Optimization of School Location-Allocation Using Genetic Algorithm

Pudy Prima, Aniati Murni Arymurthy +

Faculty of Computer Science, Universitas Indonesia, Depok, West Java, Indonesia

Abstract. The enactment of zoning policy on Indonesia New Student Admission System requires schools to accept students only from its zone. This condition needs management of school allocation to ensure the facilities suffice the student distribution. Optimization of school location-allocation aims to obtain optimal allocation such that the students travel cost is minimized. This problem can be modelled as p-median problem, and it can be solved using metaheuristic approach. This study utilized genetic algorithm with two-dimensional individual representation to solve junior school location-allocation problem in South Jakarta. The results obtained by the proposed algorithm lowers 40.55% of student travel cost compared to the previous allocation with assumption that students tend to choose favorite schools in the center of regency.

Keywords: genetic algorithm, location-allocation problem, school allocation, p-median problem.

1. Introduction

The government of Indonesia enacted a zoning policy on a New Student Admission System (Penerimaan Peserta Didik Baru) (PPDB) starting at 2017. Based on zoning policy, students may only choose schools that are in the same zone as their residence [1]. This regulation, beside encouraging equal distribution of education quality, is also expected to reduce students travel cost. This could indirectly reduce congestion, especially at school rush hour. The implementation of zoning policy to PPDB needs to be followed by the management of student allocation to schools. A good school location-allocation management can encourage optimality and prevent inefficiency use of facilities. Management of student allocation to schools could provide an evaluation of the existing schools that have excess capacity or the existing schools that have the potential to provide more services.

Students allocation to schools can be viewed as an optimization problem, which object to optimize the number of students who can be served by schools, considering the constraints that apply [2]. In this case, the constraints include the consideration of school and student's residence lie in the same zone, the consideration of number of school demand at each sub-district, and the consideration of students travel distance to school. P-median problem is well-known approach to model the facility planning problems [3]. The p-median problem is a model of locating p number of facilities to minimize the weighted average distance between demand nodes and the selected facility [4]. On our school location-allocation problem, facility represents schools while demand node represents urban villages.

Various approaches have been proposed over time to solve p-median problem which is known to be NP-Hard problem. Heuristic approaches are commonly used to solve p-median problems in general [5,6,7]. In optimization of school location-allocation problems, there are several studies that have been conducted with various approaches. Teixeira's study used a mixed integer linear programming approach and solved it using XPRESS-MP solver to optimize the redeployment of secondary school network in Coimbra, Portugal [8]. Alifi's research used ArcGIS Network Analyst tool to build location-allocation optimization model of junior schools in Bandung, Indonesia [2].

⁺ Corresponding author. Tel.: + 62 857 2241 9476

E-mail address: pudy.prima61@ui.ac.id.

Although heuristics and integer programming are widely used in facility location research, meta-heuristic approach is also considered as powerful method to solve p-median problems. Meta-heuristic is a higher-level solution strategy that use certain trade off of randomization and local search to create a process capable of escaping from local optima and performing a robust search of solution space [9]. Most meta-heuristic algorithms came from the adaptation to nature-inspired behavior, such as Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization, and Simulated Annealing. Metaheuristic does not guarantee that the best solutions can be found, but it performs improvement at each iteration so that the algorithm tend to converge to near-optimal value [9]. This research used genetic algorithm to solve p-median model of school location-allocation problem. Specifically, this study aims to use genetic algorithm with two-dimensional individual representation approach to solve school location-allocation problem in South Jakarta Junior Public Schools.

2. Problem Formulation

There are 65 junior public schools spread across 65 urban villages in 10 sub-districts in South Jakarta regency. In 2017, the demand for junior school in South Jakarta is 44,003. This value is calculated based on the number of population aged 13-15 years in South Jakarta, taking into account 95.8% as school participation rates and 58.21% as ratio of the population choosing public schools by 2016 [10].

