PCC170 - Projeto e Análise de Experimentos Computacionais

Marco Antonio M. Carvalho

Departamento de Computação
Instituto de Ciências Exatas e Biológicas
Universidade Federal de Ouro Preto

UFOP
Universidade Federal de Ouro Preto

decom
departamento de computação

Conteúdo

Aviso

Este material é baseado em minha experiência em pesquisa.

Particularmente, os textos apresentados consistem em sugestões aos alunos e refletem tão somente minha visão pessoal.

Licença

Diretrizes para o relato de experimentos

Organização

A seção de experimentos computacionais pode ainda ser organizada da seguinte maneira:

5.0 Texto introdutório da seção/capítulo;
5.1 Instâncias;
5.2 Experimentos preliminares;
5.3 **Limitantes inferiores e/ou superiores**;
5.4 **Comparação com o estado da arte**;
5.5 **Análises adicionais**.

Limitantes

Opcionalmente, pode haver uma seção ou capítulo indicando o cálculo de limitantes inferiores e superiores para o valor da solução:

- Definição dos limitantes, com citação da origem;
- Descrição breve do método, com citação da origem, ou referência ao texto anterior;
- Ambiente de codificação e compilação;
- Indicação de disponibilização de resultados detalhados.

5.2. Lower bounds

The RCPMS has two trivial lower bounds. First, the value generated by the general model of parallel machines scheduling, which disregards setup times and resource constraints (Pinedo, 2008). Second, the *mold dominance*, i.e., the largest sum value of the processing times of all jobs requiring a specific mold, which may be equal to or greater than the sum of the processing times of all other jobs.

However, these lower bounds are not expected to be tight enough to support an accurate analysis of the new set of instances. Thus, in order to provide optimal solutions and tighter lower bounds, mathematical models from the literature were also considered. Nonetheless, given the limitations of the mathematical models listed in Section 3, we have selected three recent models designed for related problems and used them in addition to the general model of Pinedo (2008) to produce different levels of relaxation to the original RCPMS.

fonte: Soares & Carvalho. *Application of a hybrid evolutionary algorithm to resource-constrained parallel machine scheduling with setup times.* Computers & Operations Research. 2021.

Diretrizes para o relato de experimentos

Beezão et al. (2017) presented two models for job scheduling on identical parallel machines with setup times, a relaxed version of the RCPMS which disregards resource constraints. Among the models presented, the model named M_1 showed the best results and was therefore selected. Addressing unrelated parallel machine scheduling with resource constraints and setup times, Bitar et al. (2021) presented a mixed-integer programming (MIP) model to minimize the weighted sum of completion times using time-indexed formulations. To be applied to the RCPMS, the model was adapted to consider identical machines and the objective function was switched to minimization of

fonte: Soares & Carvalho. *Application of a hybrid evolutionary algorithm to resource-constrained parallel machine scheduling with setup times*. Computers & Operations Research. 2021.

the makespan. In a similar context, Ozer and Sarac (2019) presented a MIP to minimize the total weighted completion time for job scheduling on parallel machines with resource constraints, machine eligibility and setup times. This model does not use time-indexed formulations. In order to use this model for the RCPMS, the objective function has changed to minimization of the makespan and machine eligibility constraints have been removed.

All models were implemented using Python 3.7.4, solved using the Gurobi solver 9.1.2 and run with a time limit of 10800 s. The reference values considered in Section 5.6 were defined as solution value or best lower bound value generated by any of the models for each instance. Individual results of each model are available in the Supplementary Material.

fonte: Soares & Carvalho. *Application of a hybrid evolutionary algorithm to resource-constrained parallel machine scheduling with setup times*. Computers & Operations Research. 2021.

