<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Examining the impact of neutrality on genetic algorithm population evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Hill, Seamus; O'Riordan, Colm</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2015-11-12</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>SciTePress Digital Library</td>
</tr>
<tr>
<td><strong>Link to publisher's version</strong></td>
<td><a href="https://dx.doi.org/10.5220/0005594301960203">https://dx.doi.org/10.5220/0005594301960203</a></td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/14677">http://hdl.handle.net/10379/14677</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td><a href="http://dx.doi.org/10.5220/0005594301960203">http://dx.doi.org/10.5220/0005594301960203</a></td>
</tr>
</tbody>
</table>
Examining the Impact of Neutral Theory on Genetic Algorithm Population Evolution.

Seamus Hill, Colm O’Riordan
College of Engineering & Informatics, National University of Ireland Galway, Ireland
{seamus.hill, colm.oriordan}@nuigalway.ie

Keywords: Neutral Theory, Genetic Drift, Neutrality, Genotype, Phenotype, Genetic Algorithms

Abstract: This paper examines the introduction of Neutral Theory as proposed by Kimura (Kimura, 1968) into the genotype-phenotype mapping of a Genetic Algorithm (GA). The paper examines the evolution of both a simple GA (SGA) and a multi-layered GA (MGA) incorporating a layered genotype-phenotype mapping based on the biological concepts of Transcription and Translation. Results illustrate that the MGA population’s evolution trajectory is quite different to that of the SGA population over dynamic landscapes and that the introduction of Neutral Theory implicitly maintains genetic diversity within the population primarily through genetic drift in association with selection.

1 INTRODUCTION

Sewall Wright (Wright, 1932) noticed that many random changes in the frequency of alleles occurring in a population were not related to selection. His observations indicated that this genetic drift was an important component in the evolutionary process. Neutral theory as proposed by Kimura (Kimura, 1968), offered an alternative to the Darwinian view, stating that the mutations involved in the evolutionary process are neither advantageous nor disadvantageous to the survival of an individual and that most mutations are caused not by selection, but rather by random genetic drift. However, in (Kimura, 1983), Kimura pointed out that although natural selection does play a role in adaptive evolution, only a tiny fraction of DNA changes are adaptive. The vast bulk of mutations are phenotypically silent.

By adopting the principal of Darwinism, simple genetic algorithms (SGA), can be viewed as implementing the process of evolution without containing any explicit neutral mutations. In other words each mutation is either an advantage or a disadvantage to the individual in terms of fitness, with selection then propagating the fitter individuals. As the search progresses, the exploration and exploitation ratio decreases as the population converges. If we are to implement a genetic algorithm (GA) based on the principles of neutral theory then neutrality needs to be introduced. Neutrality can be viewed as a situation where a number of different genotypes can represent the same phenotype.

Traditionally, GAs are evaluated and compared in relation to performance statistics. In this paper the authors examine the populations dynamics associated with a SGA and a multi-layered GA (MGA). The motivation is to develop a tunable, synonymous, non-trivial GA representation which incorporates Neutral theory, in order to gain an understanding of the effects of neutrality on population dynamics. The contributions are as follows:

- an analysis of the impact of neutrality on the population evolution.
- an examination of the impact of neutrality on population variation.
- an illustration of the impact of neutrality on phenotypic variability.

The paper is laid out as follows: Section 2 gives a brief background to Neutral theory and the use of neutrality in GAs. Section 3 outlines the MGA used in the paper, while Section 4 describes the experiments undertaken. Section 5 outlines and analyses the results and finally, Section 6 concludes.

2 BACKGROUND

Neutral Theory as discussed by Kimura (Kimura, 1968), argues that mutation, not selection, is the main
force in evolution. He describes how a mutation from one gene to another can be viewed as being neutral if it does not affect the phenotype, as the number of different genotypes which store genetic information is far greater than the number of phenotypes. This implies that the representation from genotype to phenotype must incorporate an element of redundancy and neutral mutations are possible. Mutations can be viewed as neutral if they change the genotype but don’t impact on the phenotype.