Before zoning policy is enacted, students tend to choose favorite schools, causing less-favorite schools difficult to attract more students. This has impact on unbalance distribution of education quality. In addition, many students travel long distances to go to school because their home is far from their chosen school. This could potentially increase the level of congestion on the road, especially during peak hours. If it is assumed that students live in the center of urban villages, and favorite schools are located at the center of the regency, travel cost can be formulated as follows.

$$TravelCost = \sum_{j=1}^{N} d_j P_j \tag{1}$$

where

N represents number of urban villages

 d_j represents distance between center of urban village j to center of regency

 P_j represents number of school demand in urban village j

Based on formula (1) and number of South Jakarta school demand in 2017, the travel cost of South Jakarta junior student without zoning policy is about 222,857. This travel cost calculation takes the assumption that each student chooses a favorite school located in the center of the regency. With the enactment of zoning policy, students will be distributed to schools located relatively close from where they live. Each school will receive allocation of students from the same zone. By knowing the distance between schools and students' residence, the calculation of travel cost in formula (1) can be modified as follows.

$$TravelCost = \sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij} d_{ij} z_{ij}$$
(2)

subject to

$$\sum_{i=1}^{M} a_{ij} z_{ij} = P_j \tag{3}$$

where

M represents number of schools

N represents number of urban villages

- a_{ij} represents number of students from urban village *j* allocated to school *i*
- d_{ij} represents distance between center of urban village *j* to school *i*

 z_{ij} represents zoning status between school *i* and urban village *j*

 $z_{ij} = 1$ if school *i* and urban village *j* are in the same zone, $z_{ij} = 0$ otherwise

 P_j represents number of students (school demand) in urban village j

3. Methodology

Genetic algorithm (GA) was introduced by Holland in 1975 [11] and further described by Goldberd in 1989 [12]. GA is a stochastic metaheuristic search method inspired by natural selection process. It is widely used as a robust and effective search method for many optimization problems. In GA, each solution is represented by a chromosome. Each chromosome will go through the selection, crossover, and mutation process for a number of iterations. At the end of iterations, the best chromosome will be chosen as solution based on the calculation of fitness evaluation. The flow of GA is shown in Fig. 1.



Fig. 1: Genetic algorithm process flow.

3.1. Individual Representation

This study used two-dimensional chromosome representation. A chromosome c_k is encoded as $M \times N$ matrix, where M represents the number of schools and N represents the number of urban villages. Each element $c_k(i, j)$ represents the gene value located at (i; j), where $1 \le i \le M$ and $1 \le j \le N$. The gene value represents number of students at urban village j allocated to school i.

	Urban Village 1	Urban Village 2	Urban Village 3	Urban Village 4
School 1	40		20	10
School 2		50	10	
School 3	20	25	15	40

Table 1: Example of school allocation

Consider students from four urban villages that need to be allocated to three schools. Table I shows the possible allocation arrangement, in which School 1 receives 40 students from Urban Village 1, 20 students from Urban Village 3, and 10 students from Urban Village 3, School 2 receives 50 students from Urban Village 2 and 10 students from Urban Village 3, and School 3 received 20 students from Urban Village 1, 25 students from Urban Village 2, 15 students from Urban Village 3, and 40 students from Urban Village 4. Empty cell means there is no student allocation between the associated urban village and school. This might be because the associated urban village and school lie in different zone. This allocation can be represented as the following 3×4 matrix:

r 40	0	20	ן10	
0	50	10	0	
20	25	15	40	

In this study, constraint (3) is taken into consideration during the population initialization. Based on constraint (3), all students in one urban village must be allocated to the associated schools. This means that

in a chromosome matrix, the allocation summation of each column j must suffice the student population in urban village j. The chromosome generation process is described as follows.

Step 1. Initialize $M \times N$ empty matrix.

Step 2. Pick a random column to generate.

Step 3. Exclude all rows in the column that do not have same-zone relationship.

Step 4. Randomly generate float value in range [0,1] to each row in the column such that the summation of the column is 1. This will be the allocation portion value.

Step 5. For each row in the column, multiply the portion to the target value, in this case the number of students in the associated urban village.

Step 6. Repeat the same process for the rest of matrix columns.

3.2. Fitness Evaluation

Fitness function of an individual represents its closeness to optimal solution [14]. In school locationallocation problem, the solution we are trying to achieve is the optimal allocation that minimize students travel cost. Based on this goal, our fitness evaluation below comes from the objective function.

$$min\sum_{i=1}^{M}\sum_{j=1}^{N}a_{ij}d_{ij}z_{ij}$$

$$\tag{4}$$

3.3. Selection

Selection is the process of randomly select individuals for reproduction [14]. The fitness value determines the chance of individuals to be selected. There are several selection techniques commonly used in GA: tournament selection, roulette wheel selection, rank-based selection, and stochastic universal selection. In this study, we used the stochastic universal sampling (SUS) method proposed by Baker [15]. The basic idea is to pick population sample at evenly spaced intervals using a single random value.

