References

Cullis, C. A. (1988). Control of Variation in Higher Plants. In Ho, Mae-Wan & Fox, Sidney W. Eds. *Evolutionary Processes and Metaphors* John Wiley & Sons Ltd., Chichester.

Davis, L. (1991) Bit-Climbing, Representational Bias, and Test Suite Design. *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo.

de Garis, H. (1990) Genetic Programming. *Machine learning: proceedings of the seventh international conference*, 132-139.

Gould, S. J. (1991). Of Mice and Mosquitoes. *Natural History*, July, 12-20.

Grossberg, S. (1982). *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*, Reidel Press, Boston.

Harp, S. A., Samad, T. & Guha, A. (1989) Towards the Genetic Synthesis of Neural Networks. *Proceedings of the Third Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo.

Louis, S. J. (1991) Designer Genetic Algorithms: Genetic Algorithms in Structure Design. *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo.

Martí, L. J. (1992) Genetically Generated Neural Networks I: Representational Effects, Submitted for publication.

Merrill, J. W. L., Port, Robert F. (1991) Fractally Configured Neural Networks. *Neural Networks*, **4**, 53-60.

Miller, G. F., Todd, P. M. & Hegde, S. U. (1989) Designing Neural Networks using Genetic Algorithms. *Proceedings of the Third Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo.

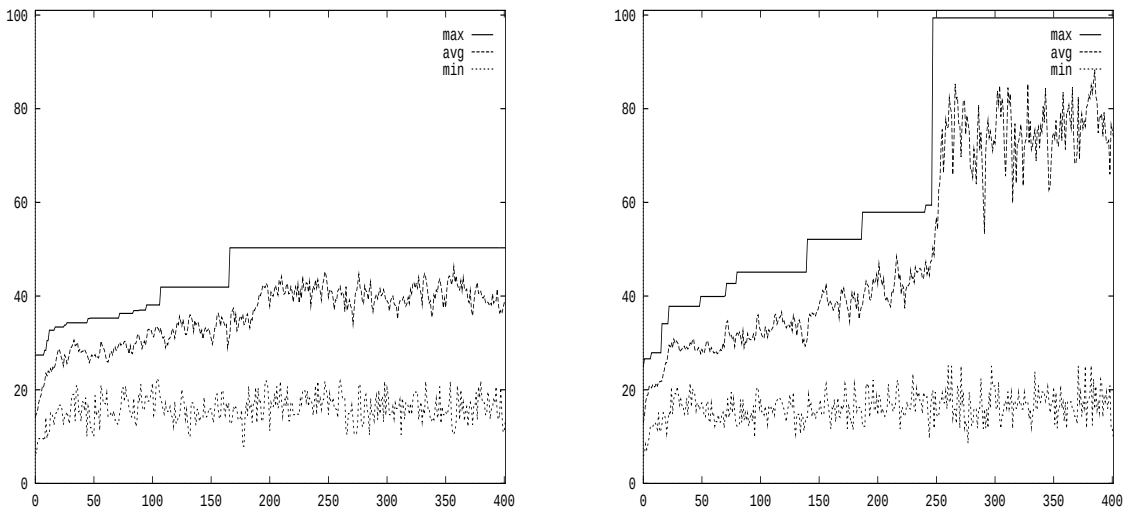

Figure 4: Fitness of best, worst, and average population members of two population runs of the Inner Genetic System over 400 generations. The left run is the result of an early outer individual, and the right run is the result of a latter individual. Fitness of 100 is optimal.

VI. Conclusions

The purpose of this paper is not to find the optimal representation for a genetic system. It rather intends to show that such an optimization does not have to be done heuristically by the system designer, but that it can be aided by the genetic system itself.

Variants of the system presented may be suggested. The choice of fitness functions can be changed. The inner system may use a linear function instead of the inverse function used. The outer system may use an average or the best result over many runs for the optimal individual, instead of the last one obtained. As is, the outer system can specify that more than one connection descriptor is located in the same position. By the same token, it can render useless areas of the inner genome. The choice of binary encoding may be modified with the use of gray encoding.

At a more fundamental level, the organization of the genetic material can also be modified. Perhaps a more biologically based approach would be to use a chaotic system to determine the shape of the network or location of each node (Merrill & Port, 1990). Ultimately the two systems can be combined into one genome.

