Taxonomy Genetic Algorithm For Implementation Partially Mapped Crossover In Travelling Salesman Problem

SM Hardi\textsuperscript{1,}\textsuperscript{*}, M Zarlis\textsuperscript{2}, S Effendi\textsuperscript{3}, M S Lydia\textsuperscript{4}

\textsuperscript{1,}\textsuperscript{2,}\textsuperscript{3,}\textsuperscript{4}Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan 20155 USU, Medan, Indonesia.

E-mail: vani.hardi@usu.ac.id

Abstract. Genetic algorithm has various processes that are carried out to get optimal results in solving the problem of the travelling salesman problem (TSP) as known as combinatorial NP problem. Salesman given a map that he has accomplished all cities only once by minimizing the total of distance and he has to return to the first city. By describing the taxonomy in genetic algorithm, it can avoid confusion in understanding the various classifications that exist in the genetic algorithm operator. This paper explains one of the parts from genetic algorithm is crossover. Crossover being an important step in genetic algorithm. Commonly mechanism crossover replaces two selected chromosome which part from two parent's individual randomly position, generate random number interval 0-1. Crossover occur for each individual by determining the probability of the crossover. If the probability crossover is smaller than random number, there is no crossover process. Partially mapped crossover is part of the taxonomy of genetic algorithms whose implementations can be applied in a variety of problem solving including in the travelling salesman problem. The test results of this research using pc 0.25 and pm 0.1 with tsp eil51 data, it was found that the optimal route average value was 4069.34 and the best fitness average was 2.46E-04, while for the eil76 test data the optimal route average value was 6844.4 and the average best fitness value of 1.46E-04.

1. Introduction

Travelling salesman problem (TSP) as known as combinatorial NP problem [1]. Salesman given a map that he has accomplished all cities only once by minimized the total of distance and he has to return to the first city [3]. Travelling salesman problem divided into asymmetric and symmetric. In Symmetric TSP, the distance between cities X and Y are equal with distance between cities Y and X. Whereas Asymmetric TSP, there are different distance between two cities [2]. There had been so many methods and algorithm for solving travelling salesman problemin some decades, such as ant colony [4], neural network[5], swarm intelligence [6], and genetic algorithms [3,7,8]. Travelling salesman problem with a large search space can be solved using genetic algorithm for obtain optimized solution[9]. Various literature using genetic algorithm held by many researchers for solving travelling salesman problem. Genetic algorithm based on principle biological evolutionary process which originated work by John Holland and Goldbiergh which inspired theory of Charles Darwin. Based on biological process, genetic algorithm having term which like nature biological process. Create an initial population, mutation, and crossover are terms in genetic algorithm. They are called operator in genetic algorithm[3]. The following
is taxonomy on genetic algorithms that can be seen in Figure 1. The purpose of taxonomy is to convey something that contains various types of information to avoid similarities or differences in giving rules to a subject. In figure 1 can be seen a collection of parts that exist on the genetic algorithm[13][14].

![Figure 1. Taxonomy of genetic algorithm](image)

Based on figure 1 can be seen genetic algorithm consist of encoding, function offitness, selection process, crossover and mutation. This paper explain some of the parts from genetic algorithm is crossover. Crossover being important step in genetic algorithm. Commonly mechanism crossover replace two selected chromosome which part from two parent’s individual randomly position, generate random number interval 0-1. Crossover occur for each individual by determining the probability of the crossover. If the probability crossover is smaller than random number, there is no crossover process. If the probability is greater than or equal to the random number, then the crossover operator is applied[11]. Various crossover has proposed by many researchers. There was one-point crossover, two-point crossover, order crossover, cycle crossover and partially mapped crossover. Partially mapped crossover (PMX) modified by formulatwo-point crossover, the main things PMX is the crossing of two points plus some additional procedures[10]. This research proposed implemented PMX in travelling salesman problem to get route optimal using data benchmark of tsp.

2. Method

In genetic algorithms there are various stages or processes carried out. This research focuses on implementing one of the processes namely the crossover process for solving travelling salesman problem by getting best fitness. Genetic algorithm having parameter which used to controlling operator for optimizing result. They are population size (N), probability of crossover (Pc) and probability of mutation (Pm). For knowing, number of population including in the crossover process in each generation is Pc x N, which value of Pc range interval [0..1]. Number of populations in every generation which possibility occur mutation is Pm * N * L.

Based on figure 1, can be seen that there were so many method in crossover process. In this research used is partially mapped crossover with a pc(probability crossover) value of 0.25. Research data is using eil51 and eil76 tsp. Crossover is process crossing gene two parent of chromosomes to produce a child chromosome that called offspring or protochildren. Protochildren/offspring will inherit some of the characteristics of the parent. Following are procedure in partially mapped crossovers:
1. Determine position of two chromosome in a random rule with same position. substrings in these two positions are called mapping regions.

