

Aprendizado por Reforço Aplicado ao Problema de Alocação de Berços no Porto de Paranaguá e Antonina

César Eduardo Maciel

Pós-Graduação em Métodos Numéricos em Engenharia
Universidade Federal do Paraná
Curitiba, Brasil
(cesarmaciel@ufpr.br)

Eduardo Tadeu Bacalhau

Luciana Casacio

Campus Pontal do Paraná - Centro de Estudos do Mar
Universidade Federal do Paraná
Pontal do Paraná, Brasil
(bacalhau@ufpr.br; lucianacasacio@ufpr.br)

Resumo—Este estudo propõe a aplicação de uma estratégia de Aprendizado por Reforço (Reinforcement Learning – RL) para o Problema de Alocação de Berços (PAB) na Associação de Portos de Paranaguá e Antonina (APPA). A APPA é um terminal portuário complexo, com berços multipropósito e diversas restrições operacionais que dificultam o processo de atracação dos navios. O PAB é modelado como um processo de decisão sequencial, no qual os estados representam os tempos remanescentes de serviço em cada berço e as ações correspondem às possíveis alocações de navios com diferentes tipos de carga. Utiliza-se a estratégia de *Q-learning* para otimizar as decisões de alocação, visando minimizar os tempos de espera dos navios e maximizar a utilização do cais. Estudos de casos são conduzidos para demonstrar o desenvolvimento da abordagem por RL e validar os resultados aplicados a instâncias elaboradas a partir de dados históricos de atracamentos no sistema portuário estudado. Os resultados demonstram que a abordagem por RL reduz atrasos e melhora a ocupação do cais em relação às estratégias de atracação tradicionalmente usadas.

Palavras-chave—Aprendizado por Reforço, Problema Alocação de Berços, *Q-Learning*

I. INTRODUÇÃO

De acordo com [1], a logística portuária continua desempenhando um papel fundamental no comércio global, respondendo por mais de 80% do volume total de mercadorias transportadas internacionalmente. Em 2023, o comércio marítimo global cresceu 2,4%, apesar dos desafios geopolíticos e climáticos que impactam a cadeia logística. Essas pressões resultaram em um aumento na média da distância percorrida por tonelada de carga, que alcançou 5.186 milhas, elevando custos e emissões de carbono.

Os portos atuam como pontos estratégicos que conectam cadeias de suprimentos, facilitando o fluxo de importações e exportações. Contudo, grandes terminais enfrentam desafios operacionais significativos, como congestionamentos, longos tempos de espera, subutilização de berços e custos elevados [2]. Com o aumento contínuo da demanda, esses problemas tendem a se agravar, tornando a eficiência operacional ainda mais crucial para a competitividade econômica e sustentabilidade ambiental, já que atrasos geram maiores emissões e custos [3]–[5].

A Associação de Portos de Paranaguá e Antonina (APPA), no sul do Brasil, é um dos principais centros logísticos do país, movimentando milhões de toneladas, com destaque para grãos, fertilizantes e contêineres. Em 2024, o volume total atingiu 66 milhões de toneladas, um aumento de 2,1% em relação ao ano anterior [2]. Apesar dessa demanda crescente, a APPA encara desafios operacionais relevantes, especialmente durante picos como a safra de grãos. A complexidade operacional é ainda maior devido à heterogeneidade de cargas, berços multipropósitos e forte sazonalidade [6]–[8]. Esses fatores reforçam a relevância da pesquisa apresentada e a necessidade de futuros estudos.

O Problema de Alocação de Berços (PAB) trata da atribuição ótima de navios aos berços disponíveis, considerando restrições como horários de chegada, duração da atracação, tipo de carga, prioridades comerciais e características físicas dos navios, como calado e comprimento. O PAB pode ser modelado como problema discreto, contínuo ou híbrido [9]–[14]. Além disso, as abordagens para resolver o problema do PAB podem considerar cenários de chegada estática — em que as embarcações já estão no porto e podem ser atracadas a qualquer momento [11], [15] — ou chegada dinâmica, onde os navios chegam ao porto a qualquer instante [13], [16]–[18].

Métodos heurísticos têm sido amplamente utilizados para resolver o PAB [4], [19]–[21]. Por exemplo, [22] aplicou algoritmos genéticos no contexto da APPA, com ganhos operacionais importantes. Entretanto, a dinâmica e incerteza das operações reais têm impulsionado a adoção de técnicas mais avançadas, como o aprendizado de máquina.

O avanço nas técnicas é fundamental, já que modelos clássicos, como o de [23] baseado em programação inteira mista (MILP), adotam formulações estáticas que não capturam adequadamente as incertezas ou prioridades operacionais reais. Por isso, técnicas como *Reinforcement Learning* (RL) têm sido cada vez mais estudadas para ambientes portuários complexos [24], assim como arquiteturas baseadas em redes neurais profundas [25], as quais demonstram boa eficiência nesses cenários.

Estudos recentes exploram com sucesso técnicas de RL adaptadas à operação portuária: [26] combinaram modelos MILP *offline* com algoritmos *Dueling Deep Q-Network* (DDQN) para otimização *online* do PAB, incluindo tempos de preparação de guindastes; [27] desenvolveram uma arquitetura baseada em *Double Dueling DQN* (D3QN) para lidar com a aleatoriedade na chegada de navios; [28] mostraram a eficácia do *Q-learning* integrado à simulação para otimização de equipamentos portuários; e [29] avançaram no controle automático de atracação combinando aprendizado supervisionado e RL.