3.4. Crossover

Crossover operation involves the process of taking two individuals from population, then exchanging some parts of the individuals (chromosomes) [16]. Some commonly used crossover techniques include onepoint crossover, multi-point crossover, uniform crossover, and substring crossover. Those techniques are widely used in GA with string individual representation. In GA with two-dimensional individual representation, the crossover operators should suffice the constraints that apply. Since this study applies column constraint in which the sum of all matrix column should suffice a certain value, vertical crossover can be utilized. In this study, we applied one-point vertical crossover. In vertical crossover, an $M \times N$ matrix is divided into two parts at a point K, $1 \le K \le N$, becomes $M \times K$ matrix and $M \times (N-K)$ matrix. The crossover point is picked randomly.

3.5. Mutation

Mutation operation is used to maintain population diversity by changing the value of a random point of an individual [16]. In GA with two-dimensional individual representation, mutation can occur at a random point of a matrix. However, since our objective utilizes a column constraint, any changes in any random cell of the matrix will affect the sum of specific column associated to the modified cell. Therefore, instead of changing single cell, we modified all cells in a randomly chosen single column. The modification takes into consideration the targeted value defined by column constraint.

4. Experimental Results

The proposed algorithm was implemented in Python on an Intel Core-i5 PC. It was run to solve junior school location-allocation optimization problem in South Jakarta. There are 65 schools and 65 urban villages in South Jakarta, hence the individual is represented in 65×65 matrix. Table II shows the summary of experimental parameters and results.

The GA program was run 50 times. That means there were 50 solutions obtained. We picked 30 as population size and 5000 as number of maximum iterations for each running. This value is obtained after running the program several times with several combinations of parameter values. Minimum average of fitness values is obtained with this parameter combination. The average time for each run is about 104 seconds.

Table 2: Summary of experimental result			
Number of running	50		
Population size	30		
Number of iterations	5000		
Running time average	104.6630200005		
Minimum fitness value	132,483.056787		
Maximum fitness value	139,902.715935		
Average fitness value	136,018.738254		

Among 50 solutions, the minimum fitness value obtained is 132,483.056787, while the maximum fitness value obtained is 139,902.715935. The average of all solutions is 136,018.738254. If we compare the travel cost obtained with the assumption of no zoning policy as calculated in section 2, i.e. 222,857, and the the minimum travel cost obtained through the proposed algorithm, i.e. 132,483, we obtained the total reduced travel cost is 90,374 or 40.55%.

No	School Name	Allocation	Current Capacity	Difference	Percentage
1	SMP NEGERI 37	571	768	197	25,65
2	SMP NEGERI 68	688	832	144	17,31
3	SMP NEGERI 85	575	768	193	25,13
4	SMP NEGERI 86	660	768	108	14,06
5	SMP NEGERI 96	1095	672	-423	-62,95
6	SMP NEGERI 226	371	768	397	51,69

Table 3: Detailed school allocation for district of Cilandak

If we look further into the detailed allocation, there are several recommendations that could be made. For example, Table III shows the detailed school allocation for district of Cilandak based on the solution with minimum fitness value. Based on that allocation, SMP Negeri 96 has lack of capacity, while the rest of the schools has over-capacity. SMP Negeri 96 lacks of more than 50% of its current capacity, hence it is highly recommended to add the capacity of SMP Negeri 96 or open a new school near SMP Negeri 96. Moreover, SMP Negeri 226 has over-capacity more than 50% than its current capacity, hence it might indicate inefficiency of facility usage.

5. Conclusion

In this study, genetic algorithm approach is developed to solve school location-allocation optimization problem. Two-dimensional individual representation is utilized along with SUS selection, one-point crossover, and column mutation method. Junior school location-allocation optimization problem in South Jakarta is used as case study. The GA program was run for 50 times and the individual with minimum fitness value was chosen as the optimum solution. The results show that the obtained solution lowers student travel cost by 40.55% compared to previous allocation which assumed that students tend to choose favourite schools located in the center of regency.

