

Pessimistic Schema Theory for Fixed Size and Shape GP [1]: A brief summary

- For a schema theory to be of any use,
 - The effects of search operators must be included, and;
 - Any expression must be ‘easy’ to estimate.
- In order to provide a schema theory for GP Poli and Langdon propose a series of constraints.
 - GP “don’t care” symbols are limited to single genes
 - Figure 2 [1].
 - Schema can only represent tree fragments that always include the root node.
 - Rooted Tree schema [2]
- *Definition 1* – GP Schema [1]
 - “A GP schema is a rooted tree composed of nodes from the set $F \cup T \cup \{=\}$, where F and T are the function set and the terminal set used in a GP run, and the operator = is a polymorphic function with as many arities as the number of different arities of the elements of $F \cup T$, the terminals in T being 0-arity functions.”
- This provides the basis for introducing the concepts of order and length.
- *Definition 2* – Order [1]
 - “The number of non-= symbols is called the order $O(H)$ of a schema H .”
- *Definition 3* – Length [1]
 - “The total number of nodes in the schema is called the length $N(H)$ of a schema.”
- *Definition 4* – Defining Length [1]
 - “The number of links in the *minimum tree fragment* including all the non-= symbols within a schema H is called the defining length $L(H)$ of the schema.”
 - A minimum tree fragment is the number of links necessary to provide a path between all non don’t care symbols.
- Thus, definitions for order and length are independent of the shape and size of programs in the population.
 - Figure 3 [1]

- Rosca’s concept of a “rooted tree” schema provides the basis for schema position within an individual [2].
 - This provides the basis for Hyperspaces and Hyperplanes in the action of schema building blocks.
- *Defintion 5* – Hyperspaces and Hyperplanes [1]
 - “A schema G is a hyperspace if it does not contain any defining node (i.e. if $O(G) = 0$). A schema H is a hyperplane if it contains at least one defining node. The schema $G(H)$ obtained by replacing all the defining nodes in a hyperplane H with “don’t care” symbols is called the hyperspace associated to H .”
 - Figure 4 provides a set view [1];
 - Figure 5 provides a geometric view [1];

Defining Search Operators

- Require that the effect of crossover and mutation be computable!
- Point Mutation
 - Modelled on action of GA mutation.
- 1 point Crossover
 - Modelled on action of GA crossover.
 - Figure 6 (a) [1]
 - Aligns trees and utilizes a common crossover point.
 - Two scenarios,
 - Both parents have the *same* size and shape
 - Figure 6 (b) [1]
 - Both parents have the *different* size and shape
 - Figure 6 (c) [1]
- Above search operators effectively specify a “fixed size and shape” GP.
- Now in a position to derive a schema theory
 - Minimal bound only (consider exact schema theory later);
 - I.e. No constructive effects of search operators considered.
 - Introduce effect of,
 - Selection;
 - Crossover;

- Mutation;

Fitness Proportional Selection

- What is the probability that program, h , samples schema, H ?
 - $P(h \in H)$
- Exactly the same as GA case,

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

$$= \frac{m(H,t)f(H,t)}{M\bar{f}(t)}$$

1-point Crossover

- Let $D_c(h)$ denote disruptive effect of schema H when program $h \in H$ is combined with crossover on program \hat{h} .

Case 1 – $D_{c1}(h)$

- One of the parents, \hat{h} , does not have the same shape and crossover disturbs schema H .
- Let $G(H)$ denote the hyperspace of all programs utilizing the same shape as schema H .
 - Figure 7 (a) [1].
 - Probability of such an event requires that,
 - Second parent is does not sample schema H , or $P(\hat{h} \notin G(H))$ and;
 - Crossover point does not re-create schema H , or $P(D_c(h) | \hat{h} \notin G(H))$

- Note,

- Both h and \hat{h} are equally likely events, thus $P(\hat{h} \notin G(H)) \equiv P(h \in G(H))$
- $P(\hat{h} \notin G(H)) = 1 - P(h \in G(H))$
 - So, $P(\hat{h} \notin G(H)) \equiv 1 - \frac{m(G(H),t)f(G(H),t)}{M\bar{f}(t)}$
- $P(D_c(h) | \hat{h} \notin G(H))$ recognizes that not all crossovers of differing shape result in disruption to schema H .
 - Figure 8 [1].

- Let $P(D_c(h) \mid \hat{h} \notin G(H))$ denote the *fragility* of schema *shape*.
 - Hereafter let, $P_{diff}(t) \equiv P(D_c(h) \mid \hat{h} \notin G(H))$

Case 2 – $D_{c2}(h)$

- Parents h and \hat{h} have the same shape, but crossover disturbs schema H .
 - Figure 7 (b)
- Probability of such an event requires that,
 - One of the parents, \hat{h} , does not share all defining nodes (non ‘=’ nodes), or $P(\hat{h} \notin H, \hat{h} \in G(H))$ and;
 - Crossover point selected such that child $\notin H$, $P(D_c(h) \mid \hat{h} \notin H, \hat{h} \in G(H))$
- Note,
 - (i) For $P(\hat{h} \notin H, \hat{h} \in G(H))$,
 - $H \subset G(H)$
 - I.e. we’re interested in the region outside H , or the difference between the two sets,
 - $P(\hat{h} \notin H, \hat{h} \in G(H)) = P(\hat{h} \in G(H)) - P(\hat{h} \notin H)$
 - Where each term on the RHS are events defined by fitness proportional selection.
 - (ii) For the event $\{D_c(h) \mid \hat{h} \notin H, \hat{h} \in G(H)\}$ to exist, crossover must lie between non ‘=’ nodes of ‘ H ’.
 - Table 1
 - Let $B(H)$ denote the precondition.

