DETERMINATION OF SELECTION METHOD IN GENETIC ALGORITHM FOR LAND SUITABILITY

Asti Dwi Irfianti[1], Retantyo Wardoyo[2], Sri Hartati[2], EndangSulistyoningsih[3]

[1]Department of Information System, Faculty of Industrial Technology
University of Pembangunan Nasional "Veteran"Jawa Timur, Surabaya, Indonesia, [2]Department ofComputer Science and Electronic
Faculty of Mathematic and Natural Science, Gadjah Mada University, Yogyakarta,Indonesia
[3]Faculty of Agriculture Gadjah Mada University, Yogyakarta Indonesia
E-Mail: asti.ilkom2000@gmail.com

ABSTRACT

Genetic Algoirthm is one alternative solution in the field of modeling optimization, automatic programming and machine learning. The purpose of the study was to compare some type of selection methods in Genetic Algorithm for land suitability. Contribution of this research applies the best method to develop region based horticultural commodities. This testing is done by comparing the three methods on the method of selection, the Roulette Wheel, Tournament Selection and Stochastic Universal Sampling. Parameters of the locations used in the test scenarios include Temperature = 27°C, Rainfall = 1200 mm, hummidity = 30%, Cluster fruit = 4, Crossover Probabiitiy (Pc) = 0.6, Mutation Probabilty (Pm) = 0.2 and Epoch = 10. The second test epoch incluides location parameters consist of Temperature = 30°C, Rainfall = 2000 mm, Humidity = 35%, Cluster fruit = 5, Crossover Probability (Pc) = 0.7, Mutation Probability (Pm) = 0.3 and Epoch 10. The conclusion of this study shows that the Roulette Wheel is the best method because it produces more stable and fitness value than the other two methods.

Keywords: Method Selection, Genetic Algoritm, Roulette Wheel, Tournament Selection, Stocastic Universal Sampling

INTRODUCTION

Genetic Algorithm (GA) is a stochastic optimization algorithms block adopting Darwin's theory. GA has been applied to solve problems such as optimization and modeling, automatic programming and machine learning (Goldbeg, 1999). There are six basic GA need to be considered which are the issue of the chromosome (genome) representation, the selection function, genetic operators such as mutation and crossover, the initial population creation, termination criteria and the evaluation function *(*Zadeh and Ghazalian, 2010*)*.

Search on GA begins with a set of the population *(*Blickle and Thiele, 1995*)*. The population consists of chromosomes. Each chromosome is a problem-solving solution. Populations will produce new offsprings which are expected to be better than the previous population. Good natural population will have an opportunity to continue to be developed in order to produce good offspring.

In order for GA to run, first, fitness function stating the success rate of a population needs to be defined. By calculating the fitness function, then it will be determined population to be set to produce the next generation. This process is commonly referred to as the selection process. This process is one of the stages set out in an iterative process *(*Blickle and Thiele, 1995).

The selection process is a process that plays an important role in GA *(*Huang and Wang*,* 2011*).* Examine the importance of the selection process will be used as a reference for determining the chromosomes quality. The best generation is the generation that can move on to the next selection method (Mitchell, 1997).

Selection process GA is defined as a fitness function which states the success rate of the population, calculation of fitness function will get the population to maintain dan

produce the next generation. This process is a stages series of the iteration process.

There are various techniques that can be used in the selection process on the issue. Selection techniques used depends on the problem to be solved. There are several methods for parent selection, such as the rank-based Centre Task, Roulette Wheel Selection, Stochastic Universal Sampling, Selection of local, cutting Selection, Selection Tournament *(*Mohebi *et al.,* 2009*)*. In this paper, the selection process used Roulette Wheel Selection (RWS), Selection Tournament and Stochastic Universal Sampling (SUS). Parameters used in the process selection called fitness. Fitness is used to determine how well the chromosome will survive *(*Jebari and Madiafi, 2013*)*.

