K-Means Clustering and Firefly Algorithm for Shortest Route Solution based on Crime Hotspots

Nurlela Pandiangan

Master of Information System
Diponegoro University
Semarang - Indonesia

Rahmat Gerowo

Department of Physics
Diponegoro University
Semarang - Indonesia

Vicensius Gunawan

Master of Information System
Diponegoro University
Semarang - Indonesia

ABSTRACT

Routing problem becomes a problem that is often discussed. Minimize the time, vehicle, or shipping costs become the main focus. In this study, Firefly Algorithm (FA) to forming the optimal route for police patrol problems. The patrol route becomes important to be optimally formed to be effective in cracking down on crime and maintaining security. By adopting a multi-agent police patrol, this study uses two steps: the grouping of Hotspots of Criminal Case using the K-Means Algorithm to limit the Police Agent's patrol and Firefly Algorithm patrol areas as an algorithm to establish an optimal route. As a result, users still get optimal route results even with different parameter values. Based on the convergence result, the three scenarios with parameter values are bigger than other scenarios, are Alpha value: 0.3, Beta: 1,5, and Gamma: 0.45 are superior to make optimal route on two clusters A4 and A5 with total time the average execution is less that is 558862.4 microseconds.

General Terms

Shortest Route Solution Based On Crime Hotspots

Keywords

Firefly Algorithm, K-Means Clustering.

1. INTRODUCTION

Routing Problem or finding the shortest path to the problem that is often discussed, creating the shortest route that can minimize the time, vehicle or shipping costs to be the main focus in the problem [1]. Various methods have been implemented to solve the problem of routing problems, such as heuristic methods and metaheuristic methods. The heuristic method solves the problem of routing problems quickly but does not guarantee finding the optimal solution [2]. Meanwhile, metaheuristic methods are much better and flexible in finding an optimal route solution with several different criteria and a good computation time [3]. One metaheuristic method introduced by Yang in 2008 that was inspired by the way fireflies live in nature is the Firefly Algorithm (FA)[4].

Firefly Algorithm (FA) is one of metaheuristic method that can solve the optimum problem [5]. FA is able to handle optimization problems efficiently [6]. This algorithm is able to solve several problems such as problems in finding the shortest route solution, Visual tracking [7], or combined with other algorithms such as K-Means and support vector regression (SVR) for forecasting export trade of taiwan [8] and there is also only in combination with K-Means for Data clustering [9]. In this case, the FA for routing problems with case studies of the Indonesian National Police Patrol Routes of Papua Province Merauke. Patrol is one form of supervision by the police to maintain security and public order. One

effective form is to patrol the Hotspots of Criminal Case or the location of frequent crimes marked by the occurrence of criminal cases [10]. The obstacles, require the collaboration of patrol officers in order to cover the entire Hotspots of Criminal Case [11]. In this study, there are two steps to solve the problem, the grouping of K-Means to create a range of patrol area coverage in applying cooperation among patrollers or commonly called multi-agent patrols used to reduce the possibility of crime [12]. K-means Clustering is capable of being applied in the division of patrol areas. It affects the dynamic environment quickly, effectively, in solving the problem of patrol division tasks, as well as improving agent performance [13] and FA as an algorithm to create an optimal police patrol route each officer. This study is expected to develop a system that can advise the Merauke resort police about patrol routes and the duration of time required to carry out patrols that can reach all hotspots of criminal case.

2. RELATED WORK

FA is an algorithm introduced by Yang in 2008, this algorithm is inspired based on flashing behavior of fireflies that live in nature [4]. This algorithm has been implemented to find the shortest route. The author modifies the light absorption coefficient and uses the attraction matrix to select the next permutation node. Final results, FA can be applied in various coordinate types, FA also helps in modifying the optimal route by providing second route option in case of delay in suggested path so as to help management to control the path taken [1].

In different cases, FA is combined with Ant Colony System (ACS) for Vehicle Capacity Routing Problems (CVRP) and Vehicle Routing Issues with Time Window (VRPTW) combinatorial issues. ACS builds customer tours taking into account capacity, window time and other additional restrictions and using FA to establish optimum permutation solutions. The tour produced by the ACS is initialized as a firefly population and calculated light intensity based on objective function, where the smallest intensity value is considered the best firefly. The best solution for FA was found to be used to update pheromone experiments on the ACS for the next iteration [14].

