C7.5 Speciation using tag bits

William M. Spears

Naval Research Laboratory

Abstract

This article considers novel methods for evolving species in a standard generational evolutionary algorithm. Unlike other methods, these methods replace the concept of distance between individuals with tag bits that identify the species to which an individual belongs.

C7.5.1 Introduction

In a traditional generational evolutionary algorithm (EA), N individuals reproduce according to their fitness (given by some objective function) and create a set of N offspring via the application of genetic operators. The N parents are then replaced by the N offspring to produce the next generation. One method for evolving species in a generational EA is through the use of sharing and restricted mating. With sharing, the similarity of individuals within the population is used to dynamically modify the fitness of those individuals. This dynamic modification is internal to the EA, thus what is changing is the EA's perception of the objective function, not the objective function itself. The intuition is simple - a peak in the space is treated as resource that a species can exploit. Thus individuals near one peak (forming a species) have to share the resource of that peak. Overcrowding on one peak implies that a resource is overused (there are too many individuals of that species). In this case the perceived fitness of the peak goes down, reducing selective pressure in that area of the space. On the other hand, peaks with few individuals have their perceived fitness increased, increasing selective pressure in those areas. The net effect is to apportion individuals in rough proportion to the relative height of the peaks in the space. Furthermore, the similarity metric is also used to restrict mating (recombination) to those individuals that are most similar.

Although sharing and restricted mating are rather general ideas, the implementation of those ideas that has received the most attention is by Goldberg and Richardson (1987). In their implementation, sharing and restricted mating together work very well, creating stable species on many peaks within the search space. Furthermore higher peaks get more individuals than lower peaks, in rough proportion to the relative heights of the peaks, as we would expect with the sharing mechanism. Goldberg's implementation makes two assumptions. The first is the number of peaks in the space. The second is that those peaks are uniformly distributed throughout the space. However, no sensitivity study indicates how well these assumptions must be met. Finally, the implementation (as described by Goldberg and Richardson (1987)) is also expensive. The similarity of each pair of individuals is measured, resulting in $O(N^2)$ similarity comparisons.

An alternative to the use of a similarity (distance) metric is to use "labels" to help identify individuals. The motivation for this alternative comes from nature. For example, this author does not decide he is Portuguese because he is in some sense similar to a few of his cousins. Rather, he decides he is Portuguese because his ancestors were labeled as Portuguese. Although this does not mean that one genetically inherits such "labels", they certainly can be culturally inherited. What if each EA individual has a label? Similarity then becomes simply a matter of seeing if two individuals have the same label. Of course, this implies that everyone with the same label is equally similar. The results of Spears (1994) suggest that the added precision of the distance metric is often not needed.

© 1995 IOP Publishing Ltd and Oxford University Press

Labels, then, can replace the distance metric. By implementing restricted mating and sharing with labels, the efficacy of restricted mating and sharing can be simultaneously accomplished in an efficient manner. Two similar methods for achieving this are outlined below.

C7.5.2 The method by Spears

The mechanism proposed by Spears (1994) is to use "tag bits" to label individuals. The tag bits are used to restrict mating and to perform fitness sharing. In standard fitness proportional selection the expected number of offspring for an individual is:

$$\frac{f_i}{\bar{f}}$$

where f_i is the fitness of the ith individual. The average fitness of all the individuals, $\bar{f}$, is given by:

$$\bar{f} = \frac{\sum_i f_i}{N}$$

where N is the population size.

In Spears (1994) fitness is shared. Suppose there are k different sets of tag bit values at a particular generation of the EA. We denote these sets as $\{S_0, \ldots, S_{k-1}\}$. The sets are numbered arbitrarily. Each individual belongs to one S_i and all individuals in a particular S_i have the same tag bit values. For example, suppose there is only one tag bit and that some individuals exist with a tag bit value 0 and that the remainder exist with tag bit value 1. Then we can (arbitrarily) assign the former set of individuals to S_0 and the latter set to S_1.

If we use $||\ ||$ to denote the cardinality of the sets, then with sharing the perceived fitness, F_i, is a normalization of the objective fitness f_i:

$$F_i = \frac{f_i}{||S_j||}, i \in S_j$$

where $||S_j||$ is the size of the species that individual i is in.