Embora esses avanços sejam importantes, as abordagens baseadas em aprendizado profundo ainda enfrentam desafios práticos em ambientes portuários reais, como a necessidade de grandes volumes de dados e a dificuldade em incorporar regras operacionais específicas e restrições logísticas. Isso limita sua aplicação em cenários como o da APPA, onde decisões devem refletir prioridades dinâmicas em tempo quase real.

Diante disso, este trabalho apresenta uma abordagem baseada em *Q-learning*, adequada para lidar com as incertezas operacionais decorrentes de atrasos na carga e descarga, com foco especial em cargas como fertilizantes, que possuem particularidades logísticas. Para acelerar o treinamento, propõe-se agrupar os berços em *clusters* segundo suas características de carga, reduzindo o espaço de estados e tornando o aprendizado mais eficiente. O problema é formulado como um Processo de Decisão Sequencial (SDP — *Sequential Decision Process*), no qual as transições de estado seguem um processo estocástico que descreve os tempos restantes de serviço nos berços, modelados por distribuição binomial. O estado representa o tempo de serviço restante, e as ações correspondem às alocações restritas ao tipo de carga e berço. Essa abordagem mostrou bom desempenho em ambientes dinâmicos, como o da APPA, especialmente para cargas logísticas complexas.

II. FORMULAÇÃO DO PROBLEMA

O PAB é abordado como um problema discreto, em um conjunto finito de berços com tamanhos fixos. Além disso, trata a restrição temporal de forma dinâmica, simulando a natureza operacional do sistema portuário estudado, a APPA.

Esse sistema portuário é caracterizado por um cais composto por 24 berços que contemplam diferentes preferências de carga, as quais devem ser estritamente obedecidas. Algumas particularidades também devem ser mencionadas, como a retirada de um navio da fila. Este caso não é considerado no presente estudo, pois caracteriza outro problema a ser tratado, relacionado à taxa de sobreestadia, que é cobrada pelos portos aos contratantes do navio. A Fig. 1 ilustra o cais completo da APPA, destacando os complexos de Paranaguá e Antonina.

Como mencionado anteriormente, no PAB os berços têm tamanhos fixos e recebem mercadorias específicas, o que limita a flexibilidade na alocação. A Tabela I apresenta as características de cada berço, indicando, em metros, o comprimento (Comp.), a profundidade (Prof.) e a preferência de mercadoria (Pref.) no sistema portuário da APPA. As preferências são abreviadas pelas seguintes siglas: GL (Granéis

Líquidos) – combustíveis, produtos químicos e óleos alimentícios; GS (Granéis Sólidos) – fertilizantes, grãos, farelos e sal; CG (Carga Geral) – açúcar, veículos e passageiros; CTNR (Contêineres) – veículos e produtos; e Multi (Multipropósito) – açúcar, passageiros, grãos e fertilizantes.

Tabela I: Instalações Portuárias da APPA

Berço	Comp.	Prof.	Pref.	Berço	Comp.	Prof.	Pref.
141	210	12.1	GL	211	200	12.7	GS
142	190	10.6	GL	212	251	13.0	GS
143	220	11.9	GL	213	253	13.0	GS
144	220	8.0	GL	214	259	13.0	GS
200	190	12.5	GS	215	336	13.0	CTNR
201	225	11.8	GS	216	336	13.0	CTNR
202	200	11.3	CG	217	336	13.0	CTNR
204	190	11.8	GS	218	200	10.0	CTNR
205	160	11.0	CG	219	200	10.0	CTNR
206	225	10.6	GS	200A	235	9.2	GS
208	200	9.3	GS	113	160	9.5	Multi
209	241	12.7	GS	114	200	9.5	Multi

As principais desafios em aplicar o PAB no sistema portuário proposto está na presença de berços que atendem cargas de diversos tipos, como os berços multipropósitos, e nas cargas que são tratadas como preferências em berços distintos, como os fertilizantes. Nesse contexto, o aprendizado por reforço (RL) surge como uma alternativa promissora e, como o *Q-learning* pode ser aplicado de forma *model-free*, propõe-se formalizar o problema como um SDP.

A. Processo de Decisão Sequencial

O PAB é formulado como um SDP, que possibilita lidar com a dinâmica e as restrições operacionais do porto, refletindo a complexidade real da alocação dos berços. São apresentados os elementos fundamentais que compõem esse processo, como os estados, as ações, as recompensas e o processo de transição entre os estados.

1) *Estados (S)*: Representam a configuração atual do cais, com base na ocupação dos berços. Cada estado é representado por:

$$s^t = [s_1^t, s_2^t, \dots, s_b^t] \quad (1)$$

onde b é o número de berços e s_j^t é o número de horas restantes para a finalização do serviço no berço j no instante t . Cada estado $s \in S$ captura as condições operacionais do sistema em um determinado momento.

2) *Ações (A)*: Cada ação possível em um instante de decisão t é representada por:

$$a_k^t = (b_j, c_i, d) \quad \text{com} \quad a_k^t \in A \quad (2)$$

onde b_j é o identificador de um berço j disponível, $c_i \in \{\text{carga 1, carga 2, } \dots, \text{carga } m\}$ representa o tipo de carga m do navio i , $d \in \{0, 1\}$ indica o atraso, sendo $d = 0$ a alocação sem atraso e $d = 1$ a alocação com atraso, e o índice k representa a identificação única de uma ação dentro do conjunto A , em um determinado instante t . A ação de não realizar nenhuma alocação é representada por

$$a_{\text{zero}} = (0, 0, 0),$$



Figura 1: Organização dos berços de atracação na APPA – Fonte: www.portosdoparana.pr.gov.br

e não gera recompensa associada.

O conjunto A de ações possíveis em cada instante t é condicionado pelas restrições operacionais de compatibilidade entre o navio e os berços (dimensões, tipo de carga) e a disponibilidade atual dos berços, conforme o estado s^t .