In this research, FA find an optimal route solution by adding K-Means as a first step to forming a smaller coverage area for each patrol agent so that agents can focus on the Hotspots of Criminal Case within their respective territories. K-Means clustering is a superior algorithm with scalability, efficiency, simplicity to classify data. The process of K-Means Clustering is to set and group objects in one group with the same characteristics in one cluster and not belong to another cluster using proximity to the centroid of each cluster [15]. Previously, the FA has been combined with K-Means Clustering but in different studies that are grouping data. Fa is

assigned to find the optimal centroid center specified by the user to improve the accuracy of the grouping process [9].

3. RESEARCH METHODOLOGY

3.1 K-Means Clustering Algorithm

K-means Clustering is an easy and simple algorithm, k-means method iteratively classifies hotspots by minimizing the sum of squared distance between points with the center of the cluster. Each point will be entered randomly to all groups by calculating the euclidiean distance to the center of the cluster. Process by initializing the centroid and the results of each cluster / group, then inserting new samples to the group. The process continues until no point moves from one group to another or the value of the function indicates no significant change until the last iteration [16]. The clustering method involves two stages, the establishment of a reference pattern and the pattern matching process. The principle of K-means clustering among others [17] :

a.   Determine the initial cluster center (centroid).

b.   Calculate the proximity of the dataset to the centroid.

c.   Classification of datasets based on the closest distance to the centroid.

d.   Calculate the new cluster center, and

e.   Reclassify the dataset back with the new cluster center

K-Means Clustering is the simplest unattended learning algorithm for grouping [18]. Stages of K-Means Clustering Algorithm in Pseudocode can be seen in Figure 1.

```
1. Place randomly the K cluster centers
2. While not stop criterion do:
3. For each object do:
4. Compute distance measure to each cluster
5. Assign it to the closest cluster
6. End for
7. Recalculate the cluster centers positions (2)
8. End while
```

Fig 1. Pseudo code k-means clustering algorithm.

The distance between the points and the centroid is calculated by the equation :

$$f = \sum_{i=1}^{k} \sum_{j=1, j \in G_i}^{n} ||X_j - C_i||^2 \qquad (1)$$

After grouping points to the cluster, the new Centroid initialization of each cluster is recalculated by averaging the location of all the points grouped on the centroid, with equations:

$$C_i = \frac{1}{|G_i|} \sum_{j=1, j \in G_i}^{n} x_j \qquad (2)$$

3.2 Firefly Algorithm

Firefly Algorithm is inspired by light blinking fireflies in nature and reflects the physical formula of the light intensity of fireflies found in nature [13].

The main idea of Firefly Algorithm is as follows:
- All fireflies are unisex. So, one firefly will be attracted to another firefly regardless of gender.
- The attractiveness is proportional to their brightness. So, for two flashing fireflies, the less bright will move in a brighter direction. If not If there are no fireflies with higher light intensity, fireflies will move randomly.

- The brightness of the firefly is determined by the objective function/fitness function/objective function of the problem to be solved [7].

Stages of the Firefly Algorithm in Pseudo-code can be seen in Figure 2.

```
1. Objective function f(x), x = (x1, x2, ..., xd)^T
2. Generate initial population of fireflies x_i (i = 1,2, ..., n)
3. light intensity I_i at x_i is determined by f(x_i)
4. Define light absorption coefficient γ
5. While (t < MaxGeneration)
6.   for i = 1 : n (all n fireflies)
7.     for j = 1 : n all n fireflies (inner loop)
8.       if (I_i < I_j), Move Firefly i towards j; end if
9.       Vary attractiveness with distance r via exp [−γr]
10.      Evaluate new solutions and update light intensity
11.    end for j
12.  end for i
13.  Rank fireflies and find the current best g_*
14.end while
15.Post-processing the results and visualization
16.End
```

Fig 2. Pseudo code firefly algorithm.

