

An Effective Chromosome Representation on Proportional Tuition Fees Assessment Using NSGA-II

Farid Jauhari¹, Wayan Firdaus Mahmudy², Achmad Basuki³
^{1,2,3}Faculty of Computer Science, Brawijaya University, Indonesia
{faridjauhari¹, wayanfm², abazh³}@ub.ac.id

Received 27 November 2019; accepted 03 Desember 2019

Abstract. Proportional tuition fees assessment is an optimization process to find a compromise point between student willingness to pay and institution income. Using a genetic algorithm to find optimal solutions requires effective chromosome representations, parameters, and operator genetic to obtain efficient search. This paper proposes a new chromosome representation and also finding efficient genetic parameters to solve the proportional tuition fees assessment problem. The results of applying the new chromosome representation are compared with another chromosome representation in the previous study. The evaluations show that the proposed chromosome representation obtains better results than the other in both execution time required and the quality of the solutions.

1 Introduction

Proportional tuition fees or in the Indonesian language called Uang Kuliah Tunggal (UKT) assessment is an optimization process to find a compromise point between student willingness to pay and institution income. Institution income cannot be less than its minimum operational funds. Ministerial regulation about UKT (Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia that is abbreviated as permenristekdikti) number 22 year 2015 ruled institutions to give category I and II for minimum 5% out of total students in each study program [1]. Category I and II are the two cheapest tuition fees. Besides student willingness to pay, the assessment result cannot contain collisions of the financial ability range in each category. Those things make proportional tuition fees assessment problem more complicated.

The proportional tuition fees assessment problem is indeed a multi-objective optimization problem [2]. An algorithm that is often used to solve multi-objective problems is Non-dominated Sort Genetic Algorithm II (NSGA-II) [3]. Although [2] has produced a feasible solution, the modeling of NSGA-II still requires considerable time in finding a feasible solution. As in [4], shorter chromosome representation is expected to reduce the processing time of NSGA-II to solve multi-objective problems of proportional tuition fees assessment.

Many studies have proven that optimization problems can be solved more efficiently by an effective chromosome. A newly defined chromosome solved flexible job shop scheduling problems (FJSP) with less processing time and more quality of solutions

[4]. Genetic algorithm with chromosome representation of real numbers (RCGA) is proven outperform other methods on solving FJSP, despite using simple genetic operators [5], [6]. RCGA was also showing excellent performance in solving knapsack problems [7]. A permutation chromosome representation employed to solve multiple sequences of alignment of bioinformatics [8]. NSGA-II with a new problem-specific chromosome makes significant performance improvement in oil refinery scheduling [9] and production-distribution optimization [10]. Better performance also produced by problem-specific chromosome on the base station placement in cellular networks [11]. Variable-length chromosome outperforms fixed-length in satellite constellations [12] and road traffic coordination as a multipath optimization problem [13]. A dual chromosome is better than a single chromosome in the optimization of resource allocation in container-based clouds [14]. In this paper, we proposed a shorter problem-specific chromosome than [2], which faster and produces a better quality of the solutions.

This paper is organized as follows: Section 1: explain the background of our study. Section 2 introduces the problem formulation. Section 3 describes a chromosome representation described in [2] and our proposed chromosome representation along with the chromosome specific-initialization method. Section 4 compares the results of the chromosome representations using the same dataset as in [2]. Finally, Section 5 summarizes and analyses the strengths and weaknesses of our representation.

2 Problem Formulation

Notations that used on proportional tuition fees assessment problems are defined as in Table 1.

Table 1 List of tuition fees assessment notation

Symbol	Meaning
$k = 1,2,3,4,5,6$	Index of UKT category
$m = 1,2,\dots M$	Index students
$s = 1,2,3,\dots S$	Index of the study programs
S_m	The study program of student m
U_m	The tuition fee of student m
U_k	The tuition fee of category k
E_m	Financial ability of student m
H	Minimum operational funds
N_s	Number of students at study program s
$K_m = \{1,2,3,4,5,6\}$	UKT category for student m
NK_k	Number of students at study program s and category k

The fitness (objective) function of the multi-objective optimization of the tuition fees assessment can be seen in equations (1) and (2). Equation (1) is a fitness function for the distance between the tuition fee and the financial ability of students (willingness to pay). Equation (2) is a fitness function that represents the total amount of student's tuition fees or can be referred to as institution income.

