

Efficient selection operators in NSGA-II for Solving Bi-Objective Multi-Skill Resource-Constrained Project Scheduling Problem

Paweł B. Myszkowski, Maciej Laszczyk, Joanna Lichodij
Wrocław University of Science and Technology
Computational Intelligence Department

Wybrzeże Stanisława Wyspiańskiego, 50-370 Wrocław, Poland
email:pawel.myszkowski@pwr.edu.pl, maciej.laszczyk@pwr.edu.pl, 203370@student.pwr.edu.pl

Abstract—This paper presents multiple variances of selection operator used in Non-dominated Sorting Genetic Algorithm II applied to solving Bi-Objective Multi-Skill Resource Constrained Project Scheduling Problem. A hybrid Differential Evolution with Greedy Algorithm has been proven to work very well on the researched problem and so it is used to probe the multi-objective solution space. It is then determined whether a multi-objective approach can outperform single-objective approaches in finding potential Pareto Fronts. Additional modified selection operators and a clone prevention method have been introduced and experiments have shown the increase in efficiency caused by their utilization.

I. INTRODUCTION

SCHEDULING problem plays an important role in today's science and business. It can be met in transportation [1], production [2], project management [3], etc. The problem itself can be informally defined as the function that aims to find the lowest duration and cost of the schedule by assigning resources to tasks. Multi-Skill Resource-Constrained Project Scheduling Problem (MS-RCPSP) is NP-hard and there are no methods capable of finding an optimal solution in polynomial time [4].

The goal of the research presented in this paper is to verify how Differential Evolution hybridized with Greedy (DEGR) and Non-dominated Sorting Genetic Algorithm II (NSGA-II) approaches explore space in the context of multi-objective optimization. DEGR algorithm is a single-objective method, and potential Pareto Front (*PF*) is created by running it multiple times. Each run uses different weights values in the fitness function. Set of points resulting from all runs is considered during evaluation. Moreover, a tournament selection method is investigated in NSGA-II to boost its selective pressure and a clone prevention method is implemented to increase the diversity of the resulting potential *PF*s. Results are evaluated and compared with a set of multi-objective measures. This paper presents the transition from a single to a multi-objective approach to MS-RCPSP and introduces modified selection operators, which have proven to increase efficiency of NSGA-II.

The rest of the paper is organized as follows. Section III lines the MS-RCPSP. Sections IV and V describe implemented methods - NSGA-II and DEGR appropriately. All experiments

along with its results have been presented in section VI. Section VII contains conclusions and directions of future work.

II. RELATED WORKS

There are many different types of scheduling problem models. PSPLIB library [5] is often used as a baseline to compare methods efficiency. It doesn't support a skill extension of the problem. Additionally, it comprises only one criterion - duration of the schedule, which makes it infeasible for the purpose of this article.

A Software Project Scheduling Problem (SPSP) is another example and was first presented in [6]. It is the most similar problem to the MS-RCPSP as it contains skills and two criteria - duration and cost. Additionally, it allows for tasks to be worked on by multiple resources. It has been more thoroughly described in [7].

Due to NP-hard nature of the problem, researchers have often tackled it with metaheuristics, which often provide satisfactory solution in acceptable time. Many methods have been developed to solve scheduling problems: Differential Evolution [8], Genetic Algorithm [9], Tabu Search [10], Grasp [11] and Teaching-learning-based optimization algorithm [14]. Additionally, swarm optimization techniques have been used: Ant Colony Optimization [12] or Particle Swarm Optimization [13].

The MS-RCPSP is defined as a multi-objective problem. The goal is to minimize both duration and cost of the schedule. It is often difficult to compare different criteria, so it is desired to find a set of equally-good solutions but with different objective values. An existence of populations in evolutionary algorithms perfectly fit the need to find multiple points on the *PF*. There are very few articles that deal with multi-objective MS-RCPSP. Simulated Annealing [15] and Genetic Algorithm [16] have been used for that purpose. The best known results for a single-objective MS-RCPSP have been achieved by a DEGR [8] method and therefore it is used in this paper.

NSGA-II has been proposed in [17] and has proven its efficiency in scheduling problems [18] [19]. The algorithm is often used as a benchmark approach for multi-objective problems.

III. FORMULATION OF MS–RCPSP

In MS–RCPSP [7] the schedule can be evaluated by both its cost and duration, making the problem multi-objective.

MS–RCPSP comprises a sets of tasks, resources and skills. Each task has a skill required to work on that task, a set of predecessors that have to be completed before the task can be started. Each resource has a set of skills and a cost associated with it. Skill is described by a type and level of expertise. The goal is to create task-resource assignments in a way that satisfies given constraints and minimizes both objectives. Detailedinition can found in [7].