In the Firefly Algorithm, two important things for this algorithm are the intensity of light and attraction. The intensity of light *I(r)* is defined in relation to the inverse square law [14]. The intensity of light varies with distance *r*, the equation as follows [15] :

$$I(r) = I_0 e^{-\gamma r^2} \qquad (3)$$

$I_0 =$ intensity at the source

$\gamma =$ light absorption coefficient

r = distance between fireflies

While, the attraction of fireflies is related to the intensity of light seen by other fireflies, where β_0 is the attraction at distance = 0. The attraction is formulated as follows :

$$\beta = \beta_0 e^{-\gamma r^2} \qquad (4)$$

The distance between fireflies *i* and *j* on x_i and x_j is the

distance calculated by the Euclidean equation:

$$r_{ij} = ||x_i - x_j|| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \qquad (5)$$

$x_{i,k} =$ component of spatial coordinates x_i from fireflies *i*.

$x_{j,k} =$ component of spatial coordinates x_j from fireflies *j*.

The movement of fireflies *i* attracted to the lighter fireflies *j* is determined by the equation :

$$x_i = x_i + \beta_0 e^{-\gamma r^2} (x_j - x_i) + \alpha_1(\varepsilon_i) \qquad (6)$$

The α parameter is the attraction parameter, it is random

and Yang (2010) shows the value of this parameter is usually the range [0,1] and ε_i is a vector of random numbers drawn from either a Gaussian or uniform generally [-0.5,0.5] distribution. While γ parameter characterizes the variation of attractiveness and its value determines the convergence of the

algorithm. For most applications, γ is usually set between 0.1 and 10 [4].

4. DESIGN OF RESEARCH

4.1 Data

The research material used in this research are data of Merauke Police Chief Order Letter that is the Patrol Route / Purpose of the Police of the Republic of Indonesia of Papua Region Merauke Resort with field survey, take coordinates Location of Place of Case Event, and Coordinates of Roads listed as crime-prone areas by the Merauke Police Force. The overall coordinate data is grouped into five clusters as well as the number of Police Stations assumed the number of patrol agents in charge, grouping using the K-means Algorithm, and the cluster results are used in the FA Method process as the dataset for formed a patrol route according to the boundaries of each agent.

4.2 Information System Framework

In this research, patrol route optimization begins with the formation of territorial patrol agent boundaries based on the grouping of overall coordinate data Hotspots of Criminal Case using K-Means Algorithm and Firefly Algorithm to form the optimal route of each agent. This process can be seen in Figure 3.
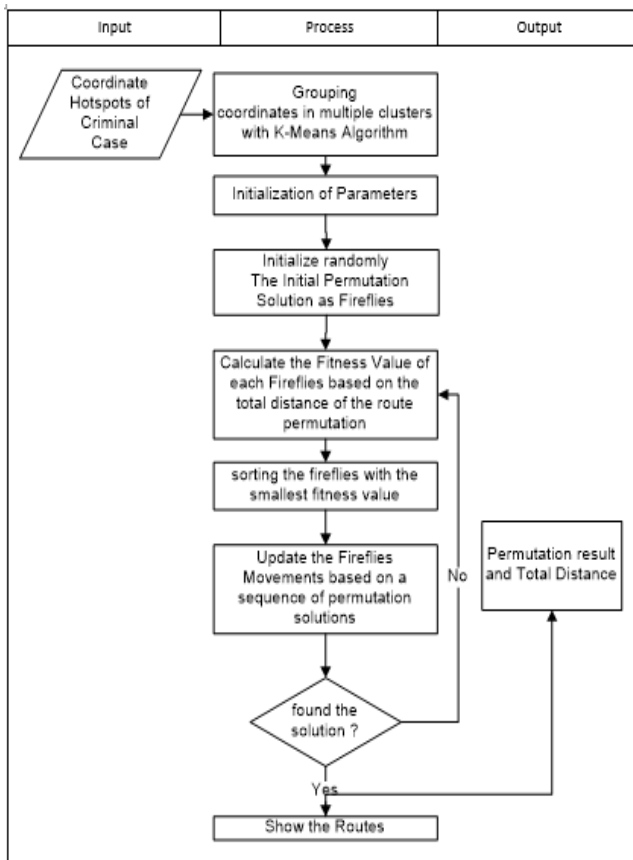

Fig 3. Information system framework

5. RESULT AND ANALYSIS

Police Patrol Strategy designed, illustrated using five patrol agents that have patrol areas against different Hotspots of Criminal Case points established using the K-Means algorithm. After that, each agent will be established a police patrol route by FA.