Table 1 – Effects of crossover on Schema sampled from program $h = (\times a b)$ when \hat{h} has same shape ($= = =$) [1]

Schema	P (B(H))	Crossover point	
		Between \times and a	Between \times and b
($= = =$)	0	Unaffected	Unaffected
($= = b$)	0	Unaffected	Unaffected
($= a =$)	0	Unaffected	Unaffected
($\times = =$)	0	Unaffected	Unaffected
($= a b$)	1	Disruption?	Disruption?
($\times = b$)	–	Unaffected	Disruption?
($\times a =$)	–	Disruption?	Unaffected
($\times a b$)	1	Disruption?	Disruption?

- Only interested in a lower bound, therefore $P(B(H))$ is merely the ratio of the total number of crossover points in the shape against the number of links between non ‘=’ genes,
 - $P(B(H)) = \frac{L(H)}{N(H) - 1}$
 - Or $P(B(H)) \equiv \textit{fragility}$ of schema node **composition**.
- Only interested in negative effects of crossover and given that $B(H)$ is a precondition, we observe that,
 - $P(D_c(h) \mid \hat{h} \notin H, \hat{h} \in G(H)) \leq P(B(H))$
 - I.e. the more exact expression has no impact on the lower bound.

Complete Relation for disruptive crossover effects

- Both $D_{c1}(h)$ and $D_{c2}(h)$ are mutually exclusive,
- $P(D_c(h)) = P(D_{c1}(h)) + P(D_{c2}(h))$

$$P(D_c(h)) \leq p_{diff}(t) \times \left[1 - \frac{m(G(H),t)f(G(H),t)}{M\bar{f}(t)} \right] + \frac{L(H)}{N(H)-1} \times \frac{m(G(H),t)f(G(H),t) - m(H,t)f(H,t)}{M\bar{f}(t)}$$

- Thus, the probability of 1-point crossover disrupting schema H is,
 - $p_{crossover} \times P(D_c(h))$

Point mutation

- Point mutation,
 - does not introduce any shape change, and;
 - is applied on a gene by gene basis.
- Schema H survives only if all defining genes are untouched
 - There are $O(H)$ defining genes.
 - Probability that a gene survives is $1 - p_{mutate}$.
 - $D_m(H) = 1 - (1 - p_{mutate})^{O(H)}$
- I.e. as per simple GA schema.

Fixed Size and Shape GP Schema Theorem

- The expected number of strings matching schema H at generation $t + 1$, $E[m(H, t + 1)]$, is now given by,

$$\begin{aligned} \text{○ } E[m(H, t + 1)] &\geq \frac{m(H,t)f(H,t)}{M\bar{f}(t)} \times (1 - p_{mutate})^{O(H)} \times \\ &\left\{ 1 - p_{crossover} \left[p_{diff}(t) \left(1 - \frac{m(G(H),t)f(G(H),t)}{M\bar{f}(t)} \right) + \right. \right. \\ &\quad \left. \left. \frac{L(H)}{N(H)-1} \times \frac{m(G(H),t)f(G(H),t) - m(H,t)f(H,t)}{M\bar{f}(t)} \right] \right\} \\ \text{○ } E[m(H, t + 1)] &\geq M P(\hat{h} \notin H) (1 - P(D_m(H))) [1 - p_{crossover} P(D_c(H))] \\ \text{○ } &\text{I.e. no account taken for positive contributions from GP.} \end{aligned}$$

Discussion

- (i) $m(G(H), t)$ and $f(G(H), t)$ summarize properties of programs in the population with the same size and shape as schema H .
- (ii) Shape fragility, $P_{diff}(t)$, modulates further the propagation of schema;
- (iii) The 1-point crossover creates two competitions,
 - a. For low generation counts, differing hyperspaces, $G(H)$, compete.
 - b. Once stable hyperspaces are established, competition focuses on the hyperplanes within hyperspaces.

General Conclusion

- Building block hypothesis does appear to hold;
- Appears that at least two types of schema are present,
 - Hyperspaces, $G(H)$ – shape of schema;
 - Hyperplanes, H – content of schema.
- Principle reasons for this uncertainty are,
 - Lower bound;
 - Bound in terms of expectation.

Reference:

1. Poli R., Langdon W. B., (1998) Schema Theory for Genetic Programming with One-Point Crossover and Point Mutation. *Evolutionary Computation*, 6(3), MIT Press, pp 231-252.
2. Rosca J.P., (1997) Analysis of complexity drift in Genetic Programming. *Proceedings of the 2nd Annual Conference on Genetic Programming*. Morgan Kaufmann, pp 286-294.