|  |  |  |  |
| --- | --- | --- | --- |
|  | 11/21/2011 |  | |
|  | |  | |
| Artificial life project  *Gene pairing grammar* | | | |
|  |  | | |  |  |  | | --- | --- | --- | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |
|  |  | | Aravind Rajesh |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Artificial life project

Gene pairing grammar

With the theory of neo Darwinism, evolution is the connection between the microevolution and macro evolution .The term Micro evolution refers to the process of change in the phenotype of the organism within the species diversified by geographic locations and adapts them to their environment .All available evidence supports the conclusion of adaptability within the species.

House sparrows have adapted to the climate of North America, crows adapted to the climate of Europe and other different cases. These micro evolutionary changes not only depends on the factors of random mutation but also should have a causal geographical changes based on which the mutation is possible to be a positive for the organism in its environment.

Let’s say from the ancestor bear descends two kind of bears, geographically distributed in different areas as brown bear(“brown in colour”)and polar bears(“white in colour”).Lets imagine we had only brown bears in polar region due to random mutation they changed to white colour. Assume now you go on a vacation to polar region and find that the ratio of brown to white bears is 70:30.Now after a year now you check its now 40:60(brown: white) now you have checked a micro evolutionary pattern: change in gene frequency new population has evolved. This could have happened because the bears that are white in colour survived to reproduce more frequently than bears with brown colouration genes, so that more white bears got into the next generation, but this seems more complex than this, polar bears we’d all surely agree that were selected for being the same colour as their environment, and not for being white, then being white is no camouflage if snow is, red. In point of Darwin evolution with no intelligence in selection. But to select for “certain traits” as opposed to “selecting” them by not having them died out - It might be obvious that being the same colour as your environment is more important than being white. Many scientific documents are available as evidence for proving microevolution, eg development of resistance against certain chemical by unicellular organism, such as virus, and bacteria’s which is by natural selection.

Macro evolution, on the other hand, covers the process responsible for the divergence among species, otherwise called as speciation. It’s is the blind process produced the specified complexity of life by mutation mainly occurring during DNA replication, through the deletion, addition or change of a single nucleotide. But if we see closely mutations are nothing but mistakes, errors that violate the rules of spelling and grammar in the language of DNA.

Genetic code which is responsible for the evolution is a set of rules by which information encoded in genetic material DNA is translated into proteins. This genome that codes for protein into an mRNA is called a gene. This code defines how sequences of three nucleotides, called codons are formed. Each gene is transcribes into a molecule of the related polymer RNA. The information available in the DNA is decoded in to RNA to make the network of polypeptides and helps in the development of the organism. We know that result of mutation which is random in gene can be “positive”, “neutral”, or “negative” in other words

Mutation in a RNA sequence can cause several different changes in sequences these changes can the function of gene, or prevent the gene from functioning or have no change. . (Sawyer, 2007 )Studies in the fly Drosophila melanogaster suggest that if a mutation changes a protein produced by a gene, this will probably be harmful, with about 70 per cent of these mutations having damaging effects, and the remainder being either neutral or weakly beneficial. Due to the damaging effects that mutations can have on genes, organisms have mechanisms such as DNA repair to remove negative mutations.( Kauffman 1993,p.29)Points that complex systems respond differently to minor changes in internal system, they may change behaviour relatively smooth or radically. Adaptive evolution involves all accumulation of all these positive changes in genotypes over a period of time .

It’s a complex Mechanism of “genotype "carried within the genetic code which states the "phenotype” of that organism developed, If it is not possible for a single gene to bring any changes.(R.Lewontin ,1974)In his work indicated that 1.a theory mapping genotype to phenotype,2.a theory relating phenotype to fitness. Changes in phenotype to the relating genotype, the closure between phenotype and genotype is obtained the drift or cumulative selection. It is a repetitive process of copying of genetic variation in a population, eventually the best among it is chosen to evolve.

Most evolutionists believes that mutation doesn’t happen at one time in a single gene

But it’s a repetitive process of mutation in multiple genes which is copied from fitter organism i.e.: one who has managed to survive and reproduce (natural selection) to the next generation.

All the research and the evidence for the macroevolution states till now, only how an organ or a new species could have developed, but no evidence supports saying, why it has changed?, Or the reason for the evolution. Why can’t macro evolutionist couldn’t give a situation or cause that would have prevailed for the speciation. The simple reason is

If then evolution is no more random and it becomes goal oriented. To me adaptation and evolution has driven objective .If Evolution is based on randomness alone then how it evolved complex organism. Because evolution is not based on complete randomness, but also involves cumulative selection.

