



# Aprendizado por reforço na avaliação do layout em problemas de empacotamento irregular

**Palavras-Chave:** Aprendizado de máquina, aprendizado por reforço, problemas de empacotamento irregular

**Autores(as):**

MURILO JOSÉ DE TOLEDO LIZIÉR, FCA – UNICAMP

Prof. Dr. WASHINGTON ALVES DE OLIVEIRA (orientador), FCA - UNICAMP

---

## INTRODUÇÃO:

Problemas de corte e empacotamento formam uma classe abrangente de desafios de otimização e possuem grande importância industrial, principalmente quando envolvem formatos irregulares (*Nesting Problems*), como observado em fabricação de roupas, cortes de chapas de metal, construção de móveis e manufatura de sapatos (BENNEL e OLIVEIRA, 2006). A Figura 1 ilustra um exemplo do layout de uma solução para esse problema.

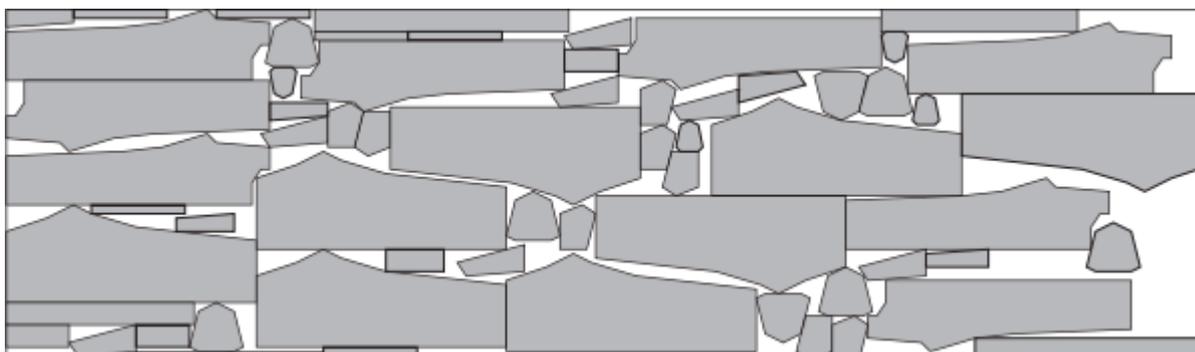


Figura 1: Exemplo de solução de problema de empacotamento irregular (LEAO, 2020)

Eles são considerados problemas NP-difíceis, ou seja, impossíveis de se resolver em tempo polinomial, na definição informal (COOK, 2000). Estratégias com alto grau de sofisticação são mais presentes na literatura de *Nesting Problems*, combinando modelos matemáticos complexos ao uso de heurísticas e metaheurísticas, além de programação linear inteira-mista, programação não-linear e programação de restrição (LEAO et al, 2020).

Para este trabalho será considerado o problema de empacotamento irregular em faixas, que se caracteriza pela alocação de peças (objetos irregulares ou regulares) sobre um objeto maior de

formato retangular, com uma dimensão fixa e a outra variável, adequando-se ao layout final da solução.

Para garantir que uma solução seja factível, é necessário garantir que os objetos menores estejam completamente internos ao objeto onde serão alocados e que, entre as peças, não haja nenhuma sobreposição (LEAO et al., 2020). Essas duas restrições configuram a maior dificuldade inerente aos problemas de empacotamento irregular, haja vista que tais delimitações são muito sensíveis à geometria das peças envolvidas.

Mais recentemente, estratégias baseadas em inteligência artificial, mais especificamente aprendizado por reforço, começaram a ser adotadas, apresentando bons resultados (FANG et al, 2023). Portanto, o presente trabalho utiliza uma variação do método *Q-learning* na busca pela melhor solução, visando minimizar a área final do retângulo que envolve o layout final.

