



A schemata theorem for trees

Antonia J. Jones

Department of Computing, Imperial College London

Abstract.

The proof of a schemata theorem for trees of bounded branching with a probabilistic crossover is sketched. The argument presented applies only to the case where the genetic information is placed at the terminal nodes of the tree although the shape of the tree itself can be considered as an additional form of genetic information.

Keywords: Trees, Schemata theorem.

DEPARTMENT OF COMPUTING
IMPERIAL COLLEGE OF SCIENCE,
TECHNOLOGY AND MEDICINE
UNIVERSITY OF LONDON
180 Queen's Gate, London, SW7 2BZ
Telephone: 071-589 5111 Ext: 5017/5096
Telefax: 071-581 8024
E-Mail: ajj@doc.ic.ac.uk

A schemata theorem for trees.

CONTENTS

Introduction	1
d-trees and primitive node substitution	1
Schemata theorem for the d-tree representation	4
References	9

List of figures

Figure 1 A quadtree and the corresponding subdivision of the unit square it represents.	2
Figure 2 Two possible parent quadtrees P and Q, and the subdivisions of the unit square which they represent.	2
Figure 3 The two children which can result from [14].	2
Figure 4 The two children which can result from [4].	2
Figure 5 The probability of schemata survival.	9

A schemata theorem for trees.

Introduction.

For the most part Genetic Algorithms have been applied to various search problems using a string representation for the genotype. However, there is no particular reason why the genotype of a structure need be a string. In many applications, particularly those related to a divide and conquer approach, a tree structure might be more appropriate. However, if one changes the data structure for a GA the immediate question becomes: is there a schemata theorem for the new representation and the associated crossover?

In this note we sketch the proof of a schemata theorem for trees of bounded branching with what seems a reasonably natural crossover operator. The argument presented applies only to the case where the genetic information is placed at the terminal nodes of the tree although the shape of the tree itself can be considered as an additional form of genetic information.

Whilst tree structures have not been frequently used as genotypes, an exception is the work of Koza [Koza 1991], in which S-expressions in LISP correspond directly to a "parse tree". Koza evolves computer programs using genetic algorithms in which the crossover creates new offspring S-expressions by exchanging sub-trees. Although the work here does not apply directly to Koza's scenario (for example, his trees contain information at all nodes), it may eventually help to explain why his *Genetic Programming Paradigm* seems to work as well as it does. Koza's crossover permits substitution of a node (and associated subtree) from one level of the donor tree to a node in the recipient at a different level, which permits the child tree to have a depth different from either parent. This does not happen in the argument presented here, in which the probabilistic crossover is the combined effect of a number of primitive substitutions each of which is at the same level. However, variable depth of the child could perhaps be arranged through a mutation operator.

Let each node of a tree have $d \geq 2$ associated edges. We shall illustrate our trees and crossover operator using $d = 4$, but the ideas are quite general and can be applied for any fixed $d > 1$.

d-trees and primitive node substitution.

The genotype for our GA will be an ordered tree, which we shall call a d-tree or, when $d = 4$, a *quadtree*, since each node has valence four.

We can give our a quadtrees a simple geometric interpretation which may be helpful.

A schemata theorem for trees.

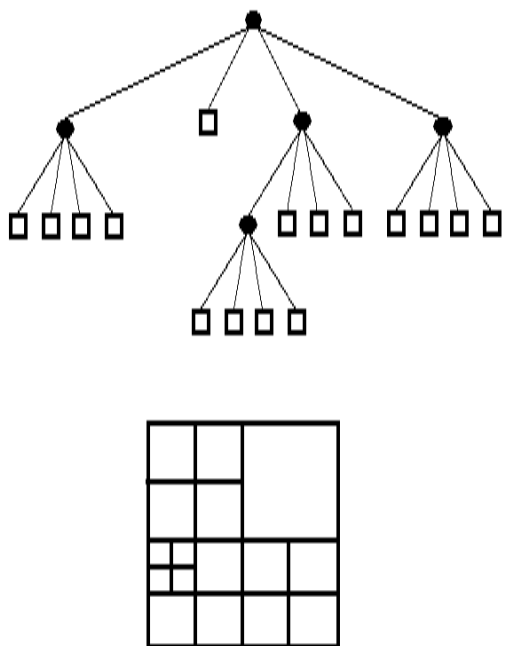


Figure 1 A quadtree and the corresponding subdivision of the unit square it represents.