Diretrizes para o relato de experimentos

Comparação com o estado da arte

Esta é uma das maiores seções de qualquer texto. Deve incluir:

- Indicação e referência aos métodos utilizados como comparação;
- Tabela com resultados médios de execuções independentes, para cada conjunto de instâncias;
- Análise textual com inclusão de indicadores novos, para cada conjunto de instâncias
  - O impacto das dimensões e atributos das instâncias;
  - O impacto de cada componente do método proposto (opcionalmente, na próxima seção).
- Análise estatística da comparação entre os diferentes métodos.
- Explicação de quaisquer fenômenos adicionais observados.

Comparação com o estado da arte

As tabelas devem incluir:

- Dimensões das instâncias e outras características;
- Número de instâncias (i);
- Solução ótima (OPT) ou melhor solução conhecida (BKS);
- Valor da solução inicial (S_0);
- Melhor solução encontrada por cada método (S^*);
- Solução média encontrada por cada método (S);

Comparação com o estado da arte

As tabelas devem incluir:

- Tempo médio de execução de cada método em segundos (T);
- *gap* (ou distância percentual) em relação ao melhor resultado conhecido;
- Desvio padrão das soluções para uma mesma instância (σ);

Adicione também uma última linha nas tabelas, com a média dos valores em cada coluna.

Além disto, indique valores em negrito quando as melhores soluções foram atingidas por algum dos métodos.

Comparação com o estado da arte

Muitos dos dados elencados podem estar disponíveis apenas para o método proposto.

Neste caso, apresenta-se uma tabela com os dados do método proposto e outra tabela com os dados disponíveis para ambos os métodos.

Quando há muitas instâncias, os dados das tabelas são agrupados por conjuntos de instâncias, ou subconjuntos com características em comum.

5.3. Comparison to the state-of-the-art

The ALNS solutions for the Small set were compared with the optimal solutions, and the results for the Grid and HB sets were compared with those generated by VFS (Pardo et al., 2013).

The ALNS average results over 10 independent runs are summarized in Table 2. Each line represents the average results for a given instance set. The columns present the optimal or best known solutions (*OPT*/*BKS*), the best results obtained by the ALNS (S^*), the average solution value (S), and the value of the initial solution (S_0). Moreover, the columns present the percentage distance between S_0 and S^* (gap_{S_0,S^*}), and between S^* and *OPT*/*BKS* ($gap_{S^*,OPT}$). The percentage distance $gap_{b,a}$ is calculated as $100 \times \frac{b-a}{a}$. The standard deviation is shown in column σ, and finally, the average running time is given in s (T).

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

Table 2
ALNS average results.

	OPT/BKS	S^*	S	S_0	gap_{S_0,S^*}	$gap_{S^*,OPT}$	σ	$T(s)$
Small	4.92	4.92	4.93	5.15	6.52	0.00	0.02	0.08
Grid	11.56	11.56	11.56	11.56	0.00	0.00	0.00	8.60
HB	311.55[a]	311.80	315.01	336.70	22.12	0.39	2.19	222.23

[a] Some solutions are not proven to have an optimal value.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

The proposed ALNS matched all the optimal solutions for the Small instance set. That was expected, as its solution is trivial owing to the dimensions of its instances. Nonetheless, the ALNS still played an important role because it closed the gap between the initial solution and the optimal solution. This is the first study in which all the optimal solutions for the Grid instance set are obtained. All the optimal solutions were obtained by the newly proposed initial solution method.

For the HB set, whose instances have the largest dimensions, the percentage distance ($gap_{S^*,OPT}$) is the lowest reported in the literature. Moreover, the ALNS performed satisfactorily, as it reduced by 21.73 percentage points the gap from the initial solution to the best known solution. The ALNS also obtained new best solutions for the instances *can*__445, *dwt*__361, *dwt*__503, *saylr*3, *str*____0, with respective values 116, 38, 127, 44, and 381. The post-processing step decreased the average gap of the HB set by 0.28 percentage points, whereas there was no possible improvement on the Grid set solutions.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to*

Diretrizes para o relato de experimentos

Regarding the standard derivation and running time, the reported values are low for the Small and Grid instance set. For the HB set, the standard deviation is still low; however, as the instances have larger dimensions, it required longer running time. The average running time for this set is strongly influenced by the instances *shl*____0, *shl*__200, *shl*__400, and *west*0655, which represent 63% of the total running time. The post-processing step increased the average running time by 5 s for the HB set and by 3 s for the Grid set.

