Journal of Algorithms and Computation

journal homepage: http://jac.ut.ac.ir

Developing a New Genetic Algorithm for Selecting Efficient Project Portfolio

Mohammad Mirabi*[1] and Mohammad Reza Zare Banadkouki †[2]

[1,2]Department of Industrial Engineering, Faculty of Engineering, Meybod University, Meybod, Iran

ABSTRACT

selecting an efficient project portfolio among available projects is a vital decision for any project manager. The main questions are which projects can have more long-term benefit for manager or organization. Due to the complexity of this field of research, todays so many approaches are developed for project selection. Calculation time and the quality of result are two main criterions that almost all researchers have considerate on them. In this research a new hybrid genetic algorithm with new heuristic mutation and cross over are developed to choosing a good portfolio of available projects. Presented algorithm is fast and effective to reach the good result in reasonable time. Finding a good point to start as initial population and using good operator a heuristic mutation and cross over are main points of our algorithm. To check

Keyword: Project Portfolio; New Genetic Algorithm; Heuristic Mutation; Heuristic Cross over.

ARTICLE INFO

Article history:
Research paper
Received 03 January 2022
Received in revised form 17,April 2022
Accepted 12 May 2022
Available online 01, June 2022

AMS subject Classification: 05C78.

*Corresponding author: M. Mirabi. Email: mirabi@meybod.ac.ir
†mr.zare@gmail.com

1 Abstract continued

the quality of results we compare developed algorithm with some recent ones in the literature and comparison studies and statistical calculation demonstrate the efficiency of the new genetic algorithm to select a good portfolio.

2 Introduction

Selecting process is an important issue in our life. Also in management science selecting a good choice between all available options is a vital expectation of a manager. A project manager always face with the selection between projects. Selection models (or guidelines) help managers pick winners from a pool of projects. Screening models are numeric or nonnumeric and should have:

- Realism: reflect constraints and organizational goals
- Capability: widely useable, robust
- Flexibility: easy to modify
- Ease of use: useable by many organizational members
- Cost effectiveness
- Comparability

The project selection decision belong to the multi decision criteria scope and there are several criterions to select the best portfolio between all available projects. In the real world this field of study called as NP hard problem based on its complexity [9, 12].
Problems can be categorized into two main classes of P and NP based on the computational theory. Problem of complexity class of P can be solved by a deterministic algorithm in polynomial time. Then it is relatively easy to solve. Minimum spanning tree, shortest path problems, maximum flow network, maximum bipartite matching, and linear programming continuous models are belonging to P class. Problem of complexity class of NP can be solved by a nondeterministic algorithm in polynomial time. A decision problem belonging to NP is NP complete if all other problems of class NP are reduced in polynomial time. NP-hard problems are optimization problems whose associated decision problems are NP-complete. Most of the real-world optimization problems and many academic popular problems are NP-hard. Project selection is an example of of NP hard problems [4, 9]. All models only partially reflect reality and have both objective and subjective factors imbedded. For NP-hard problems provably efficient algorithms do not exist and therefore meta-heuristics in pure and hybrid structure have wide applications to solve this kind of problems. There are so many approaches to project screening as:

- Checklist model
- Simplified scoring models
- Analytic hierarchy process
- Profile models
- Financial models
- Heuristics and meta-heuristics

The systematic process of selecting, supporting, and managing the firm's collection of projects. Portfolio management requires:

- decision making,
- prioritization,
- review,
- realignment,
- reprioritization of a firm's projects.

A meta-heuristic is a higher-level heuristic designed to find, generate, or select a heuristic that may provide a good solution to an optimization problem by combining intelligently different concepts for exploring and exploiting the search space especially with incomplete or imperfect information or limited computation capacity [11].

