



Ancestry Mapping in Genetic Algorithms: A Didactic Visualization Inspired by Greek Mythology

Rubelmar Maia de Azevedo Cruz Neto

Escola Superior de Tecnologia (EST) – Universidade do Estado do Amazonas (UEA)
Av. Darcy Vargas, 1.200 - Parque Dez de Novembro, Manaus - AM, 69050-020
rcruzneto@uea.edu.br

Renata da Encarnação Onety

Escola Superior de Tecnologia (EST) – Universidade do Estado do Amazonas (UEA)
Av. Darcy Vargas, 1.200 - Parque Dez de Novembro, Manaus - AM, 69050-020
ronety@uea.edu.br

ABSTRACT

This work presents a metaphorical implementation of genetic ancestry tests in a Genetic Algorithm (GA), aiming to visually map gene transmission across generations. A predictive categorical model was developed using GA, where individuals were represented by chromosomes composed of real-valued genes. The GA was applied to classify gender based on physical traits such as age, height, weight, and footwear. A novel visualization was implemented in Wolfram Language using color patterns to trace genetic ancestry and mutation history. To enhance didactic appeal, individuals were named after Greek mythological figures — from Titans to Demigods. This approach allowed the ancestry of each individual to be visually tracked, including mutation events and inheritance patterns. Notably, the fittest individual emerged not from the top ancestor, but from a specific combination of genes and beneficial mutations. The proposed visualization offers an intuitive and educational strategy for understanding gene propagation and selection in evolutionary computation.

KEYWORDS. Genetic Algorithm, Wolfram Language, Genetic Ancestry Visualization

Paper topics: MH – Metaheurísticas, AS&DS – Análise e Ciência de Dados, IA – PO e IA

1. Introduction

Bioinspired methods do not aim to model natural phenomena, the idea is to look for a metaphor, an abstraction of the central concepts, that results in some scientific-technological differential to the computing [Castro 2010]. The Genetic Algorithm (GA) introduced by John Holland in 1975 [Holland 1992] is a metaphor from *natural selection*, a process formalized by Charles Darwin in the *Theory of Evolution*. Natural selection favors those individuals that compete for the given resources most effectively, in other words, those that are adapted or fit to the environmental conditions best. This phenomenon is also known as *survival of the fittest* [Eiben and Smith 2015]. The main operators of GA are reproduction, crossover, and mutation, and these were some elements that were “extracted” from original theory. In this study, a metaphor of genetic ancestry tests was implemented in a GA. The actual tests are used to find out which groups or location an individual is originated from, it reveals its ancestors by mapping the relationship between their genes. Based on this test, color patterns were used to map the relationships between the different generations in the GA implementation.

The metaphorical nature of genetic algorithms has been extensively explored as a powerful tool for modeling adaptive systems. Koza [1992] emphasizes how genetic programming can evolve computer programs to solve complex problems, highlighting the potential of genetic algorithms in simulating natural evolutionary processes. Complementing this perspective, De Jong [2006] presents a unified approach to evolutionary computation, stressing the importance of interpretability and accessibility of these algorithms, particularly in educational contexts. Furthermore, Santos *et al.* [2025] demonstrate that interactive visualization tools can significantly enhance the understanding of genetic algorithms, making them more accessible to students and professionals in training. These approaches underscore the relevance of didactic strategies that employ visualizations and metaphors to facilitate the teaching and learning of evolutionary algorithms.

2. Methodology

In this case study, GA was used to fit a categorical model to predict gender (male or female) from four independent variables: weight, age, height, and footwear. Data were collected from people living in the State of Amazonas in Brazil, with 240 observations. This problem can be addressed in GA as a linear weighted sum:

$$Q = \sum_{i=1}^4 \sum_{j=1}^4 b_{ij} \times x_{ij} \quad (1)$$

where i and j represents the independent variable and the category, respectively. There are four continuous variables: weight, age, height, and footwear. Each variable was discretized into four categories, totalizing 16 dichotomous discrete variables, binary variables, x_{ij} . Figure 1 shows the discretization of data from the first observation of dataset. It is possible to see how the set of four binary variables ($x_{i1}, \dots, x_{i4}$) are converted from each continuous variable.

The coefficients b_{ij} are the *genes*, real numbers randomly generated between -1 and 1 ($-1 \leq b_{ij} \leq 1$). The vector formed by all genes is a chromosome (individual) of GA. The number of individuals (population) of each generation was fixed in eight. To enhance didactic appeal, the names of the individuals were taken from *Greek Mythology*. Therefore, the first generation is composed by the *Titans*, the two following are generations of *Gods*, and the final is the generation of *Demigods*. Figure 2a shows the gene values of each individual of the Titans, a single color was chosen for each *Titan* (Figure 2b). For example, all genes of *Cronos* were represented by the red color. This is an important step in the proposed methodology because these colors will be used to map the locations of the genes across the generations.