2. Swapping two substrings between parents to produce protochildren.

3. Determine mapping relationship between two mapping regions.

4. Determine offspring chromosomes refers to mapping relationships.

3. Result and Discussion

Genetic algorithm consists of several processes or stages. The following is a flowchart overview of genetic algorithms in figure 2. The data used is a benchmark data taken from TSPLIB where TSPLIB is a library of sample data for CSR problems from various sources and various types of CSR problems. The data support EDGE_WEIGHT_TYPE. EUC_2D, which is coordinate position with 2-dimensional Euclidian format. For each type of EUC_2D the coordinate point must be known for each point. In 2-dimensional Euclidean, equation of the distance between points as to b as follow: 

$$d_{(a,b)} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$$

whereas x_a means coordinate x in location city a, x_b means coordinate x in location city b, y_a means coordinate y in location city a, y_b means coordinate x in location city b.

![Figure 2. Flowchart of Genetic Algorithm](image-url)
3.1. Representation chromosome
First process in genetic algorithm are representation chromosome into path presentation. Chromosomes / individuals are collections of numbers that represent the position of cities in a series (sequence), means a chromosome / individual represents a sequence of cities visited by a salesman. So if one chromosome is $C_1 = (x_1, x_2, x_3, ..., x_n)$, it means the salesman moves from the city numbered $x_1$ to $x_2$...$x_n$.

3.2. Generate Population
Second process is random population generation. Generating a number of individuals or chromosomes randomly or through certain procedures. The size of the population depends on the problem to be solved and type of genetic operator that will be implemented. After population size is determined, then initial population is generated. If selected population size smaller, level of exploration global search space will be limited, convergence being faster. If population size bigger, time will be wasted because of large amount of data and time to convergence will be longer [10]. Formulation of Population Initialization Generation using Random Number Generation: random $(n_k, N)$, where N= population size [12]. For example: random (8,5) means that 5 chromosomes are generated in 1 population, where in 1 chromosome there are 8 genes.

3.3. Selection method: Roulette Wheel Selection
Selection method used in this research is Roulette Wheel Selection. Roulette Wheel Selection method based on value fitness of each individual, the greater value fitness of individuals chance being selected as parents for the next process. For knowing value fitness each individual, calculated fitness using this formula: Fitness = $1/D$, where D = total distance. Based on formula, can be conclusion that greater value of fitness then smaller value of distance. As known value of totally fitness from each individual, selecting chromosome will be select as parents.

3.4. Partially Mapped Crossover (PMX)
Here, processing of partially mapped crossover for recombination and crossing gene from chromosomes parents to produce protochildren/offspring can be seen in flowchart in figure 3.
Figure 3. Flowchart Partially Mapped Crossover (PMX)

For example:
Choose a position to specify substrings randomly
Parent 1: 3 8 6 | 721 | 4 5
Parent 2: 2 1 7 | 456 | 3 8
Form of mapping:
3-2-4
8-1-5
6-7
Produce: Protochildren 1: 386 | 628 | 45
Protochildren 2: 217 | 387 | 38

Test was carried out as many as 500 generations with a crossover probability of 0.25 and a mutation probability value of 0.1 by testing as much as 5 times to get the best fitness by taking the average value. Test results can be seen in Table 1 using data tsp eil51.

Table 1 shows best fitness value of PMX where intersection point is located on the chromosome in the same position, average value of best fitness after experiencing testing is 2.46E-04 and average value of the optimal route is 4069.34
Table 1. Experiment data tsp eil51.

| No | Best Fitness       | Generation | Optimal Route |
|----|--------------------|------------|---------------|
| 1  | 2.48040233869967E-04 | 444        | 4031.60400390625 |
| 2  | 2.57765273253215E-04 | 341        | 3879.49853515625 |
| 3  | 2.30385223667717E-04 | 380        | 4340.55615234375 |
| 4  | 2.61083657807542E-04 | 481        | 3830.18994140635 |
| 5  | 2.34468064888579E-04 | 444        | 4264.97314453125 |

Average Best Fitness : 2.46348490697404E-04
Average Optimal Route : 4069.3435546877

Table 2. Experiment data tsp eil76.

| No | Best Fitness       | Generation | Optimal Route |
|----|--------------------|------------|---------------|
| 1  | 1.47513033259075E-04 | 484        | 6779.06201171875 |
| 2  | 1.5314634247677E-04  | 454        | 6529.7021484375  |
| 3  | 1.50838749620417E-04 | 453        | 6629.59619140625 |
| 4  | 1.4027069617296E-04  | 486        | 7129.0725390625  |
| 5  | 1.39766295467041E-04 | 368        | 7154.80078125    |

