Simulation of Baldwin effect and Dawkins memes by genetic algorithm

Vladimír Kvasnička[1] and Jiří Pospíchal[2]

Department of Mathematics, Slovak Technical University, 812 37 Bratislava, Slovakia
email: [1] kvasnic@cvt.stuba.sk, [2] pospich@cvt.stuba.sk

Key words: Genetic algorithm, learning, fitness landscape, Baldwin effect, Dawkins memes

Abstract

Three different levels of sophistication of genetic algorithms (GAs) are described. (1) The *first* level are GAs, where the chromosome fitness is determined exactly by the chromosome position on the fitness landscape. This first level of sophistication corresponds to standard GAs, where chromosomes are directly mapped on positive real numbers. (2) In the *second* level the chromosome fitness is determined by the nearest neighborhood of the chromosome position on the fitness landscape, i.e. chromosomes are capable of learning. The role of learning in evolution theory called the Baldwin effect was first studied by GAs in 1987 by Hinton and Nowlan. (3) In the *third* level the chromosome fitness is determined not only by the nearest neighborhood of the chromosome position on the fitness landscape but also by the so-called meme that determines an information that is able to increase the fitness of chromosomes. The idea of memes was introduced to evolutionary biology by Dawkins. The concept of chromosome is enlarged to a complex of chromosome and meme, both of them determine a fitness of chromosomes itself. All the above three different levels of GAs may be formally considered as evolutionary steps of Darwin's evolution. The purpose of this communication is to present a GA simulation of learning and Dawkins' memes.

1. Introduction

Genetic algorithms (GAs) [1,2] are computational models of biological evolution. GAs are useful both for solving optimization problems and for modeling evolutionary systems. The purpose of this communication is to present a GA simulation of learning [3-5] and Dawkins memes [6] in a population composed of chromosomes that are represented by linear strings of symbols.

In 1896 J.M. Baldwin [7,8,4] proposed a hypothesis, which pointed out that if learning helps survival, then the organisms best able to learn will have the most offspring, thus increasing the frequency of genes responsible for learning. If the environment remains relatively fixed, so that the most profitable information to learn remains constant, selection can lead to a genetic encoding of a trait that originally had to be learned. Baldwin called this mechanism "organic selection", but it was latter dubbed the "Baldwin effect". We will consider Dawkins theory of memes [6] as an extension of Baldwin effect. The memes are ideas - messages broadcasted throughout the whole population and are composed of useful information how to increase fitness of chromosomes. Dawkins assumed that complexes of chromosome with meme might be important for an acceleratation of evolution. Many social traits (e.g. altruism) are discussed in the framework of population genetic [9] only with some strong restrictions and assumptions that are not acceptable by all evolutionary biologists. An application of Dawkins' memes represents a plausible possibility how to explain and interpret many features of social behaviour of animals [10].

Darwin's theory of evolution [11,12] is based on a principle of natural selection, according to which only the best adapted individuals from a population survive. By reproduction of two strong individuals such offspring are produced, which will be with high probability well adjusted for a successful survival. A detailed analysis (mainly mathematical) shows, that sufficiently effective creation of well adjusted individuals with such new features, which would substantially increase a chance for survival, requires reproduction including both genetic information exchange as well as so called mutations. The mutations influence (positively or negatively) by a random manner genetic material of individuals, influencing in long term the whole population.

In the higher degree of abstraction [1] a notion of biological individual can be replaced by a notion of chromosome, which represents a linearly ordered information content of an individual (genotype). Then we can

talk about a population of chromosomes, which reproduce with a probability directly proportional to their strength, while integral part of that reproduction are mutations. Mutations introduce into chromosomes a new information, which can increase fitness of chromosomes created by a crossover from parental chromosomes. New chromosomes push out from population chromosomes with small fitness. This basic reproduction cycle is continually repeated. After some time there is a high probability, that chromosomes with new properties substantially increasing fitness emerge in population, and push out older chromosomes without those properties.
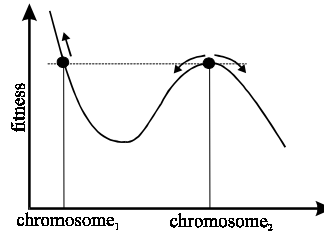
Figure 1. Diagrammatic illustration of a situation when two chromosomes are placed on a fitness landscape so that they have the same fitness. The second chromosome is situated at a local maximum and therefore it does not have a chance to increase its fitness in a nearest neighborhood. Though the first chromosome has the same fitness as the second one, it is situated on the fitness landscape in such a position that in its nearest neighborhood there exist positions with greater fitness. From the standpoint of classical Darwin's evolution theory, when learning processes are fully ignored, these chromosomes could not be distinguished and therefore would have the same chance to participate in a reproduction process. On the other hand, when the learning process is included into the evolution process, these chromosomes are distinguished, the first chromosome is evaluated by a greater fitness and natural selection prefers this chromosome.

Genetic algorithms (GAs) [1,2] are an abstraction of the above presented basic principles of Darwin's evolution theory [11,12]. According to the way of fitness calculation three different levels of GAs may be considered:

(1)   Fitness of chromosomes is determined exactly by their positions on the fitness landscape, i.e. it is determined entirely by chromosome compositions, no other effects are included in the fitness calculation. This way of fitness calculation corresponds to standard GAs, where chromosome are directly mapped on positive real numbers - fitness, without any intermediate or hidden considerations.

