Diversifying Exploration of Feature Spaces in Evolutionary Searches.

Tim Hendtlass and Howard Copland,
Center for Intelligent Systems and Complex Processes,
School of Biophysical Sciences and Electrical Engineering,
Swinburne University of Technology,
P.O.Box 218, Hawthorn, AUSTRALIA 3122.

Abstract.
Evolutionary algorithms require excellent search capabilities in order to find global minima, particularly in complex feature spaces. A means of enhancing search capabilities based upon a distributed genetic-style encoding of solution has been shown to be advantageous. Such a representation requires the use of varying gene lengths. In this paper, the effects of variable gene lengths are explored in detail.

1. Introduction.
The Evolutionary Algorithm (EA) is based on Darwin's theory of evolution by natural selection. A group of individuals in some environment have a higher probability to reproduce if their fitness (their ability to thrive in this environment) is high. As with many living organisms, offspring are created via a crossover operation, which combines characteristics of two selected parents, and are then subject to a possible mutation. Mutation is the means whereby new characteristics not present in the initial population are generated. The algorithm aims to increase the average fitness of the individuals over a number of generations.

In evolutionary algorithms the evolution of a solution is the search for a particular combination of values or set of co-ordinates. A plot of all possible combinations against the error that each combination produces can be thought of as an "imaginary landscape" with a problem specific terrain: mountains, plains, canyons and so on. The best possible combination(s) has the least error, and so the purpose of the evolutionary algorithm is to find the lowest point in the landscape.

The more complex the problem, the more complex the landscape to be explored. There may exist many shallows — local minima — that can be deceptive. As the number of dimensions to be explored increases significantly, a many-dimensional landscape can fold and twist back on itself, and tunnels or bridges linking widely separated regions can emerge. Finding the global minima on such a surface becomes increasingly difficult. Even determining which way is "down" can prove far from easy. In some circumstances, getting from a region of poor fitness to a region of better fitness may mean moving across a plain of uniform fitness or through a region of poorer fitness still — something evolutionary algorithms are reluctant to do.

By virtue of a population of solutions, evolutionary algorithms conduct a parallel search of the space they are exploring and concentrate their attention on the most likely regions in which a solution might exist. Within a few generations, individuals start to "cluster" in the lowest found regions of the landscape. Consequently, the sampling rate in different parts of the landscape depends on its elevation above the zero error plain on which it is built. As sampling rates drop off in the (seemingly) less promising regions, the diversity of the population is reduced. With a drop in diversity comes a decline in the evolutionary algorithm's ability to explore the landscape, particularly complex landscapes.

One means of solving this problem of the evolutionary process restricting the search of the feature space is to change the way information about parameters is encoded. [WR97].

The typical evolutionary algorithm uses a co-ordinate set, where each number represents some parameter to be optimized. Biological chromosomes, however, do not have this one-to-one correspondence between genetic information and the traits an individual possesses. This information is held in a distributed fashion and used repeatedly for encoding different traits. One section of a chromosome can hold multiple, overlapping genes that contribute to specific traits. The use of such an encoding technique results in radically different search behavior by an evolutionary algorithm and large increases in
diversity of the population. The traditional search path to the lowest point on the error surface requires the co-ordinate set encoded solutions to travel along a path on that surface which is determined by the sequence in which better solutions are found. Being better, the solutions bias the location of offspring about them. This leads to a clustering effect that causes a significant drop in overall diversity. A solution encoded using a genetic-style encoding scheme (overlapped and distributed) exists in a different space that is then mapped to the error surface. By existing in a different space, a much broader distribution of solutions on the error surface can be maintained. This results in a more rapid exploration of the error surface, so locating the lowest point faster.

