

K-Means Clustering and Genetic Algorithm to Solve Vehicle Routing Problem with Time Windows Problem

Adyan Nur Alfiyatin¹, Wayan Firdaus Mahmudy², Yusuf Priyo Anggodo³

^{1,2}Faculty of Computer Science, Brawijaya University, Malang 65145, Indonesia

³Data Analyst, Ilmuone Data, Jakarta 12190, Indonesia

Article Info

Article history:

Received Oct 5, 2017

Revised Apr 18, 2018

Accepted Apr 21, 2018

Keywords:

Distribution problem

Genetic algorithm,

K-means

Genetic K-means algorithm

Optimization

ABSTRACT

Distribution is an important aspect of industrial activity to serve customers on time with minimal operational cost. Therefore, it is necessary to design a quick and accurate distribution route. One of them can be design travel distribution route using k-means method and genetic algorithms. This research will combine k-means method and genetic algorithm to solve vehicle routing problem with time windows (VRPTW). K-means can do clustering properly and genetic algorithms can optimize the route. The proposed genetic algorithm employs initialize chromosome from the result of k-means and using replacement method of selection. Based on the comparison between genetic algorithm and hybrid k-means genetic algorithm proves that k-means genetic algorithm is a suitable combination method with relative low computation time, are comparison between 2700 and 3900 seconds.

*Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Adyan Nur Alfiyatin,
Faculty of Computer Science,
Brawijaya University, Malang 65145, Indonesia.
Email: wayanfm@ub.ac.id

1. INTRODUCTION

Distribution is one of the most important aspects in the industry, contributing to delivering products from producers to consumers [1]. Without a systematic distribution channel, the best promotional products and strategies will not make the product known and consumed by the end user [2]. The goal of systematic channels and distribution is to deliver products for customers on time and reduce operational costs industries. Toth and Vigo [3] revealed in Europe and North America recognize that using of computerized procedures for planning process distribution can reduce transportation costs by 5-20%. It is using mathematical programming will be more effective in the process of determining the route management of goods service distribution. Determination of goods distribution mathematical terms is Vehicle Routing Problem (VRP). VRP is a combinatorial problem that discusses the process of goods distributing from a depot (center distribution) to customers scattered in various locations with the limited vehicle, the distance between depot and customer, vehicle capacity and time [4].

VRP has been proposed since 1959 by Dantzig [5] and the VRP issues continue to grow. VRP with Time Window (VRPTW) is an important problem in the logistics and transportation system because it affects the profit company. Research conducted by Zhong and Pan [6] solve the problem VRPTW with the aim of minimizing costs and less obtained results than optimal because resulting cost not suitable with the real. In the study Zhou and Wang [7] It is mentioned that VRPTW consists of a set of customer served by one vehicle from the main depot and each vehicle is only passing once, with limited payload and time in accordance with the availability of customer.

Due to only once passing through the route, there needs to be regional grouping to serve many customers, thus making a satisfied customer. According to MacQueen [8], k-means is a method of doing

clustering will get maximum results. Setting the right center point on each cluster can provide a precise grouping of data [9]. K-Means is good method for clustering which get solution quickly. Clustering method got solution in missing value problem [10]. Based on that, k-means can give good solution for solve clustering priority areas.

Based on the observations made on the Home Industry of tempe chips. Home The industry has a wide marketing area and limited vehicles in meeting customer demand. Distribution is sometimes not meet the time set by the customer so many complaints received. The extent of marketing area, limited vehicle and limited time can be overcome by doing strategic planning such as by using route planning through which the distribution process, the capacity is loaded by each vehicle and priority areas to receive orders first. Such planning can be assisted with the help of a computer using a method in artificial intelligence like the K-Means method [8], [9], [11] to categorize priority areas. Based on the problems, the researcher proposed an area mapping method using K-Means and Genetic Algorithm for route optimization to faster time of computation than conventional current method in home industry.