The present paper represents a first step in genetically aided system design and self optimization. Biological evidence for such systems exists from research in natural genetics. As already mentioned, the appearance of homeoboxes represents one of them. Homeoboxes regulate timing and transcription of other genes. Also, the introduction of color vision in primates is believed to have been the result of duplication and later alteration in the representation of specification of retinal cells (Cullis, 1988). At a more basic level, the existence of diploidy in not all but some organisms is yet another indication of a fundamental variation in representation of similar genetic material.

Many more aspects of natural genetics remain to be successfully integrated into artificial genetic system. As more methods from natural genetics are incorporated into GAs, these systems should become more useful and find a wider range of applications.

part of the gene (the outer system) varies much slower than and independently of the rest of the genetic material.

In either case, a second and quite independent set of genetic material is needed. The information contained in this genetic material should be able to specify the location of each connection descriptor. Among all the possible manners of determining these locations a relatively simple one was chosen. Basically, the genome was used as a pointer table. A set of four alleles determines the location of each connection descriptor in a binary encoded form. For example, alleles 1 through 4 determine the location where the connection descriptor for the connection from node 1 to itself is to be relocated. Similarly, alleles 5 through 8 determine the location where the connection descriptor from node 1 to node 2 is to be relocated. This representation requires 64 alleles to determine the location of 16 connection descriptors and 4 alleles per connection descriptor. The present system was chosen for its simplicity, and found to be robust enough for the task at hand.

An example of the effect of an outer genome on the location of connections of an inner genome can be seen in Figure 2.

The fitness function for the outer genetic individuals was the fitness of the best individual of the last generation of the Inner Genetic System.

V. Results

The system was implemented on a Thinking Machines CM2 using 8K processors, and a Sun Sparcstation as the front end. This allowed a population of 90 individuals for each of the genetic systems. Therefore 8100 simultaneous ODE's were solved for each generation.

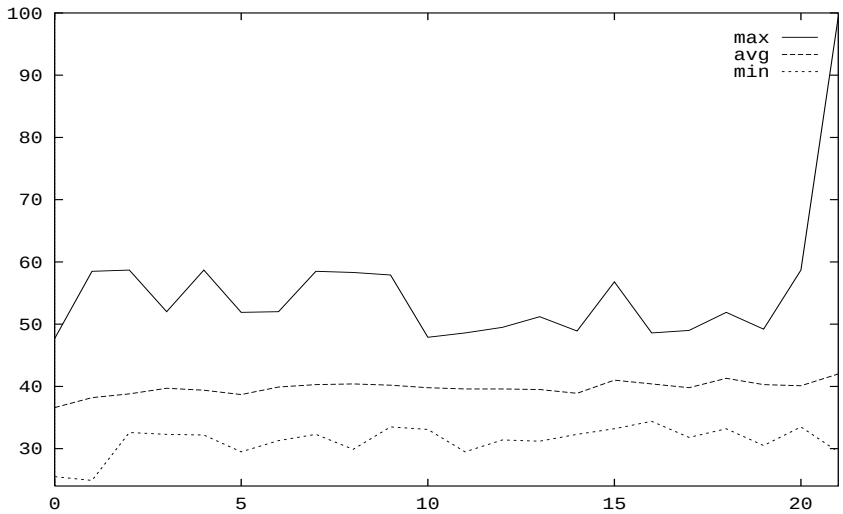

Figure 3: Fitness of best, worst, and average population members of the Outer Genetic System over 20 generations. Fitness of 100 is optimal.

Figure 3 shows the results of the outer genetic search over 20 generations. Each outer generation consisted of 20 inner generations. Whereas this is not enough to provide us with the optimal population member, partial optimization is sufficient to provide the outer system with the proper direction for the search. This can be seen by the upward direction of the curve for best and average members as shown in Figure 3. This is also seen by comparing a population run of the inner system at an early generation of the outer system with a population run of the inner system at a later generation. Examples of these two runs can be seen in Figure 4.

