Schema Theory without Expectations for GP and GAs with One-Point Crossover in the Presence of Schema Creation

Riccardo Poli
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT, UK
R.Poli@cs.bham.ac.uk

Abstract

This paper presents new theoretical results on GA and GP schemata which allow one to predict with a known probability whether the number of instances of a given schema at the next generation is going to be above a given threshold. Unlike previous results, the schema theory presented in this paper does not use expected values and accounts for schema creation.

1 Introduction

Many people in the evolutionary computation field consider the schema theorem as a useless tool for predicting the behaviour of a GA over multiple generations. One reason for this is that the schema theorem gives only a *lower bound* for the *expected value* of the number of instances of a given schema at the next generation as a function of quantities characterising the schema at the previous generation. Therefore, it is not possible to use the schema theorem recursively to predict accurately the behaviour of a genetic algorithm over multiple generations. In addition, since the schema theorem provides only a lower bound, some people argue that the predictions of the schema theorem are not very useful even for a single generation ahead.

I believe that there is some truth in these criticisms. However, I don't believe that the schema theorem is useless. I think that the schema theorem has not been fully exploited nor fully developed, and that when these is done the schema theorem will be very useful. For example, it will be possible to produce a schema based proof of convergence for GP and GAs. This is why in the last two years I have focused my research on the subject.

2 Background

In our previous research, we produced a new definition of schema and a new crossover operator (one-point crossover) for GP, which allowed us to derive a schema theorem which linked the GA theory to the GP theory (Poli and Langdon 1998). This theorem can be criticised in exactly the same ways as the GA schema theorems.

In order to understand how reliable the predictions of our theorem were, we then studied the mean and the variance of the number of instances of schemata of in the presence and in the absence of schema creation events (Poli *et al.* 1998). The result that considered schema creation events were exact predictions of the expected value and the variance of the number of schemata at the next generation. The predictions were based on quantities measured at the current generation.

At this conference I will present a poster which describes two new schema theorems based on our earlier work (Poli 1999*b*). In the theorems expectations are not present. One of the theorems in (Poli 1999*b*) is as follows:

Theorem 1 (Probabilistic Schema Theorem). *For any given constant $k > 0$*

$$\Pr\{m(H,t+1) > M\alpha - k\sqrt{M\alpha(1-\alpha)}\} \geq 1 - \frac{1}{k^2}.$$

where α is the probability that a schema H either survives or is created after selection, crossover and mutation, M is the population size, and $m(H,t+1)$ is the number of instances of H at generation $t+1$.

3 GA Theory

This theorem requires the knowledge of the total transmission probability α, i.e. it requires that not only schema survival and schema disruption events be modelled mathematically but also that schema creation events are. This may not be an easy task if one wants to stick only to the properties of the schema H (such as the number of instances of H and the fitness of H) and those of the population when expressing the quantity α. Indeed, to the best of my knowledge, none of the schema theorems presented to date in the literature (including ours) have succeeded in doing this. This is the reason why all of schema theorems provide upper bounds. However, it is possible to express exactly α (i.e. to include the effects of creation into the schema theorem) by using properties of the building blocks of the schema H. For example, for a GA with binary strings and one point crossover applied with 100% probability α is given by the following equation (which can be obtained either by simplifying the results in (Stephens and Waelbroeck 1997) or by decomposing the probability that H be created into the sum of the probabilities that this will happen for each possible crossover point):

$$\alpha = \frac{1}{N-1} \sum_{i=1}^{N-1} p(L(H,i),t) p(R(H,i),t) \tag{1}$$

where $p(x,t)$ is the selection probability of a schema x at generation t, $L(H,i)$ is the schema obtained by replacing with "don't care" symbols all the elements of H from position $i+1$ to position N, $R(H,i)$ is the schema obtained by replacing with "don't care" symbols all the elements of H from position 1 to position i, and i varies over the valid crossover points. For example, if $H =$`**1111`, $L(H,1) =$`******`, $R(H,1) =$`**1111`, $L(H,3) =$`**1***`, $R(H,3) =$`***111`.

So, provided that the total transmission probability of a schema can be expressed as a function of known properties (such as the selection probability) of some building blocks of such a schema, the schema theorem given above can be used to predict (with a known probability) whether the number of instances of a schema at the next generation is greater than a threshold. For example, in the case of a binary GA with one-point crossover applied with 100% probability, this is achieved by substituting Equation 1 into the schema theorem. I believe this is a very important result. In other work I have used this and other results to make some steps towards a convergence proof and population sizing equations for GAs (Poli 1999a). It is worth nothing that Equation1 can easily be extended to the case in which crossover is not applied to all the individuals in the population.