Table 3 shows a comparison of the best results by ALNS with the results by VFS (Pardo et al., 2013). The columns S^* present the average solution value for each set, columns *#OPT* show the number of best known solutions obtained by each method, and columns *gap* present the percentage distance between each method's best solution and the best known solution value in the literature, available at Martí et al. (2019).

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

Table 3
Comparison with state-of-the-art VFS.

	VFS			ALNS		
	S^*	*#OPT*	gap	S^*	*#OPT*	gap
Grid	12.23	59	3.25	11.56	81	0.00
HB	314.39	61	1.77	311.80	77	0.39

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

In both sets, the proposed ALNS outperformed VFS. First, on the Grid instance set, it matched all optimal solutions, whereas VFS did not match 22 optimal solutions. For the HB set, The VFS average solution values are 1.21% away from the ALNS average solution values. Additionally, regarding best known solutions, ALNS obtained 16 more solutions than VFS.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

To further analyze the performance of both algorithms, statistical experiments were conducted. The solution values over the various instances were tested whether they could be modeled according to a normal distribution and then an appropriate method for comparing paired samples was employed. The Shapiro–Wilk normality test (Shapiro & Wilk, 1965) was applied separately for the Grid and HB sets for each algorithm, which rejected the null hypothesis that the results of ALNS (p-value equal to $4.989 \cdot 10^{-5}$ and p-value less than $2.2 \cdot 10^{-16}$) and VFS (p-value equal to $4.873 \cdot 10^{-6}$ and p-value less than $2.2 \cdot 10^{-16}$) could be modeled according to a normal distribution. Subsequently, given that the populations are not normally distributed, the non-parametric Wilcoxon signed-rank test (Rey & Neuhäuser, 2011) was applied to investigate whether or not there is a significant difference between the solution values generated by ALNS and VFS. For the Grid set, the test indicates ($V = 0$ and p-value equal to $1.923 \cdot 10^{-5}$) that there are statistical differences between the two methods. The value $V = 0$ also indicates that all the ALNS solution values are less than or equal to those generated by VFS. For the HB instance set, the test indicates ($V = 28.5$, p-value equal to $5.714 \cdot 10^{-5}$) that ALNS is significantly different from VFS. However, three VFS solution values are better than those generated by ALNS, and 24 ALNS solution values are better than those generated by VFS. Therefore, it may be concluded that ALNS performed better than VFS on the HB set.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

Table 4
Results for the IPMTC-II set of instances.