$$\max f_1 = -\frac{\sum_{m=1}^M |U_m - E_m|}{M} \quad (1)$$

$$\max f_2 = \sum_{m=1}^M U_m \quad (2)$$

In addition to the fitness function, there are also optimization constraints. The first constraint is to ensure the total tuition fees of students are not less than the minimum institution's operational funds. The first constraint can be formulated in formula (3). The number of students in each tuition category is formulated in equation (4). The second and third constraints ensure that category I and category II not less than 5% of the total number of students accepted at each study program. The second and third constraints are entered into limits based on the Regulation of the Minister of Research, Technology, and Higher Education of the Republic of Indonesia (PERMERISTEKDIKTI) number 22 year 2015. The second and third constraints can be seen in formula (5) and (6), respectively. Formula (7) ensures that the maximum financial ability of students (E_m) with category k in the study program s must be greater than the minimum financial ability of students (E_m) with category $k-1$ (the tuition category below). From defining the fitness function (1) and (2), and the constraint function (3), (5) - (7) it can be formulated as a whole optimization function can be seen in formula (8).

$$g = \sum_{m=1}^M U_m - H > 0 \quad (3)$$

$$NK_{ks} = \sum_{m=1}^M ((K_m = k) * (S_m = s)) \quad (4)$$

$$h = NK_{1s} - (5\% * N_s) \geq 0 \quad (5)$$

$$i = NK_{2s} - (5\% * N_s) \geq 0 \quad (6)$$

$$j = \text{Min}E_{ks} - \text{Max}E_{k-1s} > 0 \quad (7)$$

$$\begin{aligned} \max F &= (f_1, f_2)^T \\ \text{untuk } g &> 0 \\ h &\geq 0 \\ i &\geq 0 \\ j &> 0 \end{aligned} \quad (8)$$

3 Chromosome Representation and Initialization Method

3.1 Old Chromosome Representation (K1)

Old chromosome as in [2] represented as M (number of students) array of integer. Gen represents students, and the value of gen represents the tuition fees category. Chromosome representation for M students is shown in Fig. 2.

4	3	2	6	1	3	2	1	...	3
1	2	3	4	5	6	7	8	...	M

Fig. 1 Old chromosome representation (K1) for M students

3.2 Proposed Chromosome Representation (K2)

The proposed chromosome representation (further called K2) uses real values representation. The genes on the chromosomes represent the upper limit of the financial ability of students, as can be seen in Fig. 2. The chromosome representation for one

study program and the number of UKT categories is five can be seen in Fig. 2. So for the assessment of student's UKT in a faculty with three study programs, the chromosome representation can be seen in Fig. 3.

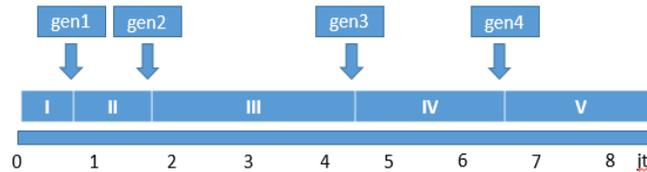


Fig. 2 Illustration of proposed chromosome (K2)

0.75	1.8	5.5	6.7	0.5	1.2	3	5	0.5	1.8	3.5	5.4
Study Program A				Study Program B				Study Program C			

Fig. 3 Proposed chromosome representation (K2) for 3 study programs

3.3 Initialization Method Of Proposed Chromosome Representation (K2)

A special initialization method is needed for K2. This initialization method utilizes Pseudo-Random and it is named NCPR-Seq (Normalized Cumulative Pseudo-Random Sequence) initialization. NCPR-Seq steps can be seen in **Error! Reference source not found.** NCPR-Seq initialization method for one study program is done by generating a number of K random numbers (Pseudo Random which can be abbreviated as PR). Then the cumulative series of K numbers are formed based on the K random numbers that have been generated. The cumulative series formed is then normalized. The normalization process is carried out by dividing the value of the cumulative series by its highest value (the total value of the random K number). The final process is to multiply the normalized cumulative series with U6s, which is the highest tuition fee price in the study program. The NCPR-Seq used for forming chromosomes is only a number of K-1. The NCPR-Seq generation process is repeated in S study programs form a complete chromosome. So the length of the chromosome is $S * (K-1)$.

