Find Best chromosomes of Goats

INFO6205_201
Final Project

Team Member:

Yumeng Chen (001409547)

Dezhi Yang (001447703)

1 Introduction

1.1 Genetic Algorithm

A genetic algorithm (GA) is a search algorithm used in computational mathematics to solve optimization problems. It is a type of evolutionary algorithm. A genetic algorithm is usually implemented as a computer simulation. For an optimization problem, a certain number of candidate solutions (called individuals) can be abstractly represented as chromosomes, allowing the population to evolve towards a better solution. Traditionally, the solution is represented in binary (a string of 0's and 1's), but other representations can also be used. Evolution starts with a population of completely random individuals, and it happens from generation to generation. Evaluate the fitness of the entire population in each generation, randomly select multiple individuals from the current population (based on their fitness) and generate new life populations through natural selection and mutation. This population becomes the current population in the next iteration of the algorithm.

1.2 Idea

Genetic algorithms are well-suited to problems with huge solution spaces and are useful for problems that have no analytical solution. In this project, we want to find genes that is the best in the reproduction of the goats.

1.3 Function

In the process of constructing the goats' reproduction process, we first built the initial population, the offspring, genotypes, expressions, phenotypes and the environment in the genetic algorithm. Then the calculation algorithm of the fitness, elimination algorithm, crossover algorithm and mutation algorithm are formulated. Additionally, the details in the algorithm, such as the interval of the propagation algebra, and the individual who is too old to participate in reproduction.

2 Overview

2.1 Description

Northern goats are also called hanging sheep, wild goats, etc., with a body length of 105-150 cm, a tail length of 12-15 cm, a shoulder height of about 100 cm, and a weight of 40-60 kg, but the maximum weight can reach 120 kg. The ibex inhabits the bare rocky plateau and mountainside rubble ridges at an altitude of 3500-6000 meters. It does not migrate to low places in winter, so it can be called one of the mammals with the highest habitat. Very good at climbing and jumping, the hoof is extremely solid, the elastic sacroiliac joints and toes like pliers. Distributed in northern India, Afghanistan and Mongolia, in China in Xinjiang and northwestern Gansu, northwestern Inner Mongolia and other places.

In Biezhentao Mountain in Wenquan County, Xinjiang, China, after entering August, the northern goats began herding after going to the winter rams and ewe grouping. After the mating was completed at the end of October, they regrouped again. When the people congregate in August, the rams will live from their own living territory to the ewe's territory for a period of time. This process is the process of migration.

2.2 Definition

2.2.1 Individual

An individual is a goat born from its female goat. The production of new individuals requires the reproduction of two other individuals. The offspring they produce inherit the characteristics of their parents and will be added to the next generation.

2.2.2 Genotype

Genotypes are determined early in the formation of an individual. For the first generation of the initial population, genotypes are randomly generated. For the offspring of the first generation, their genotype was generated by their parents. In this project, each person's genotype will determine their decision at each step in the future.

2.2.3 Expression

This is a genotype-to-phenotype mapping. Under the grouping of individuals' genes, the expression algorithm will obtain the gene sequence of a specific location and map the gene sequence to the traits of goat movement.

2.2.4 Phenotype

The characteristics of the candidate affect the quality of the solution. In this project, phenotype refers to the direction of movement of each person. There

are three types of movement: forward, clockwise and counterclockwise split.

2.2.5 Environment

In this project, the entire environment includes correct roads, misleading roads, and roads blocked by beasts. For goats, they need to understand their behavior before performing each step in this predefined map (this is because goat behavior is determined by the individual's genes). Of all types of roads in the map, the correct path allows the goat to find an exit and score, but the wrong path will hinder the goat's movement.

2.2.6 Fitness

Fitness is the degree to which the organism adapts to the environment. In this case, the fitness is the length of the correct path the goat traverses, which is obtained by the fitness algorithm.

3 Parameter

3.1 Group size

1000 goats

3.2 Chromosome length

Total 128 genes

3.3 Mutation probability

0.05

3.4 Crossover probability

The crossing must occur during reproduction, but the starting position of the intersection is entirely random.

3.5 Probability of death

Based on the goat's fitness to the environment, we divide them into two parts from the middle. The first part has a dead probability of 20% and the second part has a dead probability of 80%.