3) *Recompensa (R)*: A função de recompensa $R(a_k^t, s^t)$ avalia o impacto da ação $a_k^t \in A$ no estado atual do sistema s^t . Ela é definida por:

$$R(a_k^t, s^t) = \begin{cases} -1, & \text{se a ação resultar em violação} \\ & \text{de restrição de ocupação do berço;} \\ 0, & \text{se } a_k^t = (0, 0, 0); \\ 2 \cdot \sum_{j=1}^b s_j^t, & \text{se a alocação for feita com atraso;} \\ \sum_{j=1}^b s_j^t, & \text{caso contrário.} \end{cases} \quad (3)$$

A função proposta visa incentivar alocações que aumentam a utilização dos berços (refletida pela soma dos tempos de ocupação), penaliza decisões inválidas com recompensa negativa e atribui um peso maior às ações realizadas com atraso.

4) *Processo de Transição de Estados*: A transição de estados é um processo estocástico que descreve como os tempos restantes de serviço nos berços podem ser alterados ao longo do tempo. Contudo, para reduzir o número de estados presentes no processo de aprendizagem, os tempos foram discretizados em múltiplos de três horas, resultando em uma representação mais compacta.

A cada passo de tempo (1 hora), com probabilidade $p = 0,9$, o tempo restante de atendimento de cada berço ocupado é reduzido em uma quantidade aleatória entre 1 e 3 horas. O estado atual é então ajustado de forma compatível com a discretização adotada: primeiro se converte o valor discreto para o tempo real (multiplicando por 3), aplica-se a redução, e em seguida o valor resultante é novamente convertido ao espaço discreto (via divisão inteira com arredondamento para cima).

A dinâmica de transição é expressa pela seguinte equação:

$$s_j^{t+1} = \begin{cases} \left\lceil \frac{(s_j^t \cdot 3 - \delta)}{3} \right\rceil, & \text{com probabilidade } p, \\ s_j^t, & \text{caso contrário,} \end{cases} \quad (4)$$

onde s_j^t é o estado do berço j no instante t , $\delta \in \{1, 2, 3\}$ representa a redução estocástica em horas, e s_j^{t+1} é o novo estado no instante seguinte. Esse mecanismo permite representar a variabilidade das operações portuárias, ao mesmo tempo em que mantém a complexidade computacional controlada por meio da discretização.

III. APRENDIZADO POR REFORÇO ADAPTADO

A abordagem desenvolvida emprega *Q-learning* adaptado para otimizar a alocação dinâmica de berços no complexo portuário da APPA. A proposta é composta por um processo de clusterização dos berços, pelo treinamento independente de tabelas Q para cada *cluster* e pela validação por meio de simulações baseadas em dados reais de operação portuária.

A. Clusterização dos Berços

Os berços são agrupados com base na similaridade operacional, considerando principalmente os tipos de carga atendidos. Esse agrupamento visa segmentar o espaço de decisão em subconjuntos, de modo que cada *cluster* represente uma configuração típica de operação das cargas observadas. Para isso, são utilizados dados do sistema portuário da APPA, analisando-se a distribuição de cargas por berço. Ao agrupar berços com perfis semelhantes, os *clusters* formados reduzem o espaço de busca do RL, tornando o problema mais eficiente do ponto de vista computacional.

B. Treinamento por Cluster

Para cada *cluster* identificado, constrói-se uma tabela Q independente, responsável por mapear os pares estado-ação para um valor esperado de retorno. O treinamento é realizado a partir de sequências simuladas de alocação de navios, nas quais cada decisão afeta o estado subsequente e gera uma recompensa associada, conforme definido na Seção II-A. A tabela Q é atualizada iterativamente durante o treinamento, refletindo o conhecimento acumulado sobre os efeitos das ações em diferentes contextos. A discretização do tempo de serviço em seis níveis por berço combinada com a divisão dos *clusters* reduz significativamente o espaço de estados e torna o treinamento mais eficiente e modular.

A equação de atualização segue a forma clássica do *Q-learning*.

$$Q(s^t, a^t) \leftarrow Q(s^t, a^t) + \alpha \left[R(a_k^t, s^t) + \gamma \max_{a'} Q(s^{t+1}, a') - Q(s^t, a^t) \right] \quad (5)$$

em que α representa a taxa de aprendizado, γ o fator de desconto, $R(a_k^t, s^t)$ a recompensa observada ao aplicar a ação a_k^t no estado s^t , e s^{t+1} o estado resultante. O processo é conduzido separadamente para cada *cluster*, refletindo as particularidades operacionais de cada grupo de berços.

C. Processo de Simulação

A avaliação do modelo é realizada em um ambiente que simula a dinâmica estocástica do sistema portuário, utilizando dados históricos reais. Os tempos de serviço nos berços evoluem segundo uma distribuição binomial, conforme descrito na Seção II-A, o que reflete a incerteza das operações portuárias.

O sistema opera em ciclos horários, nos quais decisões de alocação são tomadas em tempo real com base nos valores atuais da tabela Q . Para cada navio, identifica-se o *cluster* correspondente e seleciona-se diretamente a ação (alocação) que apresenta o maior valor Q entre as opções disponíveis para o estado atual.

Navios não alocados permanecem em fila, acumulando tempo de espera, permitindo avaliar o impacto de decisões subótimas.

A simulação segue um ciclo iterativo com cinco etapas principais: (i) observação do estado do sistema com chegada de navios; (ii) escolha da ação com melhor valor na tabela Q ; (iii) verificação da viabilidade da ação segundo as restrições físicas e operacionais; (iv) alocação do navio ou acréscimo no tempo de espera, conforme o caso; e (v) encerramento ao fim do horizonte, com o acúmulo dos tempos de serviço e espera.