Much previous work on evolution was based on observed changes at a phenotypic level. Having access to molecular data, Kimura was able to experiment at a level not previously possible. As a result of examining the evolutionary rates of nucleotide substitutions, Kimura identified two types of gene mutations: “replacements of one nucleotide base for another” (Kimura, 1983) and “structural changes consisting of deletions and insertions of one or more nucleotide’s bases as well as transpositions and inversions of larger DNA segments” (Kimura, 1983). The results of Kimura’s investigations led to the development of his theory of molecular evolution: “…a majority of nucleotide substitutions in the course of evolution must be as a result of random fixation of selectively neutral or nearly neutral mutants rather than positive Darwinian selection, and many of the enzyme polymorphisms are selectively neutral and maintained by the balance between mutational input and random extension.” (Kimura, 1983).

Generally, neutrality can be viewed as a situation where the size of the search space is increased, without an equivalent increase in the solution space. This results in a situation where a mutated individual, at the genotypic level, can still represent the same phenotype. Neutrality should be beneficial when neutrality changes the search bias in order to improve the probability of locating the global optimum.

Although originally viewed as being anti-Darwinian, Kimura (Kimura, 1983) stated that although natural selection is important in evolution, the number of DNA changes which are adapted in evolution are small, with the vast majority of mutations being phenotypically silent. Following Kimura, work by King and Dukes (King and Dukes, 1969) describes how much of the evolution of proteins is down to neutral mutations and genetic drift. A number of studies focused on neutral theory (Schuster et al., 1994; Huynen et al., 1996; Huynen, 1996; Schuster, 1997; Fontana and Schuster, 1998; Shipman, 1999; Reidys and Stadler, 1998) and illustrated that by introducing redundant representation and thus neutral mutation, the connectivity between fitness landscapes can be altered. In other words, when a number of genotypes represent the same phenotype, they can be viewed as a neutral set and in turn alter the way in which a population explore the search space.

Previous work on GA genotype-phenotype mapping which introduce neutrality, have produced results indicating the neutrality may prove useful in changing environments and over more difficult landscapes. Ebner et al. (Ebner et al., 2001b; Ebner et al., 2001a), outlined how high levels of mutation could be sustained by having neutral networks present. They also identified that neutral networks assist in maintaining diversity in the population, which may be advantageous in a changing environment. Similar findings were found in (Grefenstette and Cobb, 1993; Hill and O’Riordan, 2104). Toussaint and Igel (Toussaint and Igel, 2002) argued that approaches to self-adaption in evolutionary algorithms can be viewed as an example of the benefits of neutrality, because with non-trivial neutrality, different genotypes which are part of the same neutral set may have different phenotypic distributions.

3 MULTI-LAYERED GENETIC ALGORITHM (MGA)

The concepts of Variation and Variability can to be differentiated. Variation can be described as the difference between individuals in a population and can be seen as relating to a collection. Variability, on the other hand, can be described as the leaning to vary and the variability of a phenotypic trait describes the way in which it changes in response to environmental and genetic influences (Wagner and Altenberg, 1996). When exploring the phenotypic space, it is critical to gain an understanding of the variational topology in trying to determine the shape of the landscape (Toussaint, 2003b). Many evolutionary algorithms are created using a fixed variation topology, in other words you don’t need to track the neutrality to explain the evolution trajectory. However, in nature, phenotypic variation landscapes are not fixed. These non-fixed phenotypic variation landscapes can be referred to as non-trivial in terms of their genotype-phenotype map (Toussaint, 2003b). A non-trivial genotype-phenotype map can be viewed as having the following characteristics: firstly, a phenotype can be encoded by many genotypes and secondly, the phenotypic variability of a number of phenotypes will depend on their genotypes (Tou-
ssaint, 2003a). Therefore, every genotype can be described as a combination of a phenotype and any kind of neutral traits which exist within the genotype, where different neutral traits associated with the same phenotype give a different representation of the phenotype in question (Toussaint, 2003a).