METHODOLOGY

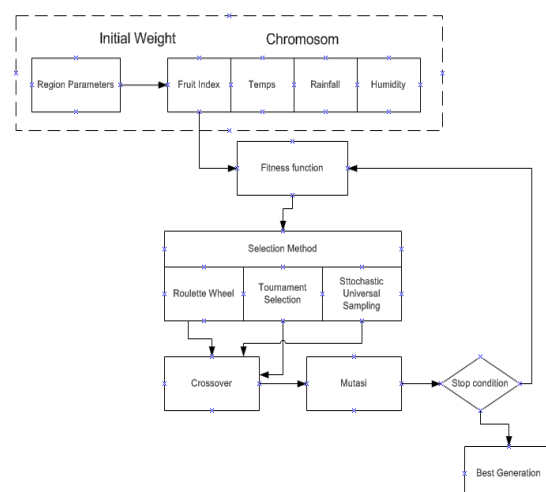Block Diagram used in this research is as follows:

Figure-1. Block Diagram Selection Method

Genetic Algorithm

Genetic Algoritm was proposed by John Holland in early 1970s, it applies some of natural evolution mechanism such as *crossover, mutation,* and survival of the fitness to optimization and machine learning. GA provides search method efficiently working on population, and has been applied to many problems of optimization and classification (Goldbeg, 1999; Mohebi, 2009). General GA processes are as follows:

1) Initiate the population of genes
2) Calculate the fitness of each individual in the population
3) Reproduce the individual selected to form a new population according to each individual's fitness
4) Perform crossover and mutation on the population
5) Repeat step (2) through (4) until some condition is statisfied.

```
Algorithm:  GA(n, χ, μ)
// Initialise generation 0:
k := 0;
P_k := a population of n randomly-generated individuals;
// Evaluate P_k:
Compute fitness(i) for each i ∈ P_k;
do
{    // Create generation k + 1:
     // 1. Copy:
     Select (1 − χ) × n members of P_k and insert into P_{k+1};
     // 2. Crossover:
     Select χ × n members of P_k; pair them up; produce offspring; insert the offspring into P_{k+1};
     // 3. Mutate:
     Select μ × n members of P_{k+1}; invert a randomly-selected bit in each;
     // Evaluate P_{k+1}:
     Compute fitness(i) for each i ∈ P_k;
     // Increment:
     k := k + 1;
}
while fitness of fittest individual in P_k is not high enough;
return the fittest individual from P_k;
```

Figure-2 Pseudocode GA (Chusadama et al., 2011)

Crossover operation swap part of genetic bit string within parents. It emulates just as crossover of genes in the real world that descendants are inherited characteristics from both parents.

Mutation operation inverts some bits from whole bit string at very low rate. In real world we can see that some mutants come out rarely. Fig 1 shows the way of applying crossover and mutation operation to genetic algorithm. Each individual in the population evolves to getting higher fitness generation by generations.
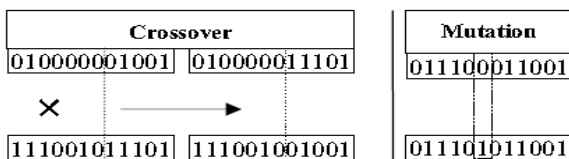
Crossover		Mutation
010000001001	010000011101	011100011001
×	→	
111001011101	111001001001	011101011001

Figure-3. Crossover and mutation

Roullete Wheel

a. Choose a solution that is useful for recombination.
b. The Fitness function provides fitness for a solution chromosomes. The fitnesss level to connect with the probability of selection of each individual chromosome.
c. If Fi is fitness for the individual i in the population, the possibility of elected if pi = fiΣNj = 1fj, where N is the number of individuals in the population.
d. The proportion is assigned to each of the wheels based on the choice of fitness. By dividing the total Gym fitness of all so that the normalization of the 1st choice is made randomly.
e. Candidate solutions with higher fitness is less likely to be eliminated, the selected fitness will eliminate the presentation of the candidates who are weak. Fitness proportionate makes it possible for weak solutions to survive. There is a benefit if the solution is weak since it may include several components that may be useful to follow the process of recombination.

Tournament Selection

```
Algorithm 1: (Tournament Selection)

Input: The population P(τ) the tournament size t ∈ {1, 2, ..., N}
Output: The population after selection P(τ)'

tournament(t, J_1, ..., J_N):
     for i ← 1 to N do
          J'_i ← best fit individual out of t randomly picked
                  individuals from {J_1, ..., J_N};
     od
     return {J'_1, ..., J'_N}
```

Figure-4. Tournament Selection (Tobias et al., 1995)

Often tournament held between two individuals (binary tournament) but in general it can change group size t which is also called the tournament size. Outline of the algorithm can be seen from the tournament selection which is implemented very efficiently.

