Aprendizado por reforço: aplicação no problema de empacotamento

Gabriel Medeiros Lopes Carneiro (19103977)

Mikaella Cristina Bernardo Vieira (18103860)

Problema de Empacotamento

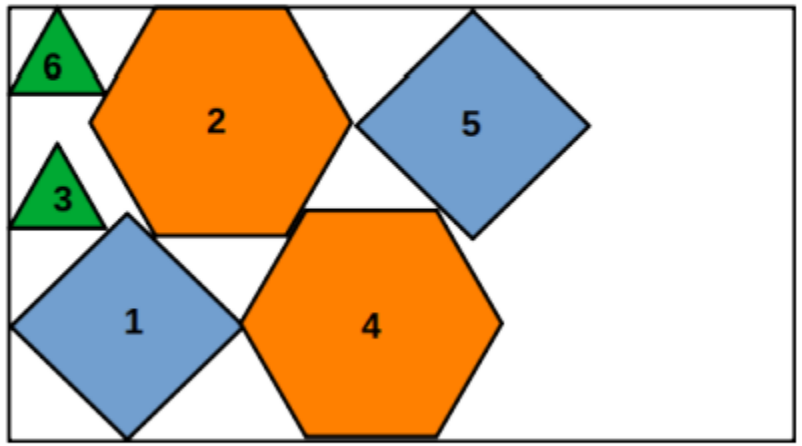
- Alocar conjunto de objetos dentro de um objeto maior.
  - Ex.: caixas em container.
- Objetos podem ser regulares ou não.
  - Quantidade de parâmetros necessários para identificação.
- O objeto maior pode ter dimensões fixas ou não.

Empacotamento de peças irregulares

- Peças irregulares bidimensionais.
- Objeto retangular de altura fixa.

Botton-left

(a) Sequência para alocação.



(b) Alocação gerada pela regra *bottom-left*.

Figura 1: Representação de uma sequência de alocação seguindo a regra *bottom-left*.

Aprendizado por reforço

- Baseado na qualidade das decisões tomadas.
  - Decisões serão analisadas e receberão recompensas e penalidades.
- Recompensas podem ser iguais independentes da qualidade da solução ou não.

Aprendizado por reforço

- Analise pode ser feita passo a passo ou ao final da solução.
  - Com alto número de repetições é possível superar as limitações de cada método.
- *Q-learning*.

Matriz de Aprendizado

- Representação do método *Q-learning*.
- $Q(n \times m)$.
  - n quantidade de tipos de peça.
  - m tamanho total da sequência.
- Q_{ij} representa o benefício do uso de uma peça do tipo i na posição j.
- A matriz é atualiza a cada solução gerada.

$$Q_{ij} = Q_{ij} + \alpha; \quad \text{se } (BW - CW) \geq 0 \text{ ou } (OW - CW) \geq 0, \tag{1}$$

$$Q_{ij} = Q_{ij} - \beta; \quad \text{caso contrário.} \tag{2}$$

Algoritmo 1 Aprendizado por reforço.

1: Dados de Entrada: OW, BW, d, m, α e β
2: **procedimento** ITERAÇÕES DE APRENDIZADO POR REFORÇO
3: $Q \leftarrow \mathbf{0}$
4: **enquanto** critério de parada não atingido **faça**
5: $S \leftarrow \emptyset$
6: **para cada** posição da solução (j)
7: $p_i \leftarrow e^{Q_{ij}}, \ \forall i : d_i \geq 0$ ▷ Calcular contribuição de cada peça factível
8: $S_j \leftarrow i$ ▷ Seleção por roleta ponderada
9: $d_i \leftarrow d_i - 1$ ▷ Atualiza a demanda da peça selecionada
10: $CW \leftarrow BL(S)$ ▷ Calcula o comprimento da solução usando a regra *bottom-left*
11: **se** $(BW - CW) \geq 0$ ou $(OW - CW) \geq 0$ **então** ▷ Atualização da matriz Q
12: **para cada** entrada da sequência
13: $Q_{ij} \leftarrow Q_{ij} + \alpha$
14: **senão**
15: **para cada** entrada da sequência
16: $Q_{ij} \leftarrow Q_{ij} - \beta$
17: **se** $CW \leq BW$ **então** ▷ Atualização da melhor solução incumbente
18: $BW \leftarrow CW$

Transferência de aprendizado

- A maioria das peças são similares.
  - Criar matriz de aprendizado e repassar para novas soluções.
- Matriz de aprendizado precisa ser redimensionada.
- Iterações puramente aleatórias
  - Adaptação ao novo exemplar.

Algoritmo 2 Aprendizado por reforço com transferência.

Dados de Entrada: Q, OW, BW, d, m, α e β
enquanto critério de parada não atingido **faça**
    procedimento ITERAÇÕES DE APRENDIZADO POR REFORÇO
        $S \leftarrow \emptyset$
        para cada posição da solução (j)
            $p_i \leftarrow e^{Q_{ij}},\ \forall i : d_i \geq 0$ ▷ Calcular contribuição de cada peça factível
            $S_j \leftarrow i$ ▷ Seleção por roleta ponderada
            $d_i \leftarrow d_i - 1$ ▷ Atualiza a demanda da peça selecionada
        $CW \leftarrow BL(S)$ ▷ Calcula o comprimento da solução usando a regra *bottom-left*
        se $(BW - CW) \geq 0$ ou $(OW - CW) \geq 0$ **então** ▷ Atualização da matriz Q
            para cada entrada da sequência
                $Q_{ij} \leftarrow Q_{ij} + \alpha$
        senão
            para cada entrada da sequência
                $Q_{ij} \leftarrow Q_{ij} + \beta$
        se $CW \leq BW$ **então** ▷ Atualização da melhor solução incumbente
            $BW \leftarrow CW$
    procedimento ITERAÇÕES COM SOLUÇÕES ALEATÓRIAS
        $S \leftarrow RS$ ▷ Cria uma solução factível aleatória com probabilidade uniforme
        $CW \leftarrow BL(S)$ ▷ Calcula o comprimento da solução usando a regra *bottom-left*
        se $(BW - CW) \geq 0$ ou $(OW - CW) \geq 0$ **então** ▷ Atualização da matriz Q
            para cada entrada da sequência
                $Q_{ij} \leftarrow Q_{ij} + \alpha$
        senão
            para cada entrada da sequência
                $Q_{ij} \leftarrow Q_{ij} - \beta$
        se $CW \leq BW$ **então** ▷ Atualização da melhor solução incumbente
            $BW \leftarrow CW$