By incorporating a number of layers into the genotype-phenotype mapping, the MGA increases the size of the search space, without increasing the size of the solution space. For example, if the binary representation of the 4-bit deceptive problem is increased by 20 extra loci, then each phenotype will be represented by $2^{24}$ points in the genotype space rather than $2^4$, to form a connected neutral network (Harvey and Thompson, 1996).

The transcription phase of the MGA maps the binary genotype to a string of characters taken from a four letter alphabet $A, C, G$ and $T$. The mapping is carried out as follows: $00 \rightarrow A; 01 \rightarrow C; 10 \rightarrow G$ and $11 \rightarrow T$. The mappings for this stage are $A \rightarrow U; C \rightarrow G; G \rightarrow C$ and $T \rightarrow A$. Following transcription, the translation phase takes place. Upon initialisation, the MGA creates a translation table, using characters taken from a four letter alphabet $A, C, G$ and $U$, based on the granularity of the neutrality selected by the user. With a 6-bit MGA representation, a $2^6$ translation table is needed. A 6-bit MGA representation requires 6 bits for each individual element of the phenotype, which we refer to as a phenotype; a 6-bit MGA representation requires 6 bits for each phene, a 6-bit MGA representation is shown in Figure 1. Missense mutation in nature is carried out at the RNA level. In relation to the MGA, Missense mutation maps as follows: $A \rightarrow U, C \rightarrow G, G \rightarrow A$ and $U \rightarrow C$. To summarise the variation operators, one-point crossover and single-point mutation occur at the genotype level prior to transcription and missense mutation takes place before translation.

4 EXPERIMENTATION

In order to examine the population evolution for both the SGA and the MGA, experiments were conducted over a 4-bit fully deceptive landscape (Whitley, 1991) which is reversed after generation 50, allowing the local optimum to become the global optimum. Table 1 and Table 2 respectively, outline the fitness values before and after the landscape changes. The 4-bit fully deceptive problem was chosen as it should provide a similar level of difficulty for both the SGA and the MGA. Neutral networks, also known as plateaus, can be viewed as a set of neighbouring solutions with the same fitness value. A neutral network can also be defined as points in the search space that are connected by neutral point mutations, where the fitness is the same for all points.

<table>
<thead>
<tr>
<th>Fitness Values - 4-bit Deceptive Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(0000) = 28</td>
</tr>
<tr>
<td>f(0100) = 22</td>
</tr>
<tr>
<td>f(1000) = 20</td>
</tr>
<tr>
<td>f(1100) = 8</td>
</tr>
</tbody>
</table>

Table 1: Fully Deceptive 4-bit Problem

<table>
<thead>
<tr>
<th>Fitness Values - Reversed Deceptive 4-bit Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(0000) = 30</td>
</tr>
<tr>
<td>f(0100) = 2</td>
</tr>
<tr>
<td>f(1000) = 0</td>
</tr>
<tr>
<td>f(1100) = 18</td>
</tr>
</tbody>
</table>

Table 2: Reversed Fully Deceptive 4-bit Problem

Figure 2 graphically illustrates the landscape of the 4-bit deceptive problem (Whitley, 1991) and Figure 3, represents the associated heat map which is designed to illustrate visually, the neutral networks (fitness plateaus).

The numerical fitness values for each plateau are also shown. To analyse the adaptive qualities of both GAs, the landscape reverses after generation 50, Figure 4, illustrates the landscape after the change and Figure 5
shows the reversed fitness plateaus, where the global optimum becomes the local optimum and visa-versa.