3.6 High quality group size

The high-quality group size is 10

3.7 Maximum number of steps

The maximum number of steps is 300

4 Algorithm Design

4.1 Expression Algorithm

4.1.1 Overview

Expression refers to the mapping between genotypes and phenotypes. By analyzing the genotype, we can get the individual's route in the maze to get the phenotype. A genotype consists of 128 genes, which have three modes of action - forward, clockwise rotation and counterclockwise rotation.

4.1.2 Working flow

First, do the following for a "chromosome" array of length 128:

We treat every two genes as a group. If the first number is 1, then +1. If the second number is 1, then +2. For example: [0,0] means 0; (1,0) means 1; [0,1] means 2; [1] means 3. Now we get an array of 64 bits in length, represented by 0, 1, 2 or 3.

Then, make the following judgments on the elements in the 64-bit array:

If it is 0, this person will not move; if the individual does not reach the exit and does not complete the maximum number of moves, the following judgment will continue:

If it is 1, this person will take a step forward. The direction of motion depends on the direction it is facing. If it is facing the wall or beyond the scope of the map, it cannot continue to move.

If it is 2, the object rotates 90 degrees clockwise and does not move forward.

If it is 3, the object is rotated 90 degrees counterclockwise and does not move forward.

In addition, which value to choose in a 64-bit array depends on the perception of the surrounding environment.

If there is a wall in front, +1; if there is a wall on the left front, +2; if there is a wall on the right, +4; if there is a wall on the left, +8; if there is a wall on the right, +16; if there is a wall behind, + 32. This will give you an integer [0,63].

4.1.3 Conclusion

Therefore, we successfully mapped the genetic map of the mobile route. Two key points are that the algorithm is highly fault tolerant. In the 128-digit gene sequence, if some sites are mutated, it does not necessarily lead to phenotypic changes; unlike "random movement", only the two types of "genotype" and "environment" and "gene expression" are the most match.

4.2 Calculation Algorithm of Fitness

Fitness is calculated based on the scores obtained by each individual stepping on the correct path. Each individual has a "fitness" attribute to store his score. In the beginning, the value of the score is 0.

We use a different number to represent each square in the entire maze. 2 means the entrance and 4 means the exit, which is unique in the maze. 0 means the road, but this way does not lead the individual to the exit. If an individual walk to a position with 0, he can move on, but he won't get a score. 1 means obstacles, and if a person encounters an obstacle, he cannot move on. 3 represents the right way. It can lead a person to the exit. If an individual goes to a square with 3 for the first time, he will get 1 point.

4.3 Elimination Algorithm

4.3.1 Overview

For individuals, elimination is based on three aspects:

1) By accident, young or with high fitness individuals may be eliminated;
2) Individuals who are too old may be eliminated;
3) Individuals with low fitness may be eliminated;

When selecting the individuals to be eliminated, we make the following judgment: first traverse the entire population, get the individual set of "age" greater than 10 years old, and put it at the end of the entire group list; according to the fitness, sort the remaining individuals in descending order.

Place these young individuals in front of older individuals and divide the entire list into two parts of the same length. Finally, the probability of death in the first part is set to 20%, and the probability of death in the second part is 80%.

4.3.2 Purposes

This can achieve the following purposes:

1) Older individuals and individuals with poor constitution are more likely to be eliminated, which is conducive to the evolution and development of the population;
2) Since we don't know the genetic relationship between new individuals and parents, we need to ensure that new individuals can compete fairly with their parents. If the gene of the newborn individual is superior to the parent's gene, we can guarantee to some extent that these excellent genes are preserved, so that the entire population can be evolved.

4.4 Parents selection algorithm

4.4.1 Purposes

All individuals who reach the age of childbearing can breed offspring. Individuals with high fitness values have a higher chance of successful reproduction.

4.4.2 Working Flow

1) A certain number of individuals are randomly selected from the population and then sorted according to the value of fitness.
2) The individuals with the highest fitness are selected as the parents from the selected individuals. In this process, the age of the individual must meet the requirements.
3) Obviously, this algorithm needs to be called twice to get both parents.

