A SIMPLE HIERARCHICAL MODEL OF NATURAL SELECTION

David J. Winter
The University of Michigan

1. Introduction

The purpose of this paper is to describe a hierarchical model of an evolutionary process resembling natural selection. In the model, one begins with an initial "population" of "genetic structures". Each genetic structure of a population reproduces "progeny" and the progeny of these genetic structures make up the next population. The degree of success of reproduction of progeny by a particular genetic structure is determined by its "fitness" in the "environment". One expects fitness of a given genetic structure to correlate highly with some of its characteristics in the environment, the characteristics of that genetic struc- ture being seen through the eyes of "descriptors". With the passage of time, one expects highly fit types of genetic structures to emerge and expects some of these highly fit types to be increasingly "stable", stability being defined with respect to descriptors and therefore representing the continuance of characteristics.

In this paper, we present only a very general model. The genetic structures them- selves might represent organisms within a physical environment or artificial structures within an artificial environment. It is hoped that readers from a wide variety of backgrounds may extend and improve this model and develop versions and variations of it which enable them to effectively predict the rate and direction of convergence of genetic structures to highly fit stable types of genetic structures of interest to them. (See Kimira and Ohta, 1971; Holland, 1973; and Holland, to appear.)

2. Genetic Structures, Genetic Operators, Populations and Fitness

We start with a set X (such as the set of all strings of letters). An element x of X is called a genetic structure. A genetic operator is a nonnegative real valued function $w(x,y)$ of two variables on X. For any two genetic structures x and y, $w(x,y)$ is the probable frequency of occurrence of y among the progeny of x. Given a set $w_1, \ldots, w_m$ of functions from X to X (which may be thought of as crossover, inversion, mutation, etc.) and a set of nonnegative real valued functions $\pi_1(x), \ldots, \pi_m(x)$ (which may be thought of as the probable frequency of occurrence at x of the $w_1, \ldots, w_m$) we obtain a genetic operator $w = (w_1, \pi_1, \ldots, w_m, \pi_m)$ defined by letting $w(x,y) = \sum_{i=1}^{m} \sum_{y=w_i(x)} \pi_i(x)$. A population is non- negative real valued function p on X which assigns to each genetic structure x its frequency of occurrance p_x in the population p. An environment is a nonnegative real valued function u on X which assigns to each genetic structure x its fitness $u(x)$ in the environment u.

A genetic operator $w(x, y)$ determines for each population p in an environment u the probable constituency of the next population p', the probable frequency of occurrance of y in p' being $p'_y = \sum_x u(x) p_x w(x, y)$.

3. Descriptors and Stability

A descriptor on a set X of genetic structures is a nonnegative real valued function $d(x, y)$ of two variables on X such that

1. $d(x, x) = 0$ for x in X ;
2. $d(x, y) = d(y, x)$ for x, y in X ;
3. $d(x, y) + d(y, z) \geq d(x, z)$ for x, y, z in X.

If $d(x, y)$ is small, x resembles y with respect to d. If $d(x, y) = 0$, x and y are indistinguishable with respect to d. Note that if $d(x, y) = 0$, then $d(x, z) = d(y, z)$ for all z. Letting $d(x, y)$ be a descriptor on X, we obtain an equivalence relation $d(x, y) = 0$ on X. The equivalence class $\{y \in X \mid d(x, y) = 0\}$ of x is denoted $\bar{x}$ and consists of all genetic structures indistinguishable from x with respect to d. The set $\{\bar{x} \mid x \in X\}$ of $\bar{x}$ is denoted by X/d or by $\bar{X}$. One sees easily that the descriptors d on X are the functions on X of the form $d(x, y) = m\big(f(x), f(y)\big)$ where f is a function from X into a metric space with metric m. If $d_1, \ldots, d_m$ are real valued functions on X, then $d(x, y) = \sqrt{\sum_1^m \big(d_i(x) - d_i(y)\big)^2}$ defines such a descriptor d which is denoted $d = (d_1, \ldots, d_m)$. If S is a finite set of real valued functions $d_1, \ldots, d_m$ on X, we let d_S denote the descriptor $(d_1, \ldots, d_m)$. If d^0 and d^1 are descriptors on X such that $d^0(x, y) \geq d^1(x, y)$ for all x, y, which we indicate by writing $d^0 \geq d^1$, then d^1 may be regarded as a descriptor on the set $X^1 = X/d^0 = \bar{X}$ of equivalence classes of X with respect to d^0, letting $d^1(\bar{x}, \bar{y}) = d^1(x, y)$ for x, y in X. If $d^0, d^1, \ldots, d^m$ are descriptors with $d^0 \geq d^1 \geq \ldots \geq d^m$, we may therefore regard d^i as a descriptor on X^i for $0 \leq i \leq m$ where $X^0 = X$, $X^1 = X^0/d^0, \ldots, X^m = X^{m-1}/d^{m-1}$. If $S_0, S_1, \ldots, S_m$ are finite sets of real valued functions on X and $S_0 \geq S_1 \geq \ldots \geq S_m$, then $d_{S_0} \geq d_{S_1} \geq \ldots \geq d_{S_m}$, so that we may regard d_{S_i} as a descriptor on $X^i = X^{i-1}/d_{S_{i-1}}$ for $i = 0, 1, \ldots, m$.