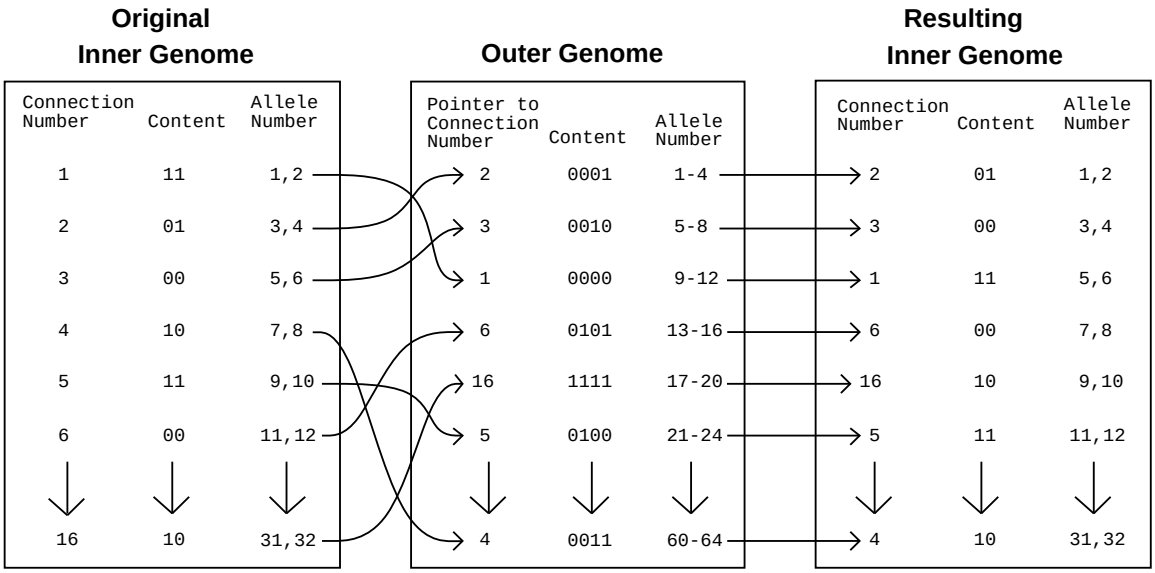

Figure 2: Example of the effect of an outer genome on the location of the connections of an inner genome.

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)(I_i + \sum f(x_g)) - (x_i + D)\sum f(x_h)$$

Where x_i is node i, i ranging from 1 to 4, A, B and D are constants set at 6.0, 5.0, 5.0 respectively, $f(x)$ is the neuron's feedback equation ($f(x) = x$; if $x > 0$), g is the set of excitatory nodes, and h the set of inhibitory nodes. The sets of inhibitory and excitatory nodes are determined by the genome. For the target circuit, g was the node itself ($f(x_i)$), and h consisted of every other node ($\sum_{h \neq i} f(x_h)$). The fitness function used was:

$$O(x) = \frac{100}{1 + \sum(K_t - y_t)}$$

where K_t is the optimal output value at time t, and y_t the actual output from the network at time t.

In this Genetic System, representation can be seen as affecting the system in at least two forms. First, the form the table takes, has been shown to heavily influence the outcome of the system (Martí, 1992). The possible values that this table can take are quite few, and can be studied analytically and exhaustively and not treated here.

Another form in which representation plays a role, is in the location of the description of each connection. It is difficult to examine the effect of altering the location of the description of each connection by testing all possible location combinations within the genome. This will be examined with the Outer Genetic System.

IV. Outer Genetic System

In order to test the effects of placing the description of each connection in different locations, an Outer Genetic System was introduced. This Outer Genetic System can be viewed in two different ways. First, it can be seen as an entirely separate genetic system, where the fitness of each individual depends upon the effectiveness of a certain connection specification placement. The effectiveness of this connection placement is determined by executing an inner genetic system, and observing how well it performs.

Another way of viewing this Outer Genetic System is as an additional set of alleles which determines the placement of the description of each connection. According to this view, the location description

In order to carry out the simulation within a reasonable time, a few unbiological simplifications have been made. The representation description was made as simple as possible in order to facilitate its analysis. Even with these simplifications, the system is quite complex to simulate. The representation description is referred to as the "Outer Genetic System" and implemented as a completely separate genetic engine. The function controlled by it is referred to as the "Inner Genetic System". The genome layout of the Inner Genetic System will be optimized by the Outer Genetic System.

GAs can be seen as a Boltzmann-like massively parallel, stochastic gradient descent system. As such, it is uniquely qualified to avoid local minima or sub-optimal results. The separation of genetic material and the genetic engine allow for a certain degree of application independence. It is for these reasons, that GAs are well suited to search the space of possible network topologies. This is a problem space which is too large to be searched exhaustively.