**Evolution of wings**

Initially it was believed, wings have evolved from archaeopteryx ,later newly discovered ceratosaur belonging to a group theropods.In both cases ,my question what should have turned a reptile with two front limbs usable for body balance and tearing prey turned into a wing structure, did it wanted to fly or its survival tactics wanted them to get wings

In other way random mutation suddenly caused develop its wings, which gave it a advantage over others, this helped it to escape from predators and reproduce effectively passing his genes to his successors thus creating a new species birds. Problem A (for survival and escape from predators), random mutation developed a solution.

**Infinite monkey theorem**

*“The infinite monkey theorem states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type a given text, such as the complete works of William Shakespeare”()*

If two events are statically independent to each other, then the probability of both happening together is equal to the product of the probability of each one happening independently.

We know that for evolution to be success full mutation should happen to more than one gene at the same time

E.g.: If the probability of flipping first gene from a set of 20 gene is 0.05 and the next set of gene is 0.008 then the chance of both happening at the same generation is

0.05 \*0.008=.0004

For easy assumption

Let’s take for instance the 6 letter sequence MONKEY is to be changed to HUMANS

Suppose the alphabets has 26 letters excluding the possibilities of it being uppercase or lowercase, if we assume the mutation is random and independent, then the chance that the first letter mutate to “H” is 1/26,and the chance that the second letter mutate to “U” is also 1/26.

Therefore the probability of MONKEY randomly mutating to HUMANS is

(1/26)\* (1/26)\* (1/26)\* (1/26)\* (1/26)\* (1/26)= (1/26)^6=1/308915776

Less than one in 3 billion.

From the above, the chance of not mutating HUMANS in a given block of 6 letters (generation) is 1 − (1/26)6. Because each block is mutated independently, the chance Xn, of not typing HUMANS in any of the first n mutation is

Xn= (1-1/ (26) ^6) ^n.

This low probability value clearly shows cases that evolutions is not quite possible with such high precision, by just random mutation.

Goal of the work is to demonstrate the misconception that evolution is random mutation alone ,but actually it is random mutation with cumulative selection.

Another goal is to characterize how GA search speed changes as gene length increases.

For this project I re-created the work of Dr.Richard Dawkins ,he gave a brief description to the program in his book “THE BLIND WATCHMAKER”.

Here we take a target string which is considered as goal to reach and mutate through a set of random character and try to achieve the target string.

The set of string is considered as a genome of an organism, where each character represents a individual gene. A random gene is selected and mutated if the child is similar to the target genome it is copied to next generation ,this process continues until it finds a genome which is same as the target string.

This work clearly points that the Darwin evolution is wrong completely with just randomness, e ThTdfgfgbut works only with cumulative selection (Dawkins 1986 p.49).

Description of my program

The target string can be of any length. Of the same length a random sequence of string is generated by my program, this sequence is parent for the first generation parents gets mutated in subsequent generation based on the mutation rate , fitness value is used to check whether the genome of children are fitter. Here we check whether the offspring are closer to the target string .Other conditions which are set in code but can be changed are number of children per generation and the mutation rate of gene.

