



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
OTIMIZAÇÃO COMBINATÓRIA E META-HEURÍSTICAS

2019/2

1º TRABALHO

Marcos A. Spalenza

Clustering Search - *CS

Vitória, ES

2019

1 Introdução

Clustering Search - *CS é uma meta-heurística que compreende a aplicação de clusterização em meio aos resultados dos otimizadores. A proposta dividi-se em quatro partes: meta-heurística de busca, clusterização iterativa, módulo de análise e a busca local (OLIVEIRA; LORENA, 2004). Assim, a técnica refere-se à agrupamento de respostas similares resultantes de uma meta-heurística inicial para estabelecer processos de otimização independentes via busca local.

Diferentemente das propostas inspiradas em situações biológicas, *Clustering Search* é a aplicação dos métodos tradicionais de data mining para busca em vizinhança das principais soluções. Assim, conhecendo a distribuição dos resultados, a ideia é tornar a busca agressiva nesses pontos para potencializar as melhores amostras. Portanto, a meta-heurística *Clustering Search* estabelece melhorias em grandes concentrações de resultados. Com os clusters formados, a técnica aprimora as soluções a partir do centróide do grupo.

Com a análise de particionamento do espaço de busca, através desse método estuda-se as regiões de soluções buscando sua minimização. A ideia, então, é a busca por melhorias nas soluções centrais dos clusters, tentando forçar tais pontos a alcançar os melhores resultados de sua vizinhança.

Assim, o trabalho a seguir é apresentado em 5 seções. Na Seção 2 apresentamos um breve histórico do PSO. Na Seção 3 definimos as principais características dessa técnica e descrevemos a sua estrutura. Na Seção 4 apresentamos as técnicas que são o estado da arte da literatura. Por fim, na Seção 5 descrevemos algumas aplicações e alguns problemas que foram tratados por esta técnica.

2 Histórico

A metaheurística que hoje é reconhecida como *Clustering Search* têm uma predecessora. Focada na melhoria dos resultados populacionais de um Algoritmo Genético, o *Evolutionary Clustering Search* - ECS (OLIVEIRA; LORENA, 2004) faz a busca nos centróides da população entre as iterações, refinando sua vizinhança. Assim, ao invés de definir a busca local em todos os itens dentro da população, o ECS seleciona para uma otimização das amostras principais. Após isso, surgiram propostas para a generalização do processo especificado nos Algoritmos Genéticos. A Figura 1 apresenta esse aspecto caracterizado pelos autores através da representação espacial dessas populações em convergência para mínimos locais.

A Figura 1 representa a situação de uma população típica de um Algoritmo Genético. Nesse aspecto espacial, a expectativa é que a clusterização seja uma forma de encontrar pelos centróides uma proximidade aos mínimos. Com os centróides, a vizinhança

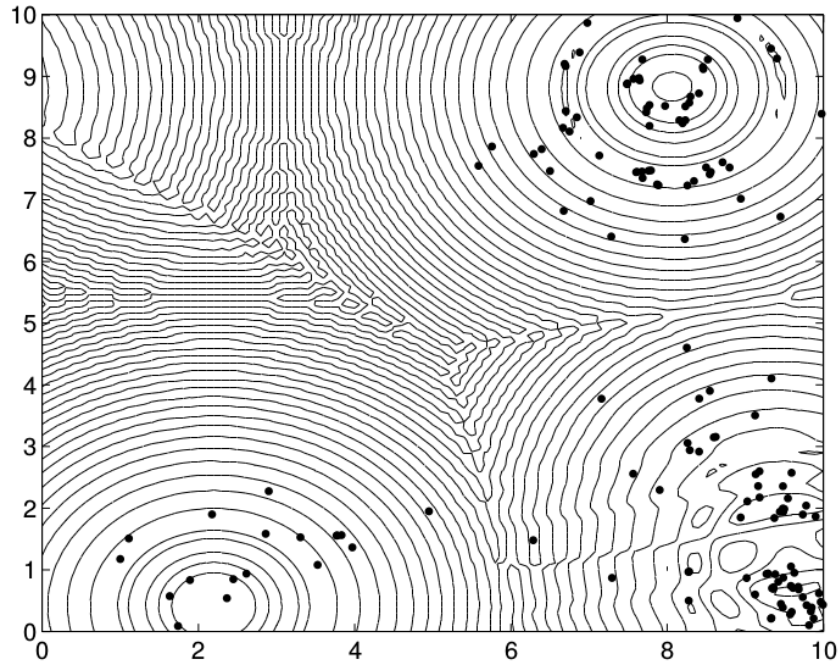


Figura 1 – Convergência de uma população para os mínimos de um determinado modelo.