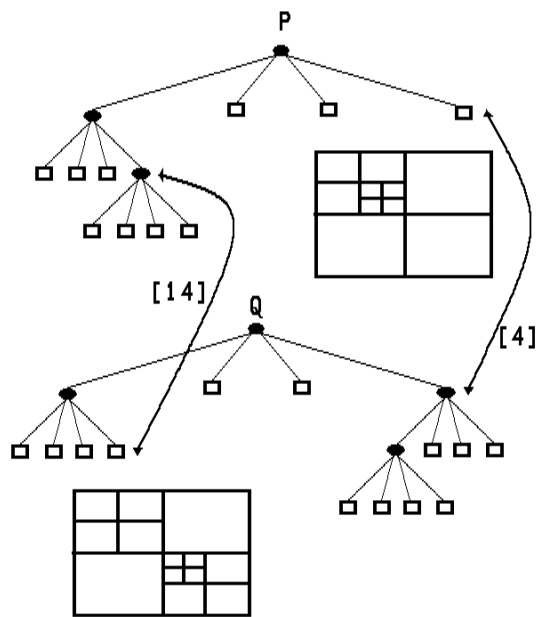


Figure 2 Two possible parent quadtrees P and Q, and the subdivisions of the unit square which they represent.

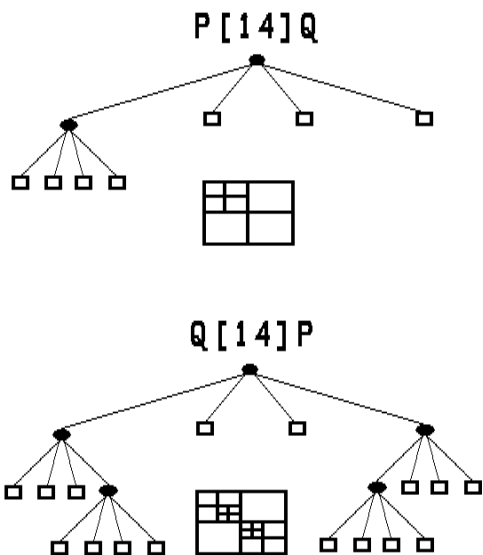


Figure 3 The two children which can result from [14].

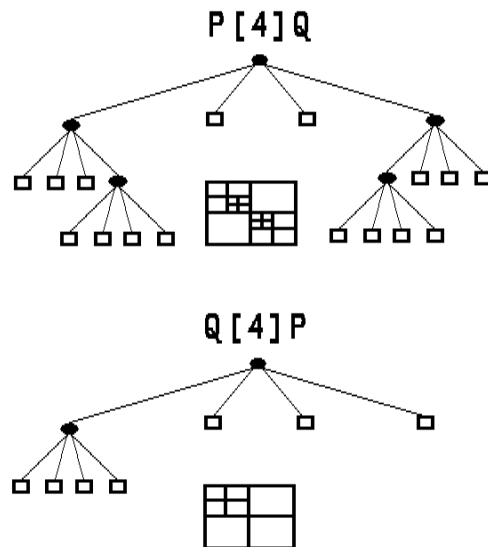


Figure 4 The two children which can result from [4].

A schemata theorem for trees.

The root node of the tree represents the unit square.

Each of the four branches from a node represent a division of the subsquare, represented by the node, into four equal subsquares.

A quadtree is not (necessarily) balanced, so subdivision of nodes for a particular quadtree may halt. An example is given in Figure 1. The convention is that the *order* of the four branches is TL, TR, BL, BR (T = Top, B = Bottom, L = Left, R = Right).

