UNIVERSIDADE FEDERAL DO MARANHÃO

UNIVERSIDADE FEDERAL DO PIAUÍ

Doutorado em Ciência da Computação Associação UFMA/UFPI

Victor Hugo Barros Silva

Aprendizado por reforço aplicado ao problema de alocação de berços graneleiros com controle de estoque

Orientador: Prof. Dr. Alexandre César Muniz de Oliveira

São Luís - MA
Fevereiro, 2026

Victor Hugo Barros Silva

Aprendizado por reforço aplicado ao problema de alocação de berços graneleiros com controle de estoque

TESE DE DOUTORADO

Tese apresentada como requisito parcial para obtenção do título de Doutor em Ciência da Computação, ao Doutorado em Ciência da Computação, Associação UFMA/UFPI.

Orientador: Prof. Dr. Alexandre César Muniz de Oliveira

São Luís - MA
Fevereiro, 2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Silva, Victor Hugo Barros.
Aprendizado por reforço aplicado ao problema de alocação de berços graneleiros com controle de estoque / Victor Hugo Barros Silva. - 2026.
101 f.

Orientador(a): Alexandre César Muniz de Oliveira.
Tese (Doutorado) - Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI, Universidade Federal do Maranhão, São Luís, 2026.

1. Pab. 2. Controle de Estoque. 3. Aprendizado Por Reforço. 4. Incertezas. I. de Oliveira, Alexandre César Muniz. II. Título.

Documento assinado digitalmente
gov.br
VICTOR HUGO BARROS SILVA
Data: 11/04/2026 08:22:52-0300
Verifique em https://validar.iti.gov.br

Victor Hugo Barros Silva

Aprendizado por reforço aplicado ao problema de alocação de berços graneleiros com controle de estoque

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Prof. Dr. Alexandre César Muniz de Oliveira
Orientador
Universidade Federal do Maranhão

Prof. Dra. Anne Magaly de Paula Canuto
Examinadora Externa
Universidade Federal do Rio Grande do Norte

Prof. Dr. Alexandre Plastino de Carvalho
Examinador Externo
Universidade Federal Fluminense

Prof. Dr. Omar Andrés Carmona Cortés
Examinador Externo
Instituto Federal de Ciência, Tecnologia e Educação do Maranhão

Prof. Dr. Luciano Reis Coutinho
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. Areolino de Almeida Neto
Examinador Interno
Universidade Federal do Maranhão

Certificamos que esta é a versão original e final da Tese de Doutorado que foi julgada adequada para obtenção do título de Doutora em Ciência da Computação.

Documento assinado digitalmente
gov.br
ALEXANDRE CESAR MUNIZ DE OLIVEIRA
Data: 11/04/2026 14:54:32-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Alexandre César Muniz de Oliveira
Orientador

ANSELMO CARDOSO DE PAIVA:37552384387
Assinado de forma digital por ANSELMO CARDOSO DE PAIVA:37552384387
Dados: 2026.04.16 09:46:33 -03'00'

Prof. Dr. Anselmo Cardoso de Paiva
Coordenador

São Luís - MA, 26 de Fevereiro de 2026

Aos meus filhos e esposa, que sempre estiverem comigo e nunca me permitiram andar só

Agradecimentos

Ao meu orientador, Professor Dr. Alexandre Oliveira, agradeço pela amizade, pelo incentivo e pelo apoio inestimáveis à minha pesquisa. Sua orientação foi fundamental para meu desenvolvimento como pesquisador.

Ao IFMA, pelo suporte concedido, sem o qual este trabalho não teria sido realizado.

Aos meus pais, pelo apoio em todas as horas.

Meus sinceros agradecimentos aos meus colegas do Laboratório de Aprendizagem Computacional, Métodos de Otimização e Robótica (LACMOR) por todos os auxílios e pelo companheirismo durante minha jornada até aqui.

Aos meus colegas da UFMA e da UFPI.

Aos professores membros da banca examinadora.

A todos os professores e funcionários do DCCMAPI pelos ensinamentos e pela colaboração.

A Selmira, Victor e Sofia.

“A vida é um tempo.”

(Mário Quintana)

Resumo

O transporte marítimo internacional consolidou-se como o pilar da economia global, sendo responsável por movimentar cerca de 80% do volume e mais de 70% do valor total do comércio mundial de bens. Diante dessa magnitude, a otimização da eficiência das operações portuárias tornou-se estratégica para a resiliência das cadeias de suprimentos. Contudo, o cenário recente (2020-2025) foi marcado por uma sucessão de eventos disruptivos, incluindo crises geopolíticas em rotas críticas e instabilidades climáticas, que impuseram desafios sem precedentes à logística portuária. Neste contexto, o Problema de Alocação de Berços (PAB), embora tradicional na literatura, demanda novas abordagens que ofereçam maior robustez diante de incertezas. As metodologias clássicas de otimização frequentemente apresentam limitações de escalabilidade e adotam premissas simplificadoras que negligenciam a alta dimensionalidade e a volatilidade operacional de terminais reais. Esta tese propõe uma formulação do PAB para granéis sólidos, integrada ao controle de estoque, modelada sob o paradigma de Aprendizado por Reforço Profundo (*Deep Reinforcement Learning* - DRL). A investigação utiliza inicialmente o algoritmo DQN (*Deep Q-Network*) e, subsequentemente, evolui para uma arquitetura que incorpora células LSTM (*Long Short-Term Memory*), visando capturar as dependências temporais intrínsecas a problemas de tomada de decisão sequencial complexos. Os resultados experimentais demonstram que a abordagem proposta não apenas gera soluções de alta qualidade em cenários dinâmicos, mas também é eficaz na mitigação de falhas de estoque, superando as limitações das heurísticas tradicionais.

Palavras-chave: PAB, Controle de Estoques, Incertezas, Aprendizado por Reforço.

Abstract

International maritime transport has established itself as the cornerstone of the global economy, accounting for approximately 80% of the volume and more than 70% of the total value of world trade in goods. Given this magnitude, optimizing port operations efficiency has become strategic for supply chain resilience. However, the recent period (2020–2025) was marked by a succession of disruptive events, including geopolitical crises along critical routes and climatic instability, which posed unprecedented challenges to port logistics. In this context, the Berth Allocation Problem (BAP), although well-established in the literature, demands new approaches that offer greater robustness to uncertainty. Classical optimization methodologies often exhibit scalability limitations and make simplifying assumptions that neglect the high dimensionality and operational volatility of real terminals. This thesis proposes a formulation of the BAP for dry bulk cargo, integrated with inventory control, and modeled under the Deep Reinforcement Learning (DRL) paradigm. The investigation initially employs the DQN (Deep Q-Network) algorithm. Subsequently, it evolves toward an architecture incorporating LSTM (Long Short-Term Memory) cells to capture the temporal dependencies inherent to complex sequential decision-making problems. The experimental results demonstrate that the proposed approach not only generates high-quality solutions in dynamic scenarios but also effectively mitigates inventory failures, surpassing the limitations of traditional heuristics.

Keywords: BAP, Inventory Control, Uncertainties, Reinforcement Learning.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

DQN	*Deep Q-Network*
ETA	*Estimated Time of Arrival (Tempo de Chegada Esperado)*
PAB	*Problema de Alocação de Berços*
RL	*Reinforcement Learning (Aprendizado por Reforço)*
TTW	*Tidal Time Window (Janela de Tempo de Maré)*

Sumário

1 Introdução

1.1 Contextualização

O transporte marítimo desempenha um papel fundamental no comércio internacional. Cerca de 80% do volume e 70% do valor deste comércio são realizados por via marítima (United Nations Conference on Trade and Development (UNCTAD), 2023) (BINGHAM; MIKKELSEN, 2023). Este cenário ilustra bem a importância do modal marítimo no comércio de mercadorias entre países e revela, consequentemente, o motivo dos grandes esforços empregados ao longo dos últimos anos na pesquisa acadêmica voltada à eficiência nas operações portuárias.

O Brasil, um dos maiores exportadores de carga a granel, atingiu 14,6% em 2020 (WTO Secretariat, 2021). Neste ano, as exportações brasileiras atingiram a posição de segundo maior exportador de soja e de minério de ferro do mundo em 2020, obtendo, em ambos os tipos de carga, 23% do comércio total, ficando atrás apenas dos Estados Unidos da América e da Austrália, respectivamente (United Nations Conference on Trade and Development, 2021).

A movimentação total portuária de granel sólido em portos brasileiros em 2021 foi de 706,6 milhões de toneladas; de granel líquido, de 314,7 milhões de toneladas; de carga conteinerizada, de 132,9 milhões de toneladas; e de carga geral, de 60,0 milhões de toneladas, totalizando 1.214,3 milhões de toneladas, o que representa um aumento de 5,09% em relação ao ano anterior (ANTAQ, 2022). Em 2023, a movimentação total portuária de granéis sólidos foi de 790,3 milhões de toneladas, de granel líquido e gasoso foi de 325,3 milhões de toneladas, enquanto as movimentações de carga conteinerizada e de carga geral somaram 187,9 milhões de toneladas, totalizando 1.303,6 milhões de toneladas, que representam 5,9% de aumento quando comparadas ao ano de 2022. Nesta conjuntura, a logística portuária desempenha um papel fundamental para a competitividade das exportações brasileiras no comércio mundial.

O Estado do Maranhão, em particular, tem grande importância no cenário portuário por acomodar um dos principais complexos portuários do Brasil, formado por três portos: (i) Terminal Marítimo de Ponta da Madeira, (ii) Porto do Itaqui e (iii) Terminal Portuário Privativo da Alumar. O primeiro é administrado por uma mineradora multinacional e tem o minério de ferro como principal carga movimentada.

O Terminal de Ponta da Madeira, administrado por uma mineradora multinacional, que tem o minério de ferro como tipo de carga mais movimentado, é considerado o maior do Brasil em volume de cargas, tendo movimentado 167,9 milhões de toneladas em 2022.

O terminal de Santos, considerado o segundo maior do Brasil em volume de carga, movimentou 126,2 milhões de toneladas em 2022 (ANTAQ, 2023). À frente do terminal de Santos, considerado o segundo maior do Brasil em volume de carga, movimentou 126,2 milhões de toneladas em 2023 (ANTAQ, 2024).

O porto do Itaqui movimenta tanto cargas em granel sólido como em granel líquido e cargas em geral. O maior destaque vai para a movimentação de soja e milho (granéis sólidos), que, somados, totalizaram 17,8 milhões de toneladas (EMAP, 2023). Em 2022, o Porto do Itaqui foi o 9º maior porto no Brasil em movimentação de cargas, ao atingir a marca total de 33,5 milhões de toneladas (ANTAQ, 2023).