The Genetic Algorithm can solve the problem for route optimization on VRP issues [12] With fast computing time [4]. The research of Mnasri, et al [13] the genetic algorithm is used to optimize the route on the VRPTW problem by improving the process on the crossover resulting in the best route and minimum cost, then the research by Lesmawati [14] using a genetic algorithm for the distribution of frozen food, subsequent research by Philip, et al [15] using genetic algorithm for distribution route. From several studies conducted to prove that the genetic algorithm to get the optimal route with fast computing time and suitable for multi objectives optimization problem [16]. In addition to the wide space problems and complex genetic algorithms can find the most optimal solution [17]. The focus of this research is to solve the complex problem of VRPTW by combining k-means method and genetic algorithm on tempe chips distribution Malang. K-means do clustering priority areas that visiting by vehicle and genetic algorithm do scheduling vehicle which visiting the customers.

2. RELATED WORK

Various improvements to the genetic algorithm have been done to solve the VRPTW problem. As mentioned in the research of Mnasri, et al [13] There are modifications to the chromosome initialization process, generated by Solomon's insertion, random interchange, random ordering. Further modification by Chunyu and Xiaobo using a cross-order operator and the partial route reversal operator route to increase the convergent speed [18]. Research Yucenur and Demirel [19] perform clustering using nearest neighbor and genetic algorithm to solve the problem of VRP with the merger of these two methods of computing time is used more quickly. Cheng, et al [20] incorporating k-means and GA to create adaptive clusters aimed at reducing the complexity of time and space. The results show that the method is feasible and effective in conducting cluster analysis. Zhaoxia and Hui [21] using a genetic algorithm to determine the initial center of the cluster in order to obtain maximum results and prove that method is suitable for use both in small data groups and in more complex data groups. Further research conducted by Krishna and Murty [22] combines a genetic algorithm and k-means to resolve the problem of global search, k-means method is used to limit the search to the crossover operator in genetic algorithm, produce a method is name GKA (Genetic K-Means Algorithm), which includes the k-means operator on genetic algorithms operator. Differences in this study k-means clustering method in use prior to classify the area then the result of the k-means used to input on genetic algorithms for optimized route followed.

3. RESEARCH METHOD

The data used in this study come from a home industry tempe chips "XYZ" Malang and distance data measured by using the Google Maps tool. Data contained in attachment.

3.1. K-Means

For clustering process, a method that how it works simple and get optimal results is k-means [8]. The first thing done in the process of *clustering* initializing values of k is the number of *clusters*, and then determine the center point for each cluster randomly. In this study, the determination of the central point is set manually. Each object will be inserted and set to a particular cluster, in which each object is included in the same cluster have similarities. To determining the similarity of the object with the center point of the cluster is measured by *Euclidian distance* (D) shown in Equation 1. An object will enter the cluster that has the smallest D values.

$$D(x_2, x_1) = \sqrt{\sum p_j = 1(x_2 - x_1)^2}, \quad (1)$$

where;

- p = dimension data
- x1 = position of point 1
- x2 = position of point 2

The flow of k-means clustering algorithm is as follows [9] :

1. Initialize the value of k cluster and each cluster centroids.
2. Determine each object include the cluster with the closest distance based on the value of Euclidean distance.
3. To recalculate the value of the centroid of each cluster with Equation 2.

$$V_j = \frac{1}{n_j} \sum_{p=0}^n data_p, \tag{2}$$

V_j is the value of centroids from cluster j . n_j is the number of objects in cluster j . $data_p$ is vector data to p .

4. Repeat steps 2 until centroid value unchanged or beyond iteration specific.

3.2. Genetic Algorithm

The genetic algorithm is an optimization technique that are stochastic and based on Darwin's theory of evolution [23]. The process of the genetic algorithm includes the representation of the chromosome, the formation of the population, the calculation of fitness, reproduction with crossover and mutation, selection process to get new individuals.

1) Chromosome Representation

To explain the chromosome representation, a simple example is given. There are 6 customers (P11, P12, P21, P22, P31, P32) spread across 3 distinct regions, the number of vehicles to serve customers are 3 pickups (V1, V2, V3) containing the payload 625 tempe chips.