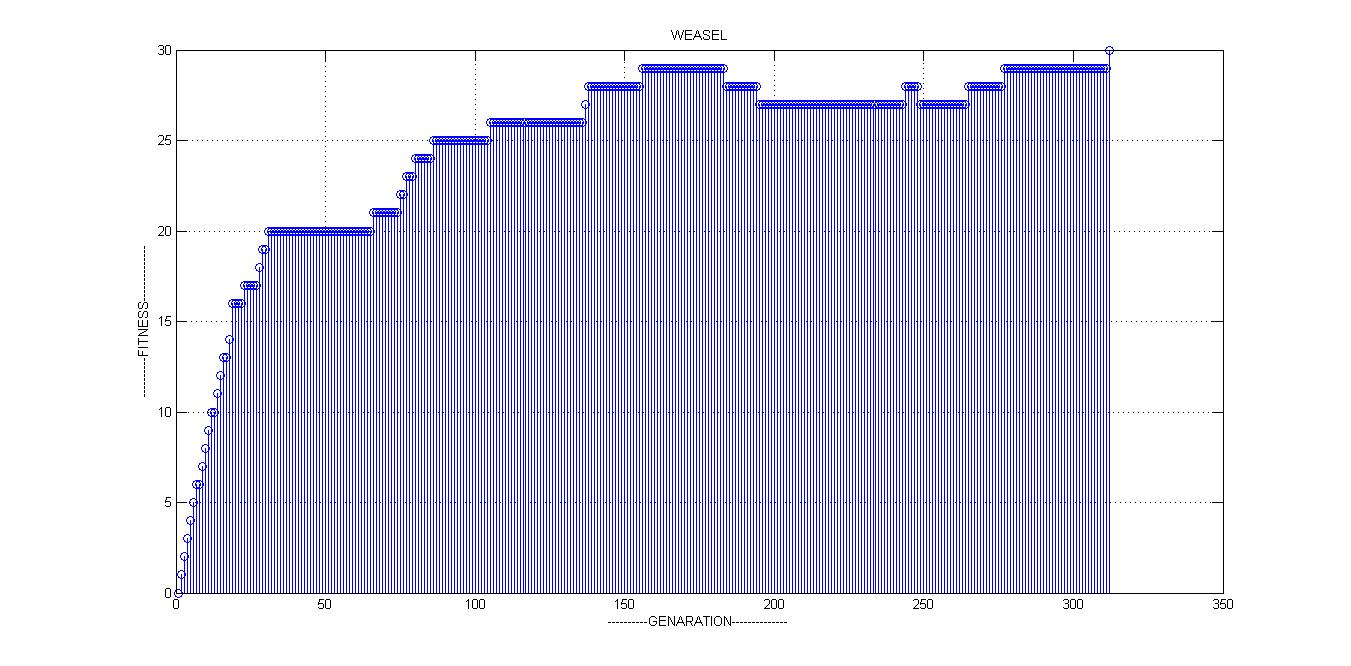
Explanation of the program

For different length of the string, program behaves differently because this program is controlled by generation size and mutation- probability rate. Here the target string which is to be achieved I reserved as” ME THINKS ITS LIKE WEASEL”. This string is 25 characters in length and to be achieved as goal by randomly generated string of same character length.

The random sequence now breeds this is the parent for first generation. In the program number of offspring’s and mutation rate are implicitly specified .After one generation 30 progenies are created with a certain random mutation. Each of the progenies are matched with the target phrase. The one which is close to the progeny is selected as the best using the fitness function.

Here the fitness function used checks whether the target and the any of the children have character at same position. Then this string sequence is selected as the parent for next generation .Thus the best string among the mutated one is selected.

The target string is achieved in 350 generations.



Here the graph reaches the peak in 150 -170 generation and again goes down and reaches the target in 340 genereation.

If there is no target string, the program runs infinitely and the inspection of the resulting progeny string has no useful text in making some sort of meaning.

When the program runs without target string it runs like a single –step evolution .We are trying to find any useful mutation (some meaning in text) which is void. According to Dawkins (1986, p.49) each new try in single step selection is a fresh one, in this way evolution would not have succeeded to this complexity of organisms. But cumulative selection set up by blind forces of nature such as fitness, reproduction, predation might be the reason only positive changes are evolved across to this complexity.

The program is created such that three main parameters can be changes to evaluate the program for different values

Target String, Number of children, Mutation rate.

After obtaining the desired result I like to explore and make few comments

This program can criticized be for the reason that it locks the character once the string matches the goal. This prevents the further mutation of that string.With out this goal function the program run infinitely without producing any useful result.

I ran this program without save function for different length of string and I have compiled the result below:

First I took just two string “ME” .Within 13 generation I could reach it .But still the goal specified.

Just a increase of three letter “ME CAT” took 6500 generation to achieve.