4 Result and Evaluation

The proposed chromosome representation (K2) will compare to the chromosome representation in the study [2], which is further called K1. Chromosome representations comparison on population size are performed using a maximum of 3000 iterations, and one-cut-point crossover and simple-random mutation with crossover rate (cr) = 0.5 and mutation rate (mr) = 0.5. Test results for popSize 40 to 200 shown at **Error! Not a valid bookmark self-reference.** shows that K1 does not produce a feasible solution up to popSize 200. It also shows that K2 has a faster time, even though it only has a small difference. The average value of f1 and f2 from K2 is better than K1. Because the average of f1 and f2 both K1 and K2 still tends to increase, to get the best popSize value, this test is repeated for popSize 400 to 2000 with the addition of 400.

Table 2.

Error! Not a valid bookmark self-reference. shows that K1 does not produce a feasible solution up to popSize 200. It also shows that K2 has a faster time, even though it only has a small difference. The average value of f1 and f2 from K2 is better than K1. Because the average of f1 and f2 both K1 and K2 still tends to increase, to get the best popSize value, this test is repeated for popSize 400 to 2000 with the addition of 400.

Table 2 Result comparison between K1 and K2 when using popSize 40-200

popSize	FeaSol		Time (s)		Avg f1		Avg f2	
	K1	K2	K1	K2	K1	K2	K1	K2
40	0	40	3.701	3.720	-4.695	-2.375	550.781	663.038
80	0	80	8.094	7.648	-4.329	-2.364	591.341	665.274
120	0	120	12.254	11.790	-4	-2.364	617.745	665.402
160	0	160	18.519	17.107	-3.891	-2.364	625.123	664.700
200	0	200	24.137	23.483	-3.731	-2.375	637.547	667.403

Table 3 Result Comparison between K1 and K2 when using popSize 400-2000

popSize	FeaSol		Time (s)		Avg f1		Avg f2	
	K1	K2	K1	K2	K1	K2	K1	K2
400	0	400	64.332	69.036	-3.336	-2.384	665.256	670.685
800	0	800	234.380	258.489	-3.022	-2.402	696.249	674.196
1200	0	1200	546.551	614.489	-2.916	-2.399	710.791	673.958
1600	0	1600	959.785	1326.240	-2.843	-2.398	722.587	674.402
2000	0	2000	1537.026	2039.133	-2.788	-2.408	726.803	676.107

Table 3 indicates that K1 still does not produce a feasible solution. Likewise, K2 still produces feasible solutions as much as popSize. As is well known that the length of the K2 chromosome is shorter than K1. The reason for K2 having a longer execution time, albeit with shorter chromosomes, is because the mapping process between students and their tuition categories (Um) is based on the limits of the financial ability of students (Em) of each category implicit in the chromosome configuration. The mapping occurs before the fitness and the constraint calculation process occurs. As for K1, the chromosome configuration has stated the UKT category of students (Km).

Table 3 shows the tendency of the increasing mean of f2 and decreasing of the mean of f1 as popSize increases. In general, the higher the popSize, the more variations it will produce in one iteration. However, the greater popSize will also add a significant

execution time. So it can be concluded that the value of popSize is good and will be used for further testing is 200.

It can be seen in Table 4, that the mean of f1 and f2 tends to increase of the maximum generation (maxGen) 10 to 100, then decrease for the remaining maxGen values. Besides that, we can see that the NSGA-II obtained a number of popSize feasible solutions even when maxGen = 9. So it can be concluded that the efficient maxGen value is 100.

Table 4 Test result of K2 in the different maximum generation (MaxGen)

maxGen	Avg f1	Avg f2	Avg time	Avg FeaSol
3000	-2.370	667.421	17.599	200
2000	-2.375	667.659	11.872	200
1000	-2.370	666.920	5.881	200
800	-2.370	666.938	5.485	200
600	-2.370	667.381	3.608	200
400	-2.364	666.220	2.522	200
200	-2.364	666.322	1.603	200
100	-2.364	667.210	1.053	200
80	-2.364	666.669	0.936	200
60	-2.364	666.293	0.834	200
40	-2.364	665.174	0.710	200
20	-2.370	665.430	0.595	200
10	-2.381	665.783	0.548	200
8	-2.375	662.871	0.534	196.8
6	-2.381	664.127	0.518	181.15
4	-2.381	662.993	0.491	91.7
2	-2.381	662.103	0.483	45.9

5 Conclusion

The test results show that the modified NSGA-II has succeeded in optimizing the problem of determining UKT. The best parameters for multi-objective optimization in determining UKT are the number of population (popSize) 200 and the maximum iteration = 100. The modified NSGA-II produced a 200 solution (the same as popSize). These solutions are vary. It starting from solutions that tend to prioritize objectives from the student's point of view (fitness 1 has the highest value). Until solutions that are very beneficial to the institution

(fitness 2 has the highest value). With the variety of solutions produced, the modified NSGA-II can assist decision-makers in choosing UKT determination solutions.