IV. NON-DOMINATED SORTING GENETIC ALGORITHM II

Non-dominated Sorting Genetic Algorithm II (NSGA-II) first presented in [17] has proven to be effective approach to multi-objective optimization. It is based on a genetic algorithm, which utilizes populations for space exploration. The use of populations fits the problem very well, as the goal is to find a set of points. NSGA-II uses a comparison operator based on a domination described in formula 1 and a crowding distance operator, which aims to maximize the smallest box which comprises only one individual.

$$i \geq_n j \text{ if } (i_{rank} < j_{rank}) \vee ((i_{rank} = j_{rank}) \wedge (i_{distance} > j_{distance})) \quad (1)$$

where i and j are two compared individuals.

NSGA-II utilizes a non-dominated sorting. Individuals created by the genetic operators are added to the same population. At the end of a generation, an entire population is sorted according to the domination operator and then it is truncated to its original size. Detailed description can be found in [17].

A. Non-dominated Tournament Genetic Algorithm

In this section, a modification to selection in NSGA-II is presented. This approach uses a tournament selection instead of sorting the population and choosing its better half.

Additional experiments have been performed to check the effectiveness of a tournament selection if an \geq_r operator, which doesn't regard crowding distance is used. It is used as:

$$i \geq_r j \text{ if } (i_{rank} < j_{rank}) \quad (2)$$

Non-dominated Tournament Genetic Algorithm (NTGA) is a method, which uses modified selection and \geq_r operator.

B. Clone prevention

A clone prevention method has been designed and introduced after the initial results of the tournament selection. The results have shown a decrease in diversity of the population in comparison to NSGA-II approach with tournament size equal to 2. It was caused by stronger selective pressure. The idea is to enforce a mutation of every newly created individual which happens to be a clone. An individual is a clone of another individual if all their genes are equal.

V. DIFFERENTIAL EVOLUTION HYBRIDIZED WITH GREEDY

DEGR has been successfully applied to the MS–RCPSP in [8]. It's an evolutionary method operating in real space. Evolution creates a real-valued phenotype, which represents a task-resource assignments. Then the greedy algorithm puts tasks on a timeline.

DEGR has been proven to be efficient in single-objective MS–RCPSP and so it is used to probe space and create a potential PF . It is compared to both regular NSGA-II and NTGA approaches. It is worth noting that DEGR is inherently not a multi-objective method.

Differential Evolutions uses a weighted fitness function (presented in 3), which allow the algorithm to focus on different parts of the solution space and potentially create a good coverage, even though this approach is Pareto blind - has no concept of the PF, but it is the simplest approach to a multi-criteria problem.

$$f(S) = w_\tau * f_\tau(S) + (1 - w_\tau) * f_c(S) \quad (3)$$

where w_τ is a weight associated with the duration of the schedule and its values can vary between [0,1], S is the schedule, f_{tau} and f_c are time and cost of the schedule, both of which are minimized. The function is implemented in a library made public on [21].

Using equation 3 a potential PF is created by running the method multiple times with different weights to allow for exploration of space and to ensure good coverage.

VI. EXPERIMENTS AND RESULTS

The goal of this paper is to present a transition from a single-objective to multi-objective approach to MS–RCPSP. Additionally novel selection methods have been proposed to further increase the efficiency of multi-objective method. The results are evaluated by a set of chosen measures. They take convergence and diversity of the found PF under consideration.

A. Measures

A set of measures [20] has been chosen to evaluate the results. The choice has been dictated by the need of evaluating both convergence and diversity of the algorithms. Selected measures are commonly used for this purpose.

An Euclidean Distance (ED) is an average Euclidean distance between the points on the potential PF and a perfect point (built by the best possible values of every criteria).

A HyperVolume (HV) is a volume of the rectangle constructed from the potential PF and a Nadir Point (point built by the worst possible values of every criteria).

A Pareto Front Size (PFS) is a number of unique points on the potential PF .

B. Dataset

For the experiments an iMOPSE [7] dataset has been used, which is located on [21]. It contains 36 data instances, all varied by the number of tasks, resources and precedence relations. 2 subsets of instances could be identified. In the

first group all instances have 100 tasks, and in the second group, they have 200 tasks. Instances in those groups have been created to preserve an average resource load and number of tasks per number of resources. The idea behind formulation of this dataset was to achieve variety between the instances.