Stochastic Universal Sampling

Parent chosen based on the fitness value, the greater the value of fitness, the greater the likelihood is selected. Stochastic Universal Sampling uses a single random value. Stochastic Universal Sampling assessed from a small random number and select the next candidate from the remaining population. Here below is the algorithm of selection methods Stochastic Universal Sampling:

```
Algorithm 6: (Stochastic Universal Sampling)
Input: The population P(τ) and the reproduction rate for each
       fitness value R_i ∈ [0, N]
Output: The population after selection P(τ)'

SUS(R_1, ..., R_n, J_1, ..., J_N):
     sum ← 0
     j ← 1
     ptr ← random[0,1)
     for i ← 1 to N do
          sum ← sum + R_i where R_i is the reproduction rate
                  of individual J_i
          while (sum > ptr) do
               J'_j ← J_i
               j ← j + 1
               ptr ← ptr + 1
          od
     od
     return {J'_1, ..., J'_N}
```

Figure-5. Stochastic Universal Sample (Tobias et.al., 1995)

RESULTS

In this paper, we have compared cluster horticulture with selection method in Genetic Algorithm. The selection method analysis are Roulette Whell, Stochastic Universal Sampling and Tournament Selection.

Parameters used in figure 6 Temperature = 27ºC, Rainfall = 1200 mm, humidity = 30%, Cluster fruit = 4, Crossover Probability (Pc) = 0.6, Mutation Probability (Pm) = 0.2 dan Epoch = 10.

The graph in figure. 6 shows fitness value from Roulette Wheel method selection. This graph below shows the difference appearance of horticulture. It can also be seen that Carrots appears more than others at 60%. That means Carrots are the most suitable plants in this location.

Figure-6. Roulette wheel (a)

The graph in figure 7 show fitness value from Tournament Selection.

Figure-7. Tournament Selection (a)

In figure 7, percentages of Carrots is 50% and chilli is30%. That means Carrots are the most suitable plants in this location.

In the figure 8 it can be seen in the graph that Orange has a frequency of occurence appears 5 times, followed by Chilli appears 3 times.

Parameters used in the experiments 2 is Temperature = 30ºC, Rainfall = 2000 mm, humidity = 35%, Cluster fruit = 5, Crossover Probability (Pc) = 0.7, Mutation Probability (Pm) = 0.3 and Epoch 10. The study produced several experiments as follows: The fitness value generated Roulette Wheel method on iteration 1 to 10 to produce the value of fitness and commodities in sequence as follows:

Figure- 8. Stochastic Universal Sampling (a)

Figure-9. Roulette Wheel (b)

The graph in figure 9 shows fitness value from Roulette Wheel (b) method selection. However, this figure was considerably percentages from Papaya 30% and Rambutan 20%. That means Papaya are the most suitable plants in this location.

Figure-10. Tournament Selection (b)

In the figure 10 seen in the graph that the frequency of occurence Broccoli appears 7 times, followed by Rambutan appears 2 times.

The graph in figure 10 shows fitness value from Tournament Selection (b) method selection. From the figure, it is worth noting the high percentages of Broccoli at 70% and Rambutan at 20%.. That means Brocolli are the most suitable plants in this location.

Figure-11. Stochastic Universal Sampling (b)

The graph in figure 11 shows fitness value from Stochastic Universal Sampling (b) method selection. Among others, Brocoli is better at 50% followed by Passion Fruit at 20%. That means Brocolli are the most suitable plants in this location

DISCUSSIONS

The Roulette Wheel (a) shows that after 10 iterations, Carrots and Orange often appear frequently. It means Carrots and Orange are the most suitable plant parameters. In the Roulette Wheel (b), it can be analyzed that after 10 iterations plants that have a higher frequency of appearance are Orange and Chilli. In the Stochastic Universal Sampling (a), it can be seen that after 10 iterations, plants that appears often are Orange and Chilli. Of three trials on crop cluster 4, it can be determined that the priority sequence is Carrots, Chili, and Lemon.

Result of the comparison of the three selection methods shown in table 1 and figure 12.

Table-1. Comparison of three methods Selection in Cluster 4.