5 RESULTS

The analysis that follows is broken into three parts: the impact of neutrality on population evolution; the impact of neutrality on variation and the impact of neutrality on phenotypic variability. However, the results begin by comparing both GAs in a conventional manner based on performance. Figure 6, illustrate the off-line (averaged best fitness) and on-line (averaged fitness) performance for both the SGA and the MGA. The results indicate that the changing 4-bit deceptive landscape initially proved equally easy for both the SGA and the MGA. However, after the landscape changes, the SGA becomes trapped on the local optimum, while the MGA succeeds in locating the global optimum. A Wilcoxon Rank sum test indicates that the results shown in Figure 6 were statistically significant.

The remaining analysis examines the evolution of both populations in an attempt to gain a better understanding of the impact on neutrality on population dynamics over the course of an evolutionary time period and to illustrate the performance of both population in a dynamic environment.

5.1 Examining the Impact of Neutrality on Population Evolution

5.1.1 Analysis Before the Landscape Change

In relation to the evolution of the SGA’s population, Figure 7, gives an overview of the population distribution over the landscape at generation 0, with the initial population of 20 individuals randomly spread over the landscape.

Figure 7: SGA - Fitness Landscape Generation 0

Figure 8, illustrates the fitness plateaus associated with each individual in the population, with each coloured shape representing an individual in the population. Identical genotypes are represented by the same colour and shape. Figure 9, shows the MGA
population distribution over the problem landscape and the associated fitness plateaus are depicted in Figure 10. The initial randomly generated population distribution for the MGA is quite similar to that of the SGA.

Examining the populations for both GAs at generation 50, which is the last generation before the landscape changes, we see that the SGA’s population has converged (apart from the impact of mutation) (Figure 11 & Figure 12). The MGA’s population has also located the global optimum, however, due to the genotype-phenotype mapping, the population doesn’t converge as much as the SGA. Figure 13, shows that many-to-one representation of the population, with different genotypes having the same fitness represented by different colours and shapes. An alternative view is shown in Figure 14, which illustrates the population distribution on each of the fitness plateaus.

Figure 11: SGA - Fitness Landscape Generation 50

Figure 12: SGA - Fitness Plateaus Generation 50

Figure 13: MGA - Fitness Landscape Generation 50

Figure 14: MGA - Fitness Plateaus Generation 50

Figure 15: SGA - Fitness Landscape Generation 51
The MGA population’s evolutionary trajectory differs considerably by not converging on the global optimum. Because of this the population consists of a number of neutral networks, which is a result of gene flow due to the presence of neutrality.

5.1.2 Analysis After the Landscape Change

Examining the population evolution through to generation 200 for both GAs. The SGA population remains trapped on the local optimum (see Figure 19 and Figure 20), while the MGA population remains clustered representation, the population is dispersed over a wider number of fitness plateaus.
around the global optimum, see Figure 21 and on relatively adjacent fitness plateaus see Figure 22. This appears to indicate that the MGA, through the genotype-phenotype mapping, implicitly maintains a level of genetic diversity within the population and is resistant to convergence, thereby offering the ability to adapt in a dynamic environment.

5.2 Examining the impact of Neutrality on Variation

As a traditional SGA maps directly from the genotype to the phenotype, the search space is identical to the solution space. Using normalised Hamming distances between individuals, Figure 23 and Figure 24, illustrates the search space for both the SGA and MGA, respectively, at generation 0. From the figures, it appears that at generation 0, there is quite a large level of diversity present in both populations. This can be explained as evolution has not yet begun and the populations have been randomly generated. Also, as neutrality is introduced the search space increases, while the MGA’s solution space (Figure 25) is quite similar to the SGA’s.

At generation 200, the SGA population remains on the local optimum (Figure 20) and contains little
diversity to allow it to escape. This situation is shown in Figure 29, reflecting the lack of diversity within the population when viewed as a function of Hamming distance.

