

Is GP Crossover a Local Search Operator?

Riccardo Poli

School of Computer Science, The University of Birmingham, B15 2TT, UK

E-mail: R.Poli@cs.bham.ac.uk, WWW: <http://www.cs.bham.ac.uk/~rmp/>

Abstract

In this short paper, backed by probabilistic calculations, a simple model and two experiments, I suggest that standard crossover in GP is a local search operator in the sense that it produces offspring which inherit most of their code from one parent most of the times. This is in strong contrast with the crossover operators used in linear GAs and might cause GP to explore only a small part of the search space and to get easily stuck in local optima.

1 Theory

Let us consider a GP system with a simple form of standard crossover in which crossover points are selected with uniform probability among all the nodes in both trees. Let us define

$$p(d) = \Pr\{\text{A crossover point at depth } d \text{ is chosen in a tree}\} = \Pr\{D = d\},$$

where D is a stochastic variable representing the depth of the crossover point,

$$P(d) = \Pr\{D \leq d\} = \sum_{x=0}^d p(x),$$

and

$$\tilde{P}(d) = \Pr\{D \geq d\} = \sum_{x=d}^{\infty} p(x) = 1 - P(d) + p(d).$$

I will refer to $p(d)$ as the *node distribution*. In an individual including S nodes, there will be $p(d)S$ nodes at depth d . So, by hypothesising that the nodes at depth d have subtrees with the same structure and size, it is possible to estimate the expected number of nodes in a subtree at depth d :

$$S(d) = \frac{\sum_{x=d}^{\infty} p(x)S}{p(d)S} = \frac{\tilde{P}(d)}{p(d)}.$$

I will refer to $S(d)$ as *size function*. It should be noted that $S = S(0)$.

Let us now consider two parents having size functions $S_1(d)$ and $S_2(d)$. If the crossover point in the first parent is at depth d_1 and the one in the second parent is at depth d_2 , then the number of nodes in the first offspring (the one obtained by replacing a subtree in the first parent with one taken from of the second) is

$$S_{o1} = S_1(0) - S_1(d_1) + S_2(d_2).$$

It is then possible to compute the expected size of the offspring as follows:

$$\bar{S}_{o1} = E[S_1(0) - S_1(D_1) + S_2(D_2)] = S_1(0) - \sum_{x=0}^{\infty} S_1(x)p_1(x) + \sum_{x=0}^{\infty} S_2(x)p_2(x) = S_1(0) - \sum_{x=0}^{\infty} (\tilde{P}_1(x) - \tilde{P}_2(x))$$

where D_1 and D_2 are two stochastic variables representing the depths of the crossover points and $p_1(x)$ and $p_2(x)$ are the probabilities that the crossover point is at depth x in the first and second parent,

respectively. Similarly one can compute the expected value of the size of the other offspring S_{o2} . It is also possible to compute the variance of the size of the offspring by expanding the following equation:

$$\begin{aligned}\text{Var}[S_{o1}] &= E[(S_{o1} - \bar{S}_{o1})^2] = S_1^2(0) - 2S_1(0)E[S_1(D_1)] + 2S_1(0)E[S_2(D_2)] \\ &\quad + E[S_1^2(D_1)] - 2E[S_1(D_1)]E[S_2(D_2)] + E[S_2^2(D_2)] - \bar{S}_{o1}^2\end{aligned}$$

2 An Example and Two Experiments

Let us suppose that two parents created with the “full” initialisation method with the same initial depth d_{max} are crossed over. If for simplicity one imagines that only arity-2 functions are used, then (for $d \leq d_{max}$):

$$p(d) = \frac{2^d}{2^{d_{max}+1} - 1}, \quad P(d) = \frac{2^{d+1} - 1}{2^{d_{max}+1} - 1}, \quad \tilde{P}(d) = \frac{2^{d_{max}+1} - 2^d}{2^{d_{max}+1} - 1}, \quad S(d) = 2^{d_{max}+1-d} - 1.$$

In this situation $\tilde{P}_1(d) = \tilde{P}_2(d) = \tilde{P}(d)$ and $\bar{S}_{o1} = S_1(0) = S(0)$. Therefore, the offspring should not be expected to grow or shrink. The variance of S_1 is

$$\begin{aligned}\text{Var}[S_{o1}] &= E[S_1^2(D_1)] - 2E[S_1(D_1)]E[S_2(D_2)] + E[S_2^2(D_2)] \\ &= 2 \sum_{x=0}^{d_{max}} S^2(x)p(x) - 2 \left(\sum_{x=0}^{d_{max}} \tilde{P}(x) \right)^2 \\ &= -8 \frac{5 \times 4^{d_{max}} - 2^{d_{max}} + 2 \times d_{max} \times 4^{d_{max}} - 4 \times 8^{d_{max}} + d_{max}^2 \times 4^{d_{max}}}{(2^{d_{max}+1} - 1)^2}\end{aligned}$$

From this formula it can be inferred that the standard deviation $\text{Std}[S_{o1}] = \sqrt{\text{Var}[S_{o1}]}$ of the size of the offspring grows more slowly than \bar{S}_{o1} and that for moderately large values of d_{max} most of the offspring are in a relatively tiny neighbourhood of their parents (size-wise). For example, if $d_{max} = 8$ then $S(0) = 511$ and $\text{Std}[S_{o1}] = 43.4$, while for $d_{max} = 12$ then $S(0) = 8191$ and $\text{Std}[S_{o1}] = 180.1$.

The calculations involved in computing the expected size and the variance of the offspring for parents with different and possibly complicated node distributions become quickly very difficult. This is why in the example given above I made many simplifying assumptions. However, it is possible to use the theory given in the previous section and perform numerical simulations. Here I will report only two simple experiments.