Assim, a simulação cria um ambiente operacional realista e dinâmico, permitindo analisar a eficácia do algoritmo desenvolvido. Nos estudos de caso a seguir, diferentes cenários operacionais são explorados para validar a abordagem proposta.

IV. ESTUDO DE CASOS

A investigação concentrou-se em dois aspectos principais: (i) a construção de uma representação operacional reduzida, obtida por meio do agrupamento de berços com base em similaridade de características operacionais; e (ii) a comparação do desempenho do modelo proposto com políticas tradicionais de operação, especificamente: *FIFO* (*First-In First-Out*), que realiza a alocação dos navios estritamente pela ordem de chegada; e *PHT* (*Priority of Handling Time*), que prioriza os navios com maior soma entre o tempo de serviço e o tempo de espera. Ambas as políticas, embora amplamente utilizadas em operações portuárias, não consideram atributos operacionais como tipo de carga ou prioridades específicas.

A. Construção das Instâncias

O conjunto de instâncias foi elaborado com base em fontes públicas, incluindo Portos do Paraná [30], *MarineTraffic* [31] e NOAA [32], reunindo informações sobre 521 navios entre abril de 2024 e janeiro de 2025. Foram consideradas variáveis

como tipo de carga, dimensões dos navios e seus respectivos tempos de serviço, além de características operacionais dos berços.

Os horários de chegada foram gerados de forma estocástica, com o objetivo de simular a imprevisibilidade do ambiente real. As alocações entre navios e berços seguem critérios de compatibilidade física e operacional e a distribuição dos horários de chegada é definida com base no número de horas estabelecido para o horizonte de simulação. Ressalta-se que, quanto maior for a proporção entre o horizonte de simulação e o número de navios, mais espaçados tendem a ser os tempos de chegada. Essa estrutura contribui para representar de forma mais realista a dinâmica operacional do terminal.

A Tabela II apresenta um exemplo de instância composta por cinco navios, construída com base nos dados descritos. São destacados o nome do navio, comprimento e calado (em metros), tipo de carga, horário de chegada e tempo do serviço (em horas).

Tabela II: Instância 5 Navios

Nome	Compr.	Calado	Carga	Chegada	Serviço
NAVIO37	203.2	10.7	FERTILIZANTES	10	11
NAVIO356	199.9	7.0	COMBUSTÍVEIS	11	13
NAVIO82	199.99	8.2	VEÍCULOS	11	9
NAVIO298	179.97	6.6	GRÃOS	11	15
NAVIO439	199.9	7.3	AÇÚCAR	12	6

B. Análise de Complexidade

Antes da definição dos experimentos, foi realizada uma análise de complexidade do espaço de estados associado à alocação de navios nos berços da APPA. Essa etapa teve como objetivo demonstrar a inviabilidade de aplicação de métodos exatos, como a Programação Dinâmica [33], e justificar a adoção de técnicas de redução de dimensionalidade.

O problema envolve 24 berços operacionais (Tabela I), cada um com possíveis tempos de espera discretizados entre 0 e 15 horas, como fora observado nos dados históricos coletados. Considerando 16 níveis discretos por berço, o total de combinações possíveis de estados seria da ordem de 16^{24} , o que representa aproximadamente $3,0 \times 10^{28}$ soluções.

Essa explosão combinatória é uma manifestação da *maldição da dimensionalidade* (do inglês, *curse of dimensionality*), conceito introduzido por Bellman [34], que descreve o crescimento exponencial do espaço de estados com o aumento do número de variáveis de decisão, tornando inviável a aplicação de métodos exatos. Em problemas de decisão sequencial como o presente, essa complexidade compromete diretamente a viabilidade do treinamento e da convergência de algoritmos por RL.

Diante dessa explosão combinatória, procedeu-se à redução da dimensionalidade por meio de duas estratégias complementares: (i) a clusterização dos berços em grupos funcionalmente similares — os 24 berços foram agrupados entre 5 e 8 *clusters*. Por exemplo, o *cluster* 5, que representa um subconjunto com 5 agrupamentos, ainda apresentava complexidade significativa. Mesmo com discretização de 16 níveis, um *cluster* com 7

berços resultaria em $16^7 = 268.435.456$ possíveis estados; (ii) adota-se uma discretização do tempo de serviço em seis níveis: nível 0 (sem espera) e cinco níveis adicionais, representando faixas de 3 horas cada (1–3h, 4–6h, 7–9h, 10–12h e 13–15h). Essa escolha reduz o número de estados por berço de 16 para 6, resultando em $6^7 = 279.936$ combinações, porém, um valor ainda elevado, mas manejável para abordagens em RL. Como destaque, o *cluster* 1, composto por 7 berços, resultou em uma tabela Q com 279.936 estados e 29 ações, totalizando mais de 8 milhões de pares (s, a) . O treinamento desse *cluster* foi realizado ao longo de cerca de 162,4 milhões de épocas e consumiu aproximadamente 1,3 h de processamento em uma máquina com Intel Core i5-13450HX (2,4 GHz) e 24 GB de RAM. Outros *clusters* variaram entre 5 e 7 berços, com tabelas Q menores e número de épocas ajustado à sua complexidade, sendo o número mínimo definido em 5 milhões.

