

# Optimization of autonomous agents by means of learning and evolution

Vladimir RED'KO<sup>a,1</sup>

<sup>a</sup>*Scientific Research Institute for System Analysis, Russian Academy of Sciences,  
Moscow, Russia*

**Abstract.** Interaction between learning and evolution in populations of autonomous agents is investigated. Any agent of the population has both the genotype (the genome) and the phenotype. The genotype and the phenotype are chains of binary symbols. The initial phenotype (at the moment of the agent birth) is equal to the agent genotype. There is a certain optimum; namely, there is the optimal chain that is searched for by means of learning and evolution. Genotypes are optimized by evolution; phenotypes are optimized by learning. The final phenotype (at the moment of the end of the agent life) determines the agent fitness. Three mechanisms of interaction between learning and evolution are investigated. 1) The mechanism of the genetic assimilation of the acquired features during a number of generations of Darwinian evolution is analyzed. It is shown that the genetic assimilation takes place as follows: the phenotype distribution moves towards the optimum at learning and further selection; subsequently the genomes of selected organisms also move towards the optimum. 2) The mechanism of the hidden effect is studied; this effect means that strong learning can ensure finding the optimal phenotype independently on the agent genotype in some situations; consequently, strong learning can inhibit the genotype optimization. 3) The mechanism of influence of the learning load on investigated processes is also analyzed. It is shown that the learning load leads to a significant acceleration of the genetic assimilation.

**Keywords:** speed and efficiency of evolutionary search, Baldwin effect, genetic assimilation, hiding effect, learning load

## Introduction

After the appearance of the Darwinian theory of evolution, many researchers asked the following question. The evolutionary process is a result of mutations and further selection. So, are random mutations able to ensure finding very non-trivial useful features of living organisms? In the XIX century, the concepts, suggesting that interaction between learning (or other processes of the acquisition of organism features during the life of the organism) and the evolutionary process is possible, appeared (Baldwin, 1896; Morgan, 1896; Osborn, 1896). Moreover, according to these concepts, learning can contribute significantly to the evolutionary process. This type of influence of learning on the evolutionary process is often called the Baldwin effect. According to this effect, initially acquired features can become inherited during a number of generations. The

---

<sup>1</sup> Vladimir Red'ko, E-mail: vgrekko@gmail.com.

evolutionary “re-invention” of useful features, initially obtained by means of learning, is often called the genetic assimilation (Waddington, 1942).

A number of works attempted to model and analyze interactions between learning and evolution by means of computer simulations (Belew & Mitchell, 1996; Turney, Whitley, & Anderson, 1996; Ackley & Littman, 1992; Hinton & Nowlan, 1987; Mayley, 1997; Red’ko, Mosalov, & Prokhorov, 2005). In particular, Hinton and Nowlan (1987) demonstrated that learning can guide the evolutionary process to find the optimum. Mayley (1997) investigated different aspects of interaction between learning and evolution and demonstrated that the hiding effect can take place if learning is sufficiently strong. The essence of the hiding effect is as follows: if learning is strong enough to optimize the phenotype of the organism, and the organism is selected at the evolution in accordance with the phenotype, then the selection can be independent on the genotype. The hiding effect significantly reduced the role of genotypes at the evolutionary selection, and the genetic assimilation becomes less pronounced. In addition, Mayley (1997) took the learning load into account. The learning load means that the process of learning has an additional load for the organism and the fitness of the organism is reduced under influence of the load.

Red’ko, Mosalov, and Prokhorov (2005) modeled interaction between learning and evolution for the case of neural network control systems of autonomous agents. The genetic assimilation of acquired features of agents was observed during several generations of evolution. In addition, it was demonstrated that learning could significantly accelerate the evolutionary optimization. However, it was difficult to analyze detailed mechanisms of interaction between learning and evolution in that models, as these mechanisms were «hidden» in the dynamics of numerous synapse weights of neural networks of agents.