For simple problems, evolutionary algorithms — compared to almost any other method of finding a solution — tend to take a long time, which is why they are not often used. However, as the computation required increases more slowly with problem complexity than other methods, they come into their own with complex problems. Computation time, however, is always a significant concern. Increasing the population size allows for greater diversity and so potentially fewer generations necessary to find a good solution, but at an increased computational cost. Another possible problem lies in the size of a co-ordinate set. Each additional co-ordinate (i.e. parameter to the problem) dramatically increases the size of the space and therefore the number of possible paths through the space. Genetic-style encoding, however, allows for increasing numbers of parameters to be represented in the same sized space.

In previous work [CH98] we reported the results of a hybrid genetic-conventional representation and showed that this had advantages in complex problem spaces over the more traditional encoding methods. The traditional encoding of, for example, a four parameter solution (each parameter represented by a separate gene containing single real number) would be as shown in figure one.

Biologically, DNA is made from four acid bases, represented by the letters G, C, A and T. Rather than implementing such an abstract representation, the “artificial DNA” used in this previous work was binary in form. A chromosome, then, was simply a long binary string with markers that indicated the boundaries of a particular gene. Two markers specify each gene: one defines the start position of the gene (initiator), the other the end (terminator). An example that would produce the same four-parameter solution as that of figure one is shown in figure two below. It is considerably more complex, the purpose of the control chromosomes is to scale the genetically encoded solution to the feature space. Note that genes #3 and #4 share the same control chromosome and that all genes, to some extent, overlap.

The parameter values in both cases are the same. Take, for example, parameter two, which has the value of 6.0179. In the genetic representation, the attribute chromosome yields a value of 6 (the total number of 1's in gene 2), and the second control chromosome is used. Using successive multiplication of the control chromosome values, this yields \(-1.84 \times -13.8 \times 0.0395\) giving a value of 1.002984. The final value of 6.0179 is the product of 6 and 1.002984.

Summing the number of 1's in the attribute gene to obtain its value is not the only possibility, the gene could be evaluated as a binary string, for example. However, the effect a single bit mutation would have on the gene value would now be position dependent, which might or might not be an advantage. Similarly, successive multiplication by the control chromosome elements is not the only alternative, successive addition could easily be used. In this paper we consider the effects of the overlapped gene representation as used in the attribute chromosome in greater detail and do not further discuss explicitly the control chromosomes.
2. Gene representation and different chromosome structures.

In this work the value of a gene is defined as the sum of the bits within the gene. In the conventional artificial chromosome representation, the length of all the genes is fixed and the length of the chromosome is the sum of the lengths of the individual genes within it. In particular, the genes do not overlap. For overlapped gene chromosomes the length of a gene is not fixed. The value of the gene is the sum of the bits between the two markers (inclusive). During the creation of an offspring, two of the parents' four markers are chosen at random to act as the markers for the new child.

There are two ways to use markers. In the first case, the linear chromosome, the first marker to be encountered moving from the start of the chromosome is used as the initiator. The other marks the end of the gene. In the second case, the wraparound chromosome, the initiator and terminator are decided at the time of birth and do not depend on the marker values. If the initiator occurs after the terminator the bits to be summed go from the initiator to the end of the chromosome and then also from the start of the chromosome to the terminator. There are, therefore, two values for any given set of markers and bits that constitute an overlapped chromosome with wraparound depending which marker occurs first. The overlap distribution between genes without wraparound is normal, while wraparound it is pseudo-Gaussian.

Examples of chromosomes containing two genes are shown in figure three above. Note that for an overlapped chromosome with wraparound both possible values of each gene are shown.

As a result of summing the enclosed bits to find the value of a gene, many combinations of chromosome bit pattern and gene boundary markers result in the same gene value. The number of times any particular \( X \) \( Y \) value will result from all possible combinations of the two genes (\( X \) and \( Y \)) can be calculated. This is referred to as the total visits to that position. The visit frequency for a position is it's total visits expressed as a percentage of the total visits to the most visited position. The visit frequencies for a 16-bit chromosome that contains two genes are shown in figure four, one for each chromosome structure.