Por fim, o Terminal Portuário Privativo da Alumar está associado a um complexo industrial que produz alumina e alumínio, este último suspenso temporariamente. O terminal portuário é responsável tanto pela importação das principais matérias-primas usadas na produção, como soda cáustica, carvão e bauxita, quanto pelo escoamento da alumina e do alumínio produzidos (Alcoa Brasil, 2023). Em 2022, o Terminal Portuário Privativo da Alumar atingiu o 19º lugar em movimentação de cargas ao alcançar 15,0 milhões de toneladas (ANTAQ, 2023).

Diante do papel importante que o transporte aquaviário e a logística portuária desempenham no comércio, não apenas regional, mas também mundial, diversos trabalhos já foram apresentados com o objetivo de tornar as operações portuárias mais eficientes. Neste contexto, o uso de técnicas de otimização surge naturalmente como uma ferramenta para enfrentar esses desafios. No campo da Pesquisa Operacional (PO), por exemplo, utilizaram-se técnicas tradicionais de otimização, como a programação linear e as metaheurísticas, para a obtenção de soluções ótimas ou aproximadas.

Diversos problemas têm sido tratados no âmbito das operações portuárias, com destaque para o Problema de Alocação de Berços (PAB) e o Problema de Alocação de Guindastes (PAG) e, muitas vezes, para ambos combinados. A importância do PAB e do PAG decorre do efeito que produzem na produtividade do porto e na qualidade do serviço, relacionados ao sistema de movimentação do terminal Chang, Lin e Tsai (2024). Bierwirth e Meisel (2010) e Bierwirth e Meisel (2015) fizeram uma ampla investigação sobre ambos os problemas, na qual classificaram os tipos de problemas e os métodos empregados para solucioná-los. Mais detalhes sobre o PAB estão na Seção 2.1.1. Um exemplo de outro problema neste contexto é o chamado Problema de Planejamento de Estiva (*Stowage Planning Problem*), que consiste em maximizar a utilização dos navios e reduzir os custos de armazenamento da carga, sob várias e complexas restrições, como navegabilidade, regras de empilhamento e consumo de combustível (TWILLER et al., 2023) (STEENKEN; VOSS; STAHLBOCK, 2004).

Neste trabalho, o interesse está voltado para um caso específico do problema de alocação de berços, apresentado por Barros et al. (2011), em que os níveis de estoque

influenciam o processo de decisão de atracação dos navios. Em cada unidade de tempo (janelas de atracação discretizadas, determinadas pelas marés), os níveis de estoque no porto devem permanecer em patamares seguros. Este tipo de situação é mais comum em portos graneleiros, que eventualmente estão associados à produção ou ao consumo da carga operada. Em contraste, os terminais de contêineres apresentam o Problema de Empilhamento de Contêineres nos pátios. O objetivo básico deste tipo de problema é otimizar a ordem de empilhamento para reduzir o número de bloqueios entre contêineres, o que reduz a quantidade de reorganizações, obtendo maior eficiência e, consequentemente, economia de energia (JIN et al., 2023).

Vários trabalhos na literatura abordam a gestão de estoque associada ao PAB. Liu et al. (2016) levam em conta o uso de uma medida de desempenho de reposição e de gestão de estoque no processo de decisão em um terminal de contêineres. Aghalari, Nur e Marufuzzaman (2020) discutem como a perecibilidade de produtos nos estoques afeta a satisfação dos clientes, o que demanda maior eficiência do porto. Recentemente, Mehdi et al. (2023) modelaram uma cadeia de suprimentos a nível global que integra PAB, gestão de estoque e agendamento de produção. A gestão de estoque é realizada em pontos de estocagem ao longo da cadeia. Belov et al. (2020) tratam de um problema complexo que envolve o fornecimento de carvão de fontes e com diferentes características, para o qual deve ser realizada uma mistura a fim de garantir a qualidade mínima exigida pelo cliente. Estes e outros trabalhos demonstram que a gestão de estoque, entre outros elementos no âmbito das operações portuárias, tem demandado maiores esforços dos administradores portuários na alocação de navios.

Nos últimos anos, o comércio marítimo sofreu impactos significativos em decorrência da pandemia de COVID-19 e de crises geopolíticas. Em 2022, o volume do comércio marítimo sofreu uma queda de 0,4%, seguindo uma tendência de 2021. Congestionamentos devido à alta demanda sustentada e à guerra na Ucrânia prejudicaram a eficiência portuária. A partir do segundo semestre de 2022, a dinâmica de oferta e demanda normalizou-se devido à flexibilização dos bloqueios decorrentes da pandemia e à regularização dos estoques das empresas dos Estados Unidos da América e da Europa, em particular. Como consequência, a projeção para 2023 indica crescimento (WTO Secretariat, 2023). Todavia, toda essa conjuntura tem ampliado o sentimento de incerteza entre os diversos agentes do comércio marítimo. O resultado disso é o aumento dos esforços para tornar toda a cadeia de suprimentos mais resiliente a novas configurações.

No contexto do PAB, segundo Lv et al. (2024), a resiliência é a capacidade de um porto de suportar as operações durante eventos imprevistos, inclusive disruptivos, e de recuperar-se rapidamente à normalidade. As incertezas decorrentes de eventos imprevistos relacionam-se a atrasos de embarque, falhas de máquinas e congestionamentos nas filas de atendimento, entre outros. A maioria dos trabalhos considera as incertezas no

tempo de chegada ou de operação das cargas no porto, como em Xiang e Liu (2021). Na literatura, encontram-se exemplos de abordagens que lidam com a incerteza no ambiente portuário por meio do artifício de reotimizações (BELOV et al., 2020). Outros exemplos de abordagens incorporam explicitamente elementos estocásticos ao modelo (SCHEPLER et al., 2019). Recentemente, tem-se observado um crescente interesse no uso de aprendizagem de máquina aplicada ao PAB (KOLLEY et al., 2022) (FILOM; AMIRI; RAZAVI, 2022), especialmente na predição de parâmetros incertos. O uso de aprendizado por reforço, particularmente, aproveita que a modelagem por meio de Processo de Decisão de Markov permite melhor adaptabilidade às mudanças em tempo real sem necessidade de reotimizações (LI; YANG; YANG, 2023). Além disso, o aprendizado por reforço não depende de regras específicas, como em algoritmos tradicionais de otimização, pois é capaz de aprender critérios de decisão por meio da interação contínua com o ambiente, o que lhe confere maior habilidade para lidar com dinamismo, incertezas e restrições complexas (LV et al., 2024).

1.2 Problema de pesquisa

Neste trabalho, investiga-se a abordagem do PAB com controle de estoque como um problema de aprendizado por reforço. A questão principal que se pretende responder é se a formulação do PAB como um problema de aprendizado por reforço é capaz de aprender critérios de tomada de decisão que levem à obtenção de soluções que não violem a restrição de controle de estoque.

A seguinte hipótese é verificada: a abordagem que trata o problema de alocação de berços com restrições de nível de estoque aplicando a técnica de aprendizado por reforço profundo, conhecida como DQN, é capaz de gerar soluções de qualidade, ou seja, próximas às soluções ótimas sob incertezas e que atendam às restrições de controle de estoque, quando comparadas aos resultados dos algoritmos apresentados na literatura.

1.3 Objetivos

1.3.1 Objetivo geral

Avaliar a eficácia de um arcabouço de Aprendizado por Reforço Profundo, que incorpora arquiteturas com e sem memória recorrente, para abordar o Problema de Alocação de Berços com controle de estoque em terminais portuários graneleiros, considerando cenários com e sem incerteza nos tempos de chegada dos navios.

1.3.2 Objetivo específicos

1. Modelar o Problema de Alocação de Berços (PAB) com controle de estoque sob a perspectiva do Aprendizado por Reforço, definindo adequadamente os componentes de estado, ação e recompensa no contexto portuário.
2. Propor uma arquitetura de Aprendizado por Reforço Profundo aplicada ao PAB, com controle de estoque, baseada no algoritmo *Deep Q-Network* (DQN), combinada com redes neurais recorrentes do tipo *Long Short-Term Memory* (LSTM), para tratar decisões sequenciais no PAB.
3. Desenvolver, sistematizar e disponibilizar um ambiente completo de simulação do PAB, incluindo instâncias do problema para testes e validação.

1.4 Justificativas

Mesmo que vários trabalhos na literatura tratem do problema de alocação de berços, a principal abordagem consiste no uso de modelos matemáticos de programação linear inteira, que, além de terem limitações quanto à escala, assim como o uso de técnicas meta-heurísticas, normalmente são apresentados como aproximações simplificadoras devido à dificuldade no tratamento das incertezas. Além disso, os últimos acontecimentos globais evidenciam a urgência de aumentar a resiliência nas operações portuárias para evitar colapsos graves. Diante deste cenário, é importante investigar enfoques alternativos, neste caso, aproveitando os resultados promissores obtidos recentemente com a aplicação de técnicas de aprendizado por reforço em problemas distintos.

Esta proposta pode abrir espaço para trabalhos futuros que envolvam outros cenários portuários, com conjuntos de restrições distintos. Com o uso de aprendizado por reforço, a incorporação de incertezas na resolução de problemas de alocação de berços é uma tarefa que depende basicamente da modelagem dos dados, o que pode aproximar as decisões tomadas dos cenários reais. Desta forma, é possível vislumbrar novas contribuições para o setor portuário.

1.5 Principais contribuições

Diante da discussão anterior, as principais contribuições deste trabalho são resumidas a seguir.

- Formulação de um problema de alocação de berços como um problema de aprendizado por reforço a partir de um modelo matemático de programação linear inteira;

- Software simulador do cenário operacional de um porto graneleiro cujo processo decisório depende de fatores naturais como marés e controle de estoque sobre uma gama de granéis carregados ou descarregados nos/dos pátios de estocagem;
- Base de dados pública para fins acadêmicos, incluindo instâncias de problema adequadas ao uso de sistemas de apoio à decisão baseados em otimizadores;
- Compêndio sobre critérios de decisão adequados à tomada de decisão no contexto de terminais portuários de granéis, cenários operacionais típicos do complexo portuário instalados na ilha de São Luís.

1.6 Organização do trabalho

Este trabalho está estruturado da seguinte forma: o Capítulo 3 descreve trabalhos que tratam da aplicação de técnicas de aprendizado de máquina, particularmente de aprendizado por reforço, a trabalhos relacionados ao problema de alocação de berços. O capítulo 2 apresenta uma introdução aos fundamentos teóricos dos principais métodos abordados neste trabalho. O Capítulo 4 descreve a formulação do PAB como um problema de aprendizado por reforço proposto. Dois cenários da formulação são apresentados no Capítulo 5. No Capítulo 6, são apresentadas as etapas adotadas que compõem a metodologia proposta para este trabalho. Os resultados obtidos e as respectivas análises são apresentados no Capítulo 7. Por fim, as considerações finais e o encaminhamento dos trabalhos futuros são apresentados no Capítulo 8.