An important property of the d-tree representation is that it provides a *relatively natural crossover operator* which has several desirable characteristics.

Suppose we have two d-trees and wish to create a child d-tree. A natural primitive operation is:

*Substitute a node (and its associated subtree) from one parent (the donor) into the other (the recipient) at exactly the same relative position. If the target node does not exist in the recipient then no substitution takes place.*¹

This is illustrated with two possible parents P and Q in Figure 2, and two possible subtree substitutions in Figure 3 and Figure 4. Here parent quadtrees P and Q exchange genetic material by two substitutions $[14]$ and $[4]$. The effect of $P[14]Q$ is to take parent P, the recipient, and substitute into it the node (and its associated subtree), defined by the pointer string $[14]$ from root, from parent Q, the donor. Similarly $Q[4]P$ substitutes the node defined by the pointer string $[4]$ from donor P into recipient Q. Note the operation is not commutative.

In Figure 3 the two children which can result from $[14]$ are illustrated. Similarly Figure 4 illustrates $[4]$. It will be seen that structural features from both parents are included in the children. Note some children are identical, even though they were produced by different substitutions on nodes at different levels. This is not particularly desirable but it would not appear to be a fatal characteristic.

The crossover operator.

To obtain a crossover operator which has an opportunity to combine coadapted structural features of d-trees we shall use a number of primitive node substitutions controlled in a probabilistic fashion. Therefore we need to select the probability λ_l of using a primitive substitution on each node at level l . In a conventional crossover we should expect on average 50% exchange of genetic material and we endeavour to choose the set $\lambda_1, \dots, \lambda_l$ to approximately satisfy this condition.

¹ Actually we have a choice here. If the target node does not exist in the recipient we could build a linear sequence of nodes (with associated arbitrary terminal nodes hanging off) to the target node and then perform the substitution. This would not affect the subsequent argument on schemata disruption and would increase the proportion of leaf nodes transferred from the donor. However, it raises some unnecessary complications and tends to confuse the simplicity of the following argument.

A schemata theorem for trees.

Let the maximum depth of a tree be k . For each existing node at level $l < k$ of Q let λ_l ($1 \leq l \leq k-1$) be the probability of substituting the corresponding node, and associated subtree, from donor Q to recipient P . Perform this probabilistic substitution for every node which appears both in the donor and in the recipient. Let the resulting child be $C = P \otimes Q$. Plainly this crossover is probabilistic and depends on the choice of $\lambda_1, \dots, \lambda_{k-1}$. We note that this crossover operator has the important property that $P \otimes P = P$. Moreover it can transfer structurally coadapted genes (subtrees) from one parent to the other.

We next inquire about the proportion of leaf nodes affected by crossover. In order to make this precise we shall attach weights to the leaf nodes according to their depth. A leaf node at depth l should account for a *weighted proportion* $1/d^l$ of the leaf nodes. For example, in Figure 1 the leaf node corresponding to the upper right quadrant of the unit square would account for $1/4$ of the weighted proportion of leaf nodes even though there are 16 leaf nodes altogether.

Suppose crossover begins at level 1 and proceeds down the tree. At any level $l \geq 1$ define t_l to be the *expected* weighted proportion of those leaf nodes of the recipient (so far unaffected by substitutions at higher levels) which are affected by substitutions at level l .

At level 1 of the donor the expected weighted proportion of leaf nodes affected in the recipient is $t_1 = \lambda_1$. In general at level l we have

$$t_l \leq \lambda_l(1 - t_{l-1}) \leq \lambda_l \quad (2 \leq l \leq k-1) \quad (1)$$

The lefthand inequality applies because substitution occurs with probability λ_l only if the corresponding node appears in both parents.