Average Best Fitness =1.463070234023012E-04
Average Optimal Route= 6844.44677734375

Based on Table 2, average value of best fitness after experiencing 5 tests is 1.46E-04 and average value of the optimal route is 6844.4

3.5. Swapping Mutation

Next process in genetic algorithm involves mutation operator. The aim mutation operator implemented for recovering genetic that have been lost from another process as selection and crossing process. Mutations process, new individuals can be created by modifying one or more gene values in the same individual. These mutations play a role in replacing genes which lost from population to selection which allows the reappearance of genes that do not appear at the initialization of the population. Mutation operator that implemented in this research is swapping mutation. Swapping mutation is a mutation operator by exchanging one or several gene values in a chromosome. In this research using probably mutation value (pm) 0.1.

Here, steps in mutation process.

1. Count the number of genes in the population (the length of these chromosomes is multiplied by the size of the population).
2. Randomly select the gene to be mutated.
3. Determine the chromosomes of the selected gene for mutation.
4. Take 2 genes at random and swap their positions.

4. Conclusion

Partially mapped crossover (PMX) can be implemented for solving tsp. And test results of this research using pc 0.25 and pm 0.1 with tsp eil51 data, it was found that optimal route average value was 4069.34 and the best fitness average was 2.46E-04, while for the eil76 test data the optimal route average value was 6844.4 and the average best fitness value of 1.46E-04. Taxonomy
of genetic algorithms is formed to determine the classification structure that exists in genetic algorithms, each part/process can be seen so that its implementation can be applied in various fields. Mapping on partially mapped crossover affects the variation of genes produced so that it influences the average value of best fitness. For further research, the variation of PMX can be added with the location of the mapping and random chromosome positions as well as the addition of the number of cut points which are usually done using 2 cut points, then adding 4 cut points can be added.

References

[1] E. L. Lawler, J. K. Lenstra, D. B. Shmoys, The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, NY: Wiley-Interscience, New York, 1985.
[2] Mohammad Reza Bonyadi, Mostafa Rahimi Azghadi, and Hamed Shah-Hosseini. 2008. Population-Based Optimization Algorithms for Solving the Travelling Salesman Problem. https://www.intechopen.com/books. Published: September 1st 2008. DOI: 10.5772/5586
[3] Hussain, Abid et al. 2017. Genetic Algorithm for Traveling Salesman Problem with Modified Cycle Crossover Operator: Hindawi Computational Intelligence and Neuroscience Volume 2017, Article ID 7430125, 7 pages https://doi.org/10.1155/2017/7430125.
[4] M. Dorigo and L. M. Gambardella, “Ant colony system: a cooperative learning approach to the traveling salesman problem,” IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 53–66, 1997.
[5] S. Bhide, N. John, and M. R. Kabuka, “A Boolean Neural Network Approach for the Traveling Salesman Problem,” IEEE Transactions on Computers, vol. 42, no. 10, pp. 1271–1278, 1993.
[6] J. Kennedy, R. C. Eberhart, and Y. Shi, Swarm intelligence, morgan kaufmann publishers, Inc., San Francisco, CA, USA, 2001.
[7] JiaXua, LangPeia, Rong-zhaoZhub. 2018. Application of a Algoritma Genetika with Random Crossover and Dynamic Mutation on theTravelling Salesman Problem. 8thInternational Congress of Information and Communication Technology(ICICT-2018) Procedia Computer Science 131 937–945.
[8] Deng, Yong, Yang Liu, and Deyun Zhou. 2015. An Improved Algoritma Genetika with Initial Population Strategy for Symmetric TSP. Hindawi Publishing Corporation Mathematical Problems in Engineering, Article ID 212794.
[9] Riazi A, Türker U (2018) A genetic algorithm-based search space splitting pattern and its application in hydraulic and coastal engineering problems. Neural Comput Appl 30(12):3603–3612.
[10] Goldberg D.E. “Genetic algorithms in search, optimization, and machine learning”, Addison Wesley Longman, Inc., ISBN 0-201-15767-5, 1989.
[11] J McCall. Genetic algorithms for modelling and optimisation-Journal of computational and Applied Mathematics, 2005 –Elsevier.
[12] Larrañaga, P., Kuijpers, C. M. H., Murga, R. H., Inza, I., & Dizdarevic, S. (1999). Genetic algorithms for the travelling salesman problem: A review of representations and
[13] Almutfi, Saman et all, Taxonomy of bio-inspired optimization algorithms. Journal of Advanced Computer Science & Technology, 8 (2) (2019) 23-31.
[14] Nasution, M. K. M. (2017), Ontologi, Ontologi dan Taksonomi Informasi, DOI:10.13140/RG.2.2.22463.92323