Comparativo

- Aprendizado por reforço (R).
  - 700 segundos.
- Transferência de aprendizado (T).
  - 600 segundos para geração de Q.
  - 100 segundos no algoritmo 2.

Comparativo

- 10 exemplares.
  - 5 de peças convexas (rco).
  - 5 com peças côncavas (blazewicz).

Tabela 1: Descrição dos exemplares utilizados.

Exemplar	n	d	m
rco1	7	1	7
rco2	7	2	14
rco3	7	3	21
rco4	7	4	28
rco5	7	5	35
blazewicz1	7	1	7
blazewicz2	7	2	14
blazewicz3	7	3	21
blazewicz4	7	4	28
blazewicz5	7	5	35

Comparativo

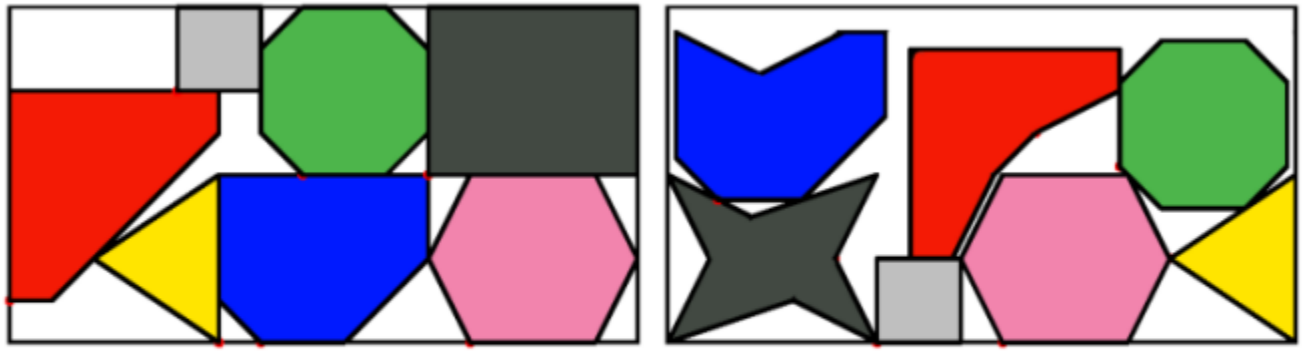


Figura 3: Peças dos exemplares *rco* (esquerda) e *blazewicz* (direita). As peças de mesma cor representam os pares de peças côncavas (*blazewicz*) e seus contornos convexos (*rco*).

- Exemplares côncavos tiveram menos iterações.
  - Maior possibilidade de posições devido aos vértices.

Resultados

Tabela 2: Resultados obtidos pelos métodos de aprendizado por reforço com e sem transferência (**T e R**) de aprendizado para os exemplares *rco* e *blazewicz*.

			Mínimo		Máximo		Mediana		Desvio Padrão	
	OW	Ref.	**T**	**R**	**T**	**R**	**T**	**R**	**T**	**R**
rco1	8,00	8,00	8,00	8,00	12,00	12,00	9,00	9,00	0,89	0,99
rco2	17,00	14,87	15,33	15,50	22,33	20,66	17,50	17,00	1,00	0,83
rco3	25,00	22,44	23,00	22,33	32,00	32,00	26,00	26,00	1,21	1,21
rco4	29,00	30,44	31,00	30,00	39,66	40,00	34,00	34,00	1,27	1,29
rco5	41,00	38,40	39,11	37,66	47,40	49,00	42,00	42,11	1,38	1,39
blazewicz1	8,00	7,40	7,40	7,43	11,85	11,50	9,00	9,00	0,87	0,85
blazewicz2	16,00	14,25	14,58	14,50	20,80	21,38	17,08	17,03	1,08	1,06
blazewicz3	22,00	21,68	21,86	22,13	30,58	29,41	24,84	24,95	1,33	1,19
blazewicz4	29,00	29,44	30,19	30,50	35,43	37,42	32,75	32,94	1,58	1,49
blazewicz5	36,00	41,05	37,36	37,47	43,97	44,81	40,27	40,15	1,79	1,79

Conclusão

- Peças côncavas exigem maior esforço para solução.
- Transferência de aprendizado traz vantagens.
  - Desempenho similar ao aprendizado por reforço, com tempo de execução menor.
- A medida que nº de exemplares cresce, a transferência encontrou melhores comprimentos mínimos.

Referências

- Bartmeyer, P., Oliveira, L., Toledo, F. e Leão, A. (2021). Aprendizado por reforço aplicado ao problema de empacotamento de peças irregulares em faixas. **LIII Simpósio Brasileiro de Pesquisa Operacional**.
- Watkins, C. J. e Dayan, P. (1992). Q-learning. Machine learning, 8(3-4):279–292.