The information in figure four below reveals the following:

---

**Figure Four.** Visit frequency as a percentage of the number of visits to the most visited location.
• The conventional chromosome can only reach a more restricted area than either of the chromosomes with overlapped genes. With fixed gene boundaries there are far fewer unique possible combinations. The most visited place is where both genes have half their bits on. The distribution about this position is approximately Gaussian.

• For the chromosomes with overlapping genes, the position reached by the largest number of combinations is much closer to the origin. The visit distribution falls away from this position more slowly than for the conventional chromosome and a significant number of combinations of bit pattern and boundary markers will produce x y values far from the origin, especially for the wraparound chromosome.

• The region of high visitation is far larger for chromosomes with overlapping genes than for conventional chromosomes, especially for the wraparound chromosome.

3. The evolutionary algorithm.

The evolutionary algorithm used for the work described in this paper is a continuous replacement version of the algorithm originally described by Holland [1175].

For all individuals:
- Initialise population of chromosomes with randomly selected bits and, if appropriate, random gene boundary markers.
- Calculate the fitness of each individual based on gene values

Loop until fitness of best individual reaches some threshold or a set number of breeding events have occurred:
- Calculate the breeding probability of each individual, which is proportional to their fitness.
- Select two parents, the probability of selection of an individual being proportional to the breeding probability of that individual.
- Breed a new individual using one point crossover and mutation.
- Calculate the fitness of the new individual.
- If the fitness of the new individual is better than or equal to the least fit member of the population replace this with the new individual. If two or more of the current population have the same minimum fitness (less than the fitness of the new individual) randomly chose which of these is to be replaced.

End Loop

The crossover is conventional and all chromosome structures can experience individual bit mutation. That is, that there exists a certain probability that each and any bit may change state. When overlapped genes are being used, the position of the markers that define the boundary of the gene can also be mutated by being moved a random amount, the only constraint being that they must remain within the bounds of the chromosome bit structure.

4. The effects of gene mutation.

The additional mutation operator for variable length genes changes the length of the gene by moving one or both of the markers. This is actually essential to maintain diversity as without this a winnowing process will take place. The limited pool of marker positions will be slowly reduced as a new chromosome that does not use this marker value replaces the last example using this particular marker value. When little variation in gene boundary marker positions remains, new individuals tend to concentrate in just a few positions. The net effect of this is shown in figure five below.

A population can lose diversity in a region of uniform fitness. Each bitmap shows the positions of 1000 new individuals after several thousand breeding events. A chromosome with overlapped or non-overlapped genes but without wraparound is shown in A, a wraparound chromosome in B, both without gene boundary mutation. In C (non-wraparound) and D (wraparound) a gene boundary mutation probability of 5% is allowed.

Figure Five.
which records the effect of mutation on a population of chromosomes in a region of uniform fitness.

All the bitmaps in figure five show the positions of all new individuals accepted into the population of 200. The population was randomly initialized and in all cases the probability of any particular bit in the chromosome changing state is 50%.

Bitmap A shows the positions of all accepted new individuals from breeding event 9,000 to 10,000 for a non-wraparound chromosome (with or without overlapped genes) when the probability of mutation of any gene boundary marker position (if appropriate) is set to zero. Note the loss of diversity. The position of concentration will vary on successive runs but the convergence to one region is common.

Bitmap B shows the positions of all accepted new individuals from breeding event 19,000 to 20,000 for a wraparound chromosome, also with zero gene boundary marker mutation probability. Although the boundary markers are order dependent, which encourages diversity, diversity is slowly lost. Eventually it is probably that the population will have only two gene boundary marker values and will converge to two distinct regions.

Bitmap C shows the positions of all accepted new individuals from breeding event 19,000 to 20,000 for a linear chromosome, this time with the probability of gene boundary marker mutation set to 5%. Some concentration is clearly evident, although the gene boundary mutation has slowed this. A gene boundary mutation of 10% appears to be sufficient to overcome the tendency for the population to concentrate.