2 Fundamentação Teórica

2.1 Problema de Alocação de Berços

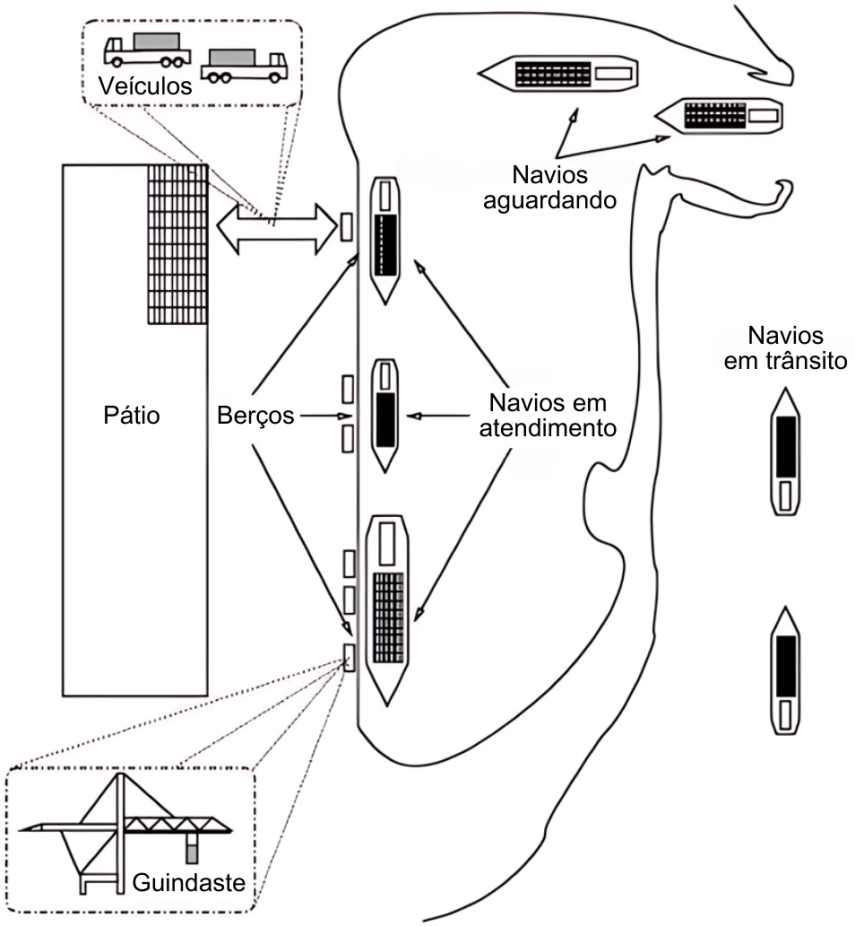
O Problema de Alocação de Berços (PAB) consiste em atribuir navios que chegam a posições de atracação (CORDEAU et al., 2005), denominadas berços, ao longo de um horizonte de planejamento com o objetivo de minimizar custos ou maximizar o desempenho em operações portuárias. Lim (1998) foi um dos primeiros trabalhos a descrever o problema de alocação de berços. Eles demonstram que o problema pode ser considerado NP-difícil, representam-no por meio de grafos e propõem uma heurística para resolvê-lo.

Ao longo das últimas três décadas, observou-se um crescente interesse pelo tema, em razão dos impactos econômicos já mencionados na Seção 1.1. Todo esse interesse resultou, naturalmente, na evolução da abordagem ao problema, tanto do ponto de vista da definição quanto da resolução do PAB. Os vários cenários portuários distintos, com estruturas e operações diferentes, geraram uma ampla variedade de problemas. Na Figura 1, apresenta-se um esquema com a estrutura geral do PAB, que permite visualizar alguns destes cenários.

A maneira como os berços operam altera completamente a abordagem a ser adotada no PAB. Estas posições de atracação podem ser discretas, contínuas ou híbridas. Um PAB com berços discretos significa que um berço pode atender exatamente a um navio por vez e pode ser visto como um problema de escalonamento de máquinas, em que navios são tarefas e berços são máquinas. Por outro lado, PAB com berços contínuos permite que os navios sejam atendidos em quaisquer posições disponíveis ao longo do cais de atracação e pode ser modelado como um problema de empacotamento bidimensional (RODRIGUES; AGRA, 2022), com o tempo e o espaço de atracação no cais como dimensões. PAB híbrido enquadra-se em portos que permitem que mais de um navio ocupe um determinado berço ao mesmo tempo ou que um único navio ocupe mais de um berço ao mesmo tempo. Na Figura 2 são ilustrados os *layouts* discreto, contínuo e híbrido mencionados.

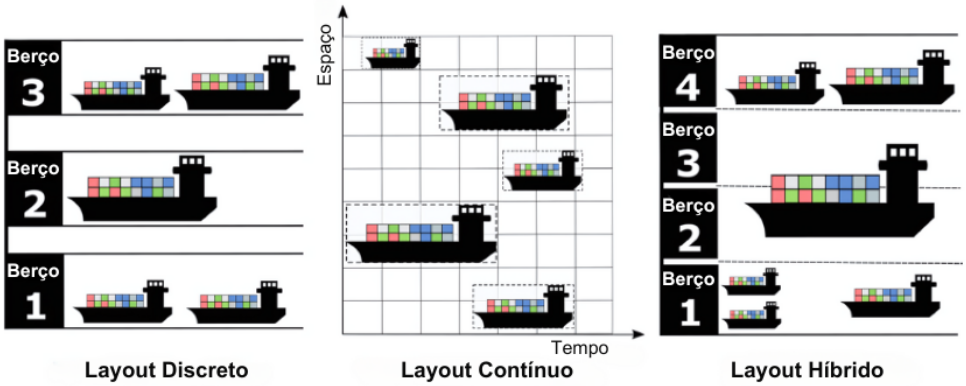
O problema normalmente é classificado como dinâmico ou estático, a partir da observação dos navios que são considerados no horizonte de planejamento. O PAB dinâmico surge quando o conjunto de navios considerados no horizonte de planejamento compreende também navios que ainda não chegaram ao porto (área de fundeio), ou seja, que ainda estão em trânsito. Imai, Nishimura e Papadimitriou (2001) fazem uma contribuição importante ao PAB dinâmico ao proporem uma solução heurística combinada

Figura 1 – Ilustração da estrutura geral do problema de alocação de berços.



Fonte: Adaptado de Cordeau et al. (2005).

Figura 2 – Layouts de berços: discreto, contínuo e híbrido.



Fonte: Adaptado de Rodrigues e Agra (2022).

com relaxação lagrangeana eficiente. O PAB estático, um caso específico do PAB dinâmico, considera apenas navios que já chegaram, como em descrito em (IMAI; NAGAIWA; TAT, 1997).

O tipo de carga operado pelo porto e, consequentemente, transportado pelos

navios influencia o problema. Basicamente, os portos podem ser graneleiros, ou seja, operam cargas em granel, terminais de contêineres ou ainda podem operar cargas em geral. Um PAB graneleiro, especificamente, impõe restrições quanto ao uso dos berços devido ao tipo de carga transportada pelo navio, enquanto terminais de contêineres comumente associam o problema de alocação de guindastes ao de alocação de berços, tendo em vista que o PAG é decisivo para os tempos de atendimento dos navios.

Estas características ajudam a descrever o PAB tratado, porém, outras peculiaridades de cada cenário portuário produzem uma grande variedade de problemas, assim como métodos de solução propostos. Na seção seguinte é descrito o PAB específico considerado neste trabalho.

2.1.1 Descrição do Problema de Alocação de Berços

Neste trabalho, utiliza-se um modelo específico, baseado em um porto localizado na região portuária do complexo de São Luís, que apresenta condições de maré restritivas para as operações de atracação e desatracação.

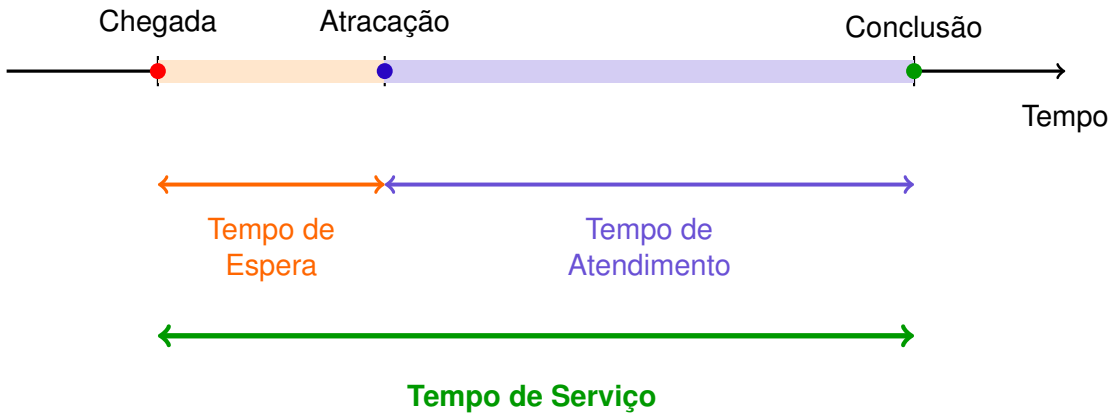
O porto considerado é graneleiro e possui berços heterogêneos, ou seja, berços capazes de operar navios e tipos de carga em tempos distintos. Cada berço é discreto e pode atender exatamente um navio por vez. Além disso, é dinâmico, ou seja, considera navios que ainda estão por chegar.

O porto considerado está associado a uma planta industrial, que depende dos recursos trazidos pelos navios para a fabricação de seus produtos. Desta forma, as decisões de atracação devem considerar os níveis de estoque atuais para evitar faltas de estoque. Os níveis de estoque são afetados não apenas pelos descarregamentos realizados pelos navios, mas também pelo consumo do estoque realizado regularmente na produção industrial.

O objetivo do problema consiste em definir o sequenciamento ótimo de navios para os berços, considerando as restrições mencionadas. O critério adotado para a minimização é o tempo de serviço total. O tempo de serviço é a soma do tempo de espera (período entre a chegada e a atracação) e do tempo de atendimento (período em que o navio está atracado sendo servido). A Figura 3 ilustra o cálculo do tempo de serviço. Para resolvê-lo, o modelo é formulado como um problema discreto e dinâmico, no caso *offline*.

Na Seção 2.1.2, é apresentada a modelagem matemática proposta em Silva (2021), utilizada como referência neste trabalho e como base para a implementação do ambiente de simulação. Esta modelagem matemática é uma evolução do modelo apresentado em Barros et al. (2011), que considera o mesmo cenário, porém, trata os berços como homogêneos, atendendo, portanto, todos os navios com o mesmo tempo de atendimento.