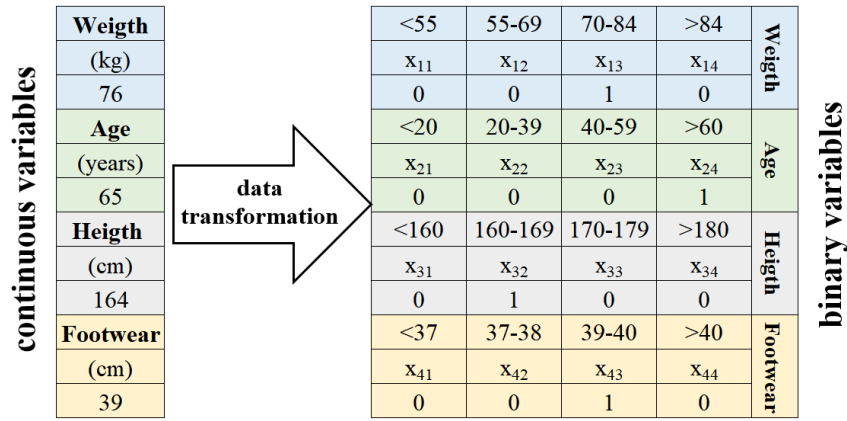


Figure 1. Data transformation - binary variables converted from numeric.

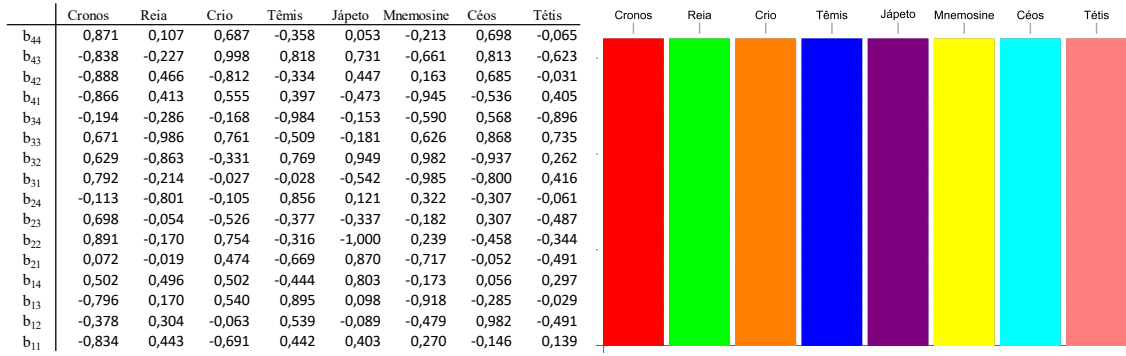


Figure 2. Model coefficients (genes) randomly generated (left) and the colors used to represent the genes of individuals (chromosomes) of first generation (right).

It was defined that when the value of Q is greater than 0, it is assumed that the predicted category is male, otherwise, the predicted category is female, this logical sentence is described in the follow equation:

$$prediction = \begin{cases} \text{if } Q > 0 \rightarrow \text{male} \\ \text{if } Q \leq 0 \rightarrow \text{female} \end{cases} \quad (2)$$

The fitness of GA is the accuracy of the model, *i.e.*, the number of correct predictions divided by the number of total observations.

In the *fitness proportional* or *Roulette Wheel* (RW) selection method, the probability of selecting a chromosome is directly proportional to its fitness value [Castro, 2006]. For example, Céos obtained the highest fitness value, consequently, he also has the largest *slice of the pizza*. The RW is constructed from the *relative fitness* (ratio of individual fitness and the total fitness) of each individual. The RW is usually represented as a pie chart (Figure 3) where the area occupied by each individual is proportional to its relative fitness, which is why Céos has the largest slice. For the parents' selection, numbers are generated randomly between 0 and 1 and stored where they fall in RW. Conceptually this is the same as repeatedly spinning a one-armed roulette wheel, where the sizes of the holes reflect the selection probabilities [Eiben and Smith 2015]. Instead of holes, there are the slices, and the more times the spinning roulette wheel, the greater the probability of selecting the fittest parent.