é explorada para minimização de pontos distribuídos no espaço de busca na tentativa de encontrar o mínimo global. Isso inibe o problema de movimentação local do presente nos algoritmos evolutivos clássicos, identificando os itens em potencial para tal detalhamento.

Apresentado em (OLIVEIRA; LORENA, 2007), os autores divulgaram melhorias para essa metaheurística adotando a população de resultados de uma segunda metaheurística como modelo para delimitar as k zonas de interesse. Desta forma, todos os conjuntos de resultados sob controle, não necessariamente sob aspecto populacional, foram inclusos em modelo híbrido em *Clustering Search*. Através desse modelo, o algoritmo é simplificado, e parte de um intermediário do Algoritmo Genético para a busca local da população. Inclusive, em um formato independente, tal minimização característica desse algoritmo é aplicada diretamente nos centróides de uma população gerada aleatoriamente. Apesar disso, a *Clustering Search*, pelo seu intuito inicial, é comumente encontrada em aplicações híbridas.

3 Definição e Descrição

Clustering Search é um algoritmo que envolve a clusterização de uma população controlada para minimização de centróides. Essa ideia considera o aspecto espacial dos mínimos ao entorno dos itens desta população, independentemente da forma ao qual esta foi criada. Assim, a clusterização busca o padrão de respostas dadas para o problema em determinadas partições do espaço para designar qual solução deverá ser trabalhada.

Mesmo atrelada como um processo após a otimização de uma outra meta-heurística,

o aprimoramento das soluções comumente se dá pelo método da Busca Local. Essa técnica visa escolher os itens que representem agrupamentos de soluções para designar a melhor solução do subespaço ao qual pertencem. Portanto, os autores descreveram em 4 módulos essa meta-heurística, listados abaixo:

- Meta-heurística de Busca - SM;
- Componente de Clusterização Iterativa - IC;
- Módulo de Análise - AM;
- Busca Local - LS.

A SM é responsável pela modificação da população. Para não ser uma busca aleatória no espaço, essa primeira meta-heurística é responsável pela convergência aos mínimos. A clusterização iterativa realiza o particionamento espacial e designa os centróides de busca. O módulo de análise verifica a cada cluster se o mesmo é promissor, geralmente em relação a sua densidade. Por fim, a busca local refina os centróides dos clusters mais promissores em busca de mínimos. A Figura 2 representa o esquema de busca dessa meta-heurística.

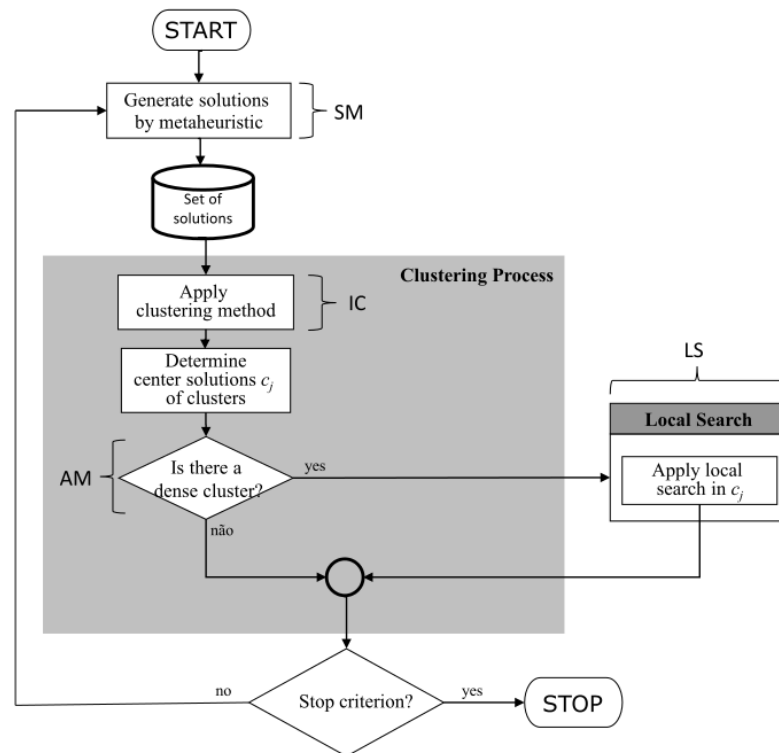


Figura 2 – Fluxograma do processo e busca com *Clustering Search*.