Figura 3 – Tempo de serviço de um navio. O tempo de serviço é a soma do tempo de espera e do de atendimento. O critério de decisão seleciona o navio que minimiza esse tempo total.



Fonte: Elaborada pelo autor.

2.1.2 Modelo matemático

O problema foi modelado como um problema de Programação Linear Inteira (PLI), em que os parâmetros de entrada do modelo são relacionados na Seção 2.1.2.1, as variáveis de decisão, na Seção 2.1.2.2, a função objetivo, na Seção 2.1.2.3 e, por fim, o conjunto de restrições, na Seção 2.1.2.4.

2.1.2.1 Parâmetros de entrada

- N: conjunto de navios;
- M: conjunto de marés;
- L: conjunto de berços;
- K: conjunto de cargas operadas;
- a_i: maré de chegada esperada do navio i (*Expected Time Arrival* - ETA);
- v_l: vazão de trabalho do berço l;
- e_k: estoque inicial da carga k;
- c_k: taxa de consumo da carga k;
- h_{il}: tempo de atendimento ou tratamento para o navio i no berço l;
- q_{ik}: quantidade de carga k transportada pelo navio i.

2.1.2.2 Variáveis de decisão

A solução para o problema é o sequenciamento de navios em cada berço, ou seja, o conjunto de decisões de atracação de navios nos berços disponíveis ao longo das janelas de atracação (*Tidal Time Window* - TTW).

Desta forma, y_{ijl} é a variável de decisão binária definida pelos conjuntos N, M e L, que indica se o navio $i \in N$ será atribuído à TTW $j \in M$ e ao berço $l \in L$ ($y_{ijl} = 1$) ou não ($y_{ijl} = 0$).

$$y_{ijl} = \begin{cases} 1, & \text{se o navio } i \text{ é alocado à TTW } j \text{ e ao berço } l \\ 0, & \text{caso contrário} \end{cases}$$

2.1.2.3 Função objetivo

Como mencionado anteriormente, o critério de decisão adotado nesta modelagem é o tempo de serviço total, ou seja, a soma dos tempos de serviço de cada navio. Dessa forma, o objetivo é minimizar essa soma, o que se representa pela função objetivo definida na Equação 2.1.

$$\min \sum_{i=1}^{|N|} \sum_{j=1}^{|M|} \sum_{l=1}^{|L|} (j + h_{il} - a_i) \times y_{ijl} \tag{2.1}$$

O tempo de serviço como função-objetivo busca maior eficiência no atendimento aos navios, mas também equilíbrio com o tempo de espera.

2.1.2.4 Restrições

O conjunto de restrições do problema é definido no modelo matemático por meio das equações 2.2, 2.3, 2.4 e 2.5.

A Equação 2.2 garante que os navios não podem atracar em TTW anterior à sua chegada esperada no porto. Para isso, as atracações nessas TTW são definidas como 0.

$$\sum_{j=1}^{a_i - 1} \sum_{l=1}^{|L|} y_{ijl} = 0, \quad \forall i \in N \tag{2.2}$$

A Equação 2.3 garante que os navios serão atracados em exatamente um berço e em uma TTW, a partir da TTW de chegada esperada.

$$\sum_{j=a_i}^{|M|} \sum_{l=1}^{|L|} y_{ijl} = 1, \quad \forall i \in N \tag{2.3}$$

A Equação 2.4 não permite que dois navios atraquem no mesmo berço e que TTW vão ocasionar, em certo momento, sobreposição de navios. O produto $|N|.|M|$ funciona como uma chave lógica: se o navio i foi atribuído ao berço l, nenhum outro navio n pode atracar no mesmo berço durante o descarregamento de i.

$$\sum_{n=1,n\neq i}^{|N|} \sum_{m=j,m\leq|M|}^{j+h_{il}-1} y_{nml} \leq (1-y_{ijl})|N||M|, \quad \forall i \in N, j \in M, l \in L \tag{2.4}$$

A Equação 2.5 garante que não haverá falhas de estoque. Esta restrição mantém o equilíbrio entre o consumo das cargas pelo porto (ou pela planta industrial associada ao porto) e o descarregamento dos navios somado ao estoque inicial. A cada TTW, esta relação deve estar acima do nível de estoque mínimo determinado, neste caso, 0.

$$\sum_{i=1}^{|N|} \sum_{l=1}^{|L|} \sum_{z=a_i}^{j} \frac{\min(j-a_i+1, h_{il})}{h_{il}} q_{ik} \times y_{izl} \leq j \times c_k + e_k, \quad \forall j \in M, k \in K \tag{2.5}$$

Desta forma, o navio não descarrega toda a carga na TTW de atracação. Uma fração linear da carga q_{ik}, determinada por h_{il}, é descarregada a cada TTW até o final da operação.

2.2 Aprendizado por Reforço

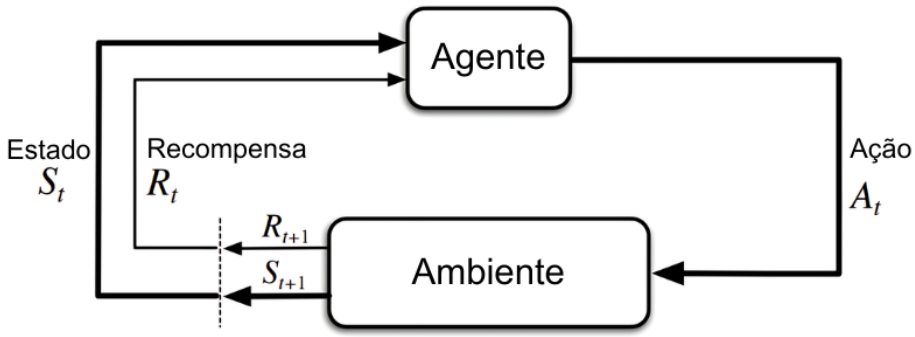
2.2.1 Conceitos fundamentais

Aprendizado por Reforço (*Reinforcement Learning* - RL) é um dos paradigmas do aprendizado de máquina, que consiste em aprender o que fazer, ou seja, mapear situações para ações, de modo a maximizar uma recompensa numérica. As três características mais importantes presentes em RL e que as difere de aprendizado supervisionado e não-supervisionado são: (i) ser essencialmente uma malha fechada, onde as ações do sistema de aprendizado têm influência nas entradas posteriores; (ii) não ter instruções sobre quais ações tomar, em vez disso, ele deve descobrir quais ações produzem as melhores recompensas; e (iii) as ações podem afetar não apenas a recompensa, mas também a próxima situação e, por conseguinte, as recompensas subsequentes (SUTTON; BARTO, 2018).

Outra característica importante do aprendizado por reforço é a interação com o ambiente durante o treinamento. O agente e o ambiente interagem em cada passo de uma sequência de passos discretos, $t = 0, 1, 2, 3, \ldots$, onde o agente recebe uma representação do estado do ambiente $S_t \in \mathcal{S}$, e seleciona uma ação $A_t \in \mathcal{A}(s)$. No próximo passo, $t+1$, o agente recebe uma recompensa numérica $R_{t+1} \in \mathcal{R} \subset \mathbb{R}$ e um

novo estado S_{t+1}. Por meio da Figura 4, é possível observar a dinâmica entre esses componentes.

Figura 4 – Representação da interação entre o Agente e o Ambiente.



Fonte: Adaptado de (SUTTON; BARTO, 2018).

O objetivo do aprendizado por reforço é, basicamente, definir a maneira como um agente se comporta em cada estado, ou seja, estabelecer um conjunto de mapeamentos de estados observados para novos estados quando determinada ação é tomada. Este mapeamento é conhecido como política e pode ser representado por π. O objetivo da política é conduzir o estado atual ao estado desejado, com a maior recompensa acumulada possível. A recompensa acumulada é obtida pela chamada *função estado-valor*, que pode ser entendida como o valor esperado que o agente acumulará no futuro a partir do estado atual ao seguir a política π, e é dada por meio da Equação 2.6, em que γ é um fator de desconto, $0 \leq \gamma \leq 1$, que representa o nível de importância que recompensas futuras têm para $v_\pi(s)$.

$$v_\pi(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t \middle| S_t = s\right] \tag{2.6}$$

A política ótima π^*, que fornece a maior recompensa acumulada para cada estado, é dada, portanto, por 2.7.

$$v^*(s) = \max_\pi v_\pi(s), \forall s \in \mathcal{S} \tag{2.7}$$

A função *valor-ação* para a política π, $q_\pi(s, a)$, mapeia um valor de recompensa esperado para cada ação a aplicada ao estado s ao seguir a política π.

$$q_\pi(s, a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t \middle| S_t = s, A_t = a\right] \tag{2.8}$$

Da mesma forma, a política ótima π^* determina a função ação-valor ótima, q^*.

$$q^*(s,a) = \max_{\pi} q_\pi(s,a), \forall s \in \mathcal{S}, \forall a \in \mathcal{A} \tag{2.9}$$

2.2.2 Aprendizado por Reforço Profundo

Apesar de um relativo sucesso no uso de algoritmos de aprendizado por reforço tradicionais, sua aplicação é limitada a domínios pequenos, que podem ser resolvidos com estratégias mais simples, ou a domínios com espaços de estado de baixa dimensão totalmente observados (MNIH et al., 2015). O aprendizado por reforço profundo busca explorar a capacidade das redes neurais profundas de extrair representações abstratas a partir de entradas de alta dimensão para contornar esses desafios. Combina-se uma rede neural profunda como agente do ambiente de aprendizado por reforço.

A ideia básica por trás de muitos algoritmos de aprendizado por reforço é aproximar a função ação-valor, usando a Equação de Bellman como uma atualização iterativa, $Q_{i+1}(s,a) = \mathbb{E}\Big[r + \gamma \max_{a'} Q_i(s',a') \Big| \; s,a\Big]$, de modo que Q_i convirja para Q^* ao passo que $i \to \infty$. Todavia, nesta estratégia não há generalização, tendo em vista que a função ação-valor é estimada para cada sequência. Como alternativa, utilizam-se aproximadores de função para estimar a função ação-valor, $Q(s,a;\theta) \approx Q^*(s,a)$, que são aproximadores de função linear ou não linear, como redes neurais (MNIH et al., 2013).