C. Procedure

A DEGR approach has been used to establish a baseline. It's been executed multiple times with weights values from 0.0 to 1.0, incremented by 0.1 and resulting points been collected. Parameters chosen for DEGR: population size (Ps) = 100, 500 generations (gen), mutation probability (Pm) = 0.1, crossover probability (Px) = 0.1 and a one-to-one selection.

Four different variants of NSGA-II have been checked. The following parameters have been chosen for NSGA-II: ps = 50, 500 gen, Pm = 0.01, Px = 0.6, tournament size (ts) = 5. Parameters chosen for NTGA: ps = 50, 500 gen, Pm = 0.002, Px = 0.9, ts = 4. We compared classic NSGA-II approach with its modifications: a tournament selection, modified comparison operator, which disregards crowding distance and clone prevention method. The goal was to increase the convergence and diversity of the method. All procedures has been repeated 30 times and results were averaged.

D. Results

An experiment has been performed to check the influence of a tournament size on chosen measures. The best values of both *ED* and *PFS* are achieved for tournament size equal to 4. Interestingly higher values improve the *HV*. High *PF* Size means high diversity of the population and also suggest good coverage of *PF*. Therefore *PFS* has been chosen as the most important measure.

A clone removal method have been introduced to increase low diveristy of the population. Since clone removal increases the distance between the individuals, a crowding distance has become redundant. Another approach has been investigated with modified comparison operator, which considers only the rank of the individual. Due to a huge volume of the table, standard deviations have been omitted and only classical NSGA-II, best variant of NTGA and DEGR approaches have been presented - table II. Additionally averaged results have been gathered and are presented in table I.

The increased size of the tournament improved all measures and standard deviations. Clone prevention method has a positive effect on average results but increases standard deviation. Checking crowding distance is not crucial when used with clone prevention method. This approach resulted in a better

convergence, represented by *ED*, a bit lower diversity, represented by *HV*, but average *PFS* has dramatically increased. Interestingly NSGA-II(t6,pc) has achieved the highest *PFS* for most instances, but NTGA(t6,pc,r) has achieved the best *PFS*. It is caused by the fact, that their results were very close, but the latter has achieved a huge lead on a couple of instances. DEGR approach achieved relatively low *ED* and the worst *PFS* of all investigated methods but at the same time the best *HV*. It is caused by the fact, that it was executed multiple times with various weight values, which allowed for searching different parts of solution space and is connected with the fact that DEGR achieved the best edge values.

VII. CONCLUSIONS AND FUTURE WORK

This paper presents how a single and a multi-objective approaches are capable of exploring the space of MS-RCPSP. It's been shown that a classic Pareto approach (NSGA-II) has a much better efficiency concerning convergence, however, lack diversity of DEGR - it's worth noting that DEGR has been executed multiple times with different weights. A modified selection operators and a clone prevention method have been presented and experiments have shown that they are capable of further increasing efficiency of NSGA-II.

Two potential directions for future work can be considered. On the one hand, an initial population, which better covers the solution space could improve achieved results. On the other hand, the selection method, which rewards better spread of individuals could occur the more diverse PF.

The used DEGR approach is hybridized with a Greedy Algorithm, which potentially is a bottleneck for the method. A very promising direction would be to remove the Greedy Approach and let an Evolutionary Algorithm take its place. In this co-evolutionary approach, there would be populations communicating with each other - one would assign the task to resources, while other would assign timestamps to tasks. DEGR has proven to be an effective method for single-objective optimization. As an extension of these works, DEGR could be introduced dominance relation and PF concept to compete with existing multi-objective methods even better.

REFERENCES

- [1] Samà, M., et al. "Lower and upper bound algorithms for the real-time train scheduling and routing problem in a railway network." IFAC-PapersOnLine 49.3 (2016): 215-220.
- [2] Varas, Mauricio, et al. "Scheduling production for a sawmill: A robust optimization approach." Inter.J.I of Prod. Econ.s 150 (2014): 37-51.
- [3] Kerzner, Harold. "Project management: a systems approach to planning, scheduling, and controlling." John Wiley & Sons, 2013.
- [4] Blazewicz, J., Lenstra J.K., and AHG Rinnooy Kan. "Scheduling subject to resource constraints: classification and complexity." Discrete Applied Mathematics 5.1 (1983): 11-24.
- [5] Kolisch R., Sprecher A.m "PSPLIB – A project scheduling problem library", Eur. J. of Oper. Res. (96), pp. 205–216, 1996.
- [6] Luna, Francisco, et al. "The software project scheduling problem: A scalability analysis of multi-objective metaheuristics." Applied Soft Computing 15 (2014): 136-148.
- [7] Myszkowski P. B., Skowroński M. E., Sikora K.; "A new benchmark dataset for Multi-Skill Resource-Constrained Project Scheduling Problem", Proc. of FEDCSIS