Como representado na Figura 2, a otimização e a clusterização se relacionam diretamente em ciclos, até o critério de parada. Os hiperparâmetros de clusterização em

geral são definidos *a priori*. Com o processo bem definido para clusterização, é relevante que a análise de consistência dos clusters formados sejam repetidamente analisados pela influência de *outliers* no algoritmo. Além disso, a análise de densidade designa os pontos de refinamento. Esse processo é descrito em detalhes no pseudocódigo da Figura 3.

```

algorithm CS
  create the initial clusters of CS
  { metaheuristic }
  while termination condition not satisfied do
    generate a solution ( $s_k$ ) by metaheuristic
    { clustering process }
    find the most similar cluster  $c_j$  to the solution  $s_k$ 
    insert  $s_k$  into  $c_j$  (  $\delta_j \leftarrow \delta_j + 1$  )
     $c_j \leftarrow \text{Assimilation}(c_j, s_k)$ 
    if  $\delta_j \geq \lambda$  then
       $\delta_j \leftarrow 0$ 
      { local search }
      find the best neighborhood of  $c_j$ 
    end if
    update the center  $c_j$ 
  end while
end algorithm

```

Figura 3 – Pseudocódigo do *Clustering Search*

O código apresentado na Figura 3 detalha elementos tradicionais das meta-heurísticas. A criação da população s_k aqui é acompanhada da criação dos clusters c_j . Após as modificações na população, os itens são associados a clusters especificando as principais soluções como os centróides de c_j . Para que esses itens sejam considerados válidos, a densidade do cluster σ deve estar acima de um coeficiente λ previamente definido. Nos clusters acima desse coeficiente, os centróides passarão pela busca local diante de sua vizinhança. Essa vizinhança também passa por uma perturbação denominada Assimilação ao agrupar os novos itens resultantes da meta-heurística de busca. Essa perturbação é analisada para evitar a influência dos *outliers*. Se representar um novo local de busca em relação com os clusters já formados é criado um novo cluster para essa solução.

4 Estado-da-Arte

A clusterização em si foi uma proposta de melhoria para as meta-heurísticas (OLIVEIRA; LORENA, 2007). Essa combinação representou combinações com a *Clustering Search* em vários híbridos. (OLIVEIRA; CHAVES; LORENA, 2013) destaca as aplicações com Algoritmos Genéticos, como sua origem, *Iterated Local Search*, *Variable Neighborhood Search*, *Simulated Annealing*, GRASP e Busca Tabu. Ainda sobre estas aplicações, o autor destaca uma variação de problemas ao qual o mesmo foi empregado. Tais problemas são listados abaixo.

- Otimização Contínua

Unconstrained Numerical Optimization;

- Problemas de Agendamento

Pattern Sequencing Problem;

Permutation Flow Shop Scheduling Problem;

M-Machine No-Wait Flow Shop Scheduling Problem;

Assembly Line Worker Assignment and Balancing Problem;

- Problemas de Localização

Capacitated p -Median Problem;

Probabilistic Maximal Covering;

Location-Allocation Problem;

Capacitated Centered Clustering Problem;

Hub Location Problem;

Capacitated Hub Location Problem;

- Problemas de Roteamento

Traveling Tournament Problem;

Prize Collecting Traveling;

Salesman Problem;

Berth Allocation Problem;

Dial-a-Ride Problem;

Workover Rig Routing Problem;

Podemos ver por essa lista a capacidade de adaptação que essa meta-heurística dispõe. Apesar de ser bem recente, as principais modificações se dão no seu método de clusterização. Sabendo que o principal modo de avaliação de clusters aplicado nesse método é por meio da densidade do cluster, algumas proposta adotam técnicas que se utilizam dessa medida. Um exemplo é o *DBSCAN* (ARAÚJO; CHAVES; LORENA, 2019), técnica de clusterização por densidade aonde os itens são agrupados enquanto pertencentes a um determinado coeficiente de distancia dos itens das extremidades.