4 GP Theory

In order to do the same for GP with one-point crossover we need to introduce a new, more general definition of schema for GP. The definition is as follows:

Definition 2 (GP hyperschema). *A* GP hyperschema *is a rooted tree composed of nodes from the set $\mathcal{F} \cup \mathcal{T} \cup \{=, \#\}$, where $\mathcal{F}$ and $\mathcal{T}$ are the function set and the terminal set used in a GP run, and the operators = and # are polymorphic functions with as many arities as the number of different arities of the elements of $\mathcal{F} \cup \mathcal{T}$, the terminals in $\mathcal{T}$ being 0-arity functions. The operator = is a "don't care" symbols which stands for exactly one node, while the operator # stands for any valid subtree.*

This definition presents some of the features of the schema definition we used for our GP schema theory (Poli and Langdon 1998): hyperschemata are rooted trees, and they include don't care symbols which stand for one node only. However, Definition 2 also includes one of the features of the schema definitions proposed by other authors (O'Reilly and Oppacher 1995, Rosca 1997): hyperschemata also include don't care symbols which stand for entire subtrees. An example of hyperschema is `(= (+ x =) #)`. This hyperschema represents all the programmes with the following characteristics: a) the root node is any binary function belonging to the function set, b) the first argument of the root node is +, c) the first argument of the + is the variable `x`, d) the second argument of the + is any valid node in the terminal set, f) the second argument of the root node is any valid subtree. The reason for the name "hyperschema" is that a hyperschema can represent a group of schemata in essentially the same way as a schema presents a group of programme trees.

With the definition of hyperschema it is possible to express the total transmission probability for any of our GP schemata, in the presence of one-point crossover applied with a 100% probability, in exactly the same way as in Equation1, i.e.

$$\alpha = \frac{1}{N(H)-1} \sum_{i=1}^{N(H)-1} p(L(H,i),t) p(U(H,i),t) \tag{2}$$

where $p(x,t)$ is the selection probability of a schema x at generation t, $N(H)$ is the number nodes in the schema H, $L(H,i)$ is the hyperschema obtained by replacing with = nodes all the nodes on the path between crossover point i and the root node and with # nodes all the subtrees connected to the nodes replaced with = (L stands for "lower"), $U(H,i)$ is the hyperschema obtained by replacing with a # node the subtree below crossover point i, and i varies over the valid $N(H)-1$ crossover points. For example, if $H =$`(= (+ x =) y)` and the crossover points are numbered as in Figure 1 then $L(H,1) =$`(= (= x #) #)`, $U(H,1) =$`(= (+ # =) y)`, $L(H,3) =$`(= (+ x =) #)`, $U(H,3) =$`(= # y)`.

So, Equation 2 allows one to compute the exact total transmission probability of the GP schema as a function of a

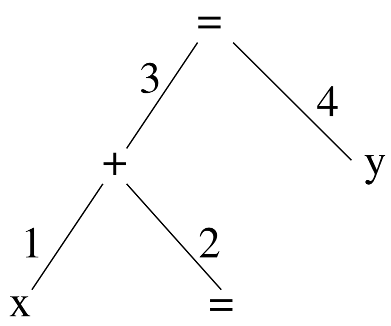

Figure 1: Possible numbering of the crossover points in a schema.

group of $2(N(H) - 1)$ building blocks (hyperschemata). Substituting the result in to Theorem 1 one can obtain precise information on how likely number of instances of a schema will grow.

Acknowledgements

The author would like to thank the members of the EEBIC (Evolutionary and Emergent Behaviour Intelligence and Computation) group at Birmingham for useful discussions and comments.

References

O'Reilly, Una-May and Franz Oppacher (1995). The troubling aspects of a building block hypothesis for genetic programming. In: *Foundations of Genetic Algorithms 3* (L. Darrell Whitley and Michael D. Vose, Eds.). Morgan Kaufmann. Estes Park, Colorado, USA. pp. 73–88.

Poli, Riccardo (1999*a*). Probabilistic schema theorems without expectation, left-to-right convergence and population sizing in genetic algorithms. Technical Report CSRP-99-3. University of Birmingham, School of Computer Science.

Poli, Riccardo (1999*b*). Schema theorems without expectations. In: *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference* (W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela and R. E. Smith, Eds.). Morgan Kaufmann. Orlando, Florida, USA. Forthcoming.

Poli, Riccardo and William B. Langdon (1998). Schema theory for genetic programming with one-point crossover and point mutation. *Evolutionary Computation* **6**(3), 231–252.

Poli, Riccardo, William B. Langdon and Una-May O'Reilly (1998). Analysis of schema variance and short term extinction likelihoods. In: *Genetic Programming 1998: Proceedings of the Third Annual Conference* (John R. Koza, Wolfgang Banzhaf, Kumar Chellapilla, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max H. Garzon, David E. Goldberg, Hitoshi Iba and Rick Riolo, Eds.). Morgan Kaufmann. University of Wisconsin, Madison, Wisconsin, USA. pp. 284–292.

Rosca, Justinian P. (1997). Analysis of complexity drift in genetic programming. In: *Genetic Programming 1997: Proceedings of the Second Annual Conference* (John R. Koza, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max Garzon, Hitoshi Iba and Rick L. Riolo, Eds.). Morgan Kaufmann. Stanford University, CA, USA. pp. 286–294.

Stephens, C. R. and H. Waelbroeck (1997). Effective degrees of freedom in genetic algorithms and the block hypothesis. In: *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)* (Thomas Bäck, Ed.). Morgan Kaufmann. East Lansing.