This study has not included consideration of current school capacity. This could be applied as improvement for future studies. Moreover, different method for parameter tuning can be applied to the proposed GA to show its efficiency and applicability. Different metaheuristic or hybrid metaheuristic method can also be examined and compared to the proposed GA.

6. Acknowledgements

The authors would like to thank DKI Jakarta Department of Population and Civil Registration for providing the data of South Jakarta residents aged 13-15 years for this research. This research is supported by the Indonesia Endowment Fund for Education (LPDP) as scholarship provider from the Ministry of Finance, Indonesia.

7. References

- Kepala Dinas Pendidikan Provinsi Daerah Khusus Ibukota Jakarta. (2017). Keputusan Kepala Dinas Pendidikan Provinsi Daerah Khusus Ibukota Jakarta Nomor 494 Tahun 2017 Tentang Penetapan Zona Sekolah Dalam Rangka Pelaksanaan Penerimaan Peserta Didik Baru Tahun Pelajaran 2017/2018.
- [2] M. R. Alifi, H. Hayati, and S. H. Supangkat. "Optimization of school network using location-allocation analysis: Case study: Bandung." in IEEE Region 10 Symposium (TENSYMP), 2017. http://doi.org/10.1109/TENCONSpring.2017.8070099
- [3] F. Ndiaye, B. M. Ndiaye, and I. Ly."Application of the p-Median Problem in School Allocation." in American Journal of Operations Research, 2012, 2, pp. 253-259. http://dx.doi.org/10.4236/ajor.2012.22030
- [4] M. S. Daskin, and K. L. Maass. "The p-Median Problem." in Location Science, Switzerland: Springer International Publishing, 2015, pp. 2145. http://doi.org/10.1007/978-3-319-13111-5
- [5] M. B. Teitz and P. Bart, "Heuristic methods for estimating generalized vertex median of a weighted graph." Oper Res, 1968, 16, pp. 955-961.
- [6] K. Ghoseiri and S. F. Ghannadpour. "An efficient heuristic method for capacitated P-Median problem." in International Journal of Management Science and Engineering Management, 2009, Vol. 4, No. 1, pp. 72-80.
- [7] Sin C. Ho. "An iterated tabu search heuristic for the Single Source Capacitated Facility Location Problem." in Applied Soft Computing, 2015, 27, pp. 169178.
- [8] J. Teixeira, A. Antunes, and D. Peeters. "An optimization-based study on the redeployment of a secondary school network." Environment and Planning B: Planning and Design, 2007, 34(2), pp. 296315. http://doi.org/10.1068/b31173
- [9] Xin-She Yang. Nature-Inspired Metaheuristic Algorithms Second Edition. United Kingdom: Luniver Press, 2010.
- [10] Kementerian Pendidikan dan Kebudayaan. "Indonesia Educational Statistics in Brief 2015/2016." Internet: http://publikasi.data.kemdikbud.go.id/uploadDir/isi AA46E7FA-90A3-46D9-BDE6-CA6111248E94 .pdf, 2016 [Dec. 21, 2017].
- [11] J.H. Holland. "Adaptation in natural and artificial system", University of Michigan Press, Ann Arbor, 1975.
- [12] D.E. Goldberg. "Genetic algorithms in search, optimization and machine learning", Addison-Wesley, Reading, MA, 1989.
- [13] J. Bhadury, J. Jaramillo, and R. Batta. "On the Use of Genetic Algorithms for Location problems." Computers and Operations Research, 2002, Vol. 29, pp. 761-779.
- [14] M. Mangla, R. Akhare, and S. Ambarkar. "Implementing Genetic Algorithm to solve Facility Location Problem" in International Research Journal of Engineering and Technology (IRJET), 2015, Vol. 02, Issue: 05, pp. 781-784.
- [15] J. E. Baker. "Reducing Bias and Inefficiency in the Selection Algorithm" in Proceedings of the Second International Conference on Genetic Algorithms and their Application, Hillsdale, New Jersey: L. Erlbaum Associates, 1987, pp. 1421.
- [16] M. W. Tsai, T. P. Hong, and W. T. Lin. "A Two-Dimensional Genetic Algorithm and Its Application to Aircraft Scheduling Problem" in Mathematical Problems in Engineering, 2015, Hindawi Publishing Corporation, Volume 2015, Article ID 906305. http://dx.doi.org/10.1155/2015/906305