5 Exemplos de Aplicações

Por seu caráter de análise distributivo, *Clustering Search* comumente é associada a problemas de posicionamento espacial. Exemplos incluem o *p -Median Problem* (CHAVES; CORREA; LORENA, 2007), o *Point-Feature Cartographic Label Placement Problem*

(RABELLO et al., 2014; ARAÚJO; CHAVES; LORENA, 2019) e *Capacitated Centered Clustering Problem* (CHAVES; LORENA, 2010). Outros como o *Job Shop Scheduling Problem* (ALTOÉ et al., 2018), o *Berth Allocation Problem* (OLIVEIRA; MAURI; LORENA, 2012) e *Traveling Tournament Problem* (BIAJOLI; LORENA, 2007) evidenciam sua adaptação em outros tipos problemas.

O problema de posicionamento de rótulos, é responsável por alocar descrições textuais em locais da imagem evitando sobreposição. Isso pode ser visto frequentemente em gráficos e mapas, aonde o ideal é priorizar a legibilidade da imagem. Em certos casos, isso ainda é dificultado pelas linhas em modelos cartográficos. Assim, para além dos gráficos, o posicionamento em mapas e suas representações de relevo, estradas e divisas aumentam a complexidade do problema e devem ser levados em consideração nesse posicionamento. A aplicação de *Clustering Search* neste problema (ARAÚJO; CHAVES; LORENA, 2019), portanto, visa identificar as melhores posições centralizando-os nos locais de referência. Através disso, a minimização estabelece evitar as linhas da imagem e a alta taxa de sobreposição. Essa situação pode ser vista na Figura 4 apresentando o mapa de São Paulo.

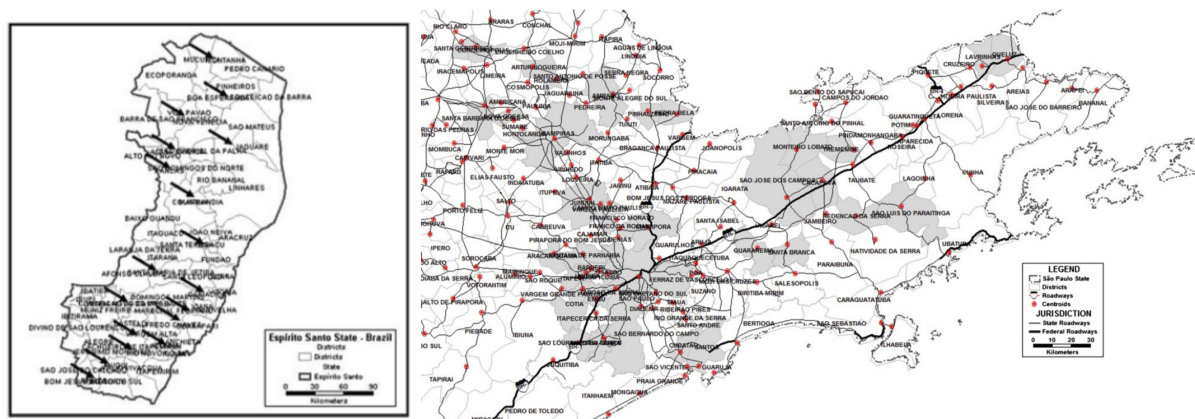


Figura 4 – Mapas dos estados do Espírito Santo e São Paulo e a sobreposição entre as linhas e anotações.

Como apresentado na Figura 4, os centróides especificados representam melhores posicionamentos dos itens. Para o mapa de São Paulo fica evidente esse problema então os autores aplicaram tal técnica na identificação dos melhores posicionamentos. Os autores incorporaram o processo de clusterização com o *BRKGA*. Junto disso, o *Clustering Search*, emprega o *DBSCAN* para analisar todas as amostras que estão acima de um coeficiente de distância. Dentro deste coeficiente as amostras entram no mesmo cluster, independente de definir o número de clusters.

Nesse artigo, o *DBSCAN* estabelece como é a relação de vizinhança, modificada pelo *Local Search 1-opt* e *2-opt*. Nesses modelos, o *Local Search* realiza 1 ou 2 trocas simultaneamente na vizinhança na minimização da função. Com os resultados de clusterização, os centróides de cada cluster formado pela análise de densidade intrínseca do *DBSCAN*.

Os aprimoramentos dão resultados melhores do que os inicialmente gerados pelo *BRKGA* como apresentado na Figura 5.