Finally, we need the concept of stability, which indicates recurrance of a characteristic in successive generations. Fixing a descriptor d on X and small positive numbers δ and ϵ, we say that a genetic structure x is stable with respect to d if $\sum_{d(x, y) \geq \delta} w(x, y) / \sum_y w(x, y)$ $< \epsilon$, that is, if the ratio of the number of probable progeny of x not resembling x with respect to d to the number of all probable progeny of x is small.

4. The Evolutionary Process

We now describe the evolutionary process of the present model, which proceeds at one level, then passes to a next and so on. We begin with an 8-tple $(X, w, u, S, \delta, \epsilon, n, p(0))$ where X is a set of genetic structures, w is a genetic operator on X, u is an environment for X, S is a finite set of real valued functions on X, δ and ϵ are small positive real numbers, n is a positive number and $p(0)$ is an initial population. Within the present level, the initial population gives rise to successive populations $p(t)$ $(t = 1, 2, \ldots, i)$ as specified by the probabalistically applied genetic operator. The genetic structures in the populations $p(t)$ are examined for fitness and for stability with respect to the descriptors d_T where T ranges over all subsets of S which are obtained from S by removal of a single element of S. As soon as n stable genetic structures are found, that T for which fitness of the stable genetic structures is maximal is chosen. At this point, a transfer is made to the next level of the process by passage to the next δ-tple $(\bar{x}, \bar{w}, \bar{u}, \bar{S}, \bar{\delta}, \bar{\epsilon}, \bar{n}, \bar{p}(0))$. The set $\bar{X}$ is taken to be X/d_T. The set $\bar{S}$ of functions on $\bar{X}$ is the set of functions on $\bar{X}$ induced by the functions in T. The values δ, ϵ, n can be taken as before or modified in accordance with the rate of convergence. For each $\bar{x}$ in $\bar{X}$, $\bar{w}(\bar{x}, \bar{y})$ is taken to be $\bar{w}(\bar{x}, \bar{y}) = \dfrac{1}{\bar{x}} \dfrac{1}{\bar{y}} \displaystyle\sum_{\substack{x \in \bar{x} \\ y \in \bar{y}}} w(x, y)$ and $\bar{u}(\bar{x})$

is taken to be $\bar{u}(\bar{x}) = \dfrac{1}{|\bar{x}|} \displaystyle\sum_{x \in \bar{x}} u(x)$, where $|\bar{x}|$, $|\bar{y}|$ are the number of elements of $\bar{x}, \bar{y}$ respectively. Finally, the initial population $\bar{p}(0)$ of the new level is defined in terms of the population $p(i)$ at the end of the preceding level, the definition being $\bar{p}(0)_{\bar{x}} = \displaystyle\sum_{x \in \bar{x}} p(i)_x$.

A drawback to the present model is the systematic search through all of the T. An improved model might be found by regarding the descriptors d_T with T any subset of S as genetic structures and producing a suitable genetic operator and environment so that these descriptors could quickly evolve into descriptors which are effective in locating highly fit and stable genetic structures in the original system.

References

Holland, J. (1973), "Genetic Algorithms and the Optimal Allocation of Trials", SIAM J. Comp. 2, 88-105.

Holland, J. (to appear) Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor (1974).

Kimira, M. and T. Ohta (1971), Theoretical Aspects of Population Genetics, Princeton University Press, Princeton, Ch. 5.