Its chromosome representation uses a permutation representation with 2 segments. The first segment is the distance between customers in different regions and the second segment is the number of vehicles used for the distribution process. Figure 1 illustrative representation of chromosomes.

Segmen 1						Segmen 2		
P11	P12	P21	P22	P31	P32	V1	V2	V3
1	2	3	4	5	6	1	2	3

Figure 1. Representation Chromosomes

In Figure 1, the blue colour indicates customers in each region and the yellow colour containing three genes which means that the gene 1 in the column V1 serve customers in the regions P11 and P12, gene 2 in the column V2 serve customers in the regions P21 and P22 as well as the gene 3 in column V3 serve P31 and P32. There are 50 stores for destinations and 3 pick ups so each vehicle gets 16 destinations with assumed 6 working days. Time duration to drop of goods is 30 minutes. If serving passes deadline time, then calculated penalty count.

2) Fitness Calculation

After the generation of value chromosome then calculating the fitness value. This calculation shows the ability of individuals to survive and continue the next process [23]. The fitness function is shown in Equation 3.

$$fitness = (1 / (1 + tt)) + (tp * (-1)), \tag{3}$$

tt is total travel time, tp is the total number of penalty/ violations. Penalty is calculated from the sum of waiting time and the over time to drop products based on time window.

3) Crossover Process

Crossover process is one method of reproduction to produce new individuals (*offspring*). The method used is the *one cut point*. New individual is generated by multiplying the number of population (pop size) and the crossover rate (cr) is determined. On this issue 0.6 cr value and pop size 5. So the results from multiplying $0.6 \times 5 = 3$ offspring. Then pick two individuals at random as a *parent* for the reproduction process. Individuals selected are P1 and P3. Figure 2 represent individuals selected.

	Segmen 1						Segmen 2		
	P11	P12	P21	P22	P31	P32	V1	V2	V3
P1	1	2	3	4	5	6	1	2	3
P3	1	4	3	2	5	6	3	2	1

Figure 2. Crossover process

After parent selected. The next step is to choose randomly cut-off the gene for the crossover process. Figure 3 shows the crossover process.

	Segmen 1						Segmen 2		
	P11	P12	P21	P22	P31	P32	V1	V2	V3
P1	1	2	3	4	5	6	1	2	3
P3	1	4	3	2	5	6	3	2	1
C1	5	6	1	4	3	2	1	2	3

Figure 3. Offspring Crossover Result

Offspring in segment 1 obtained from parent 1 with 2 genes are P31 and P32 then the reminder is obtained from the parent 2 that is not in the offspring. In the segment 2 the selected genes are the two genes so that between parent 1 and parent 2 only exchange the gene to one and the last while the central gene remains.

Mutation Process

Mutation process in the genetic algorithm is optional, depending on the problem resolved. In this study, mutation method used is the exchange mutation. It works of selecting 2 genes randomly to exchange each selected gene. The parent used in the mutation process is only one parent and randomly selected. For example, has been selected parent 5. Figure 4 shows the process mutation.

P5	1	5	3	6	2	4	1	2	3
C4	1	4	3	6	2	5	1	2	3

Figure 4. Offspring Mutation Result

4) Selection Process

The selection process to establish a new population use replacement selection, because of the problems of multi-trip VRPTW by Anggodo, et al replacement selection gives maximum results [17]. replacement selection method works by replacing a chromosome, chromosome reproduction will replace its parent chromosome when it has a higher fitness value.

4. EXPERIMENTAL RESULT AND DISCUSSION

Testing method used in genetic algorithms to solve the problem of VRPTW tempe chips distribution by doing 3 testing there are testing of population, testing of generation, testing of combination crossover rate (*cr*) and mutation rate (*mr*). Each testing of parameters in the test is five times. Because the genetic algorithm is a stochastic method to produce a different solution each time running, therefore the fitness of each test is calculated from the average value of fitness [24].