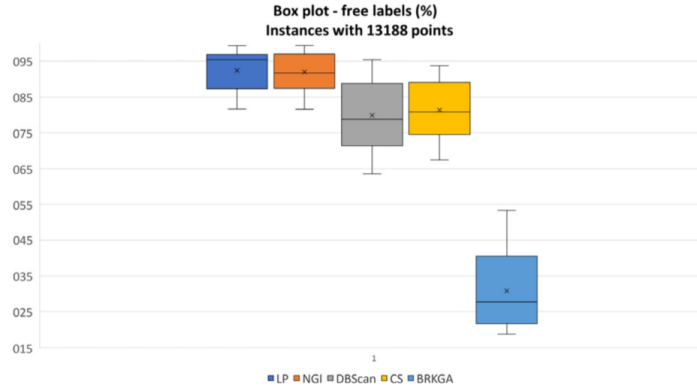


Figura 5 – Minimização do número sobreposições com propostas de *Clustering Search* no *BRKGA*

Como a Figura 5 destaca, os autores testaram as melhorias com *Clustering Search* no Algoritmo Genético *BRKGA* com *DBSCAN*, *Natural Group Identification - NGI* e *Label Propagation - LP*. Dentre os testes em 505, 5046 e 13188 pontos, os resultados são expressivos, chegando em melhorias na escala de 50% sobre as soluções do *BRKGA* neste último.

Outro problema, com características distintas é uso desta técnica aplicada no *Berth Allocation Problem - BAP* (OLIVEIRA; MAURI; LORENA, 2012). Esse é considerado o maior problema de terminais marítimos. A proposta envolve minimizar o tempo de carga e descarga de contêineres dos navios. Para isso foi utilizado um híbrido de *Simulated Annealing* com o *Clustering Search*. O tempo que deve ser minimizado é representado na Figura 6.

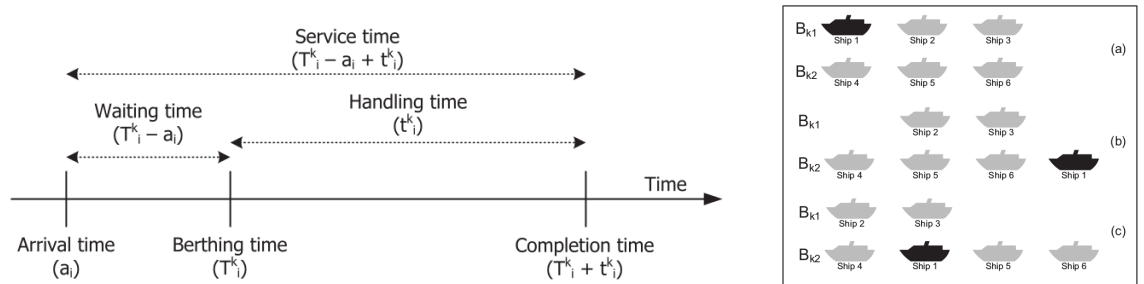


Figura 6 – Problema de Alocação de Terminais Marítimos. Improving the Clustering Search heuristic: An application to cartographic labeling

Como a Figura 6 apresenta, cada navio deve estar alocado aos terminais (*berths*). O problema tem de lidar com o tempo de serviço do navio, considerando otimizar a espera e a atividade dele no terminal. Para os testes foram utilizados dados com 30 instâncias com 60 navios e 13 *berths*, extraídos do porto de Gioia Tauro (Itália). Em geral, o artigo

apresenta os como *baseline* os resultados exaustivos adquiridos com o *CPLEX*. Assim, a aplicação de meta-heurísticas visa reduzir o tempo de resposta maximizando a qualidade das soluções. As comparações envolvem o uso de *Simulated Annealing with Re-Annealing (SA+RA)*, *Tabu Search (TS)*, *Population Training Algorithm with Linear Programming (PTA-LP)* e *Generalized Set-Partitioning Problem (GSPP)* para essa redução. A Figura 7 apresenta a perda em relação ao *CPLEX* das meta-heurísticas.

IMPROVEMENT (%)				
CPLEX	TS	SA + RA	PTA/LP	GSPP
-	0.42	0.00	0.00	0.00
51.61	0.16	0.00	0.00	0.00
55.98	0.88	0.00	0.00	0.00
70.09	0.08	0.00	0.00	0.00
54.83	0.08	0.00	0.00	0.00
-	0.08	0.00	0.00	0.00
55.70	0.00	0.00	0.00	0.00
74.91	0.00	0.00	0.00	0.00
-	0.00	0.00	0.00	0.00
-	0.00	0.00	0.00	0.00
-	0.73	0.00	0.07	0.00
58.67	0.00	0.00	0.00	0.00
-	0.00	0.00	0.00	0.00
-	0.00	0.00	0.00	0.00
72.28	0.00	0.00	0.00	0.00
68.43	0.80	0.00	0.07	0.00
-	0.00	0.00	0.00	0.00
63.46	0.07	0.00	0.00	0.00
43.04	0.22	0.22	0.00	0.00
-	0.00	0.00	0.00	0.00
-	0.37	0.00	0.00	0.00
61.99	0.45	0.00	0.00	0.00
-	0.00	0.00	0.00	0.00
74.11	0.08	0.00	0.00	0.00
30.96	0.22	0.07	0.00	0.00
47.70	0.90	0.00	0.00	0.00
60.70	0.00	0.00	0.00	0.00
-	0.44	0.07	0.07	0.00
73.38	0.16	0.00	0.00	0.00
-	0.52	0.00	0.00	0.00