44517 : 3 : MHWCAE

44518 : 2 : MHWCVR

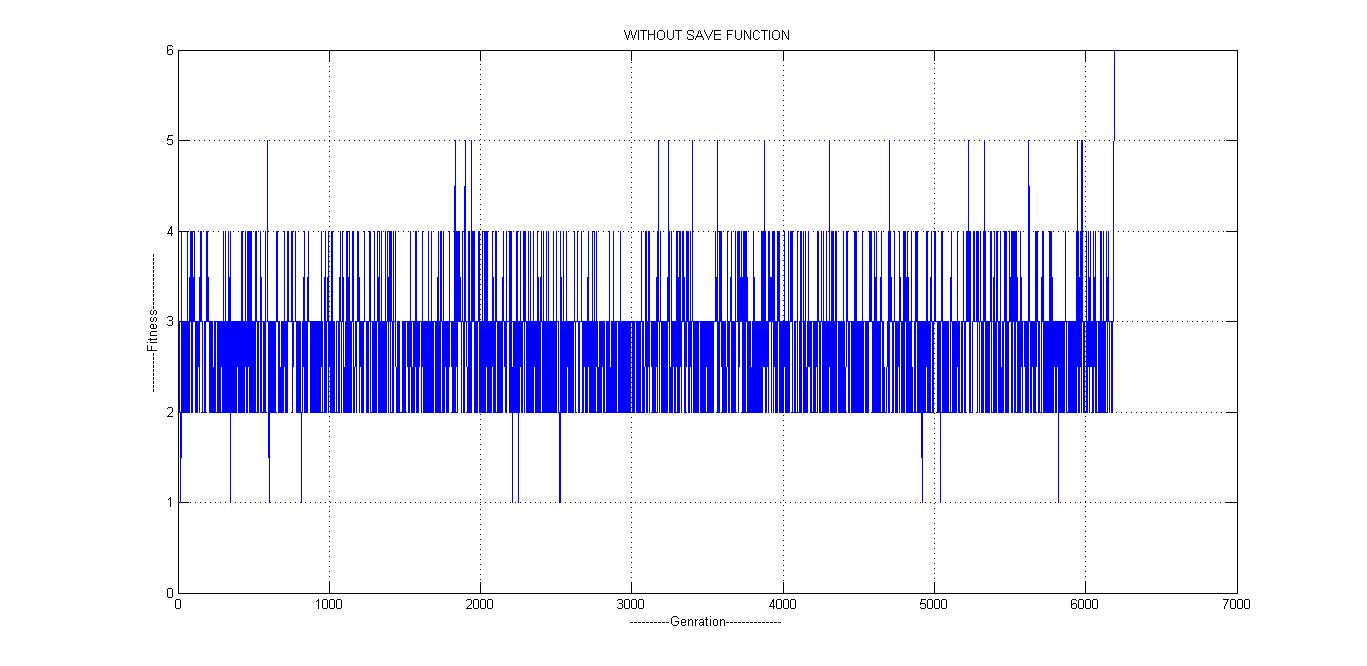
44519 : 2 : MQOCFB

44520 : 2 : MNDCKB

44521 : 3 : HNDCAT

44522 : 3 : HEQCGT

44523 : 6 : ME CAT



I ran this program for more than 1 million generation without save function and could find only 7 character match out of 27 letter characters and some characters in other generation matched less frequently.

1130096 : 4 : UAWTF YUEIXTQLLNBLRILWOHIJF

1130097 : 5 : SURTDMCQEIXASFLKCLRHLWBAFJF

1130098 : 6 : JRRTIZHCFILASFLKCERHJWQANJF

1130099 : 6 : ZHRTBJNKTNUABFLKKERMJLVGJHZ

1130100 : 5 : PHQTMJNKFCUAJFLICMROJLVRFYE

1130101 : 7 : PHQ MTNKKJI NNLIZERJPFQRSJN

1130101 : 7 : PHQ MTNKKJI NNLIZERJPFQRSJN

In addition to it I tried to include some mutation which may result in neutral or negative effect on the progeny (as discussed above –in natural mutation)making it unusable or meaning less. And found that number of generations considerably increased.