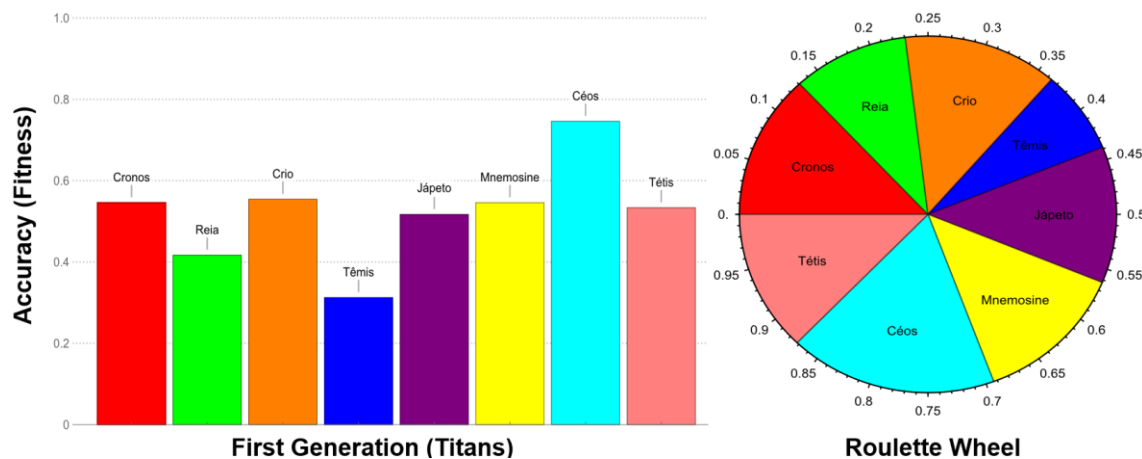


Figure 3. Model accuracy of the first generation – Titans (left) and the roulette wheel used to stochastic selection of parents (right).

After the selection step, it is necessary to choose how many genes the new individual will inherit from each parent. In this study, it was used a single crossover point that was randomly chosen. For example, in Figure 4, it can be seen that *Zeus* inherited his first four genes from *Reia* (green) and the remaining genes from *Crio* (orange). It was used the crossover probability of 80%, consequently, it may not occur crossover, and that was the case with *Deméter*, which means that *Deméter* is a clone of the first selected parent, a clone of *Jápeto* (purple). The gray color was used to represent the mutant genes and it was adopted the mutation probability of 6% which resulted, in this simulation, in one individual with two mutant genes and other without mutation (*Héstia*).

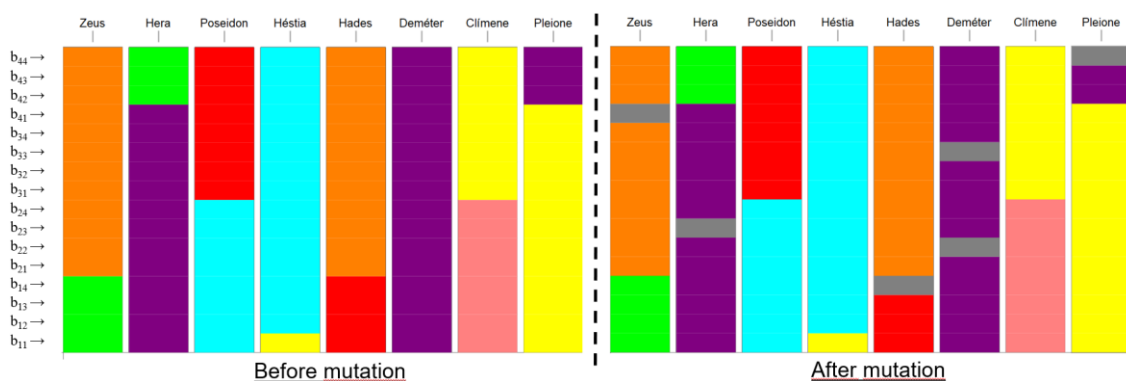


Figure 4. Chromosomes (individuals) of second generation before mutation (left) and after mutation with mutant genes in gray (right).

3. Results

In Figure 5, the upper graph shows the model's accuracy *versus* individuals, and the lower graph shows the respective chromosomes represented by color patterns. The colors were only assigned to individuals of *Titans* (first generation), allowing to map the ancestry of the follow individuals just by observing their colors, and it is also possible to note when mutations happen. For example, *Poseidon* (generation 2) was formed from the genes of *Cronos* (red) and *Céos* (light blue) without mutations. Finally, it was possible to analyze *Diana's* ancestry (*Wonder Woman*), the fittest individual of all, discovering that *Diana* was descendant of *Crio* and *Tétis* with two mutations in their original genes. Interestingly, the fittest individual, *Diana*, was not a descendant of *Céos* — the most adapted individual of the first generation — highlighting that optimal fitness may arise from synergistic gene combinations rather than direct inheritance from top ancestors. Also, it can be observed that *Diana* is almost a total clone of *Atena* with exception of one mutant gene. Therefore, since that *Atena* is the second fittest individual, we can assume that this unique

mutation increased the *Diana*'s fitness, and we can estimate that a combination of genes from *Tetis* and *Crio* may be responsible for the high fitness of *Diana*.