1) Testing of Population Size

In this test based on multiple population number of 100 to 2000 of the population, the combination of *cr* and *mr* 0.6 and 0.4 set based research Lesmawati, et al [14]. The test results are in Figure 5.

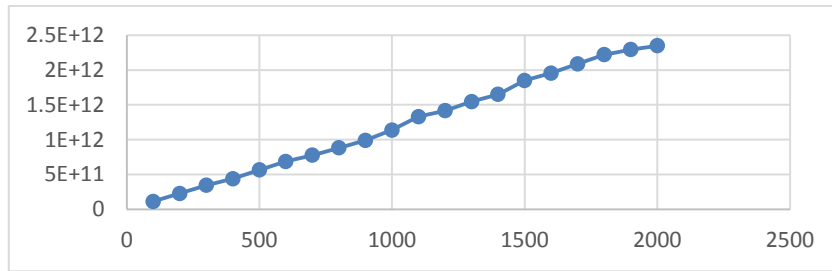


Figure 5. Population Testing

Based on the results of the test population, the fitness value which is used has the highest value in a population of 2000 with fitness 2.34728e+12. Testing discontinued in 2000 due to population fitness value generated is not limited and may not convergen and long computation time.

2) Testing of Generation/Iteration

Testing iterations/generation were performed using the best fitness results from testing a population of 2000 with a multiple of 100. For more details, contained in the Figure 6.

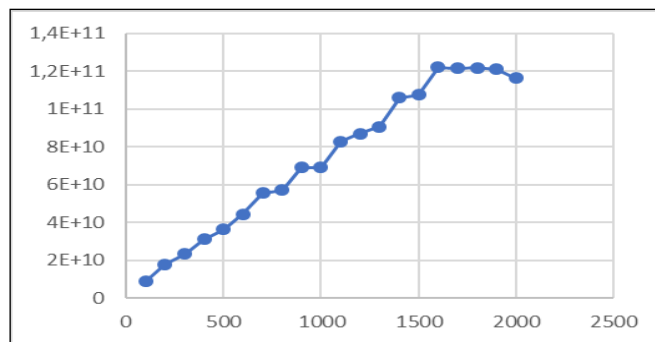


Figure 6. Iteration test multiple 2

In Figure 6 shows that the best fitness values at 1600 iterations with the fitness value 1.22059e+11 and after it convergent and the fitness values resulting lower.

3) Testing of combination cr and mr

Testing combination cr and mr with a range of 0-1. The test parameters based on the best number of population is 2000 and the best number of iteration is 1600. Figure 7 result illustration of the test combination cr and mr.

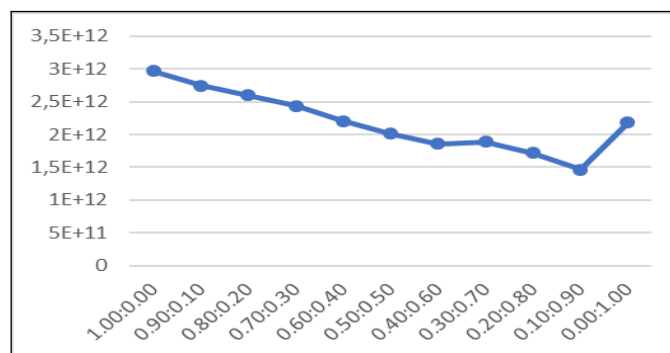


Figure 7. Combination cr and mr

From Figure 7, showing that the combination cr and mr value are the ability of the genetic algorithm in the exploration and exploitation in the search space. Cr and mr value will be different for each problem. In this study, the combination cr and mr get the highest fitness value in 1.0 and 0.0 with a value of $2.96851e+12$. The results of the test show the value of cr decreased and mr increased tend to be ugly. Large space solutions require a search operation that focused on global search. The crossover process tends to reach the outer regions compared to its nearest neighbors while the mutation process is more focused on reaching its neighbors where the mutation process only exchanges one gene in one individual. So the result of a large crossover process tends to give a good value than the mutation, although at mr value 1 the fitness results are not decreased but better than the previous mr 0.9.