No entanto, aproximadores não-lineares são conhecidos por serem instáveis e até mesmo divergirem quando são usados para aproximar uma função ação-valor Q. Três motivos importantes para isso são: as correlações presentes em uma sequência de observações; pequenas atualizações em Q podem mudar significativamente a política e a correlação entre os valores-ação Q e os valores-alvo $r + \gamma \max_{a'} Q(s',a')$. O algoritmo conhecido como Deep Q-Network, ou Deep Q-Learning (DQN), faz uso de alguns mecanismos para contornar estes problemas. Primeiramente, utiliza a chamada experiência de *replay*, que trata das correlações entre as observações ao armazenar, em um *buffer*, experiências obtidas ao longo do treinamento e, eventualmente, utilizá-las aleatoriamente novamente como entrada na rede neural. Essa estratégia também suaviza as mudanças na política. Em segundo lugar, a atualização dos Q-valores é no sentido dos Q-valores-alvo, que são atualizados periodicamente, o que reduz a correlação entre os Q-valores atuais e os Q-valores-alvo. A DQN faz uso de uma rede neural não para aproximar a função ação-valor, mas como um aproximador de função de rede neural com pesos θ. Essa rede neural pode ser treinada minimizando uma sequência de funções de perda (*loss*) $L_i(\theta_i)$ que muda a cada iteração i (MNIH et al., 2013)(MNIH et al., 2015), conforme a Equação 2.10.

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s')\sim U(D)}\Big[\Big(r + \gamma \max_{a'} Q(s',a';\theta_i^-) - Q(s,a;\theta_i)\Big)^2\Big] \tag{2.10}$$

Onde γ é o fator de desconto, θ_i são os parâmetros da rede neural na iteração i e θ_i^- são os parâmetros usados para computar o alvo na iteração i. θ_i^- são mantidos fixos por C passos; após isso, são atualizados com os valores de θ_i.

3 Trabalhos Relacionados

Filom, Amiri e Razavi (2022) realizam uma revisão sistemática da literatura sobre o uso de aprendizado de máquina em diversas áreas da indústria portuária. Com o aumento crescente de dados disponíveis, a aplicação de técnicas de aprendizado de máquina tem se mostrado uma alternativa viável para melhorar a eficiência das operações portuárias. Segundo os autores, a maioria dos trabalhos revisados utiliza o aprendizado de máquina para problemas de predição. Entretanto, cada vez mais, as técnicas de aprendizado de máquina têm sido aplicadas a problemas prescritivos, adentrando um campo tradicionalmente dominado por técnicas de otimização da pesquisa operacional e de simulação. Para problemas de alocação de berços, os autores relacionam trabalhos que envolvem o uso do aprendizado de máquina como auxílio no processo de decisão.

Em León et al. (2017), aborda-se o problema de coordenação entre a alocação de berços e as operações de movimento de cargas em granel entre os pátios e os berços. O objetivo é aumentar a eficiência na operação de carregamento e descarregamento de cargas, a fim de reduzir os tempos de atendimento aos navios e, assim, minimizar o tempo de serviço total. Usam um sistema de classificação baseado em aprendizado de máquina (KNN) para classificar algoritmos pré-definidos para resolução de cada instância.

A incerteza associada aos parâmetros utilizados em problemas de alocação de berços tem sido frequentemente abordada por meio de técnicas de aprendizado de máquina. Kolley et al. (2022) tratam especificamente da incerteza no tempo de chegada esperado (ETA) de cada navio, um parâmetro afetado pelas oscilações nas condições de navegação devido a fatores climáticos. Para lidar com essa incerteza, os autores propõem o uso de dados AIS (Sistema de Identificação Automática), transmitidos pelas embarcações e que contêm informações estáticas, como o tamanho da embarcação, bem como informações dinâmicas, como a velocidade e informações de viagem, como o destino. Esses dados são utilizados para treinar um modelo de aprendizado de máquina capaz de prever o ETA com maior precisão. Os resultados mostram que mesmo técnicas simples, como o k-NN, podem obter alta precisão nas previsões.

Zhang et al. (2022) apresentam um método para resolver o problema de alocação de guindastes em portos por meio de aprendizado por reforço. O método atua sobre um *framework* que utiliza uma abordagem hiper-heurística para selecionar heurísticas construtivas de baixo nível durante o processo de tomada de decisão, aprimorada com a técnica de Dupla DQN (DDQN). Essa técnica treina os parâmetros do *framework* com dois vetores de estado distintos. A abordagem hiper-heurística proposta é comparada a uma heurística manual e a um método gaussiano orientado por dados. Os resultados

mostram que a abordagem proposta superou os outros dois métodos em desempenho.

Cervellera et al. (2021) propõem uma estrutura baseada em aprendizado de máquina que realiza atribuições de berços em tempo real, evoluindo o ambiente após cada decisão. Essa estrutura utiliza uma política parametrizada para determinar as atribuições dos navios aos berços, que é otimizada por meio de um esquema de otimização de entropia cruzada. Os autores destacam que essa abordagem permite definir o problema em um espaço de poucos parâmetros, evitando a necessidade de lidar com um problema inteiro misto, com milhares de variáveis e restrições. Além disso, qualquer medida de desempenho pode ser considerada, sem a necessidade de projetar procedimentos de solução *ad hoc* e heurísticas. Essa estrutura também pode ser facilmente estendida a cenários de decisão mais complexos, ampliando o conjunto de resultados fornecidos pela função de política.

Poucos trabalhos têm explorado o uso direto de técnicas de aprendizado por reforço profundo em problemas de alocação de berços. Em (LI et al., 2022), o Problema de Alocação de Berços (PAB) é abordado indiretamente por meio de um escalonamento inteligente dos caminhos a serem percorridos pelas cargas dos pátios até os navios. Neste trabalho, uma ação do agente consiste na determinação de um caminho entre um pátio e um navio, e a técnica Duplo DQN foi utilizada para treinar o agente. Os autores observaram que boa parte das ações é inválida e, portanto, quando ocorrem, a recompensa para essas ações é 0 e o estado do ambiente não é alterado. Um teste de simulação foi realizado para imitar o escalonamento de uma produção real e os resultados indicaram uma maior eficiência em relação ao tempo.

Como dito anteriormente, há vários trabalhos envolvendo o problema de alocação de berços associado ao problema de alocação de guindastes. No contexto de aprendizado por reforço, a pesquisa ainda é incipiente. Recentemente, Ai et al. (2023) propuseram um método baseado em aprendizado por reforço profundo para esta combinação de problemas em um terminal graneleiro. O problema é considerado multiobjetivo, com o objetivo de melhorar os tempos de atendimento dos navios e reduzir o custo de transporte agregado. Um modelo de processo de decisão de Markov é formulado para tratar o problema de aprendizado por reforço. O espaço de estado é composto pelas informações dos navios, dos berços e dos pátios. O espaço de ações é definido pelas decisões sobre em que berço e em que pátio um navio será descarregado, sendo 1200 as ações possíveis. Desta forma, utiliza-se um mecanismo de filtragem de ações inválidas para reduzir o tamanho do espaço de estados e tornar o treinamento mais eficiente. A função de recompensa é uma linearização dos dois objetivos do problema em uma única função. Os autores propõem um método denominado PS-D3QN, que combina os pontos fortes do Double DQN e do Dueling DQN. Os experimentos realizados utilizam dados de um porto na China, e os resultados obtidos com o método PS-D3QN foram comparados aos dos

métodos Double DQN e Dueling DQN, mostrando-se superiores.

O trabalho Dai, Li e Wang (2023) também aborda o problema de alocação de berços, combinado ao de alocação de guindastes, por meio de aprendizado por reforço. O objetivo do problema é minimizar o tempo de espera total dos navios. Os autores inicialmente propõem um modelo de programação inteira mista para tratar o problema no caso *offline*, em que os tempos de chegada esperados, os tipos de carga dos navios e os tempos de configuração necessários para os guindastes são todos conhecidos. É proposto um algoritmo de inserção gulosa que, passo a passo, a partir de uma solução inicial, melhora o desempenho de alocação. O caso *online*, em que os status dos navios e dos berços são conhecidos com precisão no momento em que os navios chegam ao porto, é tratado por meio de aprendizado por reforço. O espaço de estados compreende duas matrizes (número de berços x número de tipos de carga) que representam o tempo restante para que um berço seja liberado e o tempo de processamento já consumido em um berço para determinado tipo de carga. Além desta, outras duas informações compõem o espaço de estados: um vetor com os tempos de atendimento de navio em cada berço e um escalar que representa o tipo de carga transportada pelo navio. Uma ação indica em qual berço um navio que já está aguardando no porto deve atracar. A função de recompensa é definida pelo tempo de espera. O treinamento foi realizado com um algoritmo baseado no Dueling DQN. Em ambos os casos, as abordagens propostas foram comparadas à estratégia *first-come, first-served* (FCFS). No caso online, em particular, foram usadas algumas combinações de número de berços e de tipos de carga, ambos variando entre 3 e 5, para experimentação, com 3 quantidades de navios diferentes (50, 80 e 120). Os resultados numéricos indicam um melhor desempenho do *Dueling DQN* em todos os cenários, em comparação com a estratégia FCFS.

O uso de aprendizado por reforço para tratar incertezas é feito recentemente para um problema de alocação de berços discreto e dinâmico em um terminal de contêineres em (LV et al., 2024). O trabalho concentra-se na imprevisibilidade dos tempos de chegada e de atendimento, que são modelados como distribuições exponenciais e estocásticas, respectivamente. O objetivo é minimizar o tempo de espera médio dos navios, além de tornar o porto mais resiliente, ou seja, capaz de recuperar-se mais rapidamente em situações de eventos disruptivos. O espaço de estados é composto pelo número de navios esperando, pelo número de navios já atendidos, pelas taxas de utilização de berço, pelos tipos de navios sendo atendidos nos berços, pelos tempos de atendimento de navios restantes nos berços, pelos tempos de espera dos navios em atendimento, pelos tempos de espera dos navios aguardando e pelos números de navios em diferentes níveis de espera. O espaço de ações congrega cinco regras de seleção de navio e berço, que consideram características do navio, o tempo de conclusão, o tempo de atendimento, entre outros. A função de recompensa foi projetada para minimizar o tempo de espera médio ponderado pelo status do navio a partir da aplicação de penalidades. O

treinamento do agente foi realizado com o algoritmo DQN, e os resultados mostraram-se superiores quando comparados com outros três critérios de seleção de navio usuais: (i) o primeiro que chega é o primeiro a ser servido (FCFS), (ii) o navio com maior tempo de atendimento é o primeiro a ser servido (LPT) e (iii) o navio com menor tempo de atendimento é o primeiro a ser servido (SPT).