The current work develops above-mentioned works. It uses works (Hinton and Nowlan, 1987; Mayley, 1997) as background. However, rather complex forms of the genetic algorithm (with crossovers) were used in that works, so it was difficult to analyze mechanisms of influence of learning on evolution. In contrast to works (Hinton & Nowlan, 1987; Mayley, 1997), the current work uses the clear evolution model, namely, the quasispecies model, proposed by Eigen (Eigen, 1971; Eigen & Schuster, 1979) and the quantitative estimation of the evolutionary rate and the effectiveness of evolutionary algorithms, obtained by Red’ko and Tsoy (2005). The quasispecies model considers the process of evolution that is based on the selection and mutations of genomes of organisms (without crossovers) and describes main properties of evolutionary processes. This model allows getting a better understanding of mechanisms of interaction between learning and evolution.

### **Description of the Model**

An evolving population of agents (or individuals) is considered. Similar to the work (Hinton & Nowlan, 1987) we assume that there is a strong correlation between the genotype and the phenotype of agents. We assume that the genotype (or the genome) and the phenotype of the agent have the same form, namely, they are chains (sequences of symbols); symbols of both chains are equal to 0 or 1. For example, we can assume that the genome is a modeled DNA chain, «letters» of which are equal to 0 or 1, and the phenotype is determined by the neural network of organisms, synaptic weights of the neural network are equal to 0 or 1 too. Initial synaptic weights, received at the agent

birth, are determined by the genome (more precisely, synaptic weights are equal to genome symbols). These weights are changed by means of learning during the agent life.

We assume that each agent has its own genome  $\mathbf{S}_0$ . A population consists of  $n$  agents, agent genomes are  $\mathbf{S}_{0k}$ ,  $k = 1, \dots, n$ . The agent genome  $\mathbf{S}_{0k}$  is the chain of symbols,  $S_{0ki}$ ,  $i = 1, \dots, N$ . Symbols  $S_{0ki}$  are equal to 0 or 1. We also assume that the length of chains  $N$  and the number of agents in the population  $n$  are large:  $N, n \gg 1$ . Values  $N$  and  $n$  do not change in the course of evolution. We assume that  $N$  is so large that only a small part of possible  $2^N$  genomes can be present in a particular population:  $2^N \gg n$ . Typical values of  $N$  and  $n$  in our computer simulations are as follows:  $N \sim n \sim 100$ .

The evolutionary process consists of a sequence of generations. The new generation is obtained from the old one by the selection and mutations of agents. Genomes of agents of the initial generation are random.

In order to consider learning processes, we introduce two types of sequences: 1) the genome or the initial sequence  $\mathbf{S}_{0k}$  that is received by the agent at its birth, and 2) the current sequence of the agent  $\mathbf{S}_{Tk}$ . The sequence  $\mathbf{S}_{Tk}$  is the phenotype of the agent.

Agents inherit genomes  $\mathbf{S}_{0k}$  from their parents, these genomes do not change during the agent life and are transmitted (with small mutations) to their descendants. Mutations are random changes of symbols  $S_{0ki}$ . The agent receives the genome at its birth, the current sequence  $\mathbf{S}_{Tk}$  at the birth of the agent is equal to the genome:  $\mathbf{S}_{Tk}(t=1) = \mathbf{S}_{0k}$ . The life time of any agent is equal to  $T$ . The time is discrete:  $t = 1, \dots, T$ . Duration of the generation is equal to  $T$ . The sequence of  $\mathbf{S}_{Tk}$  is modified during the agent life by means of learning.

As descendants of agents obtain the genomes  $\mathbf{S}_{0k}$  that are received by agents from their parents but not sequences  $\mathbf{S}_{Tk}$  that are optimized by learning, the evolutionary process has a Darwinian character.

It is assumed that there is an optimal sequence  $\mathbf{S}_m$  (components of which are also equal to 0 or 1), which is searched for in processes of evolution and learning. At computer simulation, the sequence  $\mathbf{S}_m$  is set to be random.