After testing the parameters then testing the best parameters and comparing the results obtained from the Genetic Algorithm and Hybrid K-Means Genetic Algorithm, the test is done with 10 times the program run then taken the average fitness value. Table 1 shows the comparison results.

Table 1. Result of Comparison

Method	Fitness	Computation Time
Genetic Algorithm	8.07E+5	3900 seconds
Hybrid K-Means and Genetic Algorithm	3.0312E+12	2700 seconds

Based on the fitness value obtained in Table 1 shows that Hybrid K-means Genetic Algorithm is a combination that matches and briefly relative computation time. The proposed method to get better results because the process of formation of travel routes have been classified each region and the route of the vehicle will not exceed the area that has been determined. The chromosomes formed in the genetic algorithm are shorter so the computation time is faster. In addition, just use genetic algorithm often gets penalty because has long rute and time distribution begin 6am until 12pm then makes genetic algorithm's fitness very low of k-means genetic algorithm's fitness.

5. CONCLUSION

In this paper, some testing has been performed using k-means and genetic algorithms. Based on three results testing has been done are population testing, iteration testing and combination cr and mr testing can be concluded that in 2000 population, 1600 iteration, combination cr 1.0 and mr 0.0 produce the best fitness value with fast computation time.

The combination of k-means method and genetic algorithms is the the right combination. K-means can cluster right and genetic algorithm can optimally optimize each cluster of routes and inter cluster relations. Based on computation time, it is better than the genetic algorithms do optimize all of the point routes. In addition, to improve the result need to be combined route optimization that focuses on finding local and tracing more deeper solutions one of them can apply auto-speed acceleration algorithm to replace crossover process which is not optimal [25], [26] and it is also possible to use many parameters i.e spatial data warehouse, geographic information system [26]. The applicability of this paper can be applied directly to the complexity of the distribution of goods, resulting in lower cost and more efficient distribution time.

REFERENCES

- [1] Z. He, T. C. E. Cheng, J. Dong, and S. Wang, "Evolutionary Location and Pricing Strategies in Competitive Hierarchical Distribution Systems : A Spatial Agent-Based Model," *IEEE Trans. Syst. MAN, Cybern. Syst.*, vol. 44, no. 7, pp. 822–833, 2014.
- [2] B. Tomoiag, M. Chindris, A. Sumper, R. Villafila-robles, and A. Sudria-andreu, "Distribution system reconfiguration using genetic algorithm based on connected graphs," *Electric Power Systems Research*, vol. 104, pp. 216–225, 2013.
- [3] D. V. Paolo Toth, *Vehicle Routing: Problems, Methods and Applications, Second Edition*. USA: Society for Industrial and Applied Mathematics Philadelphia, 2014.
- [4] R. Nallusamy, K. Duraiswamy, R. Dhanalaksmi, and P. Parthiban, "Optimization of Multiple Vehicle Routing Problems Using Approximation Algorithms," *Int. J. Eng. Sci. Technol.*, vol. 1, no. 3, pp. 129–135, 2009.
- [5] G. B. Dantzig and J. H. Ramser, "The Truck Dispatching Problem," *Manage. Sci.*, vol. 6, no. 1, pp. 80–91, 1959.
- [6] Y. Zhong, "A Hybrid Optimization Solution to VRPTW Based on Simulated Annealing," Proceedings of the IEEE International Conference on Automation and Logistics, no. 8, pp. 3113–3117, 2007.
- [7] Y. Zhou and J. Wang, "A Local Search-Based Multiobjective Optimization Algorithm for Multiobjective Vehicle Routing Problem With Time Windows," *IEEE Systems Journal*, pp. 1–14, 2014.
- [8] J. MacQueen, "Some methods for clustering and analysis of multiobservations," in Proceedings of the fifth symposium on mathematical statistic and probability, 1967, pp. 281–297.