Based on the test results, it is known that the most effective chromosome representation is the second chromosome representation (K2). It is happening because K2 achieves the best fitness value with shorter execution time. K1 is very difficult to find a feasible solution because of the vastness of the search universe or in other words, too many variations of the resulting combination. Whereas K2 further limits the search area or narrows the resulting combination by reducing the number of genes from chromosomes. However, the variation in the value of one gene is still continuous (not discrete).

In order to permit cross referencing within LNCS-Online, and eventually between different publishers and their online databases, LNCS will, from now on, be standardizing the format of the references. This new feature will increase the visibility of publications and facilitate academic research considerably. Please base your references on the examples below. References that don't adhere to this style will be reformatted by Springer. You should therefore check your references thoroughly when you receive the final pdf of your paper. The reference section must be complete. You may not omit references. Instructions as to where to find a fuller version of the references are not permissible.

References

1. MENRISTEKDIKTI Republik Indonesia, Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 22 Tahun 2015 Tentang Biaya Kuliah dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri Di Lingkungan Kementerian Riset, Teknologi dan Pendidikan. Indonesia, 2015.
2. F. Jauhari, W. F. Mahmudy, and A. Basuki, "Multi-Objective Optimization for Proportional Tuition Fees Assessment Using Non-Dominated Sorting Genetic Algorithm II (NSGA II)," 3rd Int. Conf. Sustain. Inf. Eng. Technol. SIET 2018 - Proc., pp. 292–297, 2019.
3. A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang, "Multiobjective evolutionary algorithms: A survey of the state of the art," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 32–49, 2011.
4. J. C. Tay and D. Wibowo, "An effective chromosome representation for evolving flexible job shop schedules," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3103, pp. 210–221, 2004.
5. W. F. Mahmudy, R. M. Marian, and L. H. S. Luong, "Real Coded Genetic Algorithms for Solving Flexible Job-Shop Scheduling Problem - Part I: Modelling," *Adv. Mater. Res.*, vol. 701, pp. 359–363, May 2013.
6. W. F. Mahmudy, R. M. 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, May 2013.
7. N. Rikatsih, W. F. Mahmudy, and S. Syafrial, "Hybrid Real-Coded Genetic Algorithm and Variable Neighborhood Search for Optimization of Product Storage," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, p. 166, 2019.
8. M. T. Ben Othman and G. Abdel-Aziz, "Multiple sequence alignment based on genetic algorithms with new chromosomes representation," *Proc. Mediterr. Electrotech. Conf. - MELECON*, pp. 1030–1033, 2012.
9. Y. Hou, N. Q. Wu, M. C. Zhou, and Z. W. Li, "Pareto-optimization for scheduling of crude oil operations in refinery via genetic algorithm," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 47,

- no. 3, pp. 517–530, 2017.
10. R. Rody, W. F. Mahmudy, and I. P. Tama, “Using Guided Initial Chromosome of Genetic Algorithm for Scheduling Production-Distribution System,” *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 1, p. 26, 2019.
 11. Y. Liu, W. Huangfu, H. Zhang, H. Wang, W. An, and K. Long, “An Efficient Geometry-Induced Genetic Algorithm for Base Station Placement in Cellular Networks,” *IEEE Access*, vol. 7, pp. 108604–108616, 2019.
 12. N. Hitomi and D. Selva, “Constellation optimization using an evolutionary algorithm with a variable-length chromosome,” *IEEE Aerosp. Conf. Proc.*, vol. 2018-March, pp. 1–12, 2018
 13. L. Cruz-Piris, I. Marsa-Maestre, and M. A. Lopez-Carmona, “A Variable-Length Chromosome Genetic Algorithm to Solve a Road Traffic Coordination Multipath Problem,” *IEEE Access*, vol. 7, pp. 111968–111981, 2019.
 14. B. Tan, H. Ma, and Y. Mei, “Novel Genetic Algorithm with Dual Chromosome Representation for Resource Allocation in Container-Based Clouds,” *2019 IEEE 12th Int. Conf. Cloud Comput.*, pp. 452–456, 2019.