m	n	l	BRKGA					ALNS			
			S^*	S	σ	T	$gap(\%)$	S^*	S	σ	T
3	50	30	**1336.73**	1365.00	16.03	87.75	−0.16	1338.85	1364.74	14.01	11237.29
3	50	40	1518.57	1536.11	9.97	118.57	0.84	**1505.92**	1529.15	12.31	13979.03
4	50	30	**996.30**	1019.90	12.79	40.42	−1.24	1008.85	1030.62	11.47	11705.45
4	50	40	1039.13	1057.58	10.56	53.78	0.42	**1034.80**	1059.30	12.29	14311.55
4	100	30	**2179.80**	2208.43	15.99	185.27	−5.97	2318.13	2376.55	33,72	14433.29
4	100	40	**2698.33**	2733.59	19.10	297.34	−7.41	2914.42	2984.09	43.92	14480.79
5	50	30	**809.83**	828.44	9.68	21.83	−0.55	814.32	831.25	9.12	10535.76
5	50	40	**775.18**	791.89	10.18	27.99	−0.02	775.33	792.93	9.88	14044.71
5	100	30	**1743.83**	1766.46	12.56	89.71	−6.99	1874.95	1921.41	25.92	14446.20
5	100	40	**2237.43**	2269.39	16.67	154.31	−8.12	2435.08	2494.69	35.54	14482.06
6	100	30	**1497.82**	1520.69	12.77	49.78	−6.44	1600.85	1643.21	23.63	14431.45
6	100	40	**1930.17**	1958.68	15.87	84.58	−7.91	2095.85	2158.14	35.37	14485.33
6	200	30	**2932.23**	2961.92	17.04	1780.59	−8.74	3212.97	3291.48	48.11	14524.62
6	200	40	**3539.23**	3574.86	19.09	2524.16	−8.25	3857.63	3988.87	77.46	14633.74
7	100	30	**1237.33**	1257.34	10.67	29.67	−5.99	1316.23	1346.65	17.29	14428.72
7	100	40	**1749.20**	1779.50	15.41	52.93	−4.18	1825.48	1869.58	23.43	14412.62
7	200	30	**2410.75**	2438.57	14.50	1078.50	−8.21	2626.37	2709.82	49.69	14541.81
7	200	40	**3178.30**	3217.27	20.06	1711.91	−8.27	3465.02	3586.36	68.74	14622.22
8	200	30	**2245.20**	2273.57	15.61	658.84	−8.28	2448.02	2517.24	43.00	14534.03
8	200	40	**2752.63**	2788.63	17.21	1011.76	−8.95	3023.32	3118.65	59.88	14624.44
9	200	30	**1880.13**	1902.56	12.16	410.69	−7.86	2040.53	2102.57	37.30	14531.10
9	200	40	**2346.20**	2373.00	14.27	678.48	−8.27	2557.75	2647.69	55.07	14637.45
10	200	30	**1827.12**	1849.48	12.22	297.65	−8.45	1995.75	2049.91	33.44	14528.05
10	200	40	**2059.75**	2086.52	13.96	466.93	−8.88	2260.57	2332.38	41.63	14637.09

fonte: Soares & Carvalho. *Biased random-key genetic algorithm for scheduling identical parallel machines with tooling constraints*. European Journal Of Operational Research, 2020.

Diretrizes para o relato de experimentos

Table 4
Results for the RCPMS-I set of instances.

m	n	l	Hybrid BRKGA						*Bounds*	
			S^*	S	σ	T	gap_{LB}(%)	gap_{UB}(%)	LB	UB
2	8	3	161.80	161.80	0.00	0.10	0.00	0.00	161.80	161.80
2	8	5	265.00	265.00	0.00	0.11	0.00	0.00	265.00	265.00
2	8	7	287.00	287.00	0.00	0.11	0.00	0.00	287.00	287.00
2	15	3	276.80	278.56	1.80	0.32	0.00	0.00	276.80	276.80
2	15	5	359.80	359.80	0.00	0.29	0.00	0.00	359.80	359.80
2	15	7	330.60	330.96	0.32	0.34	0.00	0.00	330.60	330.60
2	25	3	391.20	391.38	0.29	0.67	0.00	−0.05	391.20	391.40
2	25	5	444.00	444.00	0.00	0.64	0.00	0.00	444.00	444.00
2	25	7	532.60	538.50	6.58	0.75	0.00	0.00	532.60	532.60
3	15	4	201.40	203.96	1.92	0.17	0.23	0.00	201.00	201.40
3	15	5	217.40	219.00	0.92	0.33	0.17	−0.25	217.00	218.00
3	15	7	250.00	251.08	0.38	0.33	0.00	0.00	250.00	250.00
3	25	4	335.20	336.56	4.30	0.04	2.45	0.00	327.60	335.20
3	25	5	333.20	333.92	1.20	0.18	2.88	0.00	323.80	333.20
3	25	7	330.00	335.00	6.18	0.88	9.76	−0.11	301.80	330.40
4	25	5	259.80	265.12	14.09	0.30	10.71	−0.07	235.60	260.00
4	25	7	242.40	253.64	17.75	0.70	16.13	0.00	209.00	242.40
4	25	11	286.00	293.20	7.92	1.05	23.22	−0.57	236.40	287.60