II. Genetic Environment

A Genetic Algorithm makes use of a string of alleles, called a genome, where it represents the necessary information to describe an individual. This information can be tested for fitness, and the resulting fitness can be compared with that of other individuals. As in natural selection, the individuals with better fitness have a higher probability of reproducing and therefore maintaining their genetic material in subsequent generations. As better organisms appear, and are maintained, the overall fitness of the population also rises.

The genetic engine implemented here makes use of just two genetic operations: mutation and crossover. Mutation consists of randomly selecting an allele and altering its value with certain probability. In all the simulations carried out here, the mutation probability was set unchanged at 0.01. Crossover is carried out by selecting two genomes and choosing a point where the genomes will be split. At this point, the genomes will be split and recombined with the remaining section of the alternate genome. The crossover probability was also fixed and it was set at 0.85.

In addition, the genome of the best individual of each generation was copied, unchanged, for the next generation.

III. Inner Genetic System

The Inner Genetic System is the same as the one used in Martí (1992). The purpose of the Inner Genetic System is to generate Neural Networks. Each pair of alleles determines the connectivity among two nodes in a 4 node neural network. Each allele contains a binary value, and when combined determines the connectivity according to the table shown in Table 1. The location of each pair of alleles determines which connection is being specified as a 4x4 connectivity matrix. In the previous paper, this specification remained fixed. For example alleles 1 and 2 determine the connectivity from node 1 to itself, alleles 3 and 4 determine the connectivity from node 1 to node 2 and so on. This representation requires 32 alleles to determine the connectivity of all 4 nodes or 16 connections with 2 alleles per connection.

Allele pair	Connection type
00	Inhibitory
01	Disconnected
10	Disconnected
11	Excitatory

Figure 1: Table of allele representation of connections.

In order to calculate the fitness of each individual, a system of equations for the network specified was solved. The target configuration of the network was a competitive feedback circuit (Grossberg, 1982). The system of equations used was:

Genetically Generated Neural Networks II: Searching for an Optimal Representation

CAS/CNS-TR-92-015
Leonardo Martí
Boston University
Center for Adaptive Systems
111 Cummington Street
Boston, MA 02215
lmarti@cns.bu.edu

Abstract

Genetic Algorithms (GAs) make use of an internal representation of a given system in order to perform optimization functions. The actual structural layout of this representation, called a genome, has a crucial impact on the outcome of the optimization process. The purpose of this paper is to study the effects of different internal representations in a GA, which generates neural networks. A second GA was used to optimize the genome structure. This structure produces an optimized system within a shorter time interval.

I. Introduction

Though the field of natural genetics is progressing quite rapidly, understanding of the genetic process is still quite incomplete. Even so, knowledge of the natural genetic process has not been completely incorporated into the field of Genetic Algorithms (GAs). To this end, the research reported here tests some new approaches and functions to be used with GAs. In a previous paper (Martí, 1992), and in several other sources (Garis, 1990; Miller, Todd, & Hegde, 1989; Harp, Samad, & Guha, 1989) it has been shown how GAs can be used to generate optimal and novel neural network architectures. Also, it is widely understood how influential the genome representation can be in the success of the genetic search (Davis, 1991; Louis, 1991; Martí, 1992). Here, a genetic algorithm has been used to explore alternative genome representations of another genetic system.

When examined carefully, it becomes clear that natural genetics must possess this functionality in order to provide the flexible evolution that we observe today. Natural genetic evolution is capable of adding new functionality to a species represented by relatively stable genetic material. How new material is added may vary, but one method consists of duplicating a section of existing material. Once this material has been duplicated, variations of it (resulting from other genetic operations such as mutation and crossover) will result in the exhibition of the new functionality. The representation of this new functionality in the gene can be based on the representation of the original functionality. But this representation should be able to vary in order to find a more appropriate representation and to survive as stable genetic material.

How these alternative representations are generated and tested is far from being fully understood, but both representation descriptions and actual function descriptions must coexist in the same genetic material. However they do not necessarily reside on the same gene. This paper has assumed that the representation description is more stable than the actual functions. This assumption seems logical if one believes that this representation description must be older genetically than any functions controlled by it, and that older genetic material becomes more stable. An example of this can be seen in the appearance of "homeoboxes" (Gould, 1991) in natural genetic evolution.