4.5 Crossover algorithm

4.5.1 Purposes

Hybridization is the process of individual reproduction. Observe the entire population and perform the following operations for each individual.

4.5.2 Working Flow

1) Determine whether this individual is dead. If he dies, a new individual is created to replace it. Select parents randomly according to the "parents selection algorithm;
2) Generate a random number between [0,127] as the starting point of gene crossover and exchange the gene fragments of the father and mother to obtain a new gene sequence.
3) Finally, through the crossover to get a new individual, let the new individual replace the original dead individual.

4.6 Mutation algorithm

4.6.1 Overview

Mutations are a genetic level of variation rather than an individual level of variation.

4.6.2 Purposes

The genetic sequence of each individual may mutate over time throughout the population. However, this mutation does not replace the entire gene sequence with a new gene sequence, but some sites on the gene sequence are randomly replaced, and substitution of these sites does not necessarily result in a phenotypic change. In addition, the probability of a genetic mutation is relatively low, with a probability of only 5 percent. If it is initially 1, it will be replaced with 0. If it is 0, it will be replaced with 1.

Therefore, genetic changes at the individual level are achieved, which is more in line with the actual situation.

4.7 Evolve Algorithm

4.7.1 Overview

The evaluation algorithm is the integration of the above algorithms. The complete reproductive process includes:

4.7.2 Working Flow

1) Increase of age
2) Judge if an individual is dead and produce a new individual. Among them, new individuals are produced through crossover and mutation.
3) Calculate the fitness value of the population for comparison.

5 Parallel processing

First of all, we adopt 128 genotypes that correspond to different individual movements by coding. We divide each bucket into n buckets, each of which can be different. We select the organisms in parallel with other buckets based on the gene expression. When all the barrels are completed, we will merge into the individual's gene expression, which will get a special path to this separate movement.

In addition, we also use parallel processing to sort the fitness individuals in the algorithm, which is divided into three parts: young people, adults and the elderly. We use this parallel processing to sort the fitness.

6 Data Analysis

In this project, we used preset graphs to test our genetic algorithm. The size of the maze is 18 * 18. In the genetic algorithm, our initial population value is 1000 and the initial generation value is 1000. We assume that each goat will die after ten generations, and that a newly produced goat must go through one generation before mating.

In genetic algorithms, we rank all individuals in each generation, placing individuals with high adaptability in the first half of the population, and individuals with low or older adaptability in the second half of the population.

In the first case, we set the death probability in the upper half of the population to 20% and the death probability in the lower half to 80%. We get iteration results as shown below.
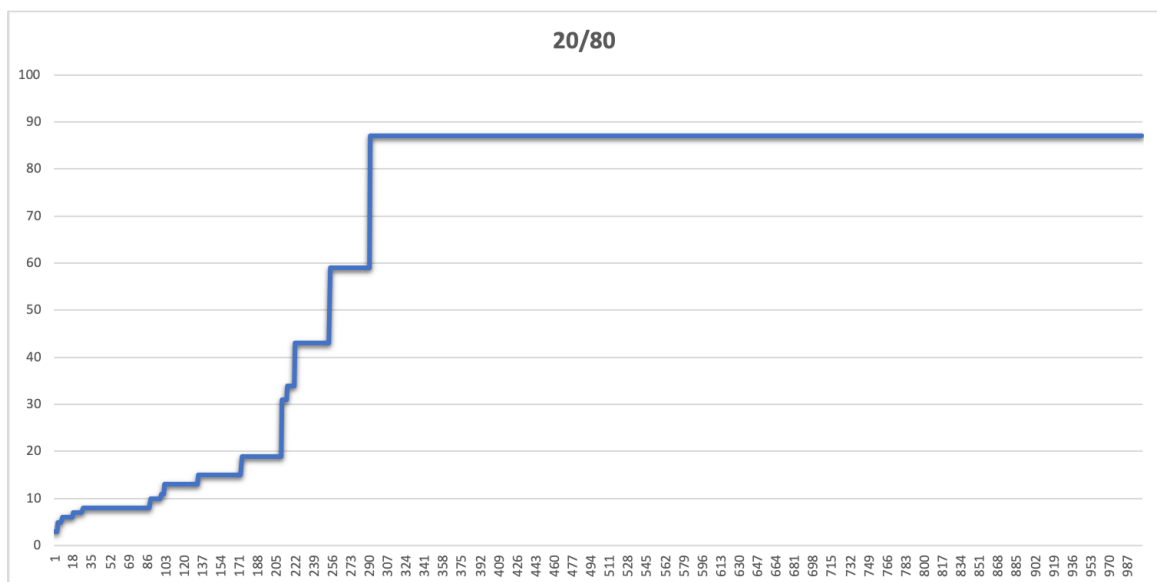