C. Experimentos realizados

1) *Experimento I - Processo de Divisão de Cluster*: Este experimento teve como objetivo principal investigar estratégias de agrupamento dos berços da APPA, visando reduzir a dimensionalidade do espaço de decisão. Foram exploradas duas abordagens: a Análise de Componentes Principais (PCA) e um Algoritmo Genético (GA), projetados para otimizar a formação de *clusters* com base em restrições operacionais.

a) *PCA*: A primeira abordagem consistiu na aplicação da técnica de *PCA*, com foco na identificação de padrões estruturais entre os berços a partir de variáveis operacionais. O processo metodológico incluiu a seleção de variáveis representativas da operação portuária, como o tipo de mercadoria movimentada (granéis líquidos e granéis sólidos). Todas as variáveis foram previamente normalizadas, garantindo que estivessem na mesma escala e evitando que aquelas com maior magnitude dominassem a análise. Para a etapa de clusterização, utilizou-se o algoritmo *KMeans*, que agrupa os dados com base na proximidade em relação aos centróides dos grupos formados.

Apesar de evidenciar certa distinção entre os tipos de carga, os agrupamentos gerados apresentaram sobreposição significativa — especialmente entre berços multipropósito e especializados. Essa limitação, ilustrada na Fig. 2, comprometeu a utilidade da segmentação como base para o modelo de RL.

Figura 2: Distribuição dos berços a partir da clusterização via PCA + KMeans.

Diante dessas limitações, optou-se por uma abordagem mais flexível e orientada a restrições operacionais, baseada em algoritmo genético, com o objetivo de formar *clusters* mais representativos do comportamento real do terminal.

A implementação baseou-se em uma população inicial de 100 indivíduos. O GA utiliza uma representação vetorial em que cada gene representa um berço, e seu valor corresponde ao *cluster* atribuído. A seleção é feita por torneio estocástico, com *crossover* uniforme (90%) e mutação aleatória (10%). A evolução ocorre por até 100 gerações, com parada antecipada caso não haja melhoria no *fitness* por 20 gerações consecutivas.

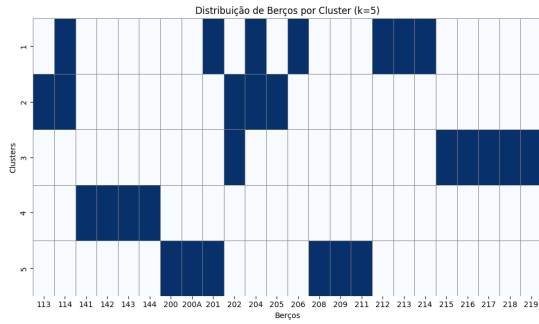


Figura 3: Mapa de calor da distribuição de berços para $k = 5$ clusters, obtida por algoritmo genético
Fonte: Autores (2025)

À medida que k aumenta, observa-se uma perda gradual de coesão entre os grupos, evidenciada pelo aumento da redundância e pela queda na representatividade.

Tabela III: Análise comparativa da clusterização por AG

k	Diversidade	Sobreposição	Faixa	Tendência
5	5,60	33,33%	4-7	↓ Diversidade
6	4,83	41,67%	3-7	↑ Sobreposição
7	4,29	50,00%	1-7	↑ Sobreposição
8	4,00	66,67%	1-7	↑ Sobreposição

Observa-se uma redução na diversidade e um aumento na sobreposição à medida que k cresce. O valor $k = 5$ representa o melhor compromisso entre eficiência e segmentação.

Sendo assim, o GA demonstrou ser uma ferramenta robusta para a clusterização de berços, superando as limitações do PCA ao incorporar explicitamente as restrições operacionais no processo de otimização. Ainda assim, algumas limitações e decisões de projeto merecem ser destacadas.

A escolha de $k \in [5, 8]$ considerou limitações computacionais e operacionais. Valores menores, como $k = 3$ ou $k = 4$, geraram aglomeração excessiva de berços em poucos grupos, ampliando drasticamente o número de estados para o *Q-Learning* e comprometendo a viabilidade do processo de treinamento. Por outro lado, valores mais altos fragmentaram os recursos, reduzindo a eficácia da segmentação. Mais do que uma etapa analítica, a clusterização via GA serve como base estrutural para o RL, permitindo um processo de treinamento mais rápido e, consequentemente, maiores ganhos de aprendizagem.

D. Experimento II - Validação do Algoritmo RL

Este experimento, composto por três testes, teve como objetivo avaliar a eficácia do aprendizado por reforço (RL) frente ao problema proposto, comparando seu desempenho com as políticas tradicionais FIFO e PHT. Foram testadas diferentes combinações de parâmetros ($\alpha \in \{0,1; 0,5; 0,9\}$ e $\gamma \in \{0,1; 0,5; 0,9\}$), bem como distintas estratégias de agrupamento de berços (entre 5 e 8 clusters), em simulações com horizontes de 24, 48, 72 e 168 horas. Cada configuração

foi executada 30 vezes para reduzir os efeitos de variações estocásticas, sendo calculadas as médias dos tempos acumulados de serviço e de espera. As simulações consideraram cenários com 40 a 90 navios (em incrementos de 10), além de casos adicionais com 150, 175 e 200 navios.

O primeiro teste teve como foco identificar a melhor combinação de α e γ para o horizonte de 48 horas, com base na análise de um gráfico de barras que evidenciou os ganhos proporcionados pela abordagem com RL. O segundo teste buscou ranquear os dez melhores resultados entre todas as combinações de parâmetros, abrangendo todos os horizontes de simulação de 24, 48 e 72 horas. Por fim, o terceiro teste concentrou-se no cenário com horizonte mais longo (168 horas), analisando o desempenho do modelo com cargas maiores, compostas por 150, 175 e 200 navios.

1) *Teste 1:* Foi analisado o tempo médio de espera e de serviço acumulado para os cenários com 40, 50, 60, 70, 80 e 90 navios, considerando o horizonte de simulação de 48 horas. Utilizou-se a combinação de parâmetros $\alpha = 0,9$ e $\gamma = 0,9$, por ter sido aquela que apresentou o maior número de melhores resultados finais.