The MGA genotypic population diversity at generation 200, is shown in Figure 30 and the phenotypic diversity is illustrated in Figure 31. The many-to-one genotype-phenotype mapping can be used to explain the level of diversity present in the population. Looking back at Figure 22, although a significant portion of the population lie on the fitness plateau associated with the global optimum, they form a neutral network, where a number of different genotypes representing the same phenotype. The level of phenotypic diversity present, relates to the individuals of the population being part of other neutral networks.

### 5.3 Examining the impact of Neutrality on Phenotypic Variability

The final examination of the effects of neutrality on the evolutionary trajectory of the MGA population relates to the impact of mutation. In order to examine the effect of single bit mutation and missense mutation, we randomly selected an individual from both the SGA and MGA populations from various generations. We then flipped all of the bits, one at a time, in sequence and measured the impact on fitness. The aim of this approach is to examine how the presence of neutrality impacts on mutation GAs which have synonymous representations. A representation is viewed as synonymously redundant if the genotypes representing the same phenotype have the same properties and are next to one another in the mutation space. With the MGA population, a genotype, when mutated, can produce either a silent or an adaptive single-bit mutation and a silent or adaptive missense mutation.

Figure 33 and Figure 32 show the impact of mutation in an SGA population. The original individual is shown in the top right fitness plateau of Figure 33 and the mutated individuals are located in neighbouring plateaus. Figure 34 and Figure 35 visualise the impact of single-bit mutation on an individual from the MGA population, the phenotypic distribution for both GAs is similar as the neutrality associated with
The MGA representation includes missense mutation which operates within the layers of the genotype-phenotype mapping and allows the phenotypic variability to differ for individuals having the same level of fitness. Figure 36 and Figure 37 illustrate the phenotypic distribution for 010000011000010000111000, showing that the phenotypic distribution for the MGA is greater than that of the SGA, as individuals are occupying more fitness plateaus.

Figures 38 and 39, illustrate the phenotypic distribution for individual 100010011110010000001010, which resides on the same neutral network with a fitness level of 30. They show that the MGA population allows individuals residing on the same fitness plateau, have different phenotypic distributions.

6 CONCLUSION

A population's ability to survive in dynamic environments often requires a level of diversity to be maintained within the population. As a GA search involves a mapping between the genotype and the phenotype, a SGA, through its one-to-one genotype-phenotype mapping, quickly eliminates diversity from the population through its selection policy and low mutation rates. The results presented, illustrate that through the implementation of Neutral theory, as proposed by Kimura (Kimura, 1968), the genotype-phenotype mapping of the MGA allows for a tunable, synonymous and non-trivial, many-to-one relationship. The contribution of this form of mapping is the implicit maintenance of related genetic diversity within the population, which allows the occupation, by the population, of a greater number of fitness plateaus. By adopting this approach convergence at a phenotypic level can be achieved, but genetic diversity is maintained at
a genotypic level. Neutral theory (Kimura, 1968), would suggest that where genetic changes spread across a population, changes may or may not change the phenotype, are a result of genetic drift. During the experiments, periods of stasis were observed within the MGA population and these were followed by rapid adaption through genetic drift. The results also indicated that neutrality, as introduced by the MGA mapping, impacts on the search space by increasing it’s size and diversity. Finally, the results showed that adaptive mutations operate in a similar manner to SGA mutations, but allow a greater number of fitness plateaus to be reached. Silent mutations, on the other hand, offer the ability to search the landscape without impacting on fitness. This increases the size of the genetic pool and impacts on gene flow.

Overall, from the results presented, neutrality, while increasing the genotype space without increasing the phenotype space, offers a number of potentially useful characteristics for GAs. Through the presence of neutrality in the genotype-phenotype mapping, a GA can maintain and alter through genetic drift, additional, highly fit, genetic material within the population. This is achievable as neutrality allows the mutation operator to produce both silent and adaptive mutations, thereby creating neutral networks of different, but related genotypes. Through an implementation of Neutral Theory the MGA implicitly maintains genetic variation within the population through selection and genetic drift, which impacts on gene flow and provides additional searching capacity.

REFERENCES