3 Literature review

Both qualitative and quantitative criteria are considered n project selection methods [10]. In the real world, there are many criteria affected to selection process. Therefore, this problem in real size can be categorized in NP hard and using heuristic and meta-heuristic approach can be effective to solve them.
One of the pioneer works in this area is due to Carazo et al. (2010) that developed a multi-objective model for portfolio project selection with the set of objectives pursued by the organization [2]. Scatter search based meta-heuristic was used to solve their problem. Genetic algorithm as a powerful meta-heuristic used by Nikkhahnasaba and Najafi (2013) to select a good portfolio and they considered the net present value of the project portfolio as an objective function [9]. They used almost a basic genetic to solve their problem. Interactive based heuristic approach developed by Nowak (2013) to solve a single portfolio in each iteration [10].
Another meta-heurist in this area is ant colony like what Fernandez et al.(2015) was used to solve multi-objective project portfolio optimization problem [6]. They also

used almost the basic approach but focused a fuzzy outranking preference model. Esfahani et al. (2016) solved project selection problem by considering a new definition of modern portfolio theory [5]. They developed a novel heuristic based on search spaces. Multi objective optimization is considered by Brester et al. (2017). They used island meta-heuristic to solve their problem [1].
Teaching-learning-based optimization based meta-heuristic developed by Kumar et al. (2018) they focused on the benefit of portfolio to select the best projects [7]. Risk has week consideration in their research. Variable neighborhood search is widely applied for heuristic and meta-heuristic approaches like what Panadero et al. (2018) developed in their research [12]. They considered simulation-optimization algorithm to solve the project selection problem. Multi-objective problem is another wide issue in this field of research. Tofighian et al. (2018) modeled project selection problem based on net profit of portfolio but tried to take multi objective into account. Genetic algorithm was a meta-heuristic that they used [13].
Multi-criteria decision-making model is used by Davoudabadi et al. (2019). They considered linear and fuzzy approaches for project selection problem [3].
To reach more studies in this field readers can refer Mohaghehgi et al. (2019) that reviewed wide researches on project portfolio selection area [8].
Keys to Successful Project Portfolio Management are

- Flexible structure and freedom of communication
- Low-cost environmental scanning
- Time-paced transition

Genetic is adaptable with above factors and as it can observed in the literature, genetic have wide consideration in project selection problem and it stimulated us to develop one agile genetic to solve the problem and reach a good portfolio.
In this research we develop one novel genetic algorithm based on novel mutation and cross over for the project selection problem.

4 Mathematical symbols and Model

At first parameters of the model must be defined.
R_i: expected return of project i
r_i: estimated risk of project i (total weight of of all kind of risks)
c_i: total cost of project i
B: in-hand budget
N_{max}, N_{min}: maximum and minimum number of projects allowed in the portfolio (if applicable).
V_{max}: upper level of acceptable risk (project manage or investor must determine)
R_{min}: lower level of expectable return (project manage or investor must determine)

n: total number of available projects
T: total number of time periods
I_i: binary decision variable
Based on the parameters, mathematical model can be defined.

$$maxR = \sum_{i=1}^{n} R_i \times I_i \qquad (1)$$

$$minV = \sum_{i=1}^{n} (r_i \times I_i) \qquad (2)$$

Subject to:

$$\sum_{i=1}^{n} c_i \leq B \qquad (3)$$

$$R \geq R_{min} \qquad (4)$$

$$V \leq V_{max} \qquad (5)$$

$$N_{min} \leq \sum_{i=1}^{n} I_i \leq N_{max} \qquad (6)$$

$$I_i = \begin{cases} 1 & if \ \ project\ i\ is\ selected \\ 0 & if\ project\ i\ is\ not\ selected \end{cases} \qquad i = 1,\ 2,\ \ldots,\ n \qquad (7)$$

Equation (1) maximize average expected of return of portfolio. Equations (2) minimize overall risk of portfolio. We assumed that investors are risk averse, meaning that given two portfolios that offer the same expected return, investors will prefer the less risky one. Thus, an investor will take on increased risk only if compensated by higher expected returns. Conversely, an investor who wants higher expected returns must accept more risk. The exact trade-off will be the same for all investors, but different investors will evaluate the trade-off differently based on individual risk aversion characteristics. Equations (3) to (5) are about maximum available budget, minimum expected return of portfolio and maximum acceptable risk of portfolio. And equation (7) is about binary parameter.