m: number of machines. n: number of jobs. l: number of molds. S^*: best solution values. S: average solution values. σ: standard deviation. T: average running times in seconds. gap: percentage distances between solution values. LB: lower bound. UB: upper bound.

fonte: Soares & Carvalho. *Application of a hybrid evolutionary algorithm to resource-constrained parallel machine scheduling with setup times*. Computers & Operations Research. 2021.

Análises adicionais

As análises adicionais se referem somente ao comportamento do método proposto, e podem ser incluídas tanto na seção anterior quanto em uma seção à parte. Devem incluir:

- Análise do impacto de cada componente do método;
- Análise estatística envolvendo componentes que competem entre si, e.g. operadores de busca local;
- Análise de convergência, ilustrada por *ttt-plots*;
- Estatísticas sobre diferentes critérios de parada, e.g., frequência;
- Estatísticas sobre diferentes critérios de aceitação, e.g., frequência;
- Descrição da oscilação de parâmetros adaptativos;
- Etc.

Análises adicionais

Em casos de estudos de caso ou de instâncias reais, também é possível realizar uma análise do impacto prático dos resultados.

Preferencialmente, esta análise deve incluir dados reais da aplicação do método.

5.4. Additional analysis

In Section 4, removal and insertion operators were presented, and it is important to evaluate which have contributed more to the quality of the results. The convergence analysis of the solution is another important analysis, as it evaluates the efficiency of ALNS as well as the required number of iterations for high-quality solutions. Thus, this section is devoted to these two types of analysis.

Over the course of all iterations the scores assigned to each operator were accumulated to obtain the final score per instance. Table 6 presents the percentage distances between the score of each operator and the largest score of any operator for each set of instances, i.e., a percentage distance that is close to zero implies a higher score on a particular set of instances.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Table 6
Percentage distances between operator score and overall best score. The highest scores have distance zero.

	Removal operators					Insertion operators			
	RR	UR	UR_E	ER_L	ER_R	RI	BI	BI_N	BII
Small	0.63	34.13	13.38	41.58	46.94	85.16	2.00	0.36	5.81
Grid	0.04	68.54	19.08	46.90	48.52	282.32	10.93	0.70	16.82
HB	0.60	85.60	48.70	56.61	59.60	383.64	17.76	0.62	34.11
Rome	0.20	46.57	20.36	51.42	56.00	101.39	6.77	0.16	10.65
Average	0.37	58.71	25.38	49.13	52.77	213.13	9.36	0.46	16.85

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

Both removal and insertion operators that achieved the best scores employed component randomness, which introduces diversity to the search. Additionally, the noise term yielded the best score among the insertion operations, which again is a diversity mechanism. Although some operators performed poorly compared to others, they are most likely important for the entire set of operators, as indicated by the irace package in the preliminary experiments.

The ANOVA test was applied to evaluate if there is significant difference among the removal operators. The p-value of $6.62 \cdot 10^{-05}$ indicates that there is such a difference. The pairwise t-test was applied to analyze where this difference lies. According to the test, UR_E significantly differs from UR, and RR significantly differs from UR, ER_L, and ER_R. It is worthwhile to mention that the new proposed operators (ER_L and ER_R) are competitive as they do not show significant difference from the UR and UR_E operators.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Likewise, the ANOVA test and the pairwise t-test were applied to the insertion operators. The ANOVA test indicated that there is significant difference among the operators (a p-value of 0.0035), and the pairwise t-test showed that the difference lies between the operator RI and all other three insertion operators.

