### Selection Operators – Fitness Proportional Selection

- ❖ Essentially all that we are attempting to model is the probability that individual,  $h$ , samples schema,  $H$ , or
  - $P(h \in H)$
- ❖ There are several ways of modeling this.
  - Consider the following two part model,
    - Probability of selection = number of instances of schema  $H$  in the population;
    - Probability of selection = average fitness of schema  $H$  relative to the average fitness of all individuals in the population.
    - Thus,

$$P(h \in H) = \frac{\text{(Number of individuals matching schema } H \text{ at generation } t)}{\text{(Population Size)}} \cdot \frac{\text{(Mean fitness of individuals matching schema } H)}{\text{(Mean fitness of individuals in the population)}}$$

$$\text{or, } P(h \sqsupseteq H) = \frac{m(H,t)f(H,t)}{M\bar{f}(t)}$$

where  $m(H, t)$  is the number of instances of schema  $H$  at generation  $t$ .

- ❖ **Lemma 1** – Under fitness proportional selection the *expected* number of instances of schema  $H$  at time  $t$  is,

$$E[m(H, t + 1)] = M \sqsupseteq P(h \sqsupseteq H) = \frac{m(H,t)f(H,t)}{\bar{f}(t)}$$

- ❖ Implication,
  - Schemas with fitness greater (lower) than the average population fitness are likely to account for proportionally more (less) of the population at the next generation.
  - Strictly speaking for accurate estimates of expectation and probability the population size should be infinite.

### Search Operators – Single point crossover

- ❖ Reproduction does nothing to improve the fitness of individuals in a population. (Single point) Crossover was the first of two search operators introduced to modify the distribution of schema in the population. In his original work, Holland concentrated on modeling the lower bound alone [1].
- ❖ Consider the following individual,  $h$ , two matching schema,  $H_1, H_2$  and crossover point between 3<sup>rd</sup> and 4<sup>th</sup> gene, or,

$h =$	1	0	1		1	1	0	0
$H_1 =$	*	0	1		*	*	*	0
$H_2 =$	*	0	1		*	*	*	*

- ❖ Observations,
  - Schema  $H_1$  will naturally be broken by the location of the crossover operator unless the second parent is able to ‘repair’ the disrupted gene.
  - Schema  $H_2$  emerges unaffected and is therefore independent of the second parent.
  - Schema with long defining length are more likely to be disrupted by single point crossover than schema using short defining lengths.
- ❖ **Lemma 2** – Under single point crossover, the (lower bound) probability of schema  $H$  surviving at generation  $t$  is,

$$P(H \text{ survives}) = 1 - P(H \text{ does not survive})$$

$$\text{or, } 1 - \frac{o(H)}{l} p_c P_{diff}(H, t)$$

Where  $P_{diff}(H, t)$  is the probability that the second parent does not match schema  $H$ ; and  $p_c$  is the *a priori* selected threshold of applying crossover.

❖ Comment,

➤ In the special case of the worst case lower bound  $P_{diff}(H, t) = 1$ .

### Search Operators – Gene wise Mutation

❖ Mutation is applied gene by gene.

❖ In order for schema  $H$  to survive, all non \* genes in the schema must remain unchanged.

❖ Probability of not changing a gene is,

$$(1 - p_m)$$

❖ Require that all  $o(H)$  non \* genes survive, or

$$(1 - p_m)^{o(H)}$$

❖ Typically the probability of applying the mutation operator,  $p_m, \ll 1$ , thus

$$(1 - p_m)^{o(H)} \approx 1 - o(H) p_m$$

❖ **Lemma 3** – Under gene wise mutation, the (lower bound) probability of an order  $o(H)$  schema  $H$  surviving at generation  $t$  is,

$$1 - o(H) p_m$$

❖ **Theorem 1** – The Schema Theorem [1, 2]

➤ The expected number of schema  $H$  at generation  $t + 1$  when using a canonical GA with proportional selection, single point crossover and gene wise mutation (where the latter are applied at rates  $p_c$  and  $p_m$ ) is,

$$E[m(H, t + 1)] \geq \frac{m(H, t) f(H, t)}{\bar{f}(t)} \left[ 1 - p_c \frac{o(H)}{l} P_{diff}(H, t) - o(H) p_m \right]$$

❖ Comments,

➤ The theorem is described in terms of expectation, thus strictly speaking is only true for the case of a population with an infinite number of members. In the case of finite population sizes the significance of population drift plays an increasingly important role [3].

➤ The above form is naturally specific to the selection and search operators under which it was derived. A more generic form for the Schema Theorem might take the form,

$$E[m(H, t + 1)] \geq m(H, t) \sigma(H, t) \{1 - \sigma(H, t)\}$$

- Where  $\sigma(H, t)$  is the ‘selection coefficient’ and  $\tau(H, t)$  is the ‘transcription error’ [4].
- This makes the various contributions more explicit. Specifically for schema  $H$  to survive then,

$$\sigma(H, t) \geq \{1 - \tau(H, t)\}$$

$$\text{or, } \frac{f(H, t)}{\bar{f}(t)} \geq \left[ 1 - p_c \frac{\sigma(H)}{1 - l} p_{diff}(H, t) - o(H) p_m \right]$$

- This is the basis for the observation that short (defining length), low order schema of above average population fitness will be favored by canonical GAs, or the *Building Block Hypothesis* [1, 2].

## Problems

- ❖ The Schema Theorem as defined by Holland represented a mile stone in the development of Genetic Algorithms in particular and latter in the development of corresponding theorems for Genetic Programming. However, it has also several significant shortcomings which have lead to more modern approaches to theorems for Genetic Algorithms. Consider the following,
  - Only the worst-case scenario is considered. No positive effects of the search operators are considered. This has lead to the development of Exact Schema Theorems [5];
  - The theorem concentrates on the *number* of schema surviving not *which* schema survive. Such considerations have been addressed by the utilization of Markov chains to provide models of behavior associated with specific individuals in the population [6].
  - Claims of “exponential increases” in fit schema i.e., if the expectation operator of Theorem 1 is ignored and the effects of crossover and mutation discounted, the following result was popularized by Goldberg [2],  $m(H, t + 1) \geq (1 + c) m(H, t)$ , where  $c$  is the constant by which fit schema are always fitter than the population average. Unfortunately, this is rather misleading as the average population fitness will tend to increase with  $t$ , thus population and fitness of remaining schema will tend to converge with increasing ‘time’.

## References

- [1] J.H. Holland, (1998) *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to biology, control and artificial intelligence*. MIT Press, ISBN 0-262-58111-6. (NB original printing 1975).

- [2] D.E., Goldberg, (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, ISBN 0-201-15767-5.
- [3] Rodgers A., Prugel-Bennett A. (1999) Genetic Drift in GA Selection Schemes, IEEE Transactions on Evolutionary Computation. 3(4), pp 298-303.
- [4] C.R. Reeves, J.E. Rowe (2003) Genetic Algorithms – Principles and Perspectives. Kluwer Academic Pub. ISBN 1-4020-7240-6.
- [5] C. Stephens, H. Waelbroeck (1998) “Effective Degrees of Freedom in Genetic Algorithms,” Physical review: E, 57(3), pp 3251-3264.
- [6] Vose M.D., Nix A. (1992) “Modeling Genetic Algorithms with Markov Chains,” Annals of Mathematics and Artificial Intelligence, 5, pp 79-98.