5 THE NEW GENETIC ALGORITHM

Developed genetic algorithm is based on heuristic mutation and heuristic cross over that make our algorithm more quick and efficient. At first one pool of initial solution is generated and each chromosome moves toward the better solution in various search

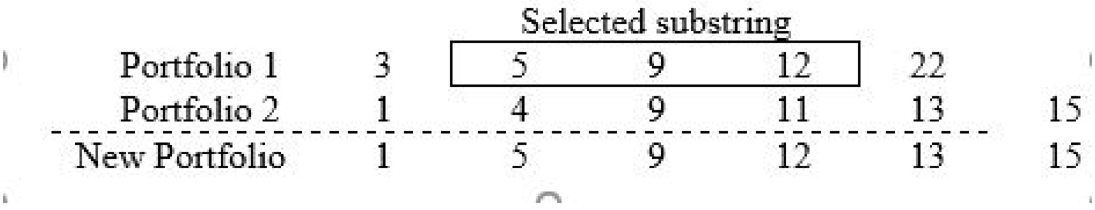

Figure 1: The order crossover operator

spaces. We used variable neighborhood search to cover approximately all spaces. Bellow steps are done to generate the initial solution.
Set I_i=0 for i=1,2,... n
For i=1: n
Used random integer generator to earn k (integer between N_{min} and N_{max})
For j=1: k
Extract one project among available ones that has more expected return of min estimated risk. Eliminate it from available projects and put it to I_i
End for
End for
Extract all I_i i=1,2,... n
After extracting all initial solution, they are evaluated by the fitness function as follow (weighted arithmetic mean of two presented objective):

$$Fitness\ Function_k\,(FF_k) = w_1 \times R - w_2 \times V \qquad (8)$$

Project manager or investor must determine w_1 and w_2 based on their preferences. It is clear that sum of them must be equal to 1.
Heuristic mutation and order crossover operations are described in figure 1 and 2. In figure 1, two portfolios (called 1 and 2) are selected by random. The number of projects in each portfolio maybe different (N_1 and N_2). One random integer called g is selected between 1 and n=min(N_1 and N_2) and in new earned portfolio, g projects are selected from portfolio 1 and others are selected from portfolio 2. It is clear that the size of portfolio is maximum size of its parent. First position of first project in portfolio 1 is selected by random but must be less that its size minus g and others are selected respectively. If one project is repeated in new portfolio it is must be replaced by random. Figure 1 illustrate this operator. In figure 1, projects numbered 5, 9 and 12 are selected from portfolio 1 and projects numbered q, 13 and 15 are selected from portfolio 2 to earn a new portfolio. It can be seen the size of new portfolio is equal to maximum size of its parents (6).
In heuristic mutation operator, one portfolio is selected by random and one random integer called n is selected between 1 and the size of portfolio. n projects are selected from portfolio by random and replaced by available projects to earn some new port-

Select three genes randomly

Portfolio	$\boxed{5}$	9	$\boxed{10}$	11	$\boxed{18}$
Portfolio 1	1	9	16	11	12
Portfolio 2	3	9	17	11	15
Portfolio 3	8	9	21	11	17
Portfolio 4	2	9	2	11	22
Portfolio 5	13	9	4	11	3
Best one	1	9	11	12	16

Figure 2: The heuristic mutation operator

folios (m portfolios). Best portfolio based on fitness function (FF) is extracted from all portfolios. Figure 2 illustrate this operator. In this figure, 3 projects are replaces to earn new portfolio. Finally first earned portfolio are the best based on FF.
Heuristic operators can enrich genetic to find better offspring in acceptable time like what done by Mirabi (2014). This performance stimulate us to use this approach.
Now we can illustrate the new genetic algorithm as follow:
m: number of initial portfolios
n: number of iterations to earn new portfolio
Generated m initial portfolios by random
Select the best portfolio based on the fitness function and called it as commander (commander = arg max (fitness))
For i=1:n
Generate some new portfolios randomly by heuristic mutation of commander or order crossover between commander and one other portfolio.
If new portfolio passes all constraints, compare it with commander based on FF. If new portfolio is better than it (based on fitness function) called it new commander and replace it with the worst portfolio in the group.
End for
Select the commander as the best for the final solution.