## METODOLOGIA:

A fim de aplicar aprendizado por reforço a problemas de empacotamento irregular, foi adotado o método *Q-learning*, que consiste em: o agente testa uma ação em um estado determinado, então calcula as consequências em termos da recompensa ou penalidade imediata e estima o valor do estado ao qual é levado, o que, ao testar todas as possibilidades repetidas vezes, faz com que o agente aprenda a melhor solução (WATKINS e DAYAN, 1992). No *Q-learning* tradicional a matriz *Q* representa em suas colunas todas as ações possíveis que o agente pode tomar e, em suas linhas, todos os estados possíveis de serem alcançados pelo agente, nesse caso as entradas da matriz representam então o valor de cada solução, ou seja, o quão vantajoso é para o agente escolher determinada ação em determinado estado. A figura 2 ilustra uma notação generalizada para a matriz *Q*.

	A1	A2	A3	...	Am
S1	$Q(S1,A1)$	$Q(S1,A2)$	$Q(S1,A3)$		$Q(S1,Am)$
S2	$Q(S2,A1)$	$Q(S2,A2)$	$Q(S2,A3)$		$Q(S2,Am)$
S3	$Q(S3,A1)$	$Q(S3,A2)$	$Q(S3,A3)$		$Q(S3,Am)$
...					
Sn	$Q(Sn,A1)$	$Q(Sn,A2)$	$Q(Sn,A3)$		$Q(Sn,Am)$

Figura 2: Exemplificação da matriz *Q*

Para a atualização da matriz *Q* é utilizada, geralmente, a equação de Bellman, descrita pela Equação (1), onde  $Q(s,a)$  e  $Q^*(s,a)$  representam, respectivamente, os valores atuais e atualizados da matriz,  $R(s,a)$  representa a função recompensa,  $\max(Q^*(s',a'))$  significa o máximo valor previsto de recompensa diante do novo estado e, por fim,  $\alpha$  e  $\gamma$  representam respectivamente a taxa de

aprendizagem e o fator de desconto, parâmetros utilizados para evitar que o agente, após encontrar uma boa solução, não explore as demais.

$$Q^*(s, a) = Q(s, a) + \alpha * [R(s, a) + \gamma * (Q^*(s', a')) - Q(s, a)] \quad (1)$$

Porém, após a modelagem do problema, concluiu-se que , resultando assim na equação 2.

$$Q^*(s, a) = Q(s, a) + \alpha * R(s, a) \quad (2)$$

A modelagem em forma de um problema de aprendizado por reforço pode ser feita de mais de uma maneira. Por exemplo, como desenvolvido por Bartmeyer et al. (2021), tomando como ação as peças a serem alocadas e como estado a posição na sequência em que serão dispostas. Porém, visando uma maior precisão e inovação metodológica, este estudo se baseará em uma discretização do domínio em que os objetos serão alocados em coordenadas (x,y). Todavia, isso dificulta a modelagem do problema, visto que, caso as posições sejam consideradas os estados (linhas da matriz), a dimensão da matriz Q se tornaria inviável e complexa para tratar do ponto de vista computacional. Portanto, ao invés de ser utilizada uma matriz Q, será utilizado um tensor T de três dimensões, em que cada submatriz se refere a um objeto a ser alocado e suas entradas (de cada submatriz) representam as posições de alocação.

Exceto pela mudança de matriz para tensor, o processo de aprendizado funciona de forma convencional, operando cada submatriz normalmente como seria feito no caso do *Q-learning*, utilizando a Equação (2), onde os valores iniciais das entradas das matrizes foram todas definidas como 1,  $\alpha$  foi definido como 0.5, a fim de acelerar o processo de aprendizado, considerando que o problema trabalhado é de grande complexidade e, por isso, uma taxa de aprendizagem baixa aumentaria ainda mais o custo computacional. Além disso, a função recompensa  $R(s, a)$  foi construída da seguinte maneira:

$$R(s, a) = 20, \text{ caso a solução encontrada seja melhor que a anterior}$$

$$R(s, a) = 10, \text{ caso contrário}$$

Devido à natureza do problema e as dificuldades envolvidas, foi decidido primeiramente realizar testes com círculos ao invés de peças irregulares, pela maior facilidade de verificação de sobreposição e, por consequência, sua factibilidade é mais simples de ser verificada. Isso é de grande importância ao projeto pois, mesmo não sendo o objetivo final, uma das estratégias de representação de peças irregulares é pelo círculo que envolve cada peça (aproximações por regiões convexas), reduzindo assim o custo computacional da solução (ROCHA et al, 2014). Para exemplificar a proposta desse estudo, na sequência apresentamos um exemplar da alocação de 5 círculos, de raios 0.1, 0.2, 0.3, 0.4 e 0.5 em um contêiner, representado por um retângulo de 5 unidades de comprimento e 100 unidades de largura (intuito de representar um retângulo semi-infinito).