Li, Yang e Yang (2023) tratam o problema de alocação dinâmica e contínua de berços em múltiplos terminais integrados, com o uso de aprendizado por reforço. O objetivo é minimizar o tempo de permanência dos navios nos portos por meio do algoritmo de aprendizado por reforço do *Dueling Double DQN* (D3QN). Inicialmente, formula-se um modelo de programação inteira mista para representar o problema e, na sequência, formula-se o problema como processo de decisão de Markov. O espaço de estados é definido por um conjunto de dez primeiras observações consecutivas, em que cada observação é dividida em três grupos. O primeiro grupo é formado pelos status dos berços atuais e dos nove momentos passados; o segundo grupo compreende variáveis dinâmicas dos navios, como status, tempo de espera e tempo restante atuais e dos últimos nove momentos passados; e o último grupo compreende informações estáticas dos navios, como comprimento, profundidade e tempos de carregamento e descarregamento. Cada ação no espaço de ações representa uma posição de atracação para um navio, além da opção de aguardar. A função de recompensa é projetada como o oposto do somatório dos tempos de espera e de transferência dos navios. Uma constante é utilizada para ajustar as recompensas a valores positivos. Os resultados obtidos por meio do treinamento com o algoritmo D3QN são comparados com os do algoritmo DQN, do Dueling DQN e do PPO, bem como com os do modelo matemático, por meio do solver comercial CPLEX.

A Tabela 1 apresenta uma análise comparativa dos principais trabalhos relacionados segundo características relevantes ao problema estudado, incluindo: capacidade de decisão direta (ou parcial), consideração de incertezas, abordagem multiobjetivo, integração com controle de estoque, tipo de problema e tipo de abordagem ao problema.

Observa-se que os trabalhos iniciais concentram-se, em sua maioria, no uso de aprendizado de máquina como ferramenta auxiliar, atuando principalmente na predição de parâmetros ou na seleção de heurísticas, sem realizar diretamente a tomada de decisão no problema. Em uma segunda linha, as abordagens híbridas passam a incorporar mecanismos de aprendizado no processo decisório.

Mais recentemente, técnicas de aprendizado por reforço profundo têm sido aplicadas para modelar o problema como um processo de decisão sequencial. No entanto, de modo geral, tais abordagens concentram-se em objetivos operacionais específicos, como a minimização do tempo de espera, e não incorporam explicitamente a dinâmica de controle de estoque, aspecto fundamental em terminais graneleiros.

Tabela 1 – Comparação dos trabalhos relacionados

Trabalho	**Decide diret.**	**Incert.**	**Multi-obj.**	**Estoque**	**Problema**	**Tipo**
León et al. (2017)	Não	Não	Não	Não	PAB	ML auxiliar
Kolley et al. (2022)	Não	Sim	Não	Não	PAB	Predição
Zhang et al. (2022)	Parc.	Sim	Não	Não	Roteamento	Híbrido
Cervellera et al. (2021)	Sim	Sim	Não	Não	PAB	Política
Li et al. (2022)	Sim	Sim	Não	Não	Escalonamento	RL
Ai et al. (2023)	Sim	Não	Sim	Não	PAB+Pátios	RL
Dai, Li e Wang (2023)	Sim	Sim	Não	Não	PAB+Guindastes	RL
Li, Yang e Yang (2023)	Sim	Não	Não	Não	PAB Multiterminal	RL
Lv et al. (2024)	Sim	Sim	Não	Não	PAB	RL
Este trabalho	**Parc.**	**Sim**	**Sim**	**Sim**	**PAB**	**RL**

Fonte: Elaborada pelo autor.

A proposta deste trabalho diferencia-se ao integrar, de forma conjunta, a tomada de decisão direta por meio de aprendizado por reforço profundo, o tratamento de incertezas operacionais, a consideração de múltiplos objetivos e a incorporação explícita do controle de estoque. Essa combinação torna a solução mais aderente às condições reais de operação em terminais portuários.

4 Formulação do PAB como um problema de Aprendizado por Reforço

Neste capítulo, apresenta-se a formulação proposta do problema de alocação de berços como um problema de aprendizado por reforço (*Reinforcement Learning* - RL). O objetivo principal desta formulação, doravante denominada BAP-RLIM (*Berth Allocation Problem as Reinforcement Learning with Inventory Management*), é evitar que decisões de atracação levem a colapsos na produção e, consequentemente, a custos adicionais decorrentes da violação de restrições de nível de estoque, por meio de critérios (filas) de atendimento. Além disso, é importante considerar o refinamento da solução viável adotada, normalmente associado aos tempos de operação e de espera dos navios. Esta formulação agrega outros elementos, como o controle de estoque, outros mecanismos de decisão e métodos de solução em comparação com trabalhos incipientes que envolvem o aprendizado de políticas para a tomada de decisões no problema de alocação de berços, como o trabalho desenvolvido por Cervellera et al. (2021).

4.1 Ambiente virtual

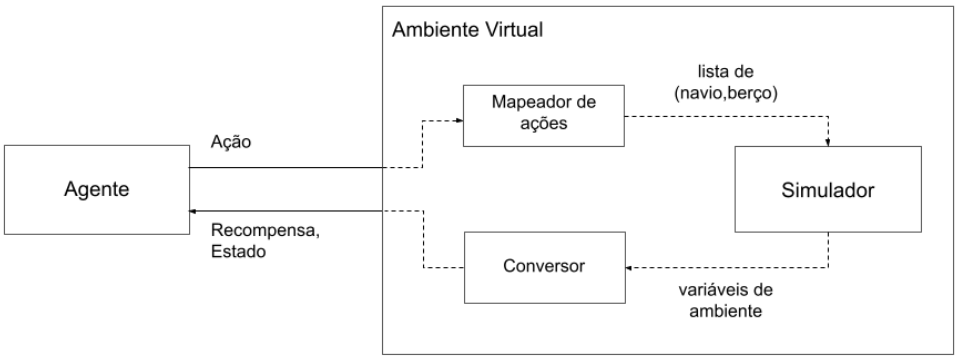
O ambiente virtual foi desenvolvido para fins de prototipação. Nesse ambiente, o simulador da dinâmica de atracação constitui um componente fundamental para o treinamento de agentes com algoritmos de aprendizado por reforço, permitindo a coleta de indicadores de desempenho a cada unidade de tempo ao longo da sequência de decisões.

Por meio da Figura 5, pode-se visualizar a interação entre o ambiente virtual, o simulador e o agente. O agente observa o estado atual e envia a ação ao ambiente virtual. No ambiente virtual, a ação é processada pelo mapeador de ações e convertida em uma lista de pares (navio, berço). O simulador atraca o(s) navio(s) e avança a simulação para o próximo estado. Por fim, o conversor extrai o estado observável e a recompensa e os envia ao agente. O Algoritmo 1 descreve estas etapas. Mais detalhes do simulador podem ser vistos no Capítulo 6.

Enquanto a Figura 5 apresenta a dinâmica de interação entre o agente e o ambiente virtual, a Figura 6 organiza esses elementos sob a perspectiva do aprendizado por reforço. Essa interação, típica do aprendizado por reforço, exige mapear o problema de alocação de berços com controle de estoque, descrito no Capítulo 2, aos elementos característicos de RL.

A Figura 6 ilustra a arquitetura básica do processo do arcabouço BAP-RLIM. Os

Figura 5 – Esquema básico do ambiente virtual. Ilustra a interação entre o agente e o ambiente virtual, bem como o processo no ambiente virtual entre os componentes mapeador de ações, simulador e conversor.



Fonte: Elaborada pelo autor.

Algorithm 1 Interação entre Ambiente Virtual, Simulador e Agente

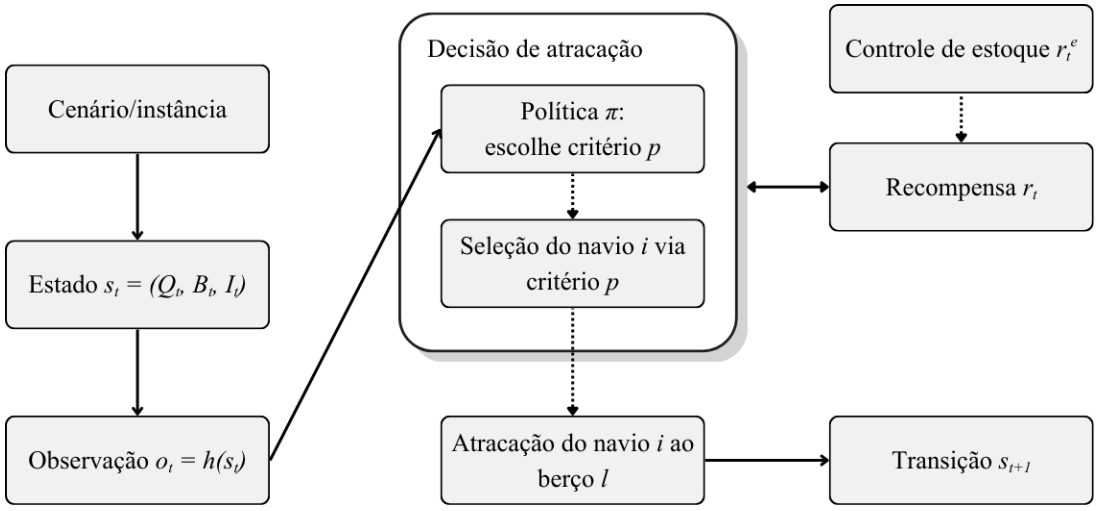
1: **Início**
2: **while** verdadeiro **do**
3: $a \leftarrow$ ação do agente
4: lista de atracações $\leftarrow$ mapear a ação no ambiente(a)
5: variáveis de ambiente $\leftarrow$ processar no simulador(lista de atracações)
6: $(s', r) \leftarrow$ extrair estado e calcular recompensa(variáveis de ambiente)
7: enviar (s', r) ao agente
8: **Fim**

componentes desta arquitetura são derivados da estrutura padrão do aprendizado por reforço, porém, com os elementos essenciais ao problema em questão. O *dataset* de treinamento e de testes é derivado do conjunto de instâncias do PAB. Os *status* dos navios, dos berços e dos estoques compõem o estado. As decisões de atracação são determinadas pela ação do agente. A recompensa orienta as decisões do agente para soluções de qualidade que também atendam às condições desejáveis. As atracações dos navios e a transição para os próximos estados são realizadas pelo simulador. O controle de estoque é um componente-chave que perpassa todos os componentes. Os demais elementos que tornam este arcabouço aplicável são detalhados ao longo deste capítulo. Nessa arquitetura, o fluxo parte do cenário ou instância, que define o estado do sistema e sua observação, segue para a decisão de atracação guiada pela política, e resulta na transição de estados e na recompensa, ambas influenciadas pelo controle de estoque.

4.2 Instância

A instância representa o cenário do PAB considerado. Como mencionado na Seção 2.1.1, vários cenários são encontrados entre os diversos portos ao redor do mundo, assim como na literatura relacionada. Diante deste fato, mesmo limitando-se a um subconjunto de problemas de alocação de berços, o arcabouço deve prover a

Figura 6 – Arquitetura do processo do arcabouço BAP-RLIM.