A Figura 4 apresenta o tempo total acumulados em horas, considerando tanto os tempos de serviço realizados quanto as horas de atraso associadas a navios que permaneceram na fila aguardando atendimento.

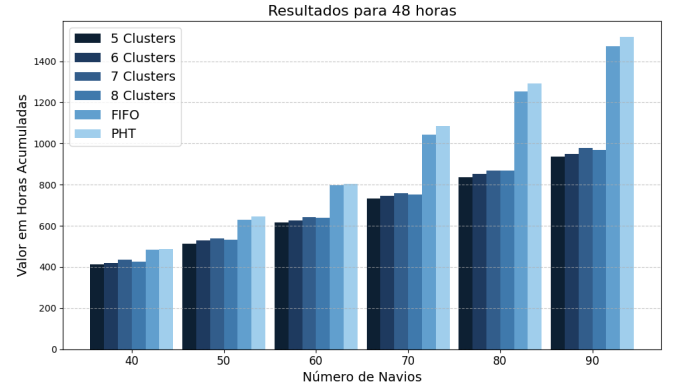


Figura 4: Resultados da comparação entre a abordagem RL e as políticas operacionais tradicionais - $\alpha = 0.9 / \gamma = 0.9$
Fonte: Autores (2025)

Observa-se que, à medida que o número de navios aumenta, as estratégias de operação FIFO e PHT acumulam maiores atrasos, evidenciando a superioridade do processo de otimização proporcionado pela abordagem RL. Nota-se também que, com o aumento do número de clusters, essa vantagem tende a diminuir ligeiramente, embora, em alguns casos, o cluster 8 tenha apresentado resultados melhores que o cluster 7. Contudo, a abordagem RL ainda mantém desempenho superior em relação às estratégias FIFO e PHT.

O desempenho superior da estratégia FIFO em relação à PHT pode ser explicado pela forma como a PHT prioriza a soma do tempo de espera com o tempo de serviço e pela aleatoriedade produzida pela geração de instâncias. Navios

com tempo de serviço menores, mesmo que estejam esperando por um longo período, não são alocados rapidamente, pois navios com tempo de serviço maior acabam passando na frente. Essa dinâmica provoca um acúmulo significativo de horas de espera.

Contudo, para o caso com 40 navios, observa-se que a abordagem por RL apresenta um ganho relativamente pequeno em comparação às demais estratégias. Isso se deve ao fato de o horizonte de planejamento ser maior que a quantidade de navios à serem alocados. Nessas condições, as instâncias geradas tendem a apresentar chegadas mais espaçadas, o que reduz a simultaneidade de atracações e permite que as estratégias FIFO e PHT sejam capazes de alocar os navios com maior eficiência.

2) *Teste 2*: Complementarmente à análise gráfica, a Tabela IV apresenta os dez cenários de melhor desempenho, com variações nos valores de α e γ , para horizontes de 24, 48 e 72 horas. A tabela mostra: o horizonte de simulação (Horas), o número de navios (Navios), o número de *clusters* (Clusters), os parâmetros α e γ , a média do tempo acumulado (Tempo Médio) e a redução percentual em relação à estratégia FIFO (Redução), considerada a mais eficaz entre as estratégias abordadas, conforme justificado anteriormente.

Tabela IV: Melhores 10 resultados observados

Horas	Navios	Clusters	α	γ	Tempo Médio	Redução(%)
48	90	Cluster 5	0.9	0.9	940.87	37.87
24	90	Cluster 5	0.5	0.9	954.27	37.08
24	90	Cluster 5	0.9	0.9	960.17	37.04
48	90	Cluster 6	0.9	0.9	957.47	36.77
24	90	Cluster 5	0.1	0.9	977.13	36.59
24	90	Cluster 6	0.9	0.9	969.1	36.45
48	90	Cluster 5	0.1	0.9	957.1	36.11
48	90	Cluster 8	0.9	0.9	968.97	36.01
48	90	Cluster 5	0.5	0.9	941.9	36.00
48	90	Cluster 5	0.9	0.5	956.7	35.97

Os resultados obtidos reforçam o bom desempenho da abordagem RL em diferentes horizontes de simulação e cenários, superando a estratégia FIFO. A proximidade entre a maior redução observada (37,87%) e a menor entre os melhores casos (35,97%) indica que o modelo é robusto mesmo em situações com alto nível de congestionamento, dado dentre os melhores resultados apenas figuraram as instâncias com 90 navios.

Em relação aos parâmetros α e γ , observa-se que a taxa de desconto permaneceu constante na maioria dos experimentos ($\gamma = 0,9$), o que sugere que o processo de aprendizagem prioriza os efeitos das decisões futuras. Esse comportamento está alinhado com a dinâmica do problema, no qual a chegada contínua de navios impacta diretamente o desempenho da alocação ao longo do tempo.

Portanto, os resultados mostram que a abordagem proposta se adaptou a diferentes configurações e mantém desempenho estável mesmo com variações nos parâmetros e formas de discretização.

3) *Teste 3*: Conforme observado, o aumento no número de navios amplia os ganhos obtidos com a abordagem RL.

Por isso, optou-se por simular um horizonte mais longo, com maior volume de navios, para demonstrar que os benefícios são acumulativos. Busca-se também ilustrar como seria o planejamento real de um operador no sistema portuário da APPA, explorando cenários mais representativos.

A Tabela V apresenta os resultados das horas acumuladas para todos os *clusters*, com os valores de $\alpha = 0,9$ e $\gamma = 0,9$ fixados, destacando a redução no tempo médio em relação à estratégia de operação FIFO.