Since $t_1 = \lambda_1$ a lower bound on the expected weighted proportion of leaf nodes affected by crossover is given trivially by

$$\sum_{l=1}^{k-1} t_l \geq \lambda_1 \quad (2)$$

In general the sum depends on the choice of $\lambda_1, \dots, \lambda_{k-1}$ (which we are not yet ready to decide) and the distribution of nodes at different levels in the population of trees. However, it transpires that for our purposes the trivial lower bound $t_1 = \lambda_1$ will suffice.

Schemata theorem for the d-tree representation.

We next sketch the argument that this combination of representation and crossover will satisfy a suitably modified schemata theorem. Apart from the power of intrinsic parallelism, our understanding of the schemata theorem (Theorem 6.2.3, p 102, [Holland 1975]) is as follows.

Firstly, consider an algorithm in which a structure from the population is selected according to the selection probabilities

A schemata theorem for trees.

$$P_i = \frac{v(s_i)}{\sum_{j=1}^N v(s_j)} \quad (3)$$

(where $v(s_i)$ is the fitness of the i th structure) and then a copy of the structure is inserted into the population, replacing some structure selected at random. Duplication is the simplest possible reproduction operator.

Suppose that $P(\xi, t)$ is the proportion of the current population which are in schemata ξ . Then at time $t+1$, when every member of the population has reproduced, it is a simple matter to show that

$$\langle P(\xi, t+1) \rangle = \frac{v(\xi, t)}{\bar{v}(t)} P(\xi, t) \quad (4)$$

where the angle brackets denote expectation and

$$v(\xi, t) = \frac{1}{N(\xi, t)} \sum_{s \in \xi} v(s) \quad (5)$$

is the *observed fitness* of schemata ξ , $N(\xi, t)$ is the number of examples of ξ in the current population at time t , and $\bar{v}(t)$ is the mean fitness of the current population at time t .

Equation (4) represents the case where reproduction is by duplication. However, in general if reproduction is by crossover, rather than duplication, then we must modify this conclusion. In this case the probability of schemata survival may be less than unity. The result becomes an inequality, rather than an equality. We write

$$\frac{v(\xi, t)}{\bar{v}(t)} P(\xi, t) \geq \langle P(\xi, t+1) \rangle \geq H \frac{v(\xi, t)}{\bar{v}(t)} P(\xi, t) \quad (6)$$

where $0 \leq H \leq 1$ is the *probability of schemata survival* (which we name after Holland). The schemata theorem is basically this inequality plus a lower bound for H .

What this inequality says is that if the observed fitness of ξ is significantly greater than the average fitness of structures across the whole population, i.e. if

$$H \frac{v(\xi, t)}{\bar{v}(t)} > 1 \quad (7)$$

($0 \leq H \leq 1$) then the proportion of the population in ξ can be expected to grow exponentially.

Thus exceptionally fit schemata *must* be expected to grow, but once the average fitness of the population rises accordingly and (7) is no longer satisfied, there is no (probabilistic) guarantee

A schemata theorem for trees.

of growth. Indeed, $P(\xi, t)$ obviously cannot grow without bound, since its maximum value is 1. So the schemata theorem can at most (probabilistically) guarantee rapid growth for a limited period. Once

$$H \frac{v(\xi, t)}{\bar{v}(t)} < 1 \quad (8)$$

the righthand inequality of (6) is less useful. Equally obvious is the observation that the nearer H is to 1 the stronger the inequality (6).

H is the probability that, if $s \in \xi$ and s is a recipient parent, the resulting child will not have its membership of ξ disrupted. This probability will depend on the nature of the genotype data structure and the crossover operator.

Thus we want to design the combination of genotype data structure and crossover between two data structures so as to satisfy two conditions.

- H should be as close to 1 as possible (to get the strongest possible schemata theorem).
- Ideally the crossover of two structures should effect an approximately 50:50 mix of the genotype information of both parents.

Now crossover should also satisfy some other conditions, e.g. $P \otimes P = P$, but the above two conditions need to be satisfied in order to have a satisfactory schemata theorem *and* a useful crossover.

In order to accomplish this we have to give an appropriate definition of schemata in this context.