Learning is performed by the following method of trial and error. Every time moment  $t$  each symbol of the sequence  $\mathbf{S}_{Tk}$  is randomly changed to 0 or 1, and if this new symbol coincides with the corresponding symbol of the optimal sequence  $\mathbf{S}_m$ , then this symbol is fixed in the  $\mathbf{S}_{Tk}$ ; otherwise, the old symbol of the sequence  $\mathbf{S}_{Tk}$  is restored. So, during learning, the current sequence  $\mathbf{S}_{Tk}$  moves towards the optimal sequence  $\mathbf{S}_m$ .

At the end of the generation, the selection of individuals in accordance with their fitness takes place. The fitness is determined by the sequence  $\mathbf{S}_{Tk}$  at the time moment  $t = T$ . We denote this sequence  $\mathbf{S}_{Fk}$ , i.e. we set  $\mathbf{S}_{Fk} = \mathbf{S}_{Tk}(t = T)$ . The fitness of the agent  $\mathbf{S}_k$  is determined by the Hamming distance  $\rho = \rho(\mathbf{S}_{Fk}, \mathbf{S}_m)$  between the sequences  $\mathbf{S}_{Fk}$  and  $\mathbf{S}_m$ :

$$f(\mathbf{S}_k) = \exp[-\beta\rho(\mathbf{S}_{Fk}, \mathbf{S}_m)] + \varepsilon, \quad (1)$$

where  $\beta$  is the positive parameter, that characterizes the intensity of the selection,  $0 < \varepsilon \ll 1$ . The role of the value  $\varepsilon$  in (1) can be considered as influence of random factors of the environment on the fitness of individuals.

The selection of individuals into a new generation is made by the well-known method of the fitness proportionate selection (or the roulette wheel selection). In this method individuals are selected into a new generation probabilistically. The choice of an individual into the next generation takes place  $n$  times, so the number of individuals in the population in all generations is equal to  $n$ . At any choice, the probability of the selection of a particular individual is proportional to its fitness.

Thus, individuals are selected at the end of a generation in accordance with their final phenotype  $\mathbf{S}_{F_k} = \mathbf{S}_{T_k}(t = T)$ , i.e. in accordance with the final result of learning, whereas initial genomes  $\mathbf{S}_{0_k}$  (modified by small mutations) are transmitted from parents to descendants.

In addition, we take into account the learning load (the cost of learning), namely, we assume that the learning process has a certain burden on the individual and the fitness of the individual may be reduced under influence of the load. For this purpose, we consider the modified fitness of individuals:

$$f_m(\mathbf{S}_k) = \exp(-\alpha d) \{ \exp[-\beta \rho(\mathbf{S}_{F_k}, \mathbf{S}_m)] + \varepsilon \}, \quad (2)$$

where  $\alpha$  is the positive parameter, which takes into account the learning load,  $d = \rho(\mathbf{S}_{0_k}, \mathbf{S}_{F_k})$  is the Hamming distance between the initial sequence  $\mathbf{S}_{0_k}$  and the final sequence  $\mathbf{S}_{F_k}$  of the individual, i.e. the value  $d$  characterizes the intensity of the whole learning process of the individual during its life. The factor  $\exp(-\alpha d)$  decreases the fitness of the individual, this decrease is related to the change of the current sequence  $\mathbf{S}_{T_k}$  at the learning process.

It should be noted that since initial sequences  $\mathbf{S}_{0_k}$  of individuals in the initial population are random, the average Hamming distance between these sequences and the optimal one  $\mathbf{S}_m$  is approximately equal to  $N/2$ . Sequences  $\mathbf{S}_k$  should overcome this distance at learning and evolution in order to reach  $\mathbf{S}_m$ .