5.1 Implementation of K-Means Clustering Algorithm

The results of grouping Hotspots of Criminal Case points for each cluster of patrol agents can be seen in Table 1 and shown in Figure 4. Five clusters are initialized by agent name Agent 1, Agent 2, Agent 3, Agent 4, and Agent 5. The results grouping of the Hotspots of Criminal Case show Agent 1 gets 30 Hotspots of Criminal Case, Agent 2 gets 16 points, Agent 3 gets 11 points, Agent 4 gets 38 points and Agent 5 gets 21 Hotspots of Criminal Case. Figure 4 shows the spread of hotspots of criminal case in the form of graphs with dots that have been grouped in each cluster with different colors.

Table 1. Results grouping hotspots of criminal case using k-menas clustering

Name	The number of points
Agent1	30
Agent2	16
Agent3	11
Agent4	38
Agent5	21

Table 2. Centroid on each cluster agent

Name	Longitude	Latitude
Centroid Agent1	-8.496741	140.398781
Centroid Agent2	-8.485589	140.388575
Centroid Agent3	-8.531604	140.426987
Centroid Agent4	-8.508995	140.404916
Centroid Agent5	-8.504232	140.377836

Table 2 shows the end result of centroid location formed in the K-Means Clustering process of each cluster agent. The centroid point of each cluster will be the starting point of the police patrol route each agent.
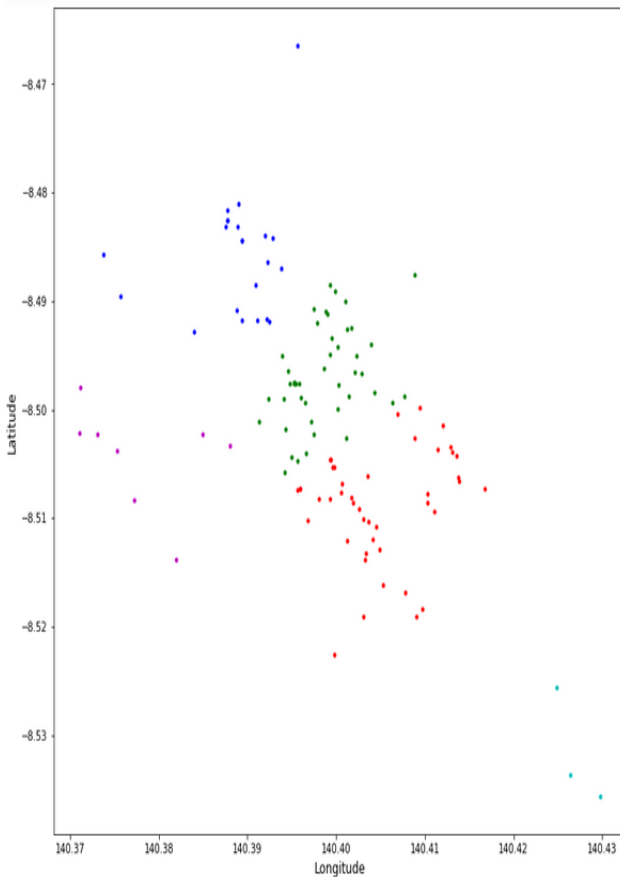
Fig 4. Results grouping hotspots of criminal case with used k-means clustering