Tabela V: Resultados para Horizontes de Simulação Maiores

Horas	Navios	Clusters	Tempo Médio	Redução (%)
168	150	cluster 5	1541.63	26.27
168	175	cluster 5	1815.90	34.79
168	200	cluster 5	2062.67	40.97
168	150	cluster 6	1565.37	25.14
168	175	cluster 6	1855.50	33.36
168	200	cluster 6	2118.17	39.39
168	150	cluster 7	1594.40	23.75
168	175	cluster 7	1877.27	32.58
168	200	cluster 7	2142.67	38.68
168	150	cluster 8	1608.53	23.07
168	175	cluster 8	1908.70	31.45
168	200	cluster 8	2162.20	38.13

Os melhores resultados foram encontrados pelo *cluster 5*, como já era esperado. Contudo, fica evidenciado o ganho acumulativo da abordagem RL, uma vez que os demais *clusters* também obtiveram excelentes resultados finais, mesmo com a sobreposição demonstrada no Experimento I.

Outro ponto a se destacar é a robustez da abordagem RL. Mesmo com um número elevado de execuções, os resultados obtidos foram consistentes em relação aos testes anteriores, o que mostra que os valores de α e γ estão bem ajustados, reduzindo a necessidade de novos treinamentos. Além disso, a variabilidade das instâncias não influenciou os resultados finais, mesmo nos cenários com grande quantidade de navios, o que reforça a capacidade de generalização da abordagem.

De modo geral, os resultados reforçam que a abordagem proposta é eficaz para o problema abordado, conseguindo lidar com a complexidade do ambiente portuário.

V. CONCLUSÃO

Este trabalho apresentou uma abordagem de Aprendizado por Reforço (RL), utilizando a estratégia de *Q-learning*, aplicada ao problema de alocação de berços em cenários reais da Administração dos Portos de Paranaguá e Antonina (APPA). O problema considera os berços em operação no sistema portuário da APPA, com dados de navios extraídos de registros históricos, e é formulado como um processo de decisão sequencial.

O estado do sistema é representado pela configuração do cais. Considerando o grande número de berços, realizou-se análise de complexidade que revelou o crescimento exponencial do espaço de estados. Diante disso, propôs-se o uso de algoritmos auxiliares — Análise de Componentes Principais e Algoritmo Genético — para agrupar os berços segundo a similaridade operacional dos tipos de carga.

Os estudos de casos demonstraram a viabilidade e os benefícios da aplicação da abordagem RL. No Experimento I, ficou evidente que a clusterização baseada em Algoritmo Genético superou o método de Análise de Componentes Principais, proporcionando agrupamentos de berços mais coerentes do ponto de vista operacional. Esses agrupamentos mostraram-se adequados e foram validados no experimento seguinte.

No Experimento II, a abordagem RL apresentou resultados consistentes, superando as estratégias FIFO (*First-In, First-Out*) e PHT (*Priority of Handling Time*), especialmente em cenários de alto congestionamento. A maior redução observada, de 37,87% no tempo total de serviço e espera dos navios, evidencia o potencial da otimização para melhorar a eficiência no uso dos recursos do sistema portuário da APPA. Além disso, em cenários com grandes horizontes de simulação, o ganho foi acumulativo, reforçando sua utilidade para o planejamento operacional a longo prazo.

Esses resultados indicam que a abordagem proposta, combinada com a técnica de agrupamento desenvolvida, oferece soluções eficazes para problemas logísticos complexos em ambientes reais, como o sistema portuário da APPA.