TABLE I: Comparison of averaged results for all methods

	<i>ED</i>		<i>HV</i>		<i>PFS</i>	
	avg	std	avg	std	avg	std
NSGA-II	0.2720	0.0059	0.5506	0.0041	138.44	15.91
NTGA	0.2677	0.0079	0.5280	0.0067	75.068	22.65
NSGA-II(t5,pc)	0.2764	0.0059	0.5668	0.0029	170.32	15.36
NTGA(t4,pc,r)	0.2575	0.0061	0.5446	0.0054	199.60	54.67
DEGR[8]	0.2743	0.0066	0.5708	0.0022	55.64	5.81

TABLE II: Summary of achieved results for all instances, for NSGA-II, NTGA(t6,pc,r) and DEGR

Dataset	NSGA-II			NTGA(t4,pc,r)			DEGR[8]		
	ED	HV	PFS	ED	HV	PFS	ED	HV	PFS
100_10_26_15	0.35271	0.54870	87.63333	0.32815	0.55718	128.23333	0.35941	0.57701	65.36667
100_10_27_9_D2	0.22198	0.50687	85.23333	0.21294	0.49899	84.50000	0.25224	0.53003	59.20000
100_10_47_9	0.31489	0.41905	101.10000	0.30121	0.42291	128.83333	0.33060	0.45468	80.93333
100_10_48_15	0.28040	0.54009	79.00000	0.26999	0.54407	156.03333	0.28915	0.55432	49.10000
100_10_64_9	0.24551	0.58652	97.76667	0.23946	0.58405	101.90000	0.26022	0.61965	63.20000
100_10_65_15	0.35004	0.44194	101.80000	0.33618	0.44119	128.03333	0.36425	0.45712	53.46667
100_20_22_15	0.18120	0.68286	50.43333	0.16505	0.69295	70.70000	0.19414	0.70781	44.76667
100_20_23_9_D1	0.14247	0.56986	48.70000	0.13125	0.57601	53.96667	0.16323	0.58871	37.26667
100_20_46_15	0.29761	0.61377	54.30000	0.26353	0.62906	67.13333	0.29743	0.64431	40.23333
100_20_47_9	0.19554	0.64914	57.86667	0.17999	0.66038	70.93333	0.22001	0.68313	49.13333
100_20_65_15	0.18821	0.66971	41.83333	0.18769	0.67120	51.36667	0.19160	0.67018	17.36667
100_20_65_9	0.28556	0.62284	61.53333	0.26752	0.63342	79.00000	0.28575	0.66536	51.13333
100_5_20_9_D3	0.31961	0.39914	211.66667	0.41975	0.39247	1436.13333	0.32627	0.40919	81.70000
100_5_22_15	0.29908	0.32875	139.40000	0.29949	0.32851	425.80000	0.29706	0.32998	37.00000
100_5_46_15	0.31338	0.21446	408.23333	0.31875	0.21189	1054.56667	0.28728	0.21874	54.20000
100_5_48_9	0.33326	0.16053	366.20000	0.31288	0.15300	298.63333	0.31593	0.16404	57.00000
100_5_64_15	0.35073	0.33845	196.70000	0.34073	0.32949	167.46667	0.35407	0.34478	65.66667
100_5_64_9	0.41493	0.45878	313.23333	0.41841	0.45375	1649.30000	0.43884	0.46434	63.73333
200_10_128_15	0.26315	0.59973	83.13333	0.25873	0.59596	81.76667	0.26908	0.61335	58.56667
200_10_135_9_D6	0.23999	0.35931	78.40000	0.23323	0.35222	65.53333	0.28168	0.38303	37.50000
200_10_50_15	0.19365	0.69882	72.50000	0.18237	0.69948	85.43333	0.20545	0.72625	65.23333
200_10_50_9	0.34110	0.56584	122.93333	0.30567	0.57388	138.96667	0.34149	0.62752	105.60000
200_10_84_9	0.23396	0.60410	90.96667	0.22055	0.61050	105.43333	0.24512	0.64830	82.53333
200_10_85_15	0.49872	0.42401	118.83333	0.47776	0.42717	113.63333	0.47468	0.46807	74.93333
200_20_145_15	0.30754	0.54307	75.36667	0.27912	0.56445	84.90000	0.28282	0.60734	60.60000
200_20_150_9_D5	0.08246	0.48148	44.53333	0.07096	0.48157	31.40000	0.17924	0.49675	27.63333
200_20_54_15	0.30901	0.55282	73.60000	0.29261	0.56239	81.10000	0.29998	0.60581	63.66667
200_20_55_9	0.23944	0.67040	74.10000	0.21237	0.69720	91.40000	0.21964	0.75396	77.43333
200_20_97_15	0.30325	0.56906	57.66667	0.28615	0.57981	59.26667	0.28157	0.61092	46.83333
200_20_97_9	0.34257	0.58709	68.10000	0.30349	0.62013	76.76667	0.29243	0.68489	58.23333
200_40_130_9_D4	0.13819	0.54772	41.50000	0.12301	0.54498	30.36667	0.18596	0.55891	22.83333
200_40_133_15	0.22678	0.63095	49.43333	0.18595	0.67492	55.93333	0.19539	0.71218	44.00000
200_40_45_15	0.26180	0.65191	52.60000	0.23289	0.68378	66.63333	0.23497	0.72835	54.26667
200_40_45_9	0.24057	0.67755	46.03333	0.20298	0.71549	64.53333	0.21762	0.75096	51.26667
200_40_90_9	0.26610	0.64150	47.16667	0.22524	0.68442	59.93333	0.22606	0.73970	52.93333
200_40_91_15	0.24036	0.67520	43.86667	0.20890	0.71228	58.03333	0.21481	0.74899	48.50000
Average	0.27266	0.53422	103.98241	0.25819	0.54336	208.43241	0.27432	0.57080	55.63981