5.2 Implementation of Firefly Algorithm

In the Firefly Algorithm implementation, the continuous FA equations are adjusted for discrete permutation problems, provided that the light intensity in the FA is assumed to be the total distance traveled in the route permutation, and the distance between the fireflies is assumed to be the total sequence of different routes between the permutation solutions. The FA results in the Hotspots of Criminal Case coordinate dataset are grouped into five clusters using K-Means Clustering and performed with three different scenarios, but with the same Firefly number of 10 Firefly and the same iteration of 100 iterations can be seen in Table 3 which shows the difference in total distance and execution time of each Agent A1, A2, A3, A4, and A5 in three different scenarios A, B, and C. The optimal route with the smallest total distance from each cluster in graphical form can be seen in Figure 5-9:

Table 3. The scenario of convergence with 10 fireflies and 100 iterations

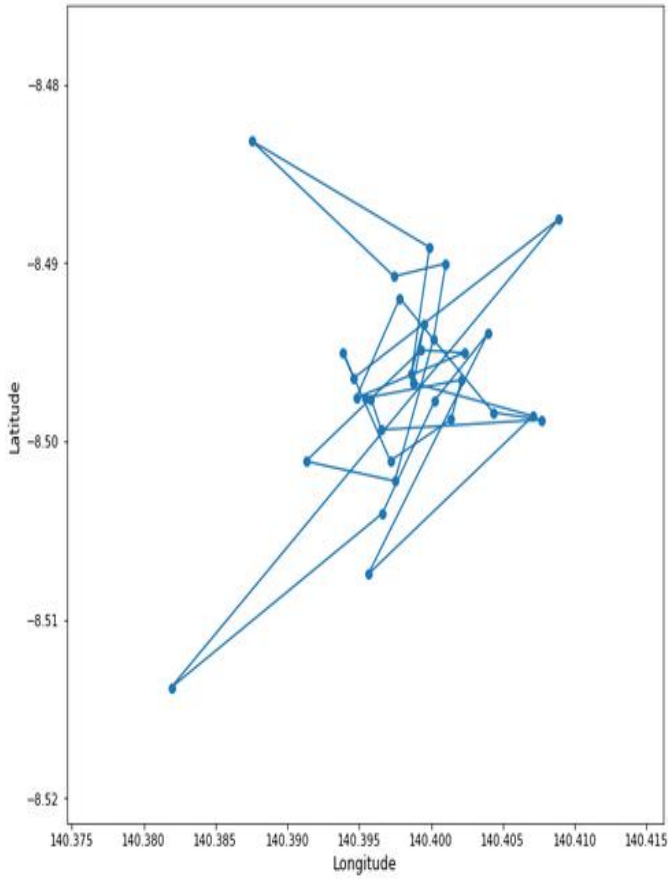| No.	Scen.	FA Params	Convergence	execution time (microseconds)
1.	A	Alfa : 0.1	A1 : 0.26157795236761944	732264
		Beta : 0.35	A2 : 0.16053914762649507	967469
		Gamma : 0.15	A3 : 0.2763106611992022	714525
		Number of firefly : 10	A4: 0.35049720477886814	46706
			A5 : 0.3381648653545979	
		Iteration : 100		971877
2.	B	Alfa : 0.2	A1 : 0.26773445528825235	307473
		Beta : 0.7	A2 : 0.150422216027077224	870939
		Gamma : 0.3	A3 : 0.2298163882650403	626237
		Number of firefly : 10	A4 : 0.3815226637389988	782237
			A5 : 0.32573726100081357	
		Iteration : 100		969080
3.	C	Alfa : 0.3	A1 : 0.271144687708027	459576
		Beta : 1.5	A2 : 0.16841416698421566	783578
		Gamma : 0.45	A3 : 0.24632481853307503	722399
		Number of firefly : 10	A4 : 0.3172608533185746	803728
			A5 : 0.3082510251491538	
		Iteration : 100		25031