We found that this population reached the limit of evolution around the 300th generation, but we set the generation to 1000. In this case, the population completed the evolution process prematurely, which is not in line with the actual situation. Then we should change the individual's probability of death to make the simulation process more realistic.

When we changed the probability of death of excellent individuals in population to 25%, and changed the probability of death of low-adaptive individuals and old individuals in population to 75%, we obtained new figure as follows:

25/75
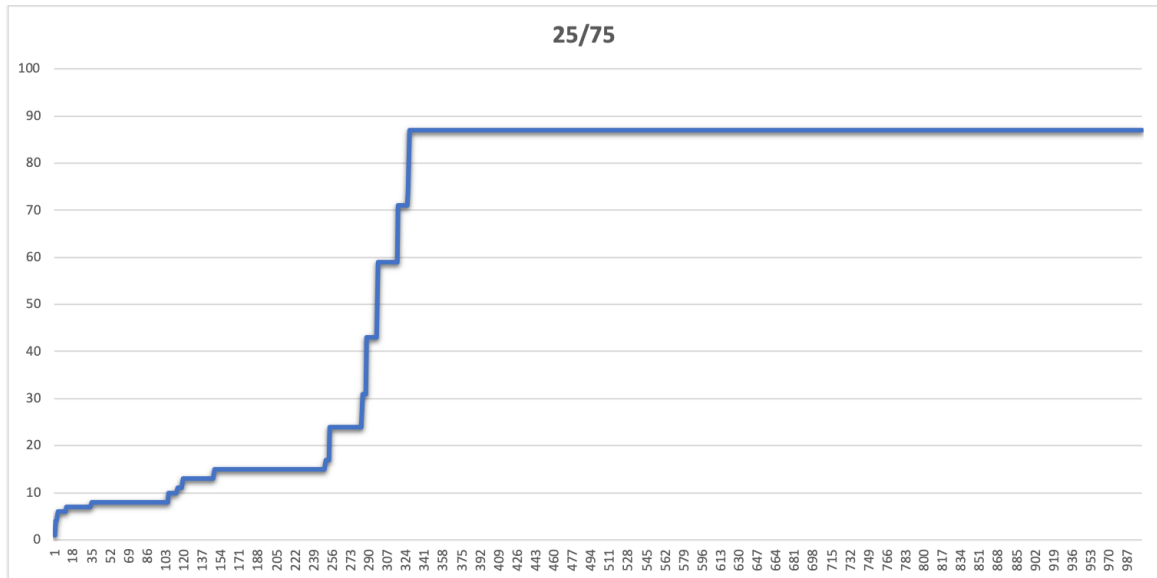
We found that the population in this chart shows that the population reached the limit around 320th generation. This result is not realistic either, so we changed another pair of the probabilities to 30% and 70%.