## RESULTADOS E DISCUSSÃO:

A fim de metrificar os resultados obtidos, uma solução inicial foi definida para que se possa comparar com as soluções encontradas. Para isso foi adotada a pior solução factível possível, ou seja, os círculos estão alocados de forma com que o menor retângulo que os envolve seja coincidente com o contêiner. Portanto, para avaliar os resultados foram utilizadas as seguintes duas medidas descritas na segunda e terceira colunas na Tabela 1: a relação entre a altura da solução encontrada e a altura da solução inicial (100 unidades); e o tempo decorrido para realizar todas as iterações de aprendizado, ou seja, quanto tempo levou para o tensor ser atualizado um número determinado de vezes.

Como os testes têm caráter aleatório, foram realizados os testes com 5, 10, 25, 100 e 1000 iterações 5 vezes cada, com função de obter resultados mais precisos, evitando que um resultado que não tenha sido satisfatório, por conta da randomicidade do processo, prejudique as análises posteriores. Os resultados obtidos foram os seguintes:

Número de iterações	Relação (média)	Tempo (total)
5	0.6442	30 segundos
10	0.5046	60 segundos
25	0.4046	143 segundos
100	0.603	475 segundos
1000	0.403	4000 segundos

Tabela 1: Resultados obtidos

Esses resultados, mesmo preliminares e ainda distantes da otimalidade, mostram que o processo de aprendizado está de acordo com o previsto, ou seja, convergindo congruentemente com o número de iterações. Esses resultados são bons indicadores de que adotar a mesma estratégia para peças irregulares pode também apresentar bons resultados.

## CONCLUSÕES:

Por se tratar de uma estratégia inovadora e ainda com muito a ser desenvolvido e melhorado, além de ser intrinsecamente mais custosa computacionalmente que a maioria dos outros métodos de aprendizado por reforço aplicados a problemas de empacotamento irregular, os resultados obtidos, ainda parciais, não são completamente satisfatórios, porém mostram um avanço e abrem possibilidades para que se possa vislumbrar estratégias de melhoria. Uma possível solução para isso seria adotar uma função recompensa mais adequada ao problema, penalizando alocações de peças

que aumentariam a área da solução final, ou seja, coordenadas ordenadas maiores, visando assim, por exemplo, direcionar mais o agente para uma solução ótima.

## **Agradecimentos:**

Por fim, gostaríamos de agradecer a Dr<sup>a</sup> Petra Maria Bartmeyer pelo grande apoio, incentivo e contribuição para esta pesquisa, validando e simplificando muito o processo.

---

## **BIBLIOGRAFIA**

BANNEL, J. A.; OLIVEIRA, J. F. **The geometry of nesting problems: A tutorial**. European journal of operational research, 184(2):397–415, 2008.

COOK, S. **The P versus NP Problem 1. Statement of the Problem**. [s.l: s.n.]. Disponível em: <<http://www.cs.utoronto.ca/~toni/Courses/Complexity/handouts/cook-clay.pdf>>. Acesso em: 22 jul. 2023.

LEAO, A.A.S.; TOLEDO, F.M.B.; OLIVEIRA, J.F. et al. **Irregular packing problems: A review of mathematical models**. European Journal of Operational Research, 282(3):803–822, 2020.

FANG, J.; RAO, Y.; ZHAO, X.; DU, B. **A Hybrid Reinforcement Learning Algorithm for 2D Irregular Packing Problems**. Mathematics, v. 11, n. 2, p. 327, 2023.

WATKINS, C.J.C.H.; DAYAN, P. **Q-learning**, Machine Learning, v.8, 279-292, 1992.

ROCHA, P.; RODRIGUES, R.; GOMES, A.M.; TOLEDO, F.M.B.; ANDRETTA, M. **Circle Covering Representation for Nesting problems with continuous rotations**. In: World Congress The International Federation of Automatic Control, 19, 2014, Cidade do Cabo. **Anais [...]**. p. 5235-5240.