(2)   Fitness of chromosomes is influenced by the nearest neighborhood of their positions on the fitness landscape. It means that chromosomes are capable of learning in which they search through nearest neighborhood on fitness landscape. As a result of this process chromosomes with the same fitness can be distinguished by a presence of higher fitness positions in their nearest neighborhood, see Fig. 1. These chromosomes are more effective for further evolution process than those ones that are not able to look for higher fitness chromosomes in the nearest neighborhood. The role of learning in evolution theory is called the Baldwin effect [7,8], its first study by GA was done by Hinton and Nowlan [3] (cf. also refs. [4,5]). However, better modeling of a learning process should also represent a quality of a learning ability by a size of a neighborhood searched around a chromosome. Really complex chromosomes then could theoretically code not only their position on a landscape, but to code their own learning algorithm as well.

(3)   Fitness of chromosomes is determined not only by the nearest neighborhood of their positions on the fitness landscape but also by the so-called memes that determine an information that is able to increase the fitness of chromosomes. The idea of memes was introduced to evolutionary biology by Dawkins in his famous book Selfish Gene [6]. The concept of chromosome is enlarged to a complex of chromosome and meme, both of them determine a fitness of chromosomes itself. Moreover, there should be introduced a notion of the so-called memetic interaction, where a chromosome-meme complex called the donor creates a message (idea). Message is broadcasted throughout a subpopulation of the so-called recipients. These recipients incorporate the broadcasted message into their memes. Incorporation of memes into GAs was already studied by [12,13].

All these three different levels of sophistication of GAs may be formally considered as evolutionary steps of Darwin's evolution itself. This view of the evolution distinguishes a complexity of fitness evaluation process. On the first lowest level the fitness is fully determined by chromosome composition only. The second level corresponds to the previous first level that is enriched by a possibility of learning, i.e. chromosomes are able of local search in their nearest neighborhoods. Finally, the third level is nothing but the second level enlarged by

memes, i.e. chromosomes are not only capable of learning but also capable of employing memes in the fitness evaluation process. We have to emphasize that in all these three levels the evolution is strictly Darwinian, there is not postulated that memes or traits acquired by learning are inherited by offspring. Offspring chromosomes are fully determined by "crossover" combination and random mutation of their parent chromosomes, no other information acquired by learning is included. The learning process is used only for evaluation of chromosomes by fitness but not for their modification. For memes is situation slightly different. In the framework of memetic interaction a special type of learning process is used for the creation of a donor meme broadcasted throughout a subpopulation of recipients, where the donor meme transforms in a random way the recipient memes. This fact is in accordance with Dawkins' concept, memes may be transformed not only in the course of a reproduction process but also in the course of their lifetime.

2. Basic concepts of GAs

One of basic concepts of GAs is a notion of *chromosome*. A chromosome x is an abstract mathematical object, which in actual applications may be identified to a string of symbols, or to a matrix composed of real or integer entries, or to a graph with evaluated vertices and edges, etc. Let X be a finite set composed of all possible chromosomes that may be constructed so that chromosome specification rules are satisfied

$$X = \{x, x', x''...\} \tag{1}$$

An *objective function* $f: X \to R$ assigns to each chromosome $x \in X$ a real number $f(x)$. We look for its global minimum over the set X

$$x_{opt} = \arg \min_{x \in X} F(x) \tag{2}$$

In GAs the notion of objective function will serve as a representation of *environment* in which chromosomes live and are evolved by Darwinian evolution. The goal of the evolution is to "construct" such chromosomes that are closely related (or even equal) to an optimal chromosome x_{opt} specified by (2).

A *mutation* is an operator $O_{mut}: X \to X$ that assigns stochastically to each chromosome $x \in X$ a new chromosome $x' = O_{mut}(x) \in X$. In a similar way, a *crossover* is an operator $O_{cross}: X^2 \to X^2$. This operator assigns stochastically to a pair $x_1, x_2 \in X$ a new pair $(x_1', x_2') = O_{cross}(x_1, x_2)$. An actual specification of both these stochastic operators depends on the form of chromosomes and on a problem to be solved by GA. Mutation and crossover operators are unified in an operator called the *reproduction* $O_{repro}: X^2 \to X^2$ determined as follows:

$$(x_1', x_2') = O_{repro}(x_1, x_2) \ , \ (\tilde{x}_1, \tilde{x}_2) = O_{cross}(x_1, x_2) \ , \ x_1' = O_{mut}(\tilde{x}_1), \quad x_2' = O_{mut}(\tilde{x}_2) \tag{3}$$

This means that in the first stage of reproduction the crossover is performed and then the resulting chromosomes are mutated.

A stochastic decision, whether the operators O_{mut} and O_{cross} will be applied or not, is based on two probabilities P_{mut} and P_{cross}, respectively. An application of the operator O_{mut} onto a chromosome x is realized so that

$$\lim_{P_{mut} \to 0} O_{mut}(x) = x \tag{4}$$

Each symbol in the chromosome string will be changed with a probability P_{mut}. The requirement (4) means that after turning off the probability P_{mut}, the result of mutation is equal to the original (nonmutated) chromosome x. In the opposite case, if the probability P_{mut} is greater and greater, then the chromosome $O_{mut}(x)$ is more and more dissimilar from the original chromosome x. The crossover operator O_{cross} is stochastically applied in the framework of the reproduction operator O_{repro}. A stochasticity of this operator is determined by a probability P_{cross}, it determines whether in the framework of the above reproduction operator the crossover process will be used or not. If no crossover process is used, then in the second relation of (3) chromosomes are simply copied.