Fig 5. Optimal route of agent 1.
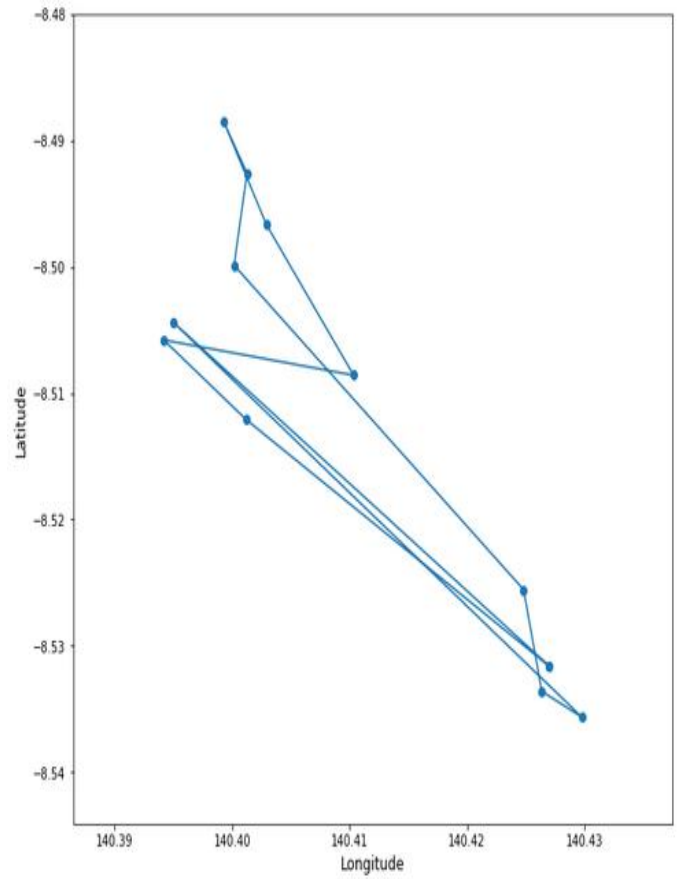
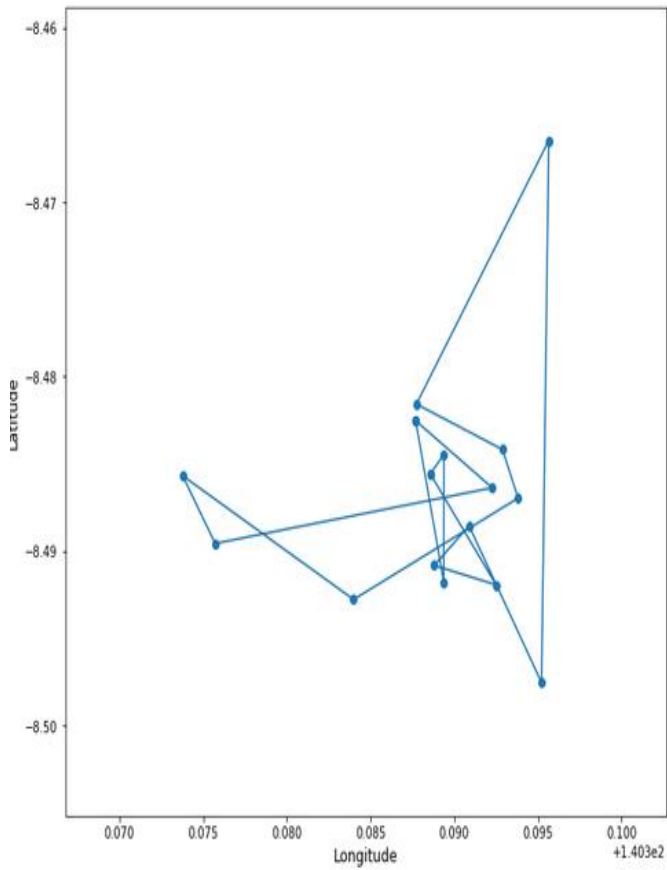**Fig 7.   Optimal route of agent 3**