The average fitness of the population, $\bar{F}$, becomes:

$$\bar{F} = \frac{\sum_{i \in S_0} \frac{f_i}{||S_0||} + \ldots + \sum_{i \in S_{k-1}} \frac{f_i}{||S_{k-1}||}}{||S_0|| + \ldots + ||S_{k-1}||}$$

which is just:

$$\bar{F} = \frac{\sum_{i \in S_0} \frac{f_i}{||S_0||} + \ldots + \sum_{i \in S_{k-1}} \frac{f_i}{||S_{k-1}||}}{N}$$

since the species sizes have to total N (recall that no individual can lie in more than one species). The expected number of offspring for an individual is now:

$$\frac{F_i}{\bar{F}}$$

Restricted mating is performed by only allowing recombination to occur between individuals with the same tag bit values. Mutation can flip all bits, including the tag bits, thus allowing individuals to change labels. Experimental results, as well as some modifications to the above mechanism can be found in Spears (1994).

C7.5.3 The method by Perry

Perry's thesis work (Perry 1984) with speciation is extremely similar to the above technique. Perry includes both species and environmental regions in an EA. Species are identified via tag bits and an environmental region is similar to an EA population.

In Perry's work, recombination within an environment can occur only on individuals of the same species (i.e., with the same tag bit values). Mutation is allowed to change tag bits, in order to

 © 1995 IOP Publishing Ltd and Oxford University Press

introduce new species. The additional use of a "migration" operator, which moves individuals from one environment to another, does not have an analog in the work described by Spears (1994).

Perry gives an example of two species in an environment - fitness proportional selection (f_i / $\bar{f}$) is performed, and the average fitness of an environment is:

$$\bar{f} = \frac{\sum_{i \in S_0} f_i + \sum_{i \in S_1} f_i}{||S_0|| + ||S_1||}$$

or:

$$\bar{f} = \frac{\sum_{i \in S_0} f_i + \sum_{i \in S_1} f_i}{N}$$

where N is the population size of the environmental niche. One can see that the main difference between the two methods is the use of sharing in the computation of fitness in Spears (1994). Thus it is not surprising that in many of Perry's experimental runs one particular species would eventually dominate an environmental niche (however, it should be noted that in Perry (1984) the domination of an environment by a species was not undesirable behavior).

C7.5.4 Other related work

The use of tag bits makes restricted mating and fitness sharing more efficient because distance comparisons do not have to be computed. Interestingly, it is also possible to make Goldberg's implementation of sharing more efficient by sampling (Goldberg *et al* 1992). In other words the distance of each individual from the rest is estimated by using a subset of the remaining individuals. It will be intriguing to find out whether this change results in behavior similar to that described by Spears (1994).

Finally, this work has similarities to the EA research performed on parallel architectures. In a parallel EA, a topology is imposed on the EA population, resulting in species. However, there are some important differences between the parallel approaches and our sequential approach. In Spears (1994), the fitness of an individual and the species size are dynamic, based on the other individuals (and species). This will concentrate effort on more promising peaks, while still maintaining individuals in other areas of the search space. This is typically not true for parallel EAs implemented on MIMD or SIMD architectures. When using a MIMD architecture, species are dedicated to particular processors and the species remain a constant size. In SIMD implementations, one or two individuals reside on a processor, and species are formed by defining overlapping neighborhoods. However, due to the overlap, one particular species will eventually take over the whole population.

References

Goldberg D E and Richardson J 1987 Genetic algorithms with sharing for multimodal function optimization *Proceedings of the Second International Conference on Genetic Algorithms* (Cambridge, Mass.: Lawrence Erlbaum) pp 41-49.

Goldberg D E, Deb K and Horn J 1992 Massive multimodality, deception, and genetic algorithms *Proceedings of the Parallel Problem Solving from Nature Conference* (North-Holland) pp 37-46.

Perry Z 1984 Experimental study of speciation in ecological niche theory using genetic algorithms *Doctoral Thesis, University of Michigan.*

Spears W 1994 Simple subpopulation schemes *Conference on Evolutionary Programming* (World Scientific) pp 296-307.

© 1995 IOP Publishing Ltd and Oxford University Press