A *population* P is a multiset with fixed finite cardinality of chromosomes from the set X, $P = \{x_1, x_2, ..., x_p\} \subseteq X$. A *fitness* of chromosomes from the population P is a positive real number determined by a mapping $f: P \to R_+$ restricted by the following condition

$$F(x_1) \leq F(x_2) \Rightarrow f(x_1) \geq f(x_2) \tag{5}$$

Each chromosome from the population P is evaluated by a positive number called the *fitness*. It reflects a smallness of objective function values $F(x)$, a chromosome evaluated by a smaller function value $F(x)$ is evaluated by a greater fitness $f(x)$ so that the condition (5) is satisfied.

The fitness values are used for the calculation of probabilities of selection of chromosomes from the population P. Let $O_{select}(P)$ be a stochastic operator - *proportional selector* - which realizes a selection of chromosomes from the population P so that the probability of this selection is equal to a normalized fitness. A chromosome with a greater normalized fitness has a greater probability to be selected by the selector $O_{select}(P)$. Formally, the term $O_{select}(P)$ is understood as a chromosome $x \in P$ selected from the population P with the probability determined as follows

$$prob(x) = \frac{f(x)}{\sum_{y \in P} f(y)} \qquad (6)$$

GA imitates the Darwinian evolution over a population of abstract chromosomes. A simple realization of GA is outlined in Algorithm 1.

```
procedure GA;
begin P:=a set composed of p randomly
         generated chromosomes;
      t:=0;
      while t<t    do
               max
      begin t:=t+1;
            evaluate all chromosomes by fitness;
            Q:=∅;
            while |Q|<p do
            begin x :=O      (P); x :=O      (P);
                   1   select        2   select
                   (x' ,x' ):=O     (x ,x );
                     1   2     repro  1  2
                   Q:=Q∪{x' ,x' };
                            1   2
            end;
            P:=Q;
      end;
      x   :=best chromosome of P;
       opt
end;
```

Algorithm 1. An implementation of simple GA. Algorithm is started by a random generation of population P. Integer variable t is a counter of epochs in GA, outer while-do cycle is repeated a prescribed number of times determined by t_{max}. Inner while-do cycle forms a new population Q, which is initialized as an empty set, it is sequentially occupied by offspring created by the reproduction operator applied on a pair of quasirandomly selected chromosomes x_1 and x_2 from the population P. Input parameters are p (number of chromosomes in population), t_{max} (maximal number of generations), and probabilities P_{mut} and P_{cross}. As an output of GA is a chromosome x_{opt}, determined as that one with smallest function value $F(x)$ over the population P.

3. GA with learning

Let a *chromosome* $x = (x_1 x_2 ... x_n)$ be a linear string of the length n composed of symbols 0, 1, and #

$$x \in \{0,1,\#\}^n \qquad (7)$$

An *optimal chromosome* x_{opt} is randomly generated bit string of the length n

$$x_{opt} = \left(x_1^{opt} x_2^{opt} ... x_n^{opt}\right) \in \{0,1\}^n \qquad (8)$$

A *population P* is a multiset composed of p chromosomes

$$P = \{x_1, x_2, ..., x_p\} \subseteq \{0,1,\#\}^n \qquad (9)$$

Its *cardinality p=|P|* is kept fixed throughout all generations of genetic algorithm.

Each chromosome $x \in P$ is evaluated by a *fitness*. This evaluation process is connected with the *learning process* and it is determined as follows: Let us define a neighborhood of a given chromosome $x \in P$

$$U(x) = \{y_1, y_2, ..., y_q\} \subseteq \{0,1\}^n \qquad (10)$$

This neighborhood is composed of q binary vectors that are randomly created from the reference chromosome $\boldsymbol{x}$ so that its hash symbols are randomly substituted either by 0 or 1

$$\boldsymbol{y} = (y_1 y_2 \dots y_n) \in \{0,1\}^n \tag{11a}$$

$$y_i = \begin{cases} x_i & (\text{if } x_i = 0 \text{ or } x_i = 1) \\ \text{random}(2) & (\text{if } x_i = \#) \end{cases} \tag{11b}$$

The cardinality $q = |U(\boldsymbol{x})|$ is called the *learning size*. For the studied case there was only one optimal chromosome $\boldsymbol{x}_{\text{opt}}$. A function $d(\boldsymbol{x})$, for $\boldsymbol{x} \in P$, corresponds to a minimal distance between the optimal chromosome $\boldsymbol{x}_{\text{opt}}$ and binary vectors from the neighborhood $U(\boldsymbol{x})$

$$d(\boldsymbol{x}) = \min_{\boldsymbol{y} \in U(\boldsymbol{x})} |\boldsymbol{y} - \boldsymbol{x}_{opt}| \tag{12}$$

A fitness of the chromosome $\boldsymbol{x}$ is determined as follows

$$f(\boldsymbol{x}) = f_{min} + (f_{max} - f_{min}) e^{-\xi d(\boldsymbol{x})} \tag{13}$$

where ξ is the so-called *slope coefficient* that determines a sharpness of the fitness function. The above specified evaluation of population chromosomes by fitness has a strong stochastic character. In particular, the neighborhood $U(\boldsymbol{x})$ is randomly constructed from the reference chromosome $\boldsymbol{x}$. This stochastic character of the fitness evaluation is suppressed if the learning size $q = |U(\boldsymbol{x})|$ is increased so that there exists a real chance to find the global minimum of (12).