- [9] S. Kapil, M. Chawla, and M. D. Ansari, "On k-means data clustering algorithm with genetic algorithm," in International Conference on Parallel, Distributed and Grid Computing, 2016, pp. 202–206.
- [10] Madhu and Nagachandrika, "A New Paradigm for Development of Data Imputation Approach for Missing Value Estimation," *Int. J. Electr. Comput. Eng.*, vol. 6, no. 6, pp. 3222–3228, 2016.
- [11] N. P. Barbosa, E. S. Christo, and K. A. Costa, "Demand forecasting for production planning in a food company," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 16, pp. 7137–7141, 2015.
- [12] X. Hao and Y. Huili, "The Genetic Algorithm on the Multiple-Depot Vehicle Routing Problem with Vehicle Sharing," Second International Conference on Intelligent Computation Technology and Automation, 2009.
- [13] S. Mnasri, F. Abbes, K. Zidi, and K. Ghedira, "A Multi-Objective Hybrid BCRC-NSGAI Algorithm to Solve the VRPTW," 13th Int. Conf. Hybrid Intell. Syst., pp. 60–65, 2013.
- [14] W. Lesmawati, A. Rahmi, and W. F. Mahmudy, "Optimization of Frozen Food Distribution using Genetic Algorithms," *J. Environ. Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 51–58, 2016.
- [15] A. Philip, A. Taofiki, and O. Kehinde, "A Genetic Algorithm for Solving Travelling Salesman Problem," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 1, pp. 26–29, 2011.
- [16] Y. Lan, "A Hybrid Feature Selection based on Mutual Information and Genetic Algorithm," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 7, no. 1, pp. 214–225, 2017.
- [17] Y. P. Anggodo, A. K. Ariyani, M. K. Ardi, and W. F. Mahmudy, "Optimization of Multi-Trip Vehicle Routing Problem with Time Windows using Genetic Algorithm," *J. Environ. Eng. Sustain. Technol.*, vol. 3, no. 2, pp. 92–97, 2017.
- [18] C. Ren and X. Wang, "Research on VRP optimizing based on hierarchy clustering and IGA under common distribution," 2006 Int. Conf. Comput. Intell. Secur. ICCIAS 2006, vol. 1, no. 4, pp. 143–146, 2007.
- [19] G. N. Yücenur and N. C. Demirel, "A new geometric shape-based genetic clustering algorithm for the multi-depot vehicle routing problem," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11859–11865, 2011.
- [20] D. Cheng, X. Ding, J. Zeng, and N. Yang, "Hybrid K-means Algorithm and Genetic Algorithm for Cluster Analysis," *Indones. J. Electr. Eng.*, vol. 12, no. 4, pp. 2924–2935, 2014.
- [21] T. Zhaoxia and Z. Hui, "Improved K-means Clustering Algorithm Based on Genetic Algorithm," *Indones. J. Electr. Eng.*, vol. 12, no. 3, pp. 1917–1923, 2014.
- [22] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 29, no. 3, pp. 433–439, 1999.
- [23] R. Malhotra, N. Singh, and Y. Singh, "Genetic Algorithms : Concepts , Design for Optimization of Process Controllers," *Comput. Inf. Sci.*, vol. 4, no. 2, pp. 39–54, 2011.
- [24] W. F. Mahmudy, M. R. Marian, and L. H. S. Luong, "Real coded genetic algorithms for solving flexible job-shop scheduling problem – Part II: optimization," *Adv. Mater. Res.*, vol. 701, pp. 364–369, 2013.
- [25] Y. P. Anggodo and I. Cholissodin, "Improve Interval Optimization of FLR using Auto-Speed Acceleration Algorithm," *Telecommunication, Comput. Electron. Control*, vol. 16, no. 1, pp. 1–12, 2017.
- [26] X. Luo, J. Tu, and L. Huang, "Optimization of Express Delivery Routing Problem," *TELKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 14, no. 3A, p. 380, 2016.