This case study illustrates how color-coded ancestry tracking can enhance understanding of gene flow, mutation impact, and crossover dynamics in evolutionary computation. Beyond academic insight, the proposed visualization may also serve as an engaging pedagogical tool to teach core GA concepts. Future work may explore scaling this method to real-world optimization problems or integrating it into interactive educational platforms.

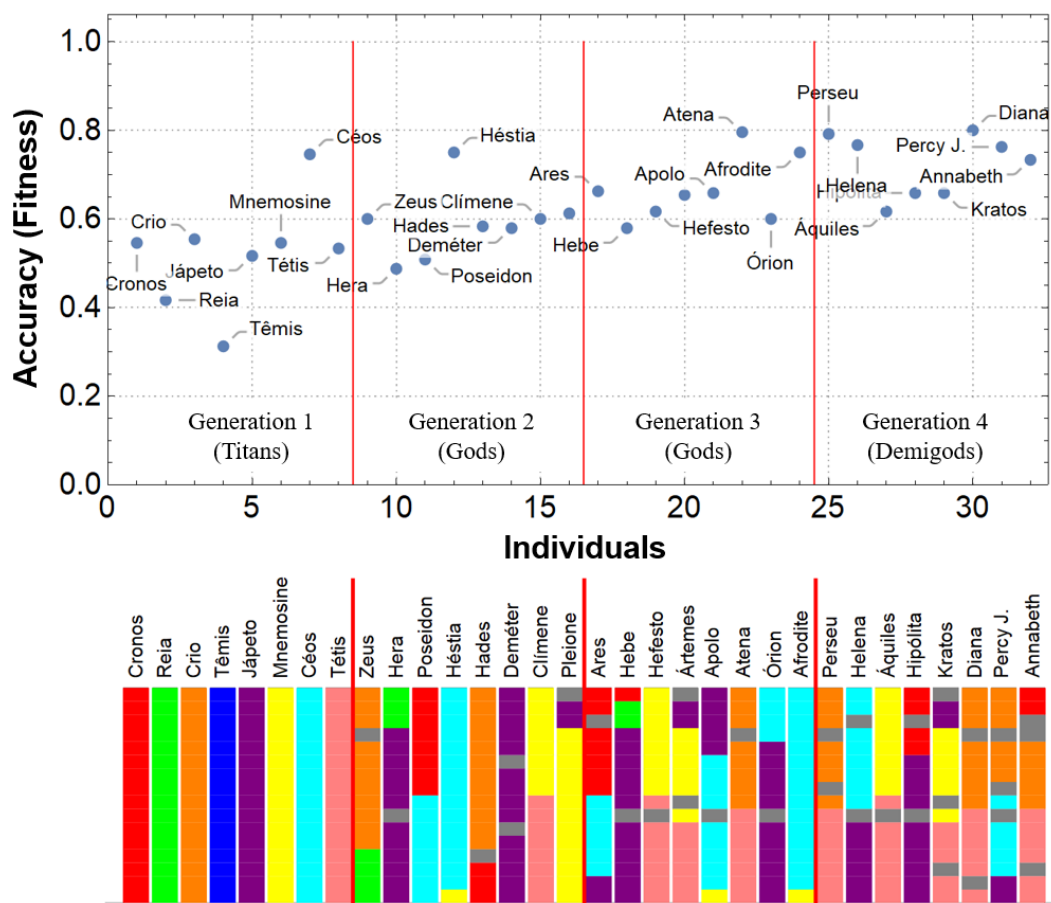


Figure 5. Accuracy versus individuals (above) and genetic ancestry map (bellow).

References

- Castro, L. N. (2006). *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*. Chapman and Hall/CRC.
- Castro, L. N. D. (2010). *Computação Natural - Uma Jornada Ilustrada*. 1. ed. São Paulo: Livraria da Física.
- De Jong, K. A. (2006). *Evolutionary Computation: A Unified Approach*. MIT Press.
- Eiben, A. E.; Smith, J. E. (2015). *Introduction to Evolutionary Computing*. 2. ed. Berlin Heidelberg: Springer-Verlag.
- Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press.
- Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.
- Santos, J. I., et al. (2025) "Interactive Visualization of Genetic Algorithm Solutions for the Traveling Salesman Problem: An Educational Tool." *The International Conference on Industrial Engineering and Industrial Management*. Springer, Cham.