1050 : 26 : ME#THINKS ITS LIKE A WEASEL

1051: 26 : ME#THINKS ITS LIKE A WEASEL

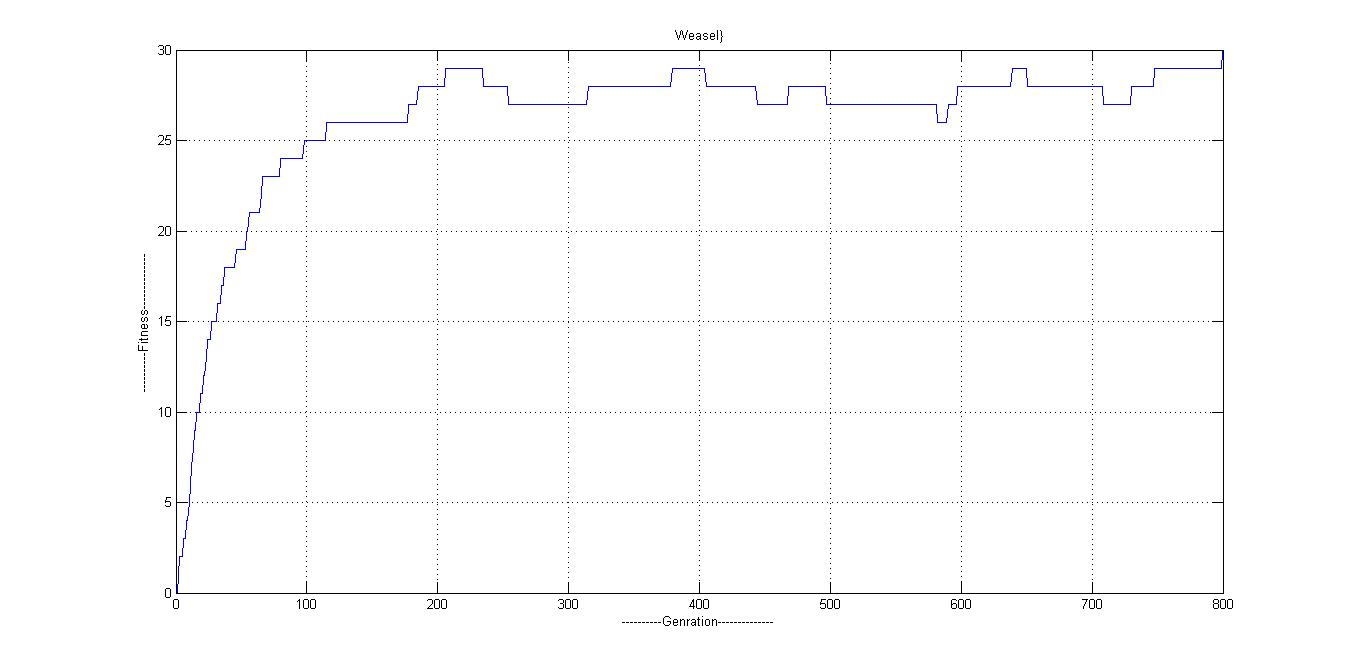
1052 : 26 : ME#THINKS ITS LIKE A WEASEL

1053 : 26 : ME!THINKS ITS LIKE A WEASEL

1054 : 26 : ME!THINKS ITS LIKE A WEASEL

1055 : 26 : ME!THINKS ITS LIKE A WEASEL

1056 : 27 : ME THINKS ITS LIKE A WEASEL



If number of children is large, reducing the probability rate, the target is achieved in fewer generations; if the mutation rate is high we never reach the target because the mutation causes copying from the parent less effective. If the mutation rate increases the target (parent) string is ingnored.From this we can clearly conclude that inheritance or cumulative selection is more important in evolution to reduce the generation.

Now to discuss about how Genetic algorithm changes with size of the target string (adapted from <http://www.matthewegbert.com/?q=node/48>)

If this program is run with larger target strings ,suppose 85 characters, children size =10000,mutation rate=30% a quite interesting result I found is the string tries to converge to the ideal target but since the mutation rate is quite higher for this length string .Mutation happens again at places where there is mismatch of string and again reduces its fitness level, multiple mutation at different points increases the probability exponentially for all them occurring exactly at same time at which it matches the target.

“Adaptive evolution in a population is a hill-climbing process” (Kauffman 1993, p.49).In this case the solution has converged to local optima. Looping around the different characters which do not match, without sacrificing local fitness. Global optimum can be achieved using different fitness function or decreasing the mutation rate.

Conclusion: The basis of this work is to prove that evolution cannot happen only with random mutation but a selection pressure and cumulative progress is required. Here the selection pressure is designed but in real world it is natural .One part where we can start extending is trying to create a model which exactly represents the natural evolution in a simulation provided it should not have any pre written goal as one in this. Other thing is to improve the fitness function or to use neutral networks to avoid the problem of getting stuck in the fitness landscape.

My work tries to address that evolution is possible through cumulative selection but it is not accepted because of the present goal ,there has been already many works done by other trying to compute without goals such as Lenski et al( no date,p.p 139-143) ,here AVIDA a language is run using the code generated by the simulation of string of letters logic functions are obtained as output .In this simulation fixed target is required to run the program initially.

Additional enhancement which can be made to this program is to make the mutation exactly like the one happening in the real world and analyse the result, we know what happen in real mutation.

1. Mutation happens only in certain strings but in this program it happens at every string in every offspring which does not occur in real.

2. Mutation is not always beneficial, so to include some negative mutation which causes the organism to be discarded.

BIBLIOGRAPHY:

Lenski, E et al.(no date)’The evolutionary origin of complex features*’*, *Digital Life Laboratory*,pp.139-143.[Online], (Accessed:23 Jan 2012).

Sawyer,A et al.(2007), ‘Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila’, *The National Academy of Sciences of the USA*, p.122[Online] doi:10.1073/pnas.0703562104(Accessed:23 Jan 2012).

Kauffman,S.A. (1993),*’The origins of order :self-organisation and selection in evolution’.* New York: Oxford University Press.

Dawkins,R.(1986),’*The Blind WatchMaker:Why the evidence of evolution reveals a universe without design’. Norton and company Press Inc*

*Lewontin,R.C.(1974).’The genetic basis of evolutionary change.New York:Columbia University Press.ISBN 0-231-033923.*