Bitmap D shows the positions of all accepted new individuals from breeding event 19,000 to 20,000 for a wraparound chromosome, also with the probability of gene boundary marker mutation set to 5%. Note that no concentration tendency is evident, a population with thus structured chromosomes that encountered a region of uniform fitness would continue to explore aggressively.

5. The performance of the chromosome types when placed in a fitness gradient.

The bitmaps in figure six show the progress of the evolutionary search as new individuals are introduced into the population when the surface being explored has a fitness increasing linearly from the top left-hand corner \( \{x: 0 \text{ and } y: 0\} \) to the bottom right-hand corner \( \{x: 512 \text{ and } y: 512\} \). The region of the surface shown covers \( \{x: 0-512\} \) and \( \{y: 0-512\} \). In all cases 25,000 new individuals are generated (but not necessarily introduced into the population) and the bitmaps show those that are accepted as they have a fitness equal to or greater than the worst member of the population. In all cases, the population is initialised within the region defined by \( \{x: 0-31\} \) and \( \{y: 0-31\} \) and the probability that any particular bit in the chromosome will change state is set to 0.05. Cases B and C also have a 0.05 probability that any boundary marker position will be altered by one position.

The conventional chromosome makes the slowest progress to the highest fitness, but does maintain limited diversity. The linear chromosomes make faster progress but fairly quickly lose diversity. The wraparound chromosomes almost reach the optimum position while also exploring the whole region.

Table one shows the results of a series of trials with the same parameters as the bitmap above, except that the convergence criterion was achieved with a fitness of 0.95 of the maximum possible fitness.

<table>
<thead>
<tr>
<th>Chromosome type</th>
<th>Average birth events</th>
<th>Standard deviation</th>
<th>Highest</th>
<th>Lowest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>87188</td>
<td>4191</td>
<td>90348</td>
<td>83195</td>
</tr>
<tr>
<td>Linear</td>
<td>52313</td>
<td>9027</td>
<td>66059</td>
<td>39014</td>
</tr>
<tr>
<td>Wrap round</td>
<td>41840</td>
<td>1724</td>
<td>45387</td>
<td>39675</td>
</tr>
</tbody>
</table>

Table One.
The wrap round chromosome structure is clearly superior. Not only does it reach the target fitness fastest, it is more consistent in performance and also explores the region far more thoroughly than either of the other structures.

6. Conclusion.

Consider the performance in a uniform gradient (figure six); the truly optimum position corresponds to the maximum possible gene length in a chromosome with all bits set to one. This is very hard to achieve since the higher the number of bits set in the chromosome, the higher number of bits that bit mutation will convert from set to clear, thus reducing the gene values. This is the reason that a control chromosome is required. It allows the attribute chromosome value to be scaled to suit the problem. Binary number attribute genes may be able to dispense with the need for control chromosomes, but the value of a bit now depends on the position of the bit within the gene(s) and that this can be dramatically changed by mutating the gene boundary markers. As a result the gene value change that could result from mutation may be of unacceptably high magnitude and mutated individuals be indistinguishable from randomly generated ones.

Continuous replacement means that a new solution must be at least as fit as the poorest solution currently in the population in order to be accepted. For the uniform gradient that produces figure six, the population is driven to produce individuals with longer genes. Unless wraparound is used, diversity is lost as the gene boundaries converge on either end of the chromosome. When wraparound is permitted the population is also driven towards long genes, but these are able to start anywhere along the chromosome and thus the diversity of gene boundary position is not lost.

Wraparound also has the effect of producing a near uniform distribution of overlap between genes, which sustains diversity of exploration as can be seen in figure six (C).

For these reasons, searching in alternative spaces, with appropriate structures and mutation operators, helps overcome the problems of an evolutionary algorithm becoming trapped within a specific region of the feature space in which either little or no fitness gradient occurs or in which a good, but not global, minima exists. This ability is particularly useful in the exploration of complex feature spaces in which such situations are more likely to occur.

7. References.


8. Acknowledgements.

It is a pleasure to acknowledge the value of discussions with our colleagues, Dr. P Cadusch and Mr. D Braendler.