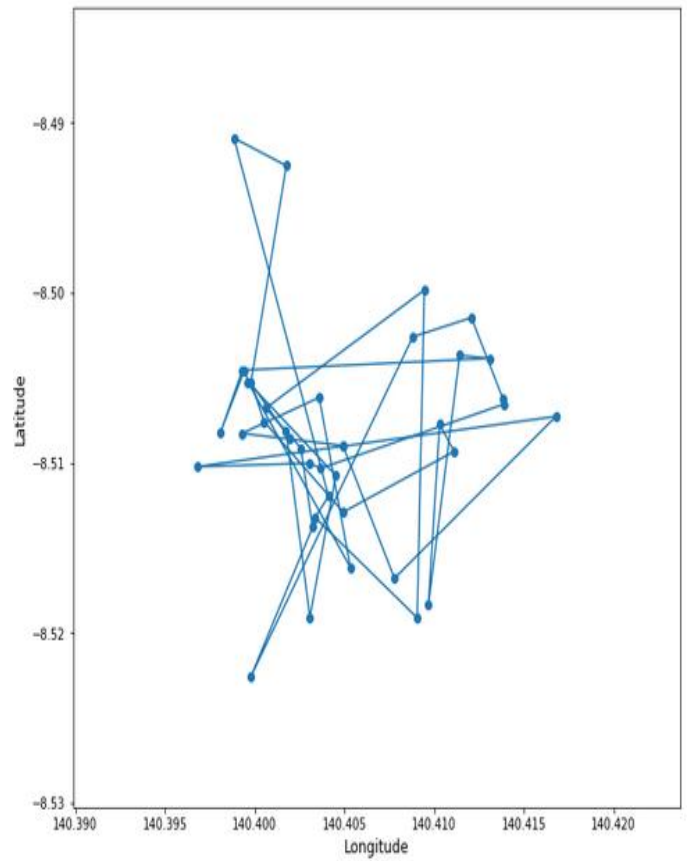**Fig 6.   Optimal route of agent 2.**

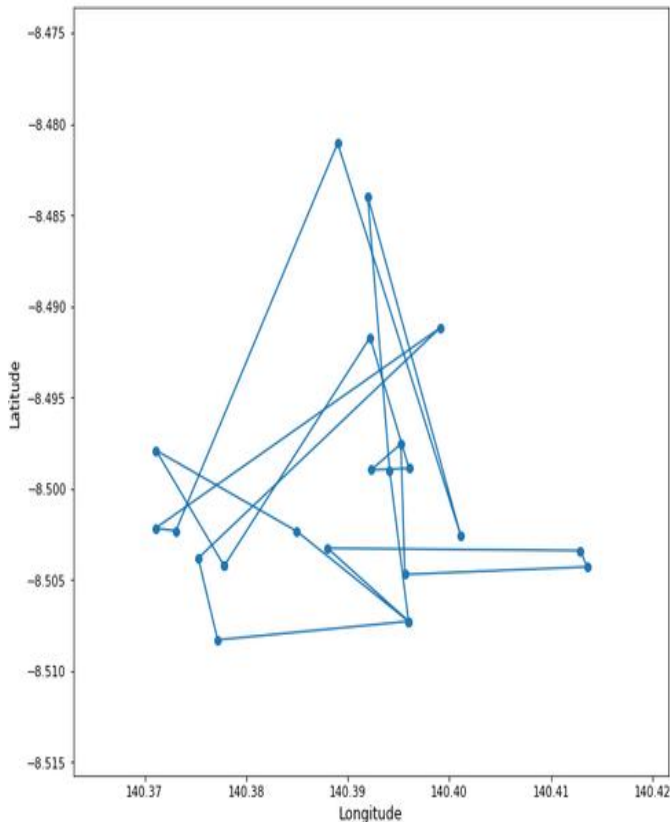**Fig 8.   Optimal route of agent 4**

Fig 9.   Optimal route of agent 5.

5.3 Result And Discussion

The result, for scenario A with the structure and value of small parameters, shows the convergence of the shortest distance value obtained in cluster A1. While scenario B is obtained on two clusters A2 and A3, whereas C scenario with structure and parameter value is highest compared to other result the shortest distance values are A4 and A5.

In A1 cluster, best scenario is scenario A, best A2 scenario cluster in scenario B, best A3 scenario is B, A4 best scenario is C, and in A5 the best scenario is scenario C. In scenario A, the average execution time generated is 686568.2 microseconds, scenario B produces 711193.2 microseconds, and the C scenario with an average execution time is 558862.4 microseconds.