6 COMPARISON STUDY

There are many researches about project selection in the literature but for comparison stud we need some researches that is adaptable with our problem. Two work are extracted from the literature. Variable neighborhood search or VN by Panadero et al. (2018) and taboo-search algorithm or TS by Kumar et al. (2018). The reason of this is about to compatibility of their research to what we developed. In this section we are going to compare our algorithm with these two capable algorithms from the literature.

100 construction projects are considered. N_{min} is set to 5, 10, 20 and 30 and also N_{max} is set to 50, 70, 80 and 100. *PM* index is used for comparison study.

$$PM = \frac{Heu_{sol} - Best_{sol}}{Best_{sol}} \qquad (9)$$

Heu_{sol} is FF of a given algorithm and $Best_{sol}$ is the most FF among all algorithms. The MATLAB is used to cade all algorithm. For fair comparison, all methods are run 10 independent times. Stopping criterion is the number of iteration. As total, 16 class of problem and 160 problem instances are obtained.

Min, Max and the average PM of each algorithm are given in Table 1. For time comparison between algorithms, the average time to solve 10 instances are given. In Table 1 "Min" columns in subscript, show the the number of instances that Heu_{sol} was equal to the $Best_{sol}$. As show in Table1, New Genetic Algorithm (NGA) have better performance based on PM value.

Table 1: comparison of PM values between all algorithms (times are in second)

Class of problem	N_{max}	N_{min}	NGA				VN				TS			
			Min	Average		Max	Min	Average		Max	Min	Average		Max
			PM	PM	Time	PM	PM	PM	Time	PM	PM	PM	Time	PM
1	50	5	0_7	0.06	1.00	0.14	0_6	0.02	1.81	0.04	0_1	0.16	1.95	0.18
2	50	10	0_8	0.03	1.10	0.11	0_5	0.07	1.82	0.15	0_2	0.10	2.76	0.14
3	50	20	0_4	0.06	2.22	0.13	0_4	0.03	2.73	0.08	0.04	0.07	5.41	0.19
4	50	30	0_5	0.06	5.01	0.08	0_3	0.03	4.08	0.05	0_2	0.05	8.42	0.12
5	70	5	0_9	0.01	1.46	0.00	0_5	0.02	2.12	0.06	0_1	0.03	2.53	0.10
6	70	10	0_9	0.02	1.58	0.09	0_3	0.08	2.85	0.15	0_2	0.08	4.80	0.13
7	70	20	0_8	0.01	3.23	0.18	0_3	0.05	2.92	0.05	0.02	0.02	9.26	0.08
8	70	30	0_6	0.03	6.08	0.06	0_5	0.07	5.12	0.16	0_2	0.12	13.34	0.22
9	80	5	0_8	0.05	1.55	0.14	0_3	0.08	2.88	0.12	0_1	0.09	3.37	0.14
10	80	10	0_9	0.03	2.58	0.15	0_4	0.03	3.50	0.07	0_2	0.13	6.55	0.21
11	80	20	0_7	0.00	4.22	0.02	0_3	0.03	4.12	0.06	0.03	0.06	9.12	0.09
12	80	30	0_8	0.01	6.35	0.02	0_6	0.04	6.21	0.10	0.02	0.06	14.17	0.20
13	100	5	0_7	0.03	2.12	0.10	0_4	0.09	3.06	0.10	0_2	0.14	3.41	0.23
14	100	10	0_6	0.07	2.68	0.19	0_3	0.08	4.45	0.12	0.03	0.07	6.75	0.15
15	100	20	0_6	0.01	3.59	0.04	0_4	0.11	5.52	0.19	0.02	0.04	7.96	0.22
16	100	30	0_7	0.01	8.59	0.09	0_6	0.03	7.68	0.08	0.05	0.06	15.17	0.23

ANOVA test is a powerful statistical test to check that differences are statistically similar or not. The ANOVA procedure tests these hypotheses:

H0: $\mu 1 = \mu 2 = \mu 3$, all results are the same

H1: two or more results are different

Table 2: ANOVA test for all methods

	Some of Square (SS)	Degree of freedom (df)	Mean Square (MS)	VR	F
Between groups(or "Factor")	437739.15	3	177321.5	79.1	2. 7
Within groups(or "Error")	2621778.55	108	21724.67		
Total	2294438.20	111			

With the $\alpha = 0.05$ significance level, computations are shown in Table 2.
In Table 2 VR=9.1>F=2. 7, as a result, differences in Table 1 are not the same. Based on information in Table 1, NGA and VN have better performance based on PM. For deep comparison between them, it must be seen that the differences between results (FF) are statistically significant or not. For this, the hypothesis that the population corresponding to the differences has mean (μ) zero can be tested; specifically, test the (null) hypothesis $\mu = 0$ against the alternative$\mu > 0$. It is assumed that the differences between solutions (FF) is a Normal variable, and choose the significance level $\alpha = 0.05$.
If the hypothesis is true, the random variable $T = (\overline{X}_1 - \overline{X}_2)/\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}$ has a t distribution with:
$\upsilon = (S_1^2/n_1 + S_2^2/n_2)^2/(\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1})$ degrees of freedom. The critical value of c is obtained from the relation Prob($T > c$)= α = 0.05. Table 3 shown the results. For example, the first row of Table 3, corresponds to the sample size=$n_1 = n_2 = 10$, μ_0=0, sample mean for NGA and VN are$\overline{X}_1 = 62.86$and $\overline{X}_2 = 63.86$ respectively. Sample standard deviation for NGA and VNS are $S_1 = 2.37$and $S_2 = 2.61$ respectively. Since $t = 1.73 > 0.9$, we conclude that the difference is not statistically significant.
Based on Table 3 NGA outperformed VN in 68.7% of all classes and all of differences are statistically significant and VN outperformed NGA in 31.3% of all classes that in all cases, differences are statistically significant except one.
For deep comparison between NGA and VN Tukey honestly significance difference test can be used. It is a strong statistical tool to check significance by computing confidence interval similarly to the confidence interval for the difference of two means, but using the q distribution which avoids the problem of inflating α:

$$\bar{x}_i - \bar{x}_j \pm q(\alpha, r, df_w)\sqrt{\frac{MS_w}{2} \times \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Table 4 summarized the outputs of this test.

Table 3: Detail comparison between NGA and VN. Ave:Average, FF: fitness function, SD:Standard deviation, Sig:Significant. Each class contains 10 independent instances

Class of problem	N_{max}	N_{min}	Ave. FF or ($\overline{X}$)		Ave. SD or (S)		T	v	t	Sig.
			NGA	VN	NGA	VN				
1	50	5	62.86	63.86	2.37	2.61	-0.90	18	1.73	NO
2	50	10	57.67	60.45	2.49	2.78	-2.36	18	1.73	Yes
3	50	20	75.02	63.44	2.16	2.28	11.66	18	1.73	Yes
4	50	30	75.5	56.89	2.69	3.49	13.36	17	1.74	Yes
5	70	5	71.24	47.94	2.95	3.23	16.85	18	1.73	Yes
6	70	10	83.65	59.92	2.64	2.74	19.74	18	1.73	Yes
7	70	20	70.44	61.47	4.30	4.92	4.34	18	1.73	Yes
8	70	30	60.64	68.65	4.59	4.40	-3.98	18	1.73	Yes
9	80	5	72	65.00	4.21	4.46	3.61	18	1.73	Yes
10	80	10	82.3	65.97	3.77	4.58	8.71	17	1.74	Yes
11	80	20	87.09	64.23	4.21	3.91	12.59	18	1.73	Yes
12	80	30	77.99	62.97	3.59	4.30	8.48	17	1.74	Yes
13	100	5	63.42	67.99	4.83	4.97	-2.08	18	1.73	Yes
14	100	10	76.85	58.52	5.18	6.43	7.02	17	1.74	Yes
15	100	20	68.3	62.17	6.27	6.96	2.07	18	1.73	Yes
16	100	30	72.7	87.12	4.58	5.39	-6.45	18	1.73	Yes

Table 4: Tukey test results for NGA and VN

	$\hat{x}_{NGA} - \hat{x}_{VNS}$	Critical qq(α,r,df$_W$)	95% Conf Interval for $\mu_{HGA} - \mu_{LA}$		Significant at 0.05?
			Min	Max	
NGA-VNS	75.34	3.65	-59.12	61.15	Yes

Table 4 demonstrated that NGA outperforms VN.