Figura 7 –

Observamos através da Figura 7 a acentuada diferença do ótimo, ressaltando a dificuldade do processo de otimização. As meta-heurísticas, porém, apresentam resultados próximos entre elas, menos de 1% de diferença na qualidade. Em contraste com a pequena variação entre as meta-heurísticas, o tempo foi determinante. O tempo 12.79 do *Clustering Search* foi extremamente baixo em relação ao *CPLEX* (3600.00), *TS* (120.00), *SA+RA* (60.26), *PTA/LP* (93.99) e *GSPP* (14.98). Assim, a independência de busca dos centróides e dos terminais portuários estudados foram definitivos para esse resultado. A busca realizada em cada centróide garantiu atingir qualidades equiparáveis enquanto

Promissor, o *Clustering Search* têm ganhado espaço dentre os métodos de aprimoramento de soluções. Aplicado a vários tipos de problema, a proposta identifica pelo particionamento espacial grupos de soluções em convergência. Nesses grupos como os artigos relatam, esse método agiliza a convergência a mínimos através da Local Search em pontos estratégicos.

Referências

- ALTOÉ, W. A. et al. A Clustering Search Metaheuristic for the Bi-Objective Flexible Job Shop Scheduling Problem. In: IEEE. *2018 XLIV Latin American Computer Conference (CLEI)*. [S.l.], 2018. p. 158–166. Citado na página 6.
- ARAÚJO, E. J.; CHAVES, A. A.; LORENA, L. A. Improving the Clustering Search Heuristic: An Application to Cartographic Labeling. *Applied Soft Computing*, v. 77, p. 261–273, 2019. Citado 2 vezes nas páginas 5 e 6.
- BIAJOLI, F. L.; LORENA, L. A. N. Clustering Search Approach for the Traveling Tournament Problem. In: GELBUKH, A.; MORALES, Á. F. K. (Ed.). *MICAI 2007: Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 83–93. Citado na página 6.
- CHAVES, A. A.; CORREA, F. de A.; LORENA, L. A. N. Clustering search heuristic for the capacitated p-median problem. In: _____. *Innovations in Hybrid Intelligent Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 136–143. Citado na página 5.
- CHAVES, A. A.; LORENA, L. A. N. Clustering Search Algorithm for the Capacitated Centered Clustering Problem. *Computers & Operations Research*, Elsevier, v. 37, n. 3, p. 552–558, 2010. Citado na página 6.
- OLIVEIRA, A. C. M.; LORENA, L. A. N. Detecting Promising Areas by Evolutionary Clustering Search. In: BAZZAN, A. L. C.; LABIDI, S. (Ed.). *Advances in Artificial Intelligence – SBIA 2004*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 385–394. Citado na página 1.
- OLIVEIRA, A. C. M.; LORENA, L. A. N. Hybrid Evolutionary Algorithms and Clustering Search. In: _____. *Hybrid Evolutionary Algorithms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 77–99. Citado 2 vezes nas páginas 2 e 4.
- OLIVEIRA, A. M.; CHAVES, A. A.; LORENA, L. A. N. Clustering Search. *Pesquisa Operacional*, scielo, v. 33, p. 105–121, 04 2013. Citado na página 4.
- OLIVEIRA, R. M. de; MAURI, G. R.; LORENA, L. A. N. Clustering Search for the Berth Allocation Problem. *Expert Systems with Applications*, v. 39, n. 5, p. 5499–5505, 2012. Citado 2 vezes nas páginas 6 e 7.
- RABELLO, R. L. et al. A Clustering Search Metaheuristic for the Point-Feature Cartographic Label Placement Problem. *European Journal of Operational Research*, v. 234, n. 3, p. 802–808, 2014. Citado na página 6.