The learning process (i.e. evaluation of chromosomes by fitness) is mainly determined by the cardinality q of the neighborhood $U(\boldsymbol{x})$ and the slope coefficient ξ. The cardinality q corresponds to a number of trial events in the learning process when we construct a binary vector y from a given chromosome $\boldsymbol{x}$ so that the binary vector y has a smallest Hamming distance from the optimal binary vector $\boldsymbol{x}_{opt}$. Increase of the cardinality q offers to the learning process a bigger chance to find the best binary vector with the smallest possible Hamming distance from $\boldsymbol{x}_{opt}$. The slope coefficient ξ is a positive real number that may be interpreted as a selector of acceptance of binary vectors that are created from the given chromosome $\boldsymbol{x}$ in the framework of the learning process. If this parameter tends to infinity, then the learning process evaluates by the maximal fitness the binary vectors identical to the optimal binary vector $\boldsymbol{x}_{opt}$, all other binary vectors are evaluated by the mimimal fitness. On the other hand, if the slope coefficients ξ approaches zero, the learning process evaluates all binary vectors by values of fitness close to the maximal value f_{max}. The above discussed properties of the fitness are formally expressed by the following limiting relations

$$\lim_{\xi \to \infty} f(\boldsymbol{x}) = \begin{cases} f_{max} & (\text{if } \boldsymbol{x} = \boldsymbol{x}_{opt}) \\ f_{min} & (\text{if } \boldsymbol{x} \neq \boldsymbol{x}_{opt}) \end{cases} \tag{14a}$$

$$\lim_{\xi \to 0} f(\boldsymbol{x}) = f_{max} \tag{14b}$$

The learning process can be interpreted as a searching in a nearest neighborhood of the given chromosome (cf. eqs. (10) to (12)), where the chromosome hash symbols are randomly substituted by 0s or 1s symbols. After Hinton and Nowlan [3] this is a "needle in a haystack" search problem, since there is only one correct solution in a space of 2^n solutions. If the learning is included, then GA can be loosely interpreted as a search method, where some one advises to the "chromosome" in which direction the needle (correct solution) is hidden. More realistic interpretation of GAs with learning is presented in Fig. 1, where the learning process is interpreted as a method that is able to give an information about the nearest neighborhood on fitness landscape, i.e. whether chromosome fitness can be potentially increased.

Finally, we have to specify GAs (outlined in the previous chapter, see Algorithm 1) applied to populations of chromosomes described in the first part of this chapter. An operation of *crossover* for two chromosomes $\boldsymbol{x}$ and $\boldsymbol{x}'$ is simply determined as a 1-point crossover. A crosspoint is randomly generated and then chromosome parts behind this point are mutually exchanged. An operation of *mutation* applied to a chromosome $\boldsymbol{x}$ is performed in such a way that going successively throughout all parts of chromosome we change with small probability P_{mut} chromosome entries to one of their complements, e.g. symbol 1 is changed randomly with the same probability either to 0 or to #.

Genetic algorithm with learning features specified in this chapter has the following set of parameters:

    (1)   *n*, chromosome length ($n=20$)
    (2)   *p*, population size ($p=500$)
    (3) P_{cross}, crossover probability ($P_{cross}=0.1$)

(4) P_{mut}, mutation probability (P_{mut}=0.01)
(5) t_{max}, number of epochs (t_{max}=200)
(6) q, learning size (q=10)
(7) ξ, slope parameter (ξ=0.1 and 1.0)
(8) f_{max}, maximal value of fitness (f_{max}=2.0)
(9) f_{min}, minimal value of fitness (f_{min}=1.0)
(10)   initial population P is randomly generated so that its chromosomes
        contains 50% symbols #, 25% symbols 0, and 25% symbols 1