30/70
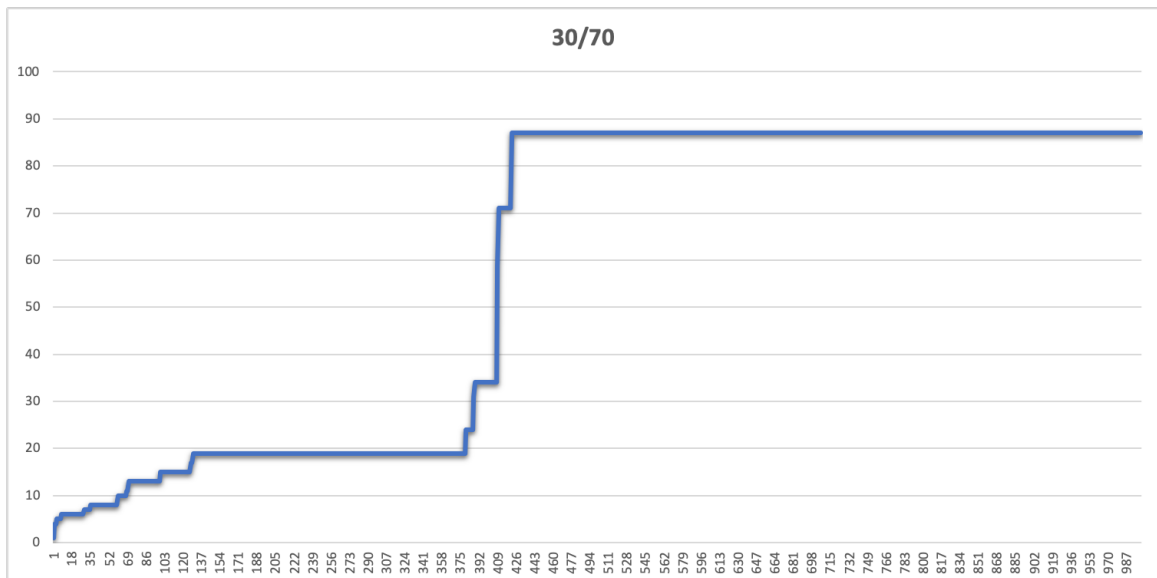
From the chart above, we can see that the evolution of population is gradually rising and does not reach the limit of evolution too fast, and at the end of the generation, the individual's evolutionary level tends to be stable and stable within a certain

interval. Such a population change process is in line with the actual situation and is in line with our expectations. As you can see from the picture:

From 1-130 generations: Individuals in the population are the most initial state, and their environmental fitness is very low.

Individuals in the 130-380 generations: population have improved their environmental fitness but are still at a lower level and the evolutionary process continues.

From 380-420 generations: Individuals in the population suddenly changed dramatically, and their environmental adaptability increased dramatically. During this time, the genes began to transform into genes that adapt to the current environment and are reflected in phenotypes.

From the 420-1000 generations: the majority of individuals in the population's environmental adaptability began to stabilize, and more and more individuals' adaptability reached about 90%.

7 Conclusion

Through this genetic algorithm, it can be concluded that during a certain time, the population always fluctuates within a certain range. So, we define different periods based on the degree of evolution. They are "primitive society" (new period), "medium-sized society" (development period) and "peak society" (peak period). These populations correspond to a portion of the result, and we can find the optimal period of the population through each phase.

In the primitive society. During this time, we generated a set of random chromosomes, and there is no direct relationship between each individual. Almost all individuals have low fitness, which means that it is difficult to reach the end of the map in primitive societies and possible to go die. Since then, after 50 generations of development, it has entered the developing period. The dominant genes and individuals constantly interact and grow rapidly. After a certain period, the evolution did not increase significantly and entered the bottleneck period.

In the medium-sized society, better genotypes are produced, and the suitability is also increasing. The group began to fluctuate. As the gene quality increases, the dominant individuals continue to increase, and the population has entered the peak of today's society. However, for a long time, the optimal genotype has not been produced, and the population has begun to have bottlenecks. Compared with the statement of the primitive society, the individuals of the medium-sized society can already get more correct squares.

In the peak society. Due to the accumulation of some high-quality genes in the early stage, the adaptability of newborn individuals has reached a peak. This is the cycle we want. After, there was no better gene production, and the adaptability of the population remains in around 90 points. Society has entered a period of stability. At this time, individuals with higher fitness appeared, and excellent individuals continued to increase. Compared with the middle society, the individuals can go to more right squares.

8   Testing

In our unit tests, we tested several different aspects of our code to check if there has a problem. Firstly, we tested the size of the genotype array as we require our genotype to match our 128-digit number to record our phenotype. The second is whether new individuals can be produced and the attributes in them are not empty. After that, we tested the fitness and tested whether the environmental suitability was in line with our expectations. Then we tested whether the death algorithm can get the effect it needs. Then we tested whether the crossover algorithm can get the corresponding effect and tested the mutation and age algorithm as well. In contrast, we can see if our algorithm meets our expectations.