Firstly, let us consider what happens when one crosses over two parents again created with the “full” method and the same maximum depths $d_{max1} = d_{max2} = 8$, using a function set including only arity-2 functions. Figure 1 (upper row) shows the functions $p(d)$, $S(d)$, $\tilde{P}(d)$ (labelled “Ptd”) for the parents and the node distribution of the offspring $p_{\text{off}}(d)$ estimated via a node histogram of the offspring obtained by crossing over the two parents 2,000 times. The difference between the node distribution of the offspring and that of the parents is very small (of the order of 10^{-3}). The measured average offspring sizes are $\bar{S}_{o1} = 510.735$ and $\bar{S}_{o2} = 511.265$. This is in agreement with the theory, as the size of the parents is 511 nodes, therefore, according to the calculations in the abovementioned example, the expected size of the offspring is also 511.

Secondly, let us consider what happens when one crosses over two parents again created with the “full” method but with different maximum depths $d_{max1} = 4$ and $d_{max2} = 8$, using a function set including only arity-2 functions. Figure 1 (lower row) shows the functions $p(d)$, $S(d)$, $\tilde{P}(d)$ for the parents and the node distribution of the offspring $p_{\text{off}}(d)$ estimated as detailed above. The difference between the node distribution of the offspring obtained by replacing a subtree of the bigger parent and

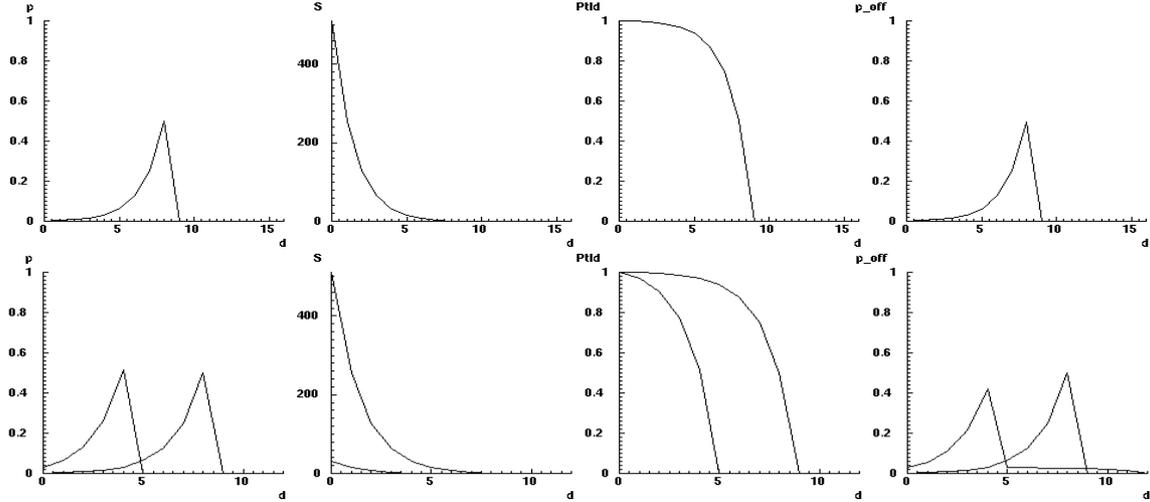


Figure 1: Plots of $p(d)$, $S(d)$, $\tilde{P}(d)$ and $p_{\text{off}}(d)$ for $d_{\text{max}1} = d_{\text{max}2} = 8$ and $d_{\text{max}1} = 4, d_{\text{max}2} = 8$.

that of such a parent is again very small. On the contrary the node distribution of the second offspring presents some differences with respect to that of the other parent (note the tail in the node distribution for $d \geq 5$). In any case the offspring distributions are not half-way between those of the parents, but rather very similar to those of one or the other parent. The measured average offspring sizes are $\bar{S}_{o1} = 35.455$ and $\bar{S}_{o2} = 506.545$. This is again in agreement with the theory which predicts $E[S_{o1}] = 34.856$ and $E[S_{o2}] = 507.144$. Both are very close to the sizes (31 and 511) of the parents.

3 Discussion and Conclusions

Admittedly the example and experiments reported in the previous section are extremely simple and these results need to be corroborated by more work with the mathematical tools presented in this paper (this is under way). Nonetheless I think some interesting comments can be made on these results.

The second experiment in the previous section suggests very strongly that crossing over two trees with different sizes does not produce offspring with an average size somewhere in the middle between the sizes of the parents. On the contrary, one of the offspring will have a size only slightly bigger than that of the smaller parent, while the other will have a size only slightly smaller than that of the larger parent. This means that crossover tends to produce offspring which inherit most of their code from one parent or the other, most of the times. The calculations in the example in the previous section seem to suggest that not only the average size of the offspring is not very different from that of the parents, but also that the probability distribution of the offspring sizes has a very narrow peak around that average.

I think that the theory presented here and the evidence obtained through it suggest very strongly that standard crossover in GP really produces offspring that, probabilistically, are not significantly different in size from their parents, i.e. that crossover is a local search operator, unlike in GAs. This might be the source of many of the problem observed in GP runs. E.g. it might cause GP to explore only a small part of the search space and to get easily stuck in local optima. A way to overcome this problem might be to bias the choice of crossover points in one or both parents so as to obtain crossover operators with a better “interpolative” behaviour (e.g. which produce offspring with $p_{\text{off}}(d)$ somehow half-way between $p_1(d)$ and $p_2(d)$).

Acknowledgements

The author wishes to thank Bill Langdon for useful comments on this manuscript.