Fonte: Elaborada pelo autor.

possibilidade de aplicação ao número máximo de variações possíveis.

A definição do cenário é a etapa do processo que permite incluir todas as características pertinentes à dinâmica do processo no âmbito do PAB com controle de estoque integrado. Esta definição é específica a cada problema; todavia, há dados e regras indispensáveis para o dimensionamento do sistema. A Tabela 2 relaciona todos os componentes essenciais do cenário. Estes componentes são divididos em conjuntos, parâmetros, condições iniciais e restrições.

4.2.1 Conjuntos

Os conjuntos de navios N, berços L e tipos de carga K compõem as entidades básicas do problema. Os berços $l \in L$ e os tipos de cargas (estoques) $k \in K$ são pré-definidos. Os navios $i \in N$ podem ser totalmente conhecidos (caso *offline*) ou revelados ao longo do tempo (caso *online*). Todavia, em ambos os casos, emprega-se uma lista de previsões (*look-ahead*) em um horizonte de planejamento rolante.

Seja $\epsilon \in \mathbb{Z}_+$ o número de navios no *look-ahead*.

Definimos $N_t^{(\epsilon)}$ como os primeiros ϵ navios não atendidos, ordenados por ETA_i crescente no instante t.

4.2.2 Parâmetros

Os parâmetros correspondem aos recursos das operações, como a infraestrutura dos navios e dos berços, o transporte e o manejo das cargas e o tempo.

Esses parâmetros podem ser caracterizados como **determinísticos** ou **estocásticos**. A compatibilidade navio-berço A_{il} e as janelas de tempo de atracação

Tabela 2 – Componentes essenciais do *cenário/instância* no arcabouço BAP-RLIM. Os itens definem a escala, as condições iniciais e as restrições.

Categoria	**Componente**	**Descrição**
Conjuntos	N, L, K	Entidades do problema: N (navios), L (berços), K (tipos de cargas ou estoques)
Parâmetros	A_{il}	Compatibilidade entre navio e berço (calado, equipamento, carga, outros aspectos). $A_{il} = 1$ se o navio $i \in N$ pode atracar no berço $l \in L$.
	$[T_i^{(s)}, T_i^{(f)}]$	Janelas de atracação por navio: início (s) e fim (f).
	$q_{i,k}$	Demanda por carga (desembarque $q_{i,k} \geq 0$, embarque $q_{i,k} \leq 0$).
	c_k	Taxa de consumo ($c_k > 0$) ou de produção ($c_k < 0$) do estoque.
	$I_{\min}^{(k)}, I_{\max}^{(k)}$	Limites inferior e superior de estoque por carga.
	ETA_i ou a_i	Tempo de chegada esperado do navio i.
	h_{il}	Tempo de atendimento do navio i no berço l.
Condições iniciais	$e_k^{(0)}$	Estoques iniciais por carga.
Restrições	Rígidas	Regras que inviabilizam operações de navios: compatibilidade, manutenção, janelas.
	Desejáveis	Custos incorporados às decisões: multas por atraso, ociosidade de navios atracados

Fonte: Elaborada pelo autor.

$[T_i^{(s)}, T_i^{(f)}]$ são parâmetros determinísticos que podem ser aplicados ao cenário. Quando omitidos, assume-se que todos os navios podem ser atendidos em quaisquer berços e que as janelas são as maiores possíveis. Por outro lado, as quantidades de carga q_{ik} transportadas pelos navios, as taxas de consumo (ou produção) dos estoques c_k e os limites inferiores e superiores são necessariamente definidos.

Alguns desses recursos determinísticos podem apresentar incertezas, sem prejuízo do arcabouço. O tempo de atendimento h_{il} é um parâmetro que, como visto no Capítulo 3, comumente é apresentado como estocástico. Também podem ser incluídos no h_{il} os tempos de início e término da operação. Do mesmo jeito, o tempo de chegada esperado ETA_i do navio i também pode ser modelado por meio de uma distribuição de probabilidade.

4.2.3 Condições iniciais

Os parâmetros de nível de estoque inicial, $e_k^{(0)} \in \mathbb{Z}_+$, e os navios em operação nos berços delimitam as condições iniciais obrigatórias.

4.2.4 Restrições

Cada cenário portuário impõe seu conjunto de restrições peculiares em razão das diferentes configurações portuárias existentes. Ainda há restrições comuns à grande maioria dos problemas de alocação de berços. Aqui, têm-se as restrições relacionadas a parâmetros, como as janelas de atracação, definidas pelo intervalo $[T_i^{(s)}, T_i^{(f)}]$ e pelo tempo de chegada esperado ETA_i. Há também as restrições referentes à compatibilidade navio-berço A_{il}. Outras restrições são triviais, como a não sobreposição de navios.

No BAP-RLIM, o controle de estoque é abordado como uma restrição essencial do problema. Intuitivamente, esta restrição é considerada uma restrição suave (*soft constraint*), ou seja, é apenas indesejada, na perspectiva da eficiência, pois os níveis de estoque não inviabilizam operacionalmente o processo de atracação. Por outro lado, as operações de embarque ou desembarque podem ser comprometidas pela falta ou pelo excesso de cargas nos estoques. Assim, o controle de estoque é tratado como uma restrição rígida (*hard constraint*) do problema.

4.3 Estados e Observações

O espaço de estados deve ser modelado de modo significativo, capaz de fornecer informações relevantes ao agente na tomada de decisão e, ao mesmo tempo, manter uma relação com a função de recompensa, a fim de garantir alguma correlação entre as partes. No BAP-RLIM, o espaço de estados é dividido entre informações de status dos navios no *look-ahead*, do atendimento dos berços e dos níveis de estoque. No arcabouço proposto, o estado no instante de decisão t é

$$s_t \;=\; (\,Q_t,\; B_t,\; I_t,\; \tau_t\,),$$

onde: (i) Q_t resume as informações relevantes dos navios *candidatos a atendimento* presentes no *look-ahead*; (ii) B_t descreve os berços (ocupação, cargas em operação e tempos remanescentes); (iii) I_t são os níveis de estoque por carga; e (iv) τ_t é a informação temporal (absoluta, Δt, progresso do episódio ou embutida nos demais componentes).

Para qualquer instanciação concreta, denotam-se por d_B, d_I, d_Q e d_τ as dimensões (número de atributos) associadas a cada componente. Se $C_t \subseteq N$ é o conjunto de candidatos no instante t, com $\epsilon = |C_t|$, então:

$$\text{número de atributos} \;=\; |L| \cdot d_B \;+\; |K| \cdot d_I \;+\; \epsilon \cdot d_Q \;+\; d_\tau.$$

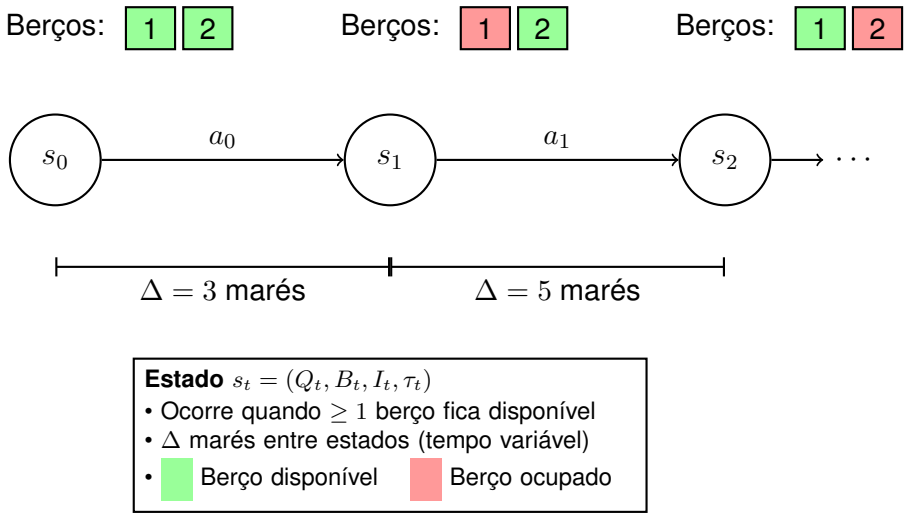
Essa contagem permite alterar ou expandir atributos sem alterar a estrutura do arcabouço.

Cada estado está associado a uma unidade de tempo (uma maré j), na qual pelo

menos um berço está disponível para atendimento. A mudança de estados ocorre da seguinte forma: no estado atual, (i) a decisão de atracação designa (através dos critérios) navios a todos os berços disponíveis; (ii) as atualizações no ambiente são recalculadas a cada unidade de tempo, representando a evolução das operações no porto, até que algum outro berço esteja disponível novamente.

Na Figura 7 é mostrado um exemplo de ocorrências de estados. Neste caso, o problema é composto por dois berços. Quando pelo menos um dos dois berços estiver disponível, ocorre um novo estado e o agente precisa tomar uma decisão a_t. É importante observar que a evolução dos estados não está proporcionalmente associada à evolução do tempo. A mudança de estado pode acontecer em uma ou várias marés, sendo o tempo, portanto, variável.

Figura 7 – Ocorrência dos estados s_t.



Fonte: Elaborada pelo autor.

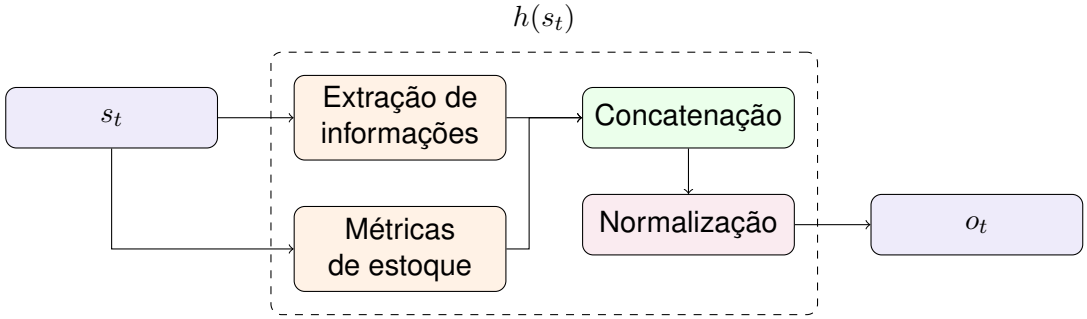
A codificação entregue ao agente é $o_t = h(s_t)$. A função $h(s_t)$ deve ser implementada conforme a arquitetura de rede neural do agente, porém, deve obedecer às regras gerais. A Figura 8 ilustra o *pipeline* dos processos gerais ϕ_Q, ϕ_B e ϕ_I, sugerido neste arcabouço, aplicados ao estado s_t, por meio da função $h(s_t)$.