- [8] Myszkowski P.B., Olech L.P., Laszczyk M. and Skowroński M.E., "Hybrid Differential Evolution and Greedy (DEGR) for Solving Multi-Skill Resource-Constrained Project Scheduling Problem", *Applied Soft Computing in review process*, 2017.
- [9] Mendes, Jorge Jose de Magalhaes, Goncalves J.F., and Mauricio GC Resende. "A random key based genetic algorithm for the resource constrained project scheduling problem." *Compu. & Op. Research* 36.1 (2009): 92-109.
- [10] Thomas P.R. and Said S. "A tabu search approach for the resource constrained project scheduling problem." *Journal of Heuristics* 4.2 (1998): 123-139.
- [11] Myszkowski P.B. and Siemieński J.J. "GRASP Applied to Multi-Skill Resource-Constrained Project Scheduling Problem." *International Conference on Computational Collective Intelligence*. Springer International Publishing, 2016.
- [12] Myszkowski P.B., et al. "Hybrid ant colony optimization in solving multi-skill resource-constrained project scheduling problem", *Soft Computing* 19.12 (2015), pp.3599-3619.
- [13] Zhang, Hong, Heng Li, and C. M. Tam. "Particle swarm optimization for resource-constrained project scheduling." *International Journal of Project Management* 24.1 (2006): 83-92.
- [14] Zheng, Huan-yu, Ling Wang, and Xiao-long Zheng. "Teaching-learning-based optimization algorithm for multi-skill resource constrained project scheduling problem." *Soft Computing* (2015): 1-12.
- [15] Abbasi, Babak, Shahram Shadrokh, and Jamal Arkat. "Bi-objective resource-constrained project scheduling with robustness and makespan criteria." *Applied mathematics and computation* 180.1 (2006): 146-152.
- [16] Cowling P., Colledge N., Dahal K., Remde S. "The Trade Off Between Diversity and Quality for Multi-objective Workforce Scheduling" [In:] Gottlieb J., Raidl G.R. (eds) *Evolutionary Computation in Combinatorial Optimization*. Springer
- [17] Deb, Kalyanmoy, et al. "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II." *International Conference on Parallel Problem Solving From Nature*. Springer Berlin Heidelberg, 2000.
- [18] Deb, Kalyanmoy, Udaya Rao N, and S. Karthik. "Dynamic multi-objective optimization and decision-making using modified NSGA-II: a case study on hydro-thermal power scheduling." *Evolutionary Multi-Criterion Optimization*. Springer
- [19] Rabiee, M., M. Zandieh, and P. Ramezani. "Bi-objective partial flexible job shop scheduling problem: NSGA-II, NPGA, MOGA and PAES approaches.", *International Journal of Production Research* 50.24 (2012): 7327-7342.
- [20] Wang, Shuai, et al. "A practical guide to select quality indicators for assessing Pareto-based search algorithms in search-based software engineering.", *Proceedings of the 38th International Conference on Software Engineering*. ACM, 2016.
- [21] <http://imopse.i.pwr.wroc.pl/>