REFERÊNCIAS

- [1] U. N. C. on Trade and D. (UNCTAD), "Review of maritime transport 2024 — un trade and development (unctad)," 10 2024, [Online; accessed 2025-05-13]. [Online]. Available: <https://unctad.org/publication/review-maritime-transport-2024>
- [2] Administração dos Portos de Paranaguá e Antonina (APPA), "Relatório anual de movimentação de cargas," Disponível em: <http://www.portosdoparana.pr.gov.br>, 2024.
- [3] M. Acciaro, H. Ghiara, and M. I. Cusano, "Energy management in seaports: A new role for port authorities," *Energy Policy*, vol. 71, pp. 4–12, 2014.
- [4] S. Karakas, M. Kirmizi, and B. Kocaoglu, "Yard block assignment, internal truck operations, and berth allocation in container terminals: Introducing carbon-footprint minimisation objectives," *Maritime Economics & Logistics*, pp. 1–22, 2021.
- [5] F. Rodrigues and A. Agra, "Berth allocation and quay crane assignment/scheduling problem under uncertainty: A survey," *European Journal of Operational Research*, vol. 303, no. 2, pp. 501–524, 2022.
- [6] CONAPORTOS, "Portos e aeroportos - indicadores portuários desempenho operacional," 03 2025, [Online; accessed 2025-03-11]. [Online]. Available: <https://www.gov.br/portos-e-aeroportos/pt-br/assuntos/transporte-aquaviario/porto-sem-papelarquivos-para-download>
- [7] Rosângela, "No paran , tempo de espera de navios com fertilizantes chegou a at  45 dias," 7 2022, [Online; accessed 2025-03-15]. [Online]. Available: <https://www.portosenavios.com.br/noticias/portos-e-logistica/no-parana-tempo-de-espera-de-navios-com-fertilizantes-chegou-a-ate-45-dias>
- [8] aen.pr.gov.br, "Nos cinco primeiros meses, portos do paran  t m alta de 14% na descarga de fertilizantes — ag ncia estadual de not cias," 6 2022, [Online; accessed 2025-03-15]. [Online]. Available: <https://www.aen.pr.gov.br/Noticia/Nos-cinco-primeiros-meses-portos-do-Parana-tem-alta-de-14-na-descarga-de-fertilizantes>
- [9] J.-F. Cordeau, G. Laporte, P. Legato, and L. Moccia, "Models and tabu search heuristics for the berth-allocation problem," *Transportation science*, vol. 39, no. 4, pp. 526–538, 2005.
- [10] K. B hrkal, S. Zuglian, S. Ropke, J. Larsen, and R. Lusby, "Models for the discrete berth allocation problem: a computational comparison," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 4, pp. 461–473, 2011.
- [11] P. Hansen and C. O uz, "A note on formulations of static and dynamic berth allocation problems," *Les Cahiers du GERAD*, vol. 30, pp. 1–17, 01 2003.
- [12] A. Lim, "The berth planning problem," *Operations Research Letters*, vol. 22, no. 2, pp. 105 – 110, 1998.
- [13] A. Imai, X. Sun, E. Nishimura, and S. Papadimitriou, "Berth allocation in a container port: using a continuous location space approach," *Transportation Research Part B: Methodological*, vol. 39, no. 3, pp. 199–221, 2005.
- [14] R. Moorthy and C. Teo, "Berth management in container terminal: The template design problem," *OR Spectrum*, vol. 28, pp. 495–518, 01 2006.
- [15] A. Imai, E. Nishimura, and S. Papadimitriou, "The dynamic berth allocation problem for a container port," *Transportation Research Part B: Methodological*, vol. 35, no. 4, pp. 401–417, 2001.
- [16] —, "Berth allocation with service priority," *Transportation Research Part B: Methodological*, vol. 37, no. 5, pp. 437–457, 2003.
- [17] C. Sahin and Y. Kuvvetli, "Differential evolution based meta-heuristic algorithm for dynamic continuous berth allocation problem," *Applied Mathematical Modelling*, vol. 40, no. 23, pp. 10 679 – 10 688, 2016.
- [18] F. Barbosa, P. C. B. Rampazzo, A. Yamakami, and A. S. Camanho, "The use of frontier techniques to identify efficient solutions for the berth allocation problem solved with a hybrid evolutionary algorithm," *Computers & Operations Research*, vol. 107, pp. 43 – 60, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0305054819300206>
- [19] J.-F. Cordeau, G. Laporte, P. Legato, and L. Moccia, "Models and tabu search heuristics for the berth-allocation problem," *Transportation science*, vol. 39, no. 4, pp. 526–538, 2005.
- [20] R. S. Banos, R. A. Rosa, G. R. Mauri, and G. M. Ribeiro, "Modelo matem tico e meta-heur stica simulated annealing para o problema de aloca  o de ber os com m ltiplas cargas," *Transportes*, vol. 24, no. 1, pp. 51–62, 2016.
- [21] M. Heli, P. Rodrigues, C. M. Dias Pereira, M. L. Santos Machado, and P. D. Lima, "Um m todo heur stico para a programa  o de aloca  o dos navios em ber os," *uem.br*, vol. Volume Number, no. Issue Number, p. Page Range, September 2012.
- [22] E. T. Bacalh u, L. Casacio, and A. T. de Azevedo, "New hybrid genetic algorithms to solve dynamic berth allocation problem," *Expert Systems with Applications*, vol. 167, p. 114198, 2021.
- [23] S. N. Nazri, K. Mokhtar, A. A. Bakar, B. C. Mclellan, and S. M. M. Ruslan, "Optimization berth allocation in container terminals: A pyomo and google colab approach," *Ocean & Coastal Management*, vol. 258, p. 107359, 2024.
- [24] C. Cervellera, C. Cervellera, M. Gaggero, M. Gaggero, D. Macci , and D. Macci , "Policy optimization for berth allocation problems," *IEEE International Joint Conference on Neural Network*, 2021.
- [25] Y. Lv, M. Zou, J. Li, and J. Liu, "Dynamic berth allocation under uncertainties based on deep reinforcement learning towards resilient ports," *Ocean & Coastal Management*, vol. 252, p. 107113, 2024.
- [26] Y. Dai, Z. Li, and B. Wang, "Optimizing berth allocation in maritime transportation with quay crane setup times using reinforcement learning," *Journal of Marine Science and Engineering*, vol. 11, no. 5, p. 1025, 2023.
- [27] Q. Zhou, X. Cao, and P. Wang, "Deep reinforcement learning for dynamic berth allocation with random ship arrivals," in *2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS)*. IEEE, 2024, pp. 799–805.
- [28] Q. Zeng, Z. Yang, and X. Hu, "A method integrating simulation and reinforcement learning for operation scheduling in container terminals," *Transport*, vol. 26, no. 4, pp. 383–393, 2011.
- [29] S. Shimizu, K. Nishihara, Y. Miyauchi, K. Wakita, R. Suyama, A. Maki, and S. Shirakawa, "Automatic berthing using supervised learning and reinforcement learning," *Ocean Engineering*, vol. 265, p. 112553, 2022.
- [30] APPA, "Administra  o dos portos de paranagu  e antonina," 01 2025, [Online; accessed 2025-01-10]. [Online]. Available: <https://www.portosdoparana.pr.gov.br/>
- [31] MarineTraffic, "Marinetraffic: Global ship tracking intelligence — ais marine traffic," 01 2025, [Online; accessed 2025-01-10]. [Online]. Available: <https://www.marinetraffic.com/en/ais/home/centerx:-12.0/centery:25.0/zoom:4>
- [32] NOAA, "Ais data for 2020," 01 2025, [Online; accessed 2025-01-10]. [Online]. Available: <https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html>
- [33] R. E. Larson, *State Increment Dynamic Programming*. American Elsevier Publishing Company, 1968.
- [34] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.