No	Roulette Wheel	Tournament Selection	Stochastic Universal Sampling
1	0.012	0.012	0.012
2	0.004	0.012	0.012
3	0.006	0.014	0.014
4	0	0.014	0.014
5	0.002	0.434	0.248
6	0	0,434	0.248
7	0.002	0.355	0.248
8	0.002	0.434	0.061
9	0.006	0.303	0.121
10	0.002	0.434	0.434

In table 1, there seems to be a little resemblance between the fitness value of Tournament Selection with Stochastic Universal Sampling selection.
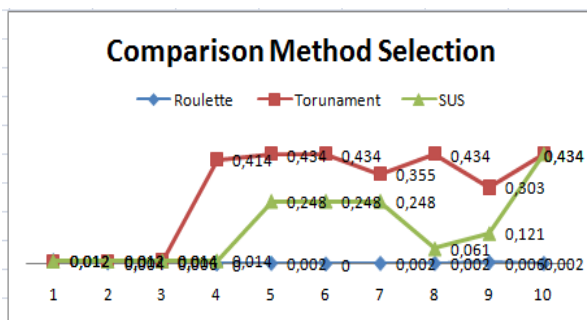

Figure-12. Comparison Fruit Cluster 4

The graph in Figure 12, illustrates comparison of three selection methods in Genetic Algorithm.. The smallest fitness value is generated by the Roulette Wheel selection method.

Table-2. Comparison of three methods selection in Cluster 5

No	Roulette Wheel	Tournament Selelction	Stochastic Universal Sampling
1	0	0,008	0,008
2	0	0,100	0,100
3	0,008	0,002	0,022
4	0	0,100	0,100
5	0,061	0,100	0,100
6	0,164	0,100	0,100
7	0,154	0,100	0,100
8	0,051	0,100	0,100
9	0,004	0,100	0,100
10	0,094	0,080	0,080

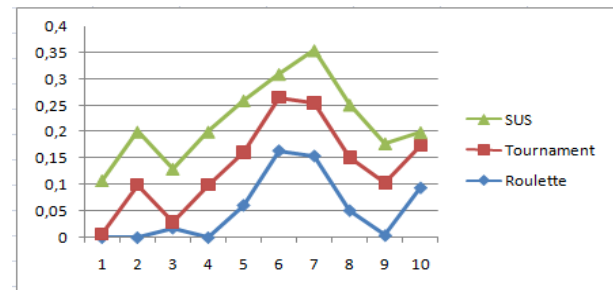The Table 2 illustrates are comparison fitness values from methods selection.

Figure-13. Comparation Fruit Cluster 5

The graphs illustrates the comparison of the three method of selection. The smalllest fitness value is generated by the Roulette wheel selection method.

CONCLUSIONS

1. The best selection method in this experiment is Roulette Whell because this selection method has small fitness values and is stable.
2. Fitness value which has been generated in each process in the genetic algorithm shows the index value of fruit.
3. Fruit index which is appeared on each iteration has demontrated suitability with the location.

ACKNOWLEDGMENTS

REFERENCES

Bolboaca S.D., Jantschi L., Balan M.C., Diudea M.V., Sestras R.E.,2010. State of Art in Genetic Algorithms for Agriculture System. Notula Botanica Horti Agrobotanic. Cluc Napoca. 51-63. ISSN 0255-965X. Electronic 1842-4309.

Blickle T. and Thiele L. 1995. A Comparison of Selection Schemes used in Genetic Algorithms.

Chudasama C., Shah S.M., Panchal,M..2011. Comparison of Parents Selection Methods of Genetic Algorithm for TSP. ,Proceeding published by International Journal of Computer Application (IJCA). International Conference on Computer Communication and Network CSI-COMNET-2011.

Goldberg D.E. and Kalyanmoy D. 1999. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms.

Goldbeg D.E. 1999. Genetic Algoritm in Search Optimization and Machine Learning. Addison-Wesley Publishing Co.inc.

Mohebi E., Bouyer A., Karimi M. 2009. An Effecient Clustering of the SOM using Rough set and Genetic Algoritm. GCC Conference & Exhibition, 5th IEEE. 17-19 March 2009. Page 1-5. ISBN 978-1-4244-3885-3.

Mitchell T. 1997. Machine Learning. Genetic Algorithm Handout. Mc Graw-Hill

Jebari K. and Madiafi M. 2013. Selection Methods for Genetic Algorithms. International Journal Emerg Science. IJES, 3(4). 333-344. Desember 2013. ISSN 2222-4254.

Huang G and Wang L. 2011. Hybrid Neural Network Models for Hydrologic Time Series Forecasting based on Genetic Algorithm. Fourth International Joint conference on Computational Sciences and Optimization, 97-0-7695-4335-2/11. DOI 10 1109/CSO 2011.146. IEEE Computer Society.

Zadeh E. A. and Ghazalian R. 2010. Modulation Classification Using Genetic Algorithm and Radial Basis Neural Network Based on the HOS. Digital Content Multimedia Technology and its Aplication (IDC). 6th International Conference, 16-18 Aug 2010, Page 375-378.. ISBN: 978-1-4244-7607-7.