If the data structure for s is a d -tree, the natural generalisation of schemata is to place a 'don't care' symbol $*$ at any node of the tree. However we need to modify this slightly. The rules for schemata in the d -tree representation are as follows:

- We are permitted at most $d-1$ symbols $*$ at any d children of a given parent node (if d stars were permitted then we might as well label the parent with a $*$ symbol).
- If any node is labelled with a $*$ symbol then it becomes a leaf node for the schemata (since the corresponding subtree is arbitrary).

We now proceed to estimate a value for H based on the d -tree structure and the probabilistic crossover operator defined in the previous section.

Let ξ be a schemata and for $l \geq 0$ let n_l be the number of nodes at level l which are *not* labelled with a symbol $*$. Then $n_0 = 1$ or 0 , and for $l \geq 1$ we have $1 \leq n_l \leq d^l$. The most general schemata has $n_0 = 0$ and includes all possible trees, in what follows we shall often assume $n_0 = 1$ without comment. Consider, for the moment, trees of maximum depth k . The

A schemata theorem for trees.

ordered tuple $[n_0, \dots, n_k]$ plays a role analogous to the *defining length* of a conventional schemata.

When crossover is performed, every node of the recipient at level $l \geq 1$ has a probability λ_l of being subjected to a primitive substitution (if the corresponding node exists in the donor and the recipient, otherwise the probability is zero). The probability that an individual node will *not* be affected is therefore at least $1 - \lambda_l$.

Since there are n_l nodes of ξ at level l which are not labelled with a symbol * the probability that none of these nodes will be affected is at least

$$(1 - \lambda_l)^{n_l} \tag{9}$$

Hence, the probability that for all l , $1 \leq l \leq k-1$, no unstarred node at level l will be affected (i.e. the probability of schemata survival) is at least

$$H \geq \prod_{l=1}^{k-1} (1 - \lambda_l)^{n_l} \tag{10}$$

The worst case analysis is that the schemata is very specific, which would happen if $n_l = d^l$ for all l , as in the righthand inequality

$$H \geq \prod_{l=1}^{k-1} (1 - \lambda_l)^{n_l} \geq \prod_{l=1}^{k-1} (1 - \lambda_l)^{d^l} \tag{11}$$

In this case there is naturally a high probability of schemata disruption. In the conventional analysis, where structures are strings, for distinct parents this probability becomes *certainty* and $H = 0$; but, of course, $P(\xi, t)$ is very small, as ξ consists of just one structure. Unlike the conventional situation for strings our crossover is more probabilistic. In our case $H > 0$ even when the schemata is totally specific, since the probability that no node is affected by crossover (i.e. that crossover reduces to duplication) is small but positive.

In fact we can do even better; we can arrange a *uniform* lower bound for H , valid for d -trees of any depth (i.e. as $k \rightarrow \infty$). Whether or not this happens depends on the choice of $\lambda_1, \dots, \lambda_l$.

In general, to get the maximum value of H we wish to take $\lambda_1, \dots, \lambda_l$ as small as possible. On the other hand to ensure that crossover affects a positive weighted proportion of genetic material and does not frequently degenerate into duplication we must make $\lambda_1, \dots, \lambda_l$ large. This is a somewhat tricky optimisation since it is not clear how to balance these conflicting objectives so as to produce the best possible algorithm. We do not address this issue directly but simply exhibit a reasonable choice of $\lambda_1, \dots, \lambda_l$.

A schemata theorem for trees.

Consideration of the righthand inequality in (11) shows that to ensure that H is large λ_1 should get smaller as l gets larger. We choose

$$\lambda_l = \rho^l \tag{12}$$

for $l \geq 1$, where $0 < \rho < 1$ remains to be chosen. We observe that

$$\lim_{l \rightarrow \infty} \left(1 - \left(\frac{1}{d} \right)^l \right)^{d^l} = \frac{1}{e} \tag{13}$$

In fact when $\rho = 1/d$ the terms of the product on the right in (11) are monotonic increasing to $1/e$, so the product behaves roughly like $(1/e)^{k-1}$, i.e. diverges to zero as $k \rightarrow \infty$. However, if $\rho < 1/d$ the infinite product is convergent (tends to a positive value as $k \rightarrow \infty$).