4.3.1 Extração de informações

O primeiro processo, *extração de informações*, é responsável pela obtenção e pelo processamento de dados para o envio de informações relevantes ao processo de aprendizado. Estes dados são extraídos dos componentes do estado $s_t = (Q_t, B_t, I_t, \tau_t)$ e ajudam o agente a tomar decisões que garantam, principalmente, o controle dos estoques e, secundariamente, a otimização ou um nível de eficiência superior nas operações. Abaixo, estão os dados de s_t importantes enviados ao agente.

Q: (i) tempos restantes de chegada; (ii) tempos de atendimento em cada berço; (iii) tipos

Figura 8 – *Pipeline* da função $h(s_t)$, com os processos gerais de transformação do estado s_t em observação o_t.



Fonte: Elaborada pelo autor.

de carga transportada; e (iv) quantidades de carga transportada.

Os navios observados estão no horizonte de planejamento rolante, ordenados crescentemente pelo ETA_i, até o limite de ϵ navios. Eventualmente, há inclusão de navios *artificiais*, gerados para compor o horizonte de planejamento (*padding*), caso não haja navios conhecidos suficientes para preencher o horizonte de planejamento no instante t.

B: (i) tempos restantes de atendimento nos berços; (ii) tipos de cargas sendo atendidos nos berços; (iii) taxas de operação de carregamento/descarregamento; e (iv) tempos de ociosidade até a próxima atracação.

Os dados dos berços são ordenados em ordem crescente, com a vazão como critério.

I: níveis de estoque (ou saldos aceitáveis).

4.3.2 Métricas de estoque

Um ponto importante no BAP-RLIM é a incorporação de métricas de estoque que visam mensurar o *status* atual do estoque. É importante notar que não se trata apenas de revelar os níveis atuais dos estoques e seus saldos em relação aos limites estabelecidos. Essas informações são triviais e podem ser obtidas pelo processo de *extração de informações* descrito anteriormente.

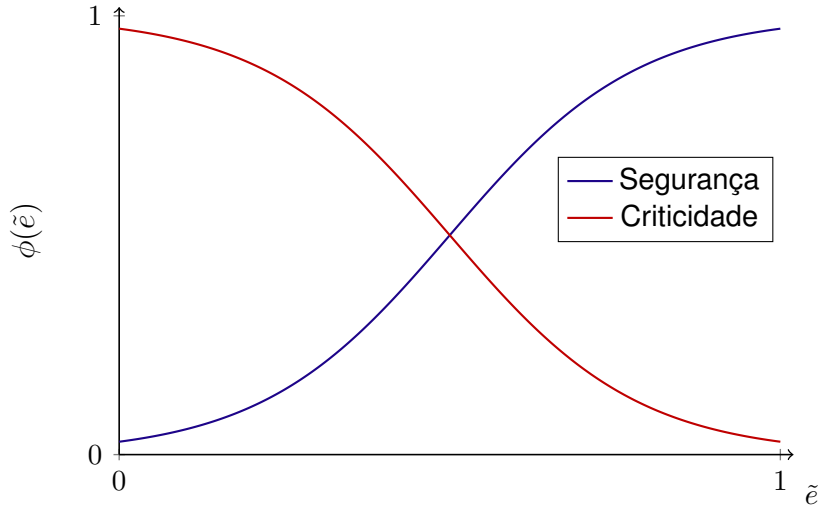
O objetivo das *métricas de estoque* é, portanto, fornecer ao agente indicadores sobre o *grau de atenção* a ser atribuído aos níveis de estoque no instante t. O grau de atenção pode ser visto de duas perspectivas diferentes:

Criticidade: quanto maior o risco de um ou mais estoques colapsarem, maior deve ser a atenção a eles.

Segurança: quanto maiores os saldos de todos os estoques em relação aos níveis mínimos de segurança, menor deve ser a atenção aos estoques.

Criticidade e segurança estão relacionadas. A criticidade máxima equivale à segurança mínima e vice-versa; entretanto, a abordagem escolhida influencia a implementação da função $\phi(\tilde{e}_t)$ subjacente, que mapeia o estoque $\tilde{e} \in [0, 1]$ para um grau de atenção no intervalo $[0, 1] \in \mathbb{R}$. A Figura 9 ilustra, de forma simplificada, a relação entre as duas perspectivas em um cenário com um estoque. $\tilde{e}$ é o saldo normalizado, considerando que 0 é o limite mínimo estabelecido no cenário de importação.

Figura 9 – Relação entre criticidade e segurança na implementação da função ϕ com uma variável.



Fonte: Elaborada pelo autor.

4.3.2.1 Vários estoques

Os portos graneleiros operam comumente com mais de um tipo de estoque. Esta característica impõe ao BAP-RLIM a implementação de métricas que levem isso em conta. Duas abordagens podem ser utilizadas para a função ϕ:

- Calcular um grau de atenção individual para cada estoque: $\phi(\tilde{e}_i)$ para $i = 1, 2, \ldots, k$. Esta abordagem trata os estoques individualmente, sem relação entre si. Neste caso, o agente pode ter uma informação precisa sobre a situação do estoque; porém, ignora qualquer informação útil dos demais.
- Ajustar ϕ para uma função de várias variáveis: $\phi : (\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_k) \to [0, 1]$. Esta abordagem, por sua vez, fornece uma visão geral e única de todos os estoques. Não há informações sobre estoques específicos. Além disso, surge a seguinte questão: *é melhor ter dois (ou mais) estoques baixos do que ter apenas um estoque ainda mais baixo?*

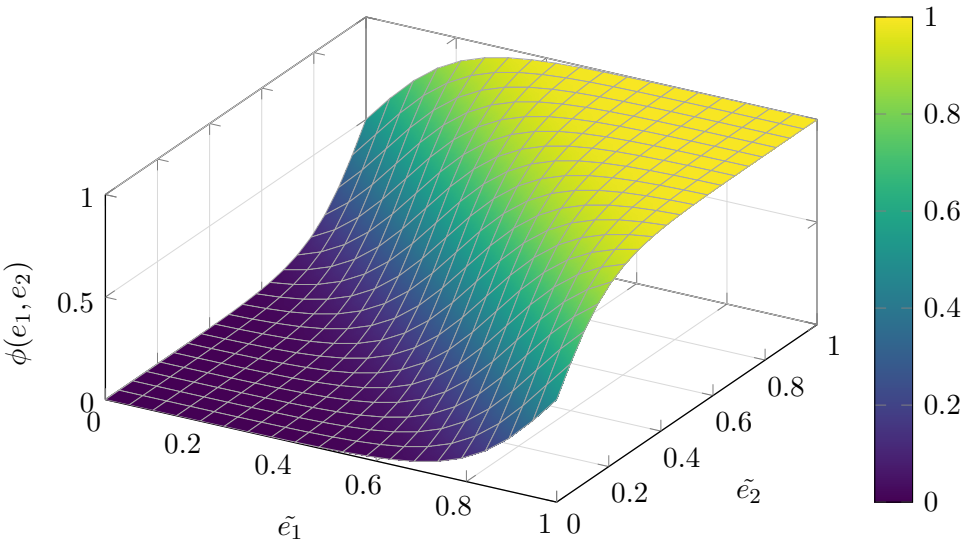
Uma função de várias variáveis é uma modelagem mais complexa devido à dificuldade de incorporar a sinergia entre diferentes estoques. Nas equações 4.1-4.3 apresentam-se exemplos de funções implementadas na perspectiva da *segurança* com dois estoques normalizados, $\tilde{e_1}$ e $\tilde{e_2}$. Na Equação 4.1, a função ϕ é uma sigmoide que cresce com o acúmulo de estoques. Em 4.2, o resultado da função ϕ é dominado pelo menor dos estoques. Neste caso, os demais estoques são ignorados. A função em 4.3 promove sinergia entre os estoques, piorando o todo quando um deles cai. As figuras 10-12 ilustram o comportamento esperado para cada tipo de função descrita.

$$\phi(\tilde{e_1}, \tilde{e_2}) = \frac{1}{1 + \exp(-a\,(\tilde{e_1} + \tilde{e_2} - 1))} \tag{4.1}$$

$$\phi(\tilde{e_1}, \tilde{e_2}) = \Big(\min\{\tilde{e_1}, \tilde{e_2}\}\Big)^a \tag{4.2}$$

$$\phi(\tilde{e_1}, \tilde{e_2}) = {\tilde{e_1}}^a {\tilde{e_2}}^a = (\tilde{e_1}\tilde{e_2})^a \tag{4.3}$$

Figura 10 – Gráfico da função $\phi(\tilde{e_1}, \tilde{e_2})$ conforme a Equação 4.1. a=10.



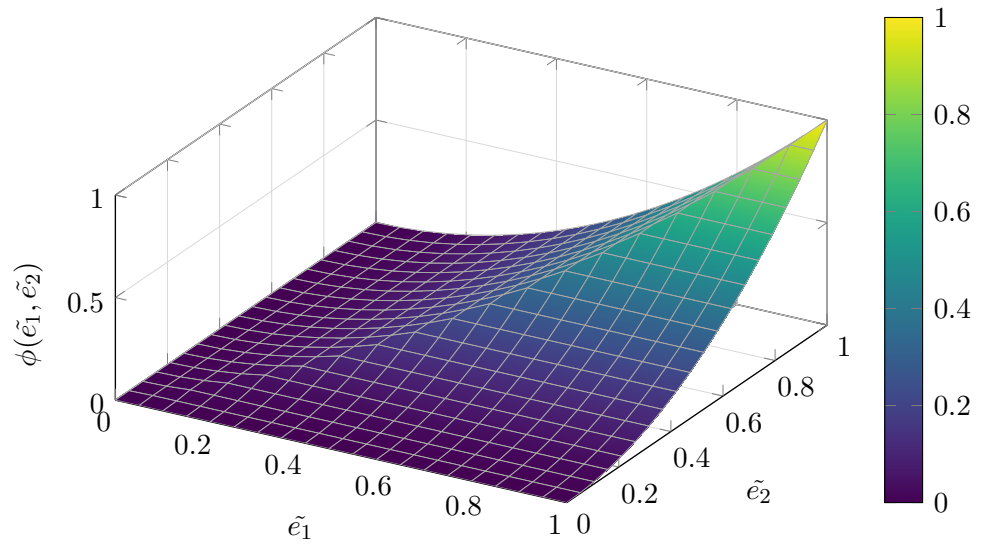
Fonte: Elaborada pelo autor.

4.3.2.2 Heurísticas

É fácil concluir que, além dos níveis de estoque, outras variáveis influenciam a decisão de priorizar navios que atendem a determinados estoques. Abaixo, são destacadas algumas dessas variáveis importantes:

