

Artificial Selection Algorithm

Kunal S Panchal

Dept of Computer Science
Thakur College of Science &
Commerce, Mumbai

Jigar K Doshi

PhD in AI
Georgia Tech, Atlanta US

Ashish Trivedi

Dept of Computer Science
Thakur College of Science &
Commerce, Mumbai

ABSTRACT

In a genetic algorithm, a population of strings (called chromosomes), which encode candidate solutions (called individuals or creatures) to an optimization problem, is evolved toward better solutions. Better solution is usually attained in form of succeeding generations. These generations are evolved using techniques inspired from the real world viz., selections, mutations and crossovers. Till date, selections were performed using the most basic technique called natural selection, in which, the fitter population has a better chance of getting selected to become the parents of the successive generations. This 'natural selection' can be directly mapped to naturalist Charles Darwin's theory of survival of the fittest. However, Darwin also mentioned another method of nature's selection process called the 'Artificial Selection'. This paper is indented towards mapping the methodology of artificial selection in Artificial Intelligence to develop a new algorithm that aims at the same goal of survival. In this method, only the required traits or combination of traits of the population will be carried forward. This is done by introducing a factor which decides which traits are needed and which are not. This factor can be implicitly included in the program or can be explicitly decided by the user input. This method differs from the existing natural selection in the way that it does not depend totally on fitness of the generations but seeks mainly the required traits or combination of traits needed to solve the problem, and also that the generations evolve according to the factor which can sometimes be user itself who decides which traits he wants to be prominent and which are to be discarded. As a result, the successive generations are a product of selective breeding by the factor. Also, the generations can completely be different in terms of traits of their parents.

General Terms

Algorithm

Keywords

Artificial selection algorithm, required traits, lesser dependency on fitness, evolution controlled by a factor, selective breeding.

1. INTRODUCTION

In the natural world, Artificial selection (or selective breeding) describes intentional breeding for certain traits, or combination of traits. The term was utilized by Charles Darwin in contrast to natural selection, in which the differential reproduction of organisms with certain traits is attributed to improved survival or reproductive ability. As opposed to artificial selection, in which humans favor specific traits, in natural selection the environment acts as a sieve

through which only certain variations can pass. Darwin noted that many domesticated animals and plants had special properties that were developed by intentional animal and plant breeding from individuals that showed desirable characteristics, and discouraging the breeding of individuals with less desirable characteristics. Humans cross breeding different species of animals for a crossbreed who has only those traits which are selected by humans is an example of selective breeding through artificial selection. This artificial selection algorithm tends to map the humans with the factor (in control of selection) and the crossbreed to the required final result population.

2. OVERVIEW

2.1 How is it different from natural selection?

In nature, artificial selection differs from natural selection by primarily one prominent point, which is, natural selection is carried out by the need of survival of the fittest. Living things that are fit to adapt to their surroundings are selected to reproduce and evolve. Whereas, artificial selection in nature is done by humans. Human select certain traits which they want to exist in the next generation. For eg, selecting a dog who has sharper sense of smell and mating it with a dog who has good fur to produce a generation of dog who has both qualities in it. Natural selection produces breeds and artificial selection produces cross-breeds.

Now, in this artificial selection algorithm, this would mean mapping the above difference one-to-one. Natural selection algorithms always select the individuals who are fitter irrespective of which traits will get carried forward and which will be lost. The end result in the form of generation is not known in natural selection and so is ended at an optimum level. In artificial selection algorithm the end resultant generation or solution has a preconceived form. So this algorithm gives prime attention on selection of only those individuals who possess either whole or part of those solution traits, and the fitter amongst such individuals are allowed to mate.

2.2 Redefining terminology

This algorithm also differs in case of terminology of representation of an individual. The standard notation of representing an individual was in the form of chromosomes of binary digits where each binary digit represent a gene. In this algorithm, we increase the granularity level by representing the genes further with the group of elements called 'alleles'. Each of the elements are called 'alleles'. Each allele represents a trait. These alleles group up to represent one gene.

Former Standard Representation of an Individual:-

(011010)

New Representation of above Individual:-

((01) (10) (00) (11) (01) (10))

Where,

0 = (01)

1 = (10)

1 = (00)

0 = (11)

1 = (01)

0 = (10)

2.3 Genes vs Alleles in nature

- A gene is a part of the DNA. Alleles on the other hand refer to different versions of the same gene.
- An interesting difference between alleles and genes is that alleles produce opposite phenotypes that are contrasting by nature.
- Alleles are basically different types of the same gene. It can be explain in this way- If your eye color was decided by a single gene, the color blue would be carried by one allele and the color green by another. Fascinating, isn't it?
- All of us inherit a pair of genes from each of our parents. These genes are exactly the same for each other. So what causes the differences between individuals? It is the result of the alleles.
- Alleles occur in pairs but there is no such pairing for genes
- Alleles determine the traits we inherit.

2.4 Algorithm explained

The algorithm will start of by generating a schemata version of the solutions which will represent an estimate or requirements of the solution. In simple words, this schemata contains the kind of traits which are required to be present in our crossbreed. Our solution will be called crossbreed solution or just 'crossbreed' and accordingly the above schemata is named 'crossbreed schemata'. We will then produce random feasible solutions or random initial population. As discussed above the population will include individuals with new representation. We will then run an optimum trait similarity search. In this search the crossbreed schemata will be taken as a reference and search will be made on whole population to see which individuals possess exact match or some part of the required traits. The outcome of this search will be those individuals who possess a part or whole of the required kind of traits. These individuals will represent artificial selection. But still the fittest amongst the individuals selected in the above search are not known. So now we will apply the fitness function on those individuals to obtain the best individuals for mating. This will ensure that we get individuals who are

comparatively fit and also possess the required kind of traits. The individuals obtained after evaluation of the fitness would be our candidates for mating and producing generations. Before crossover the individuals will be sorted according to their fitness value. We will trim the population size by deleting the lower weak and unwanted individuals whenever the population size exceeds the initial population. This trimming and deleting may not happen in every generation. Some desirable amount of individuals amongst the highest fitness valued will be secured for next generation. Sufficient amount from the leftover individuals will be chosen for crossover. Equal amount of individuals will be produced in crossover. The parents will be retained in the next generation as well. So there will be a temporary population increase after crossover and mutation. Artificial selection will again be made. This will reduce the population size and the loop will follow until the threshold condition. The idea of retaining parents in the next generation can be discarded in applications where a real world simulation is to be done, as in that case parents crossover can be unacceptable. These individuals will mate through crossover and mutate to produce generations after generations which will gradually lead to our required optimum crossbreed solutions. Each generation will undergo artificial selection as stated above. So our population will gradually decrease to produce and retain a crossbreed of optimum solutions, out of which the desirable optimum solution can be chosen. Fitness of each generation will also be tested by our fitness function. After each generation the traits of generations will also gradually start matching the required crossbreed schemata better. The production of new generations will stop as the threshold condition is reached. A final generation will represent our optimum crossbreed solutions containing the required traits.

3. ALGORITHM FLOWCHART

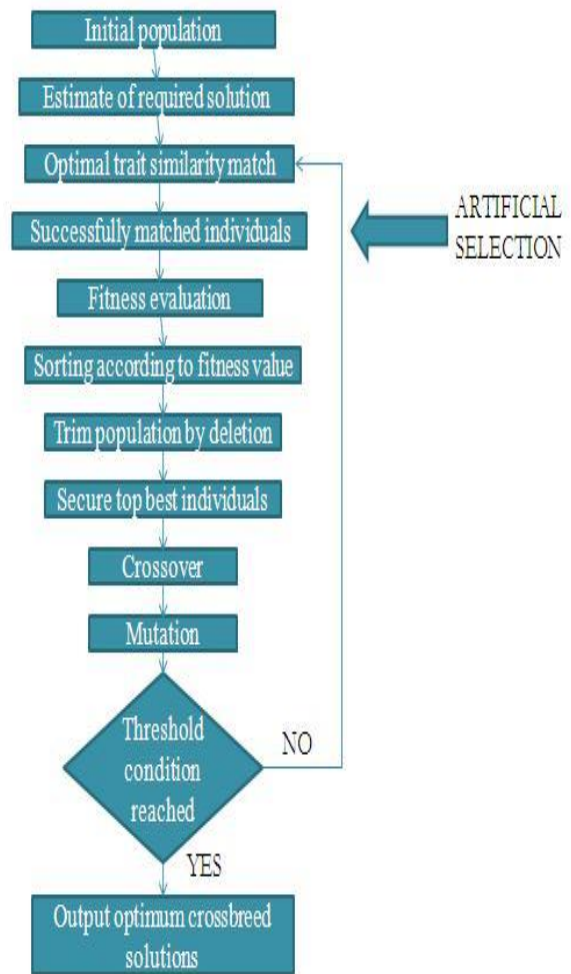


Fig 1: Algorithm Flowchart

4. SUPER ATHLETES

For portraying the capability of this algorithm we have created an abstract application.

- In this application, we create a population of athletes. Athletes range from very poor to excellent randomly.
- Each athlete will have to run on a hurdle track. The track will contain 5 types of hurdle viz., slab, pole, wire, sandpit, gate. But we do not know exactly how will the track look like. We only have an estimate that any two randomly generated hurdles will appear the most number of times on the track. On this basis we prepared a schemata containing 5 positions of the required solution. The positions of each hurdle on the 5 element schemata is fixed. The schemata marks a '1' if, that position is of the hurdle which will appear most number of time, and a don't care symbol 'z' elsewhere.
- Our athletes will overcome each obstacle with one corresponding action. A correct action performed

for a particular hurdle will give them reward and increase their fitness points. At the same time if their chromosome shows that they perform the correct action to overcome the hurdle marked as '1' in the schemata then they will be eligible for getting selected in the artificial selection.

- Our obstacle will contain a corresponding correct action to be performed for overcoming it. The correct actions are:-
 Slab = short jump
 Pole = high jump
 Sand pit = long jump
 Wire = crawl
 Gate = slide.
- The individuals are represented by a obstacle-action pair of 5 elements, each standing for one of the 5 types of obstacles and the related action performed on that obstacle by the individual.

Individual = (sand_pit: long_jump, gate: slide, pole: high_jump, wire: crawl, slab: short_jump)

The above is representation of an individual who scores maximum fitness as he does all correct actions during corresponding obstacles.

So in this way we represent an individual in form of alleles where, sand_pit: long_jump are a pair of alleles; gate: slide other pair of alleles and so on.

- The “1’s” in the schemata represent the required traits that should be present in the solutions partly or wholly.
- Then using the schemata we run an optimal trait similarity match on the whole population. The output of this match is ‘artificial selection’. It selects only those individuals who possess the correct obstacle-action pair on the place where there is a ‘1’ in schemata. There are two “1’s” in our schemata. So the schemata will also select that individual which has a correct obstacle-action pair at either of the “1” positions of the schemata. This will reduce our population.
- Now we evaluate the fitness of those individuals selected through artificial selection to obtain comparatively better individuals, and assign the fitness value to them. Our fitness function is a table which gives a particular score to the individual whenever he performs some action on the obstacles. The maximum score an athlete can get is by overcoming each obstacle with the correct action to get a +10 each time.

Table 1. Fitness function representation

	Short Jump	High Jump	Crawl	Long Jump	Slide
Sand_pit	10	-1	-5	-3	-5
Gate	-2	10	-5	-3	-5
Pole	-5	-5	10	-4	-4
Wire	-3	-3	-5	10	-5
Slab	-4	-4	-1	-4	10

- After having the fitness scores of each individual, we sort them according to their fitness scores.
- Then we trim the population if the population is exceeding the initial population, which in the first run will not happen. This step will be executed to reduce the size of the population if the generations population increases exponentially more than the initial population. At that point we will delete the lower weak and undesirable individuals obtained after sorting. This will make the generation population size same as the initial population size.
- Then it does crossover of selected individuals but we do not include the few percentage of top best individuals in crossover. This creates a generation. But the parents are not discarded in this. Parents also will become a part of the next generation. This

is because we make sure that we do not lose parents who have the required traits in this process.

- Then we mutate the alleles of the generation randomly.
- After mutation we run the loop again from artificial selection. We again select to obtain a crossbreed which is optimized than the previous generation. We again run fitness function test, sort them, trim if needed, secure best and do crossover and mutation.
- We kept a Threshold of maximum iterations. We observed that after the threshold is reached, we had a generation of athletes who retained the required traits for overcoming the schemata obstacles containing “1’s” and who had a higher fitness score than the previous generations.
- In this way we got a group of super athletes who are both possessing the required traits and are fitter than others. So our super athletes are now capable of running over a track that was estimated to contain any 2 random types of obstacles the most with an assurance of overcoming that obstacle along with overcoming other obstacles.

4.1 Advantages Limitations & Applications

Like every algorithm or technique, this algorithm also has advantages over the previous and its own limitations.

4.1.1 Advantages

- This algorithm has complete control and understanding of the kind of generations it needs to evolve.
- The needed traits of an individual are assured to be retained throughout generations.
- Increasing the granularity by representing genes in the form of groups of alleles, makes this algorithm capable of mapping terminology with the real world to a much better level with higher accuracy. This introduction to alleles may attract attention of genetics, as the representation of genes in the form of their alleles can be useful for better simulations.
- It ensures that the solution is derived from the available selected individuals efficiently.
- In the end we get a group of optimized crossbreed solutions to choose from.

4.1.1.1 Limitations

- This algorithm has considerable scope for future enhancement in terms of its complexity in execution.
- An estimate of the solution is required for the selection process of individuals from initial population.
- Careful observation is needed to retain the required trait alleles through generations, which sometimes can be tedious.
- The representation in terms of alleles of genes demands thorough understanding for interpretations and mapping of possible solutions to individuals.

4.1.1.2 Applications

Application domains can be suggested for this new algorithm on basis of its working technique.

This algorithm can be applied to domains where an estimate of the solution is known.

Genetics – Because of its granularity level of mapping which includes alleles of genes, this algorithm can attract genetics to simulate artificial selection which will closely resemble real world artificial selection.

Game playing – It can help select a better player amongst multiple players based on the previous performances, who can ultimately win the game.

5. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of this paper.

6. REFERENCES

- [1] Carl Sagan. COSMOS, Ballantine Books, 1985
- [2] David E Goldberg. Genetic Algorithms in search, Optimization & Machine Learning, Addison Wesley.
- [3] AISWeb, The Online Home of Artificial Immune Systems, <http://www.artificial-immune-systems.org/algorithms.shtml>
- [4] Potts, J.C., The development and evaluation of an improved genetic algorithm based on migration and artificial selection, IEEE Transactions on Systems, Man and Cybernetics, Vol 24, Issue 1, 1994, pp 73-86.
- [5] Leandro Nunes de Castro, Fernando J. Von Zuben, The Clonal Selection Algorithm with Engineering Applications, In Workshop Proceedings of GECCO, pp. 36-37, Workshop on Artificial Immune Systems and Their Applications, Las Vegas, USA, July 2000.