where in parentheses are given numerical values of parameters used in our illustrative calculations.
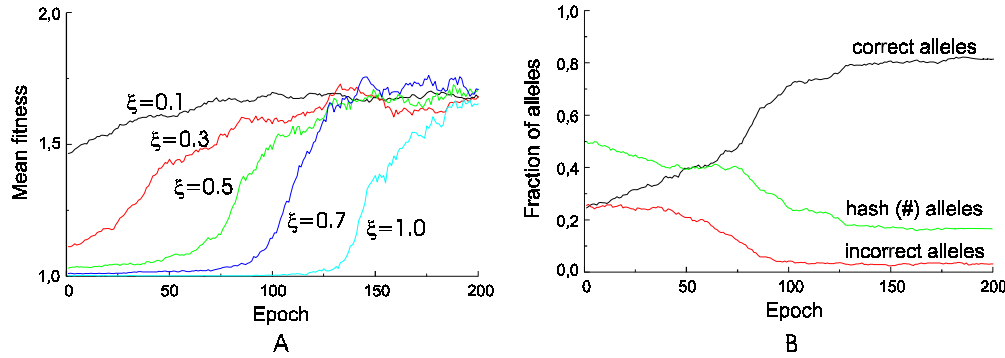
Figure 2. (A) Five plots of mean fitness vs. epochs for different values of slope parameters, from promoted learning (ξ=0.1) to suppressed learning (ξ=1.0). Other parameters of the GA are specified at the end of Chapter 3. We see that if the slope parameter ξ tends to greater values, then the learning process is more and more suppressed, the period when mean fitness is equal to the minimal value of fitness (f_{min}=1.0) is longer. If the slope parameter ξ is greater than a threshold value (say ξ=1.0), then the learning is entirely suppressed, the mean fitness in the course of whole evolution is equal to f_{min} (the optimization strength of a canonical GA is too small to show here). (B) Plot of fraction of correct alleles, incorrect alleles, and hash symbol (#) alleles for GA with learning (slope parameter is ξ=0.5). The fraction (frequency) of fixed correct alleles increased over epochs and the fraction of fixed incorrect alleles decreased almost to zero. On the other hand, the frequency of hash alleles (that can be substituted by 0s or 1s symbols in the learning process) stayed relatively very high. After Hinton and Nowlan [3] this may be caused by too insignificant selective pressure to fix all the hash alleles. This selective pressure may be substantially increased, e.g. by lowering the learning size q (it determines the cardinality of neighborhood $U(\boldsymbol{x})$, where the learning process is performed). According to our numerical experiences, when the learning size is set q=1, then fraction of hash alleles is monotonously decreasing in the course of evolution almost to zero value.

Different numerical results of our model calculations are outlined in Fig. 2. This figure is very instructive for interpretation of Baldwin effect in the framework of our GA simulation calculations. Five plots of mean fitness for different slope parameters ξ are presented in Fig. 2, diagram A . We see that the learning may be substantially suppressed by increasing the slope parameter ξ. The same effect may be also achieved by descreasing the learning size q, the learning is suppressed when the cardinality of the neighborhood is very small (i.e. $q \rightarrow 1$), i.e. the minimization problem (12) is restricted only to a few binary vectors. An evolution of fractions of alleles (assigned to 0, 1, and # symbols) is displayed in Fig. 2, diagram B. Very interesting observation is a decrease of the fraction of hash symbols in the course of evolution from the inital 50% value to the final 15% value. This means that the used GA with learning simulates properly classical Baldwin effect, it is able of a genetic encoding of a trait that originally had to be learned. This fact is manifested by monotonously decreasing value of fraction of hash symbols in the population in the course of its evolution. Hash symbols in chromosomes are introduced as learnable chromosome alleles, their descreasing fraction means that some of them were fixed by 0s or 1s alleles, i.e. learnable traits are fixed genetically.

4. GA with meme chromosomes (m-chromosomes)

In the previous chapter we have formulated simple GA with a possibility to simulate learning process in the course of the evolution process (Baldwin effect). The learning process consists in a replacing hash # symbols in chromosomes by 0s or 1s. These values are substituted randomly and in a subset of these binary vectors we look for a binary vector with minimal Hamming distance from the given optimal chromosome. The learning process assigns to each chromosome a binary vector that is the best one of all possible vectors that can be randomly constructed from the chromosome. If we compare this version of the genetic algorithm with a genetic algorithm where a learning process is considerably suppressed (e.g. slope parameter ξ is used with high numerical value or

the learning size q is of very small value), then we see that an effectiveness of the genetic algorithm with learning is greater than those one with suppressed learning, see Fig. 2.

How to intensify the genetic algorithm with already included learning process? Further possibility is to include the so-called memetic approach initially introduced into evolutionary biology by Dawkins [6]. We use very simplified model of memes so that it can be easily introduced to genetic algorithms with learning. In our approach memes are linear strings (of the same length as chromosomes) and each population chromosome is automatically accompanied by a meme. The memes are carriers of information about best local solutions that have been already achieved (see Chapter 1). Employing of memes is very important in evaluation process as a further source of information about a best solution that should be found. At the beginning of evolution the memes are empty (i.e. they have zero information content), in the course of evolution process they may receive an information in the framework of the so-called memetic interaction. Here, a quasirandomly (with high fitness) selected chromosome–donor and its meme creates a message - meme broadcasted through whole population, and some chromosomes–recipients accept this meme and incorporate it in their memes. In the reproduction process, created offspring obtain memes from one of two parental chromosomes declared as a mother (we say that offspring are educated by mothers).

A *meme* is a string composed of n symbols 0, 1, and $*$

$$\boldsymbol{m} = (m_1 m_2 \ldots m_n) \in \{0,1,*\}^n \tag{15}$$

A *width* of the meme $\boldsymbol{m}$, denoted by $|\boldsymbol{m}|$, is determined as a number of nonstar symbols, e.g. for a meme $\boldsymbol{m}=(01**01***)$ its width is $|\boldsymbol{m}|=4$. In our forthcoming considerations memes are used as an auxiliary vehicle of information used in the learning process of chromosomes. If the width of a meme is zero, then we say that this meme does not contain any useful information and its application in the framework of a learning process is irrelevant.

A *meme chromosome* (*m-chromosome*) is determined as an ordered couple of a chromosome and a meme, both of the same length

$$\boldsymbol{\alpha} = (\boldsymbol{x}, \boldsymbol{m}) = \left((x_1 x_2 \ldots x_n), (m_1 m_2 \ldots m_n) \right) \in \{0,1,\#\}^n \times \{0,1,*\}^n \tag{16}$$

where $\boldsymbol{x}$ is a *chromosome part* of $\boldsymbol{\alpha}$ and $\boldsymbol{m}$ is a *meme part* of $\boldsymbol{\alpha}$. A population P is a multiset composed of p m-chromosomes

$$P = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots \boldsymbol{\alpha}_p\} \subseteq \{0,1,\#\}^n \times \{0,1,*\}^n \tag{17}$$

The learning process of fitness evaluation of simple chromosomes, specified in the previous chapter, will be now generalized for m-chromosomes. This learning process consists of two stages.

In the *first stage*, the meme part $\boldsymbol{m}$ of $\boldsymbol{\alpha}$ is used for a repair of the chromosome part $\boldsymbol{x}$ of $\boldsymbol{\alpha}$, chromosome hash symbols are substituted by meme nonstar symbols. In particular, let $\hat{\boldsymbol{x}}$ be a chromosome created from the chromosome part $\boldsymbol{x}$ by a repair process performed with respect to the meme part $\boldsymbol{m}$

$$\hat{\boldsymbol{x}} = (\hat{x}_1 \hat{x}_2 \ldots \hat{x}_n) = O_{mod}(\boldsymbol{x}; \boldsymbol{m}) \tag{18a}$$

$$\hat{x}_i = \begin{cases} x_i & (\text{if } x_i \neq \# \quad \text{or} \quad m_i = *) \\ m_i & (\text{if } x_i = \# \text{ and } m_i \neq *) \end{cases} \tag{18b}$$

For better understanding of this repair process let us consider an m-chromosome

$$\boldsymbol{\alpha} = \left(\underbrace{(00\#\#1\#0\#)}_{\boldsymbol{x}}, \underbrace{(11***0**)}_{\boldsymbol{m}} \right) \tag{19}$$

The repaired chromosome $\hat{\boldsymbol{x}}$ is determined by (18a-b) as follows

$$\boldsymbol{x} = (0\,0\,\#\,\#\,1\,\#\,0\,\#)$$

$$\underline{\boldsymbol{m} = (1\,1\,*\,*\,*\,0\,*\,*)} \tag{20}$$

$$\hat{\boldsymbol{x}} = (0\,0\,\#\,\#\,1\,0\,0\,\#)$$

where in the sixth position the lower condition in (18b) is satisfied, i.e. hash symbol # is substituted by 0.

In the *second stage*, the repaired chromosome $\hat{\boldsymbol{x}}$ is used as a reference chromosome for construction of the neighborhood

$$U(\hat{x}) = \left\{ \hat{y}_1, \hat{y}_2, ..., \hat{y}_q \right\} \subseteq \{0,1\}^n \tag{21}$$

where binary vectors $\hat{y}_i$ are created from the repaired chromosome $\hat{x}$ so that its hash symbols are randomly substituted by 0 or 1 symbols, see eq. (11b). Applying the same procedure as in the standard GA with learning, we look for a minimal Hamming distance between binary vectors from the neighborhood and the optimal chromosome x_{opt} (cf. eq. (12))

$$\hat{d}_{min} = \min_{\hat{y} \in U(\hat{x})} \left| \hat{y} - x_{opt} \right| \tag{22}$$

Finally, a fitness assigned to an m-chromosome α is determined in the same way as in eq. (13)

$$f(\alpha) = f_{min} + (f_{max} - f_{min}) e^{-\xi \hat{d}_{min}} \tag{23}$$

Summarizing, an evaluation of m-chromosomes by fitness is performed formally by the same procedure as simple chromosomes (see previous chapter) except for the fact that meme part is used for a repair process of the chromosome part. Memes are carriers of additional information that is used for a repair process that precedes the learning process. For a usefull meme the resulting value d_{min} may be smaller than in the case when the repair process is not used (or it is used but a width of the meme part is zero).

A *memetic interaction* is a term that will be used to denote a process in which an idea - meme produced by a donor m-chromosome (quasirandomly selected) is broadcasted to a subpopulation of recipients (randomly selected). These recipients accept the broadcasted meme so that their own meme parts are modified. Let α_{donor} be m-chromosome with high fitness quasirandomly selected from the population P

$$\alpha_{donor} = O_{select}(P) \tag{24}$$

where the operator O_{select} is a proportional selector described at the end of Chapter 2. A meme from this donor m-chromosome should be extracted so that it may be used as a message composed of a useful information for other recipient m-chromosomes $P_{recip} \subset P$, where $\alpha_{donor} \notin P_{recip}$ and $r=|P_{recip}|<p=|P|$. This meme extraction is performed in a way very similar to the learning process within fitness evaluation of m-chromosomes (see initial part of this Chapter). Let $\hat{y}_{min}$ be a binary vector constructed by a minimization (22)

$$\hat{y}_{min} = \arg \min_{\hat{y} \in U(\hat{x})} \left| \hat{y} - x_{opt} \right| \tag{25}$$

where $\hat{x}$ is repaired chromosome part of the donor m-chromosome. From the binary vector $\hat{y}_{min}$ we construct a meme $\overline{m}_{donor}$ in such a way that only t randomly selected entries are taken into account, all other remaining meme entries are substituted by $*$ symbol. We say that a width of the meme $\overline{m}_{donor}$ is prescribed by an integer t ($0<t<n$) and it will be kept fixed in the course of all epochs in GA. For instance, let $\hat{y}_{min} = (00110101)$ be a binary vector determined by (25), we select randomly $t=3$ in the first, second, and fifth positions, then we get $\overline{m}_{donor} = (00**0***)$. The meme $\overline{m}_{donor}$ will be used for a modification of meme parts of all m-chromosomes from the subset P_{recip}, so that nonstar symbols in $\overline{m}_{donor}$ are introduced to their meme parts. For illustration, let $m = (10*10*0*)$ be a meme part, its modification by the message $\overline{m}_{donor} = (00**0***)$ is done by substituting the first, second, and fifth entry in m by the corresponding nonstar entries in $\overline{m}_{donor}$, we get

$$\begin{aligned} m &= (1\,0*1\,0*0*) \\ \overline{m}_{donor} &= (0\,0**0***) \\ \hline m_{mod} &= (0\,0*1\,0*0*) \end{aligned} \tag{26}$$

Formally, this modification may be expressed as a noncommutative sum of memes, $m_{mod} = m + \overline{m}_{donor}$, where single entries of m_{mod} are determined by

$$m_i^{mod} = \begin{cases} m_i & (\text{if } \overline{m}_i^{donor} = *) \\ \overline{m}_i^{donor} & (\text{if } \overline{m}_i^{donor} \neq *) \end{cases} \tag{27}$$

In order to simplify the concept of this modification process we introduce the so-called *modification operator* that modifies all m-chromosomes from recipient subset P_{recip} with respect to a meme generated from the donor chromosome
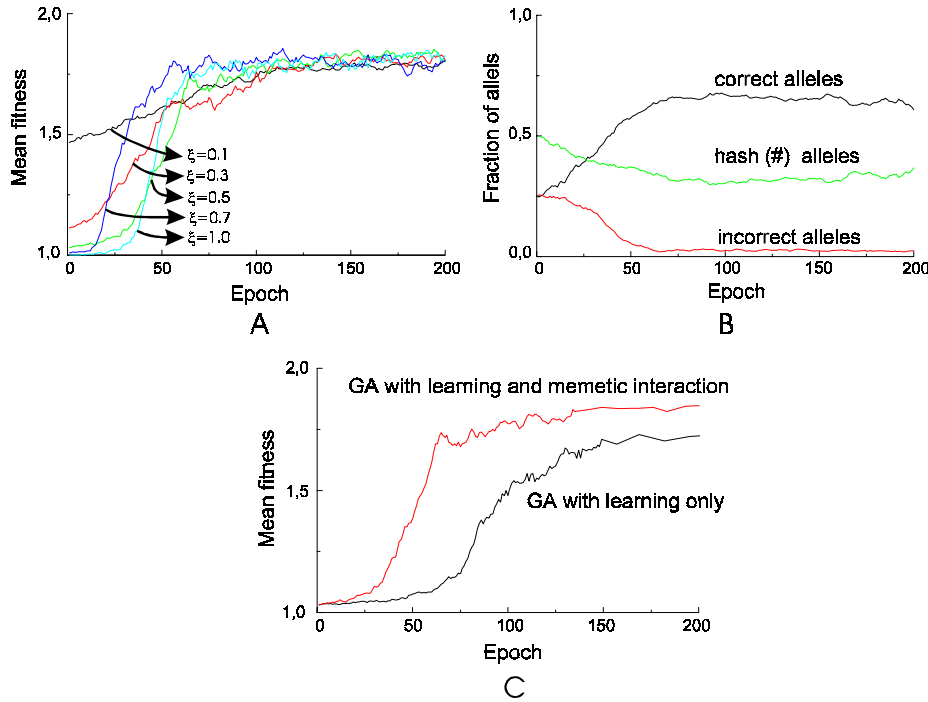
$$P'_{recip} = O_{mod}(P_{recip}; \alpha_{donor}) \tag{28}$$

where the modified recipient subset is determined as follows

$$P'_{recip} = \left\{ (x, m + \overline{m}_{donor}); (x, m) \in P_{recip} \right\} \tag{29}$$

i.e. all m-chromosomes from P_{recip} are modified in such a way that their meme parts are modified by a meme $\overline{m}_{donor}$, see example (26).

Figure 3. (A) Five plots of mean fitness vs. epochs obtained by GA with learning and memetic interaction, for different values of slope parameter, from promoted learning (ξ=0.1) to suppressed learning (ξ=1.0). Other parameters of the GA are specified at the end of Chapter 3. (B) Plot of fraction of correct alleles, incorrect alleles, and hash symbol (#) alleles for GA with learning and memetic interaction (slope parameter is ξ=0.5). (C) Plot of mean fitness for two different versions of GA with and without a memetic interaction when learning is included in both versions. Both calculations were done for slope parameter ξ=0.5. We see that GA with learning and memetic information outperforms GA with learning only.

Finally, we have to specify two GA operators crossover and mutation for m-chromosomes. The crossover operator O_{cross} applied to a couple α_1 and α_2 of m-chromosomes (that are randomly declared as mother and father) is performed so that their chromosome parts are modified by the standard 1-point crossover whereas their meme parts are determined by mother meme part. A mutation operator O_{mut} is determined for m-chromosomes such that only chromosome parts are mutated while meme part remain untouched. GA outlined by Algorithm 1 for m-chromosomes should be slightly modified, in particular a memetic interaction must be introduced at the beginning of each evolutionary epoch. Parameters specified at the end of Chapter 3 are still valid also for the present modification. Moreover, a random generation of initial population and additional two parameters used in this version of GA should be specified:

> (10') initial population P is composed of randomly generated m-chromosomes, where chromosome parts contains 50% symbols #, 25% symbols 0, and 25% symbols 1 and meme parts are of zero width (i.e. they contain only star symbols)
>
> (10) r, number of recipients in P_{recip} (r=20)
>
> (11) t, width of the donor meme $\overline{m}_{donor}$ (t=3)

Illustrative numerical results are presented in Fig. 3 . If we compare diagrams A and C in Fig. 3, then we see that an inclusion of memes into the GA with learning considerably increases its effectivity. In particular, transient phases where the mean fitness is kept to its minimal value f_{min}=1 is now substantially narrowed also for higher values of the slope parameter ξ. This fact is nicely illustrated by diagram C, where two plots of mean fitness are simultaneously displayed, one for GA with learning only and the other for GA with learning and memetic interaction. In general, we may say that an inclusion of memes to GAs means a substantial increase of the evolution effectivity, irrespective whether the simple learning (Baldwin effect) is costly or not. Plots of fractions of alleles for both versions of GA with learning only and GA with learning and memes are presented in Fig. 2, diagram B and in Fig. 3, diagram B, respectively. Comparing these plots we see that an introduction of memes decreases the selection pressure in GAs with learning and memes in comparison to GAs with learning only. In particular, a fraction of hash alleles (that are responsible for learning) for GAs with learning only asymptotically tends to 15% (see Fig. 2, diagram B) while this asymptotic value for GAs with learning and memes is substantially greater 35% (see Fig. 3, diagram B). This observation has the following simple interpretation: An

inclusion of memes decreases a selection pressure onto learnable hash symbols, in other words, memes may substitute in some extent the learning process. If the learning process is costly, the memes might be very important for acceleration of evolution, they are able to overcome of learning barriers.

5. Future plans

The population of m-chromosomes will be "geographically" structured so that it will be decomposed onto disjoint subpopulations and each subpopulation has some other subpopulations as neighborhods. We will examine the effect of erecting complete or semipermeable barriers between different subpopulations and the effect of migration. Until now, only one meme was attached to each chromosome, it will be interesting to study when this restriction is removed and a variable number of memes is attached to chromosomes. Many different strategies can be introduced which meme will be used in the process of evaluation of m-chromosomes by fitness, e.g. that one producing the best fitness or another one with greatest width, and so on. In our calculations we kept fixed the optimal chromosome x_{opt} , it means that the goal of evolution remained unchanged in the course of whole evolution. Interesting results might be obtained by slight variation of this optimal chromosome x_{opt}, in each generation it can be mutated with very low probability, e.g. $P_{mut}^{(opt)} = 0.001$. Then, we will compare an effectivity of GA with learning only and GA with learning and memes. According to our preliminary results it seems, that memes might be very important for permanent adapting of the best evolutionary solution when goal of evolution (or environment where populations exist) is slightly changing. In particular, if GA is performed only with learning, we have observed that the evolving system has a great inertia, it is unable to well adapt to a new evolution goal.

Acknowledgments

References

[1]   Holland, J.H., 1975, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, MA.
[2]   Goldberg, D.E., 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York.
[3]   Hinton, G. E., and Nowlan, S.J., 1987, How learning can guide evolution, *Complex System* **1**, 495-502.
[4]   Belew, R. K., and Mitchell, M., Eds., 1996, *Adaptive Individuals in Evolving Populations: Models and Algorithms*, Proceedings Volume XXVI, Addison-Wesley, Redwood City, CA.
[5]   *Evolution, Learning, and Instict: 100 Years of the Baldwin Effect,* 1996, special issue of journal *Evolutionary Computation* **4** (3).
[6]   Dawkins, R., 1976, *The Selfish Gene*, Oxford University Press, Oxford.
[7]   Baldwin, J. M., 1896, A New Factor in Evolution, *American Naturalist* **30**, 441-451.
[8]   Baldwin's article (Ref. 7) is available in electronic form on URL address
        http://paradigm.soci. brocku.ca/~lward/SUP/Baldwin/BALD_002.html.
[9]   Hamilton, W. D., 1964, The genetical Evolution of Social Behaviour I and II, *Journal of Theoretical Biology* **7**, 1-16 and 17-52.
[10]   Heylighen, F., Ed., 1998, *The Evolution of Complexity*, Kluwer Academic.
[11]   Darwin, C. , 1859, *The Origin of Species*, reprinted 1985, Penguin, London.
[12]   Darwin book (Ref. 11) is available in electronic form on URL address
        http://www.tbi. univie.ac.at/Origin/index.html.
[13]   Gabora, L. M., 1995, Memes and Variations: A Group Model of Cultural Evolution, in *Lectures in Complex Systems*, eds. L. Nadel, and D.L. Stein, Addison Wesley,
        available at http://www.lycaeum.org/~sputnik/Memetics/variations.txt
[14]   Hutchins, E., Hazelhurst, B., 1992, Learning in the Cultural Process, in *Artificial Life II*, eds. C. G. Langton et al. , Addison-Wesley, Redwood City, CA, pp. 689-706.