1. **Introduction**
   1. **Genetic Algorithm**

A genetic algorithm is a search heuristic that is inspired by Charles Darwin’s theory of natural evolution. Similarly, this search technique is also used in computing and Artificial Intelligence. The critical steps in this search heuristic are mutation, crossover, reproduction and selection. And this algorithm reflects the process of natural selection and evolution where the fittest individuals in the population will be selected for reproduction in order to produce offspring of the next generation. This selection process keeps on iterating and at the end, a generation with the fittest individuals will be found.

* 1. **Innovation**

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 can survive in the zebra migration. This is a natural and random process, and the set of solutions is very large, so we can use genetic algorithms to simulate this process and find the optimal solution.

* 1. **Problem realization**

In the process of constructing the zebra migration 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. Experiment to find the final map.

1. **Question Description**
   1. **Overview**

Around June each year, the grass in the Tanzanian steppe is gradually consumed, and food is becoming less and less. For food, the animals on the prairie will travel over 3,000 kilometers and stage the most spectacular animal migration on the planet. A million-headed wildebeest, hundreds of thousands of zebras form a massive team, from the Serengeti Reserve in Tanzania to the north, ending in the Masai Mara National Park in Kenya, trekking over 3,000 kilometers. At the same time, on the way, not only the lions and leopards ambush the grasslands, but also the hyenas that may be present at any time and the crocodiles gathered on the narrow Mara River. These carnivores are ready to share the upcoming feast. Of the millions of migratory teams, only 30% of the lucky ones are able to reach the destination, and with them arrive there are a large number of new born on a thrilling journey.

* 1. **Concepts in problem implementation**
     1. **Individuals**

An individual means a zebra produced by its parent zebras. The production of new individuals requires the reproduction of two other individuals. They produce the offspring which inherit the characteristics of the parents and will be added to the next generation.

* + 1. **Genotype**

Genotypes are determined at the beginning of individuals’ formation. For the first generation in initial population, genotypes are randomly generated. For the offspring of the first generation, their genotypes are produced by their parents. In this project, the genotype of each will determine their decision at each step in the future.

* + 1. **Expression**

This is the mapping of genotype to phenotype. Under an individual’s gene grouping, the expression algorithm will obtain the gene sequence of a specific part and map the gene series to the trait which is movement of the zebra.

* + 1. **Phenotype**

The trait of the candidate which affect how good a solution it is. In this project, the phenotype means the direction of movement of each individual. And there are three types of movement: split forward, clockwise and counterclockwise.

* + 1. **Environment**

In this project, the entire environment includes the correct road, misleading roads and the roads blocked by wild beast. For zebras, they need to know their behavior (which is the action of zebras are determined by the genes of the individual) before taking each step in this pre-defined map. Among all types of roads in the map, the right roads can let the zebras find the exit and get scores, but the incorrect roads will hinder the movement of the zebras.

* + 1. **Fitness**

Fitness is the degree of how well the organism is suited to the environment. In this situation, the fitness is the length of the correct path that a zebra walks through, which is derived by fitness algorithm.

1. **Parameter Settings**
   1. **Group size**

1000 zebras

* 1. **Chromosome length**

Total 128 genes

* 1. **Mutation probability**

0.05

* 1. **Crossover probability**

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

* 1. **Probability of death**

Based on the zebra'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%.

* 1. **High quality group size**

The high-quality group size is 10

* 1. **Maximum number of steps**

The maximum number of steps is 300

1. **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.

1. **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 the each individuals. 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 110 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 entire group, each individual can get 24 fitness on average, but there are also some excellent individuals and individuals who need to perform.

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 still fluctuated around 67. 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 and get an average of 67 points. According to the experiment, the best individual and can get 87 points, which means he can go through from the starting point to the end.

1. **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.