Fig. 3 shows the boxplot chart with the ALNS convergence analysis. The y axis represents the number of iterations for which the method yields its best solution value for each instance, and the x axis represents the instance sets. For each set, the value measured represents the average value for the 10 runs for each instance. The red line corresponds to the median, i.e., half of the values are located below this line, and the other half are above this line. The red crosses correspond to the outliers, i.e., values that present a great distance from the others.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

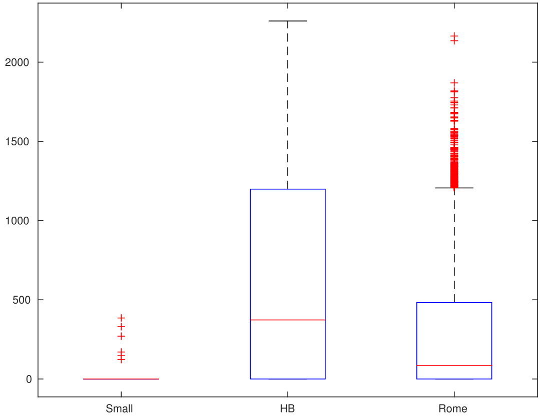

Fig. 3. Convergence analysis of ALNS for the Small, HB, and Rome instance sets.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

The ALNS convergence is not shown for the Grid instance set, as all optimal solutions are reached during the initial solution; hence, ALNS converges at the iteration zero for all of its instances. For the Small set, six instances converge after at most 385 iterations, whereas all the others have their optimal value obtained by the initial solution. The HB set has an average of 633, median of 372, and a maximum value of 2,260. The Rome Graphs set has a larger distribution of values, as it is a set with a larger number of instances and a wide range of dimensions. The set has an average equal to 273, a median of 84, and a maximum value of 2,162. This set also presents a considerable number of outliers, which are also due to its larger instances.

fonte: Santos & Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.

Diretrizes para o relato de experimentos

On average, the hybrid BRKGA required 43.41 generations to improve the initial solutions by 23.22%. The solution quality was the stopping criterion for every run, and it dominated the maximum number of generations. Among the local search procedures, job exchange contributes the most to the solution improvement. On average, job exchange, job insertion, job relocation, and 1-block accounted for 34.63%, 29.31%, 18.5%, and 17.66% of the improvements, respectively. The average running time, 0.41 s and the average standard deviation, 3.54 (equivalent to 0.87%), demonstrate the consistency of the method in generating solutions with negligible running times and low variations over independent runs.

fonte: Soares & Carvalho. *Application of a hybrid evolutionary algorithm to resource-constrained parallel machine scheduling with setup times*. Computers & Operations Research. 2021.

Leitura recomendada

- Soares, Leonardo C.R.; Carvalho, Marco A.M. *Application of a hybrid evolutionary algorithm to resource-constrained parallel machine scheduling with setup times*. Computers & Operations Research. 2021.
- Gandra, V. M. S.; Carvalho, M. A. M. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal Of Operational Research, 2021.
- Santos Gandra, Vinícius Martins; Çalık, Hatice; Wauters, Tony; Toffolo, Túlio A.M.; Moreira De Carvalho, Marco Antonio; Berghe, Greet Vanden *The impact of loading restrictions on the two-echelon location routing problem*. Computers & Industrial Engineering, v.160, p.107609, 2021.

Leitura recomendada

- Soares, Leonardo C.R.; Carvalho, Marco A.M. *Biased random-key genetic algorithm for scheduling identical parallel machines with tooling constraints*. European Journal Of Operational Research, 2020.
- Soares, Leonardo C.R.; Reinsma, Jordi Alves; Nascimento, Luis H.L.; Carvalho, Marco A.M. *Heuristic methods to Consecutive Block Minimization*. Computers & Operations Research, 2020.

Dúvidas?