6. CONCLUSION

The best scenario for the whole clusters are B and C scenarios because, the two scenarios produce two smallest total distance in the two clusters, while scenario A only produces 1 total of the smallest distance in cluster A1. In terms of computational time, scenario C is superior to the overall scenario with a computationally smaller average time of 558862.4 microseconds, this proves the greater the value of valid parameters resulting in computation time and produce better routes.

7. REFERENCES

[1] N. Ali, M. A. Othman, M. H. Misran, M. K. Nor, and H. A. Sulaiman, "Firefly Algorithm with Attractiveness Matrix Enhancement for Shortest Route Alternative Solution," no. 14ct, pp. 177–180, 2015.

[2] J. C. Ferreira, M. T. Arns Steiner, and M. Siqueira Guersola, "A Vehicle Routing Problem Solved Through Some Metaheuristics Procedures: A Case Study," *IEEE Lat. Am. Trans.*, vol. 15, no. 5, pp. 943–949, 2017.

[3] O. Dib, M. A. Manier, and A. Caminada, "A Hybrid Metaheuristic for Routing in Road Networks," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, vol. 2015–Octob, pp. 765–770, 2015.

[4] X.-S. Yang and M. Karamanoglu, *1 - Swarm Intelligence and Bio-Inspired Computation: An Overview BT - Swarm Intelligence and Bio-Inspired Computation*. 2013.

[5] I. Strumberger, "Enhanced Firefly Algorithm for Constrained Numerical Optimization," pp. 2120–2127, 2017.

[6] A. Tjahjono, D. O. Anggriawan, A. K. Faizin, A. Priyadi, M. Pujiantara, and M. H. Purnomo, "Adaptive modified firefly algorithm for optimal coordination of overcurrent relays," pp. 2575–2585, 2017.

[7] M. Gao, L. Li, X. Sun, L. Yin, H. Li, and D. Luo, "Optik Firefly algorithm (FA) based particle filter method for visual tracking," vol. 126, pp. 1705–1711, 2015.

[8] R. J. Kuo and P. S. Li, "Taiwanese export trade forecasting using firefly algorithm based K-means algorithm and SVR with wavelet transform," *Comput. Ind. Eng.*, vol. 99, pp. 153–161, 2016.

[9] T. Hassanzadeh and M. R. Meybodi, "A new hybrid approach for data clustering using firefly algorithm and K-means," *16th CSI Int. Symp. Artif. Intell. Signal Process. (AISP 2012)*, no. Aisp, pp. 007–011, 2012.

[10] L. Lei, "The GIS-based Research on Criminal Cases Hotspots Identifying," *Procedia Environ. Sci.*, vol. 12, no. Icese 2011, pp. 957–963, 2012.

[11] H. Chen, T. Cheng, and S. Wise, "Developing an online cooperative police patrol routing strategy," *Comput. Environ. Urban Syst.*, vol. 62, pp. 19–29, 2017.

[12] M. Camacho-Collados, F. Liberatore, and J. M. Angulo, "A multi-criteria Police Districting Problem for the efficient and effective design of patrol sector," *Eur. J. Oper. Res.*, vol. 246, no. 2, pp. 674–684, 2015.

[13] L. Zhiwei, Y. Xiang, and D. Yao, "A partitioning-based task allocation strategy for Police Multi-Agents," pp. 2124–2128, 2014.

[14] R. Goel and R. Maini, "SC," *J. Comput. Sci.*, 2017.

[15] S. Kapil and M. Chawla, "Performance Evaluation ofK-means Clustering Algorithm with Various Distance Metrics," pp. 4–7, 2016.

[16] Z. Jalali, "Development of slope mass rating system using K-means and fuzzy c-means clustering algorithms," *Int. J. Min. Sci. Technol.*, vol. 26, no. 6, pp. 959–966, 2016.

[17] T. Tada, K. Hitomi, Y. Wu, S. Y. Kim, H. Yamazaki, and K. Ishii, "K-mean clustering algorithm for processing signals from compound semiconductor detectors," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 659, no. 1, pp. 242–246, 2011.

[18] B. Felipe, "Ac ce pt us cr t," 2015.