## Results of simulation

### *Scheme and parameters of simulation*

Two modes of operation of the model are considered: 1) evolution combined with learning, as described above, 2) “pure evolution”, that is evolution without learning, in this case, learning doesn't take place and  $\mathbf{S}_{T_k} = \mathbf{S}_{0_k}$ . In addition, the influence of the learning load is analyzed; in this case, the fitness of an individual is calculated according to the expression (2). Analysis of the model was carried out by means of computer simulation.

The parameters of the model at simulation are chosen in such manner that the evolutionary search is effective. The experience of the work (Red'ko, & Tsoy, 2005) for the case of pure evolution (at  $\varepsilon = 0$ ) is used at this choice. We believe that the length of the chain is large:  $N = 100$ . We also set  $\beta = 1$ , this corresponds to a sufficiently high intensity of the selection, so the selection time is small, thus the duration of the evolutionary search is determined mainly by the intensity of mutations. The intensity of mutations must not be too large, in order to remove the possibility of mutation losses of already found good individuals. On the other hand, the intensity of mutations must not be too small, in order to ensure sufficiently large mutational search during the evolutionary optimization. Taking this into account, we believe that the probability to substitute any symbol in the sequence  $\mathbf{S}_{0_k}$  in one generation at mutations is  $p_m = N^{-1} = 0.01$ . At this mutation intensity  $p_m$  approximately one symbol in the genome of any individual is replaced at one generation, i.e. during one generation the Hamming distance  $\rho$  between sequences of individuals  $\mathbf{S}_k$  in the population and the optimal sequence  $\mathbf{S}_m$  changes on average by 1 by means of mutations. The selection leads to a decrease of the distance  $\rho$ . Since the intensity of the selection is large, and the Hamming distance between sequences of the initial population and the optimal sequence  $\mathbf{S}_m$  is of the order

of  $N$ , the whole process of evolution will take approximately  $G_T \sim N$  generations. This estimation of the rate of evolution is true, if the population is large enough and the neutral selection of individuals (that is the selection independent on the fitness of individuals) can be neglected. To satisfy this condition, it is enough to require that the characteristic time of the neutral selection (that is of the order of the population size  $n$  (Kimura, 1983)) should be greater or of the order of  $G_T$ , so we believe that  $n = G_T = N$ .

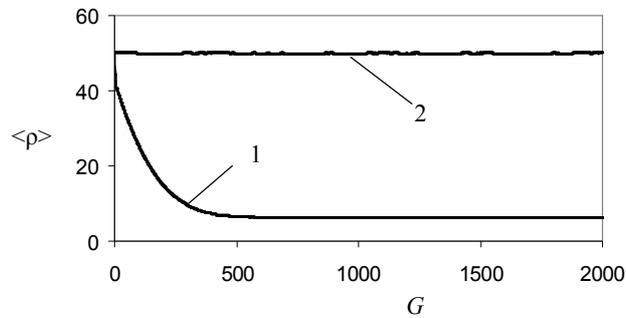
Thus, the parameters of simulation are chosen as follows:  $N = 100$ ,  $\beta = 1$ ,  $p_m = N^{-1} = 0.01$ ,  $n = G_T = N = 100$ .

We also believe that the probability of a random replacement of symbols of sequences  $\mathbf{S}_{T_k}$  during learning  $p_l$  is large:  $p_l = 1$ , the number of time moments during a generation is equal to 2:  $T = 2$ ; the parameter  $\varepsilon$  is small:  $\varepsilon = 10^{-6}$ .

### *Comparison of pure evolution and evolution combined with learning*

Figure 1 shows the dependence of the average of the Hamming distance  $\rho = \rho(\mathbf{S}_k, \mathbf{S}_m)$  between sequences  $\mathbf{S}_k$  of individuals in the population and the optimal sequence  $\mathbf{S}_m$  at the beginning of generations (i.e. in this case  $\rho(\mathbf{S}_k, \mathbf{S}_m) = \rho(\mathbf{S}_{0k}, \mathbf{S}_m)$ ) on the generation number  $G$ . Curve 1 characterizes the regime of evolution combined with learning; curve 2 characterizes the regime of pure evolution. The fitness of individuals is determined by the expression (1). The dependences are averaged for all individuals of population. Each point of the plots represents an average over 1000 simulations, each starting with a different random seed. Typical errors of values  $\langle \rho \rangle$  at the plots are smaller than 0.1.

We can see that pure evolution without learning (curve 2) does not optimize individuals  $\mathbf{S}_k$  at all even at small values of  $\varepsilon$ ; whereas evolution combined with learning (curve 1) obviously ensures the movement towards the optimal individual  $\mathbf{S}_m$ .



**Figure 1.** Dependence of the value  $\langle \rho \rangle = \langle \rho(\mathbf{S}_k, \mathbf{S}_m) \rangle$  on the generation number  $G$  for evolution combined with learning (1) and for pure evolution (2).

To understand, why pure evolution does not ensure a decrease of the value  $\rho$ , let us estimate the value of the fitness (1) in the original population. Individuals of the initial population  $\mathbf{S}_{0k}$  are far from the optimal one  $\mathbf{S}_m$ ; the Hamming distance  $\rho = \rho(\mathbf{S}_{0k}, \mathbf{S}_m)$  is of the order of  $N/2 = 50$ , therefore,  $\exp(-\rho) \sim 10^{-22}$  and  $\exp(-\rho) \ll \varepsilon$ . This means that all individuals of the population have approximately the same value of the fitness  $f(\mathbf{S}_k)$ :  $f(\mathbf{S}_k) \approx \varepsilon$ . Consequently, the selection of individuals does not occur in the case of pure evolution. Therefore, the movement towards  $\mathbf{S}_m$  occurs only in the presence of learning;

this movement leads to the decrease of  $\rho$ . The detailed analysis of dynamics of genome and phenotype distributions demonstrates that this genetic assimilation takes place as follows: the phenotype distribution moves towards the optimum  $S_m$  at learning and further selection; then the genomes of selected organisms also move towards the optimum.

Thus, learning can lead to a radical acceleration of evolutionary processes.

### *Hiding effect*

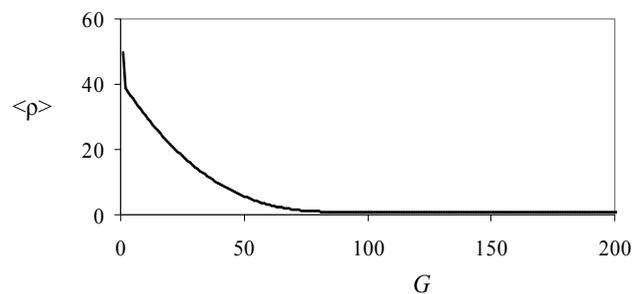
However, strong learning can not only accelerate the evolutionary search, but it can prevent a finding of an optimal genome. Curve 1 in Figure 1 shows that at large  $G$  the decrease of  $\langle \rho \rangle = \langle \rho(S_{0k}, S_m) \rangle$  is limited: the final value  $\langle \rho \rangle$  remains rather large, the asymptotic value  $\langle \rho \rangle$  is approximately equal to 6.2. This is due to the fact that at large  $G$  ( $G \sim 1000$ ), strong learning ( $p_l = 1, T = 2$ ) results in finding the optimal sequence  $S_m$  independently on the genome  $S_{0k}$ . Namely, phenotypes  $S_{T_k}$  of individuals of the population at  $t = T$  are equal to  $S_m$  or very close to  $S_m$ . Therefore, at final stages of evolution the genomes  $S_{0k}$  do not move towards the optimum  $S_m$ . So, the hiding effect (Mayley, 1997) at final stages of evolution is observed.

This effect means that strong leaning prevents the evolutionary optimization, as it increases chances of finding a good phenotype independently on genomes of individuals. In our case, the hiding effect is observed at the end of the evolutionary process.