- Thus we can choose any $\rho < 1/d$ and guarantee that $H = H(\rho, d) > 0$ in all cases.

For example

$$H \geq H_0(\rho, d) = \prod_{l=1}^{\infty} (1 - \rho^l)^{d^l} > 0 \quad (\text{for } \rho < \frac{1}{d}) \tag{14}$$

It now remains to estimate the sum in (2) given the choice of λ_l in (12). From (1) we have $t_1 = \lambda_1 = \rho$ so that

$$\sum_{l=1}^{k-1} t_l \geq \rho \tag{15}$$

For $\rho < 1/d$ the additional terms in the sum would, in any case, be of small order so it suffices to accept the lower bound ρ , for the weighted proportion of genetic material expected to be affected by the crossover.

Now if ρ or $H(\rho, d)$ were to depend on k , so that either tended to zero as k tended to infinity, the method would be *weak*, i.e. the genetic algorithm would become less effective as k became large. However, we have shown that this is not the case. Our lower bound $H_0(\rho, d)$ is uniform in k . Thus it is possible to take a specific value for $\rho < 1/d$, e.g. $\rho = 1/(2d)$, and compute a value $H_0(\rho, d)$ so that this value will apply in the schemata theorem *independently of the maximum depth of the trees*. Moreover in each crossover a weighted proportion of at least ρ leaf nodes of the recipient d -tree can be expected to be affected.

The following graph shows how $H_0(\rho, d)$ varies with ρ for different values of d .

A schemata theorem for trees.

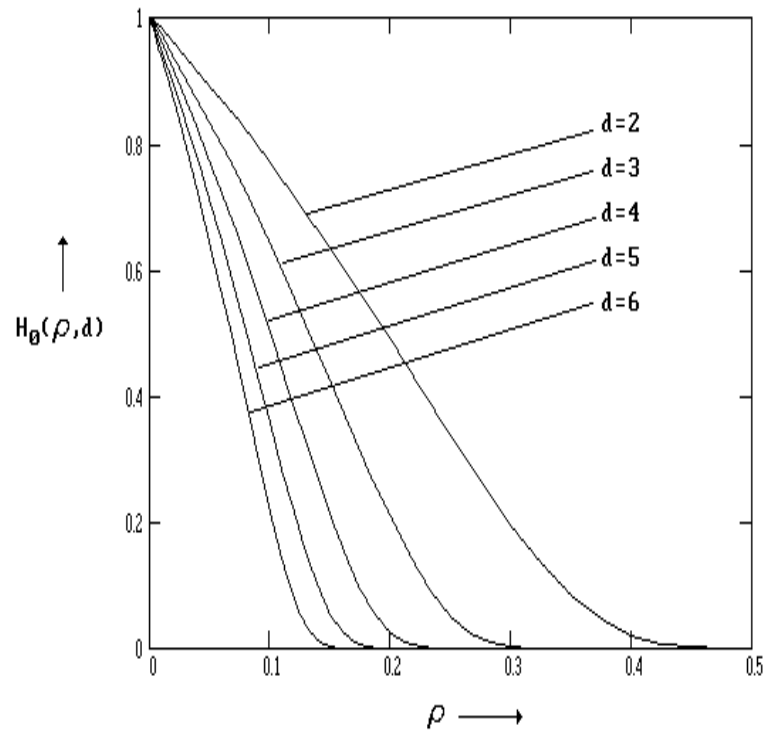


Figure 5 The probability of schemata survival.

References.

[Holland 1975] John H. Holland. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, 1975.

[Koza 1991] John R. Koza. *Evolving a computer program to generate random numbers using the genetic programming paradigm*. Proceedings of the Fourth International Conference on Genetic Algorithms. Morgan Kaufmann, San Mateo, 1991.