7 CONCLUSION

In this research we developed one new genetic algorithm by heuristic mutation and cross over for project selection problem. Presented method is based on genetic algorithm meta-heuristic therefore, we called it as new genetic algorithm (NGA). Initial population is generated and some portfolios are gained with the size between N_{min}

and N_{max} and the best portfolio called commander. Each child (earned by mutation or crossover) challenge commander to substitute with worst member of solutions and be a new commander. After finite number of iterations, the best solution (commander) is the final solution. For the verification test, we compared developed algorithm with two recent developed algorithms in the literature as VN and TS. Based on the comparison study, NGA works very competitive to portfolio selection problems.

References

[1] Brester, C., Ryzhikov, I., Semenkin E., Multi-objective optimization algorithms with the island metaheuristic for effective project management problem solving, Organizacija, 50 (4) (2017), pp. 364-373.

[2] Carazo, A.F., Gómez, T., Molina, J., Hernández-Díaz, A.J., Guerrero, F.M., Caballero, R. Solving a comprehensive model for multi-objective project portfolio selection, Comput Oper Res, 37 (4) (2010), pp. 630-639.

[3] Davoudabadi, R., Mousavi, S. M., Šaparauskas, J., & Gitinavard, H. Solving construction project selection problem by a new uncertain weighting and ranking based on compromise solution with linear assignment approach. Journal of Civil Engineering and Management, 25(3)(2019), 241-251.

[4] El-Ghazali, Talbi, (2009). Meta-heuristics, from design to implementation, published by John Wiley & Sons, Inc., Hoboken, New Jersey.

[5] Esfahani, H.N., hossein Sobhiyah, M., Yousefi, V.R., Project portfolio selection via harmony search algorithm and modern portfolio theory, Procedia-Soc Behav Sci, 226 (2016), pp. 51-58.

[6] Fernandez, E., Gomez, C., Rivera, G., Cruz-Reyes L., Hybrid metaheuristic approach for handling many objectives and decisions on partial support in project portfolio optimization, Inf Sci, 315 (2015), pp. 102-122.

[7] Kumar, M., Mittal, M., Soni, G., Joshi, D., A hybrid TLBO-TS algorithm for integrated selection and scheduling of projects, Comput Indus Eng, 119 (2018), pp. 121-130.

[8] Mohagheghi, V., Mousavi, M., Antucheviciene, J., Mojtahedi, M. Project portfolio selection problem: a review of model, uncertainty approaches, solution techniques and case studies, Technological and Economic Development of Economy, 25(6) (2019), pp. 1380-1412.

[9] Nikkhahnasaba, M., Najaf, A., Project Portfolio Selection with the Maximization of Net Present Value, Journal of Optimization in Industrial Engineering 12 (2013), pp. 85-92.

[10] Nowak, M., Project Portfolio Selection Using Interactive Approach, Procedia Engineering, 57, (2013), PP. 814-822.

[11] Osman, I.H., & Potts, C.N.. Simulated annealing for permutation flow-shop scheduling. OMEGA, The International Journal of Management Science, 17 (1989), pp. 551–557.

[12] Panadero, J., Doering, J., Kizys, R., Juan, A.A., Fito, A., A variable neighborhood search simheuristic for project portfolio selection under uncertainty, J Heuristics (2018), pp. 1-23.

[13] Tofighian, A.A., Moezzi, H., Barfuei, M.K., Shafiee, M., Multi-period project portfolio selection under risk considerations and stochastic income, J Indus Eng Int, 14 (3)(2018), pp. 571-584.