### *Influence of the learning load on the modeled processes*

We also analyzed the influence of the learning load on modeled processes. In this case, the fitness is determined by the expression (2). Simulation is performed for the mentioned parameters ( $N = n = 100, \beta = 1, p_m = 0.01, p_l = 1, T = 2, \varepsilon = 10^{-6}$ ), the value of  $\alpha$  is equal to 1.

Simulation results are represented by Figure 2 that shows the dependence of the average distance  $\langle \rho \rangle$  between sequences  $S_{0k}$  and the optimal sequence of  $S_m$  on the generation number  $G$ . Each point of the plot represents an average over 1000 simulations, each starting with a different random seed.



**Figure 2.** Dependence of the value  $\langle \rho \rangle$  on the generation number  $G$ ; the fitness of individuals is determined by the expression (2); the decrease of  $\langle \rho \rangle$  is much faster than in Figure 1.

Thus, the learning load leads to the more effective selection of individuals with the genomes of  $S_{0k}$ , which are close to  $S_m$ , and the evolution process and the genetic assimilation are significantly accelerated.

## Conclusion

Thus, the model of interaction between learning and evolutionary optimization has been constructed and investigated. Our model is based on the quasispecies model of evolution (Eigen, 1971; Eigen & Schuster, 1979). The optimization of autonomous agents by means of learning and evolution is analyzed. It is shown that learning can lead to a radical acceleration of evolutionary processes. The mechanisms of the genetic assimilation is studied. The genetic assimilation takes place as follows: the phenotype distribution moves towards the optimum at learning and further selection; subsequently the genomes of selected organisms also move towards the optimum. The mechanisms of the hiding effect is analyzed. This effect means that strong learning inhibits the evolutionary search for the optimal genome, if this learning increases the chances of finding a good phenotype independently on the genome of the individual. The influence of the learning load on evolutionary processes is studied. It is shown that the learning load leads to additional significant acceleration of the evolutionary optimization and the genetic assimilation.

## Acknowledgments

This work is partially supported by the Russian Foundation for Basic Research, Grant No 13-01-00399.

## References

- Ackley, D., & Littman, M. (1992). Interactions between learning and evolution. In: C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.). *Artificial Life II: Proceedings of the Second Artificial Life Workshop* (pp. 487-509). Redwood City CA: Addison-Wesley.
- Baldwin, J. M. (1896). A new factor in evolution. *American Naturalist*, 30, 441-451
- Belew, R. K., & Mitchell, M. (Eds.). (1996). *Adaptive Organisms in Evolving Populations: Models and Algorithms*. Massachusetts: Addison-Wesley.
- Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules, *Naturwissenschaften* 56, 465-523.
- Eigen, M., & Schuster, P. (1979). *The hypercycle: A principle of natural self-organization*. Berlin etc.: Springer Verlag.
- Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution, *Complex Systems*, 1, 495-502.
- Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.
- Mayley, G. (1997). Guiding or hiding: Explorations into the effects of learning on the rate of evolution. In: P. Husbands & I. Harvey (Eds.) *Proceedings of the Fourth European Conference on Artificial Life (ECAL 97)*. (pp. 135-144). Cambridge, Massachusetts: The MIT Press.
- Morgan, C. L. (1896). On modification and variation. *Science* 4, 733-740.
- Osborn, H. F. (1896). Ontogenetic and phylogenetic variation. *Science* 4, 786-789.
- Red'ko, V. G., Mosalov, O. P., & Prokhorov D. V., (2005). A model of evolution and learning. *Neural Networks*, 18, 738-745.
- Red'ko, V. G., & Tsoy, Yu. R. (2005). Estimation of the efficiency of evolution algorithms, *Doklady Mathematics*, 72, 810-813.

- Turney, P., Whitley, D., & Anderson, R. (Eds.). (1996). *Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect*. Special Issue of Evolutionary Computation on the Baldwin Effect (Vol. 4, No. 3).
- Waddington, C. H. (1942). Canalization of development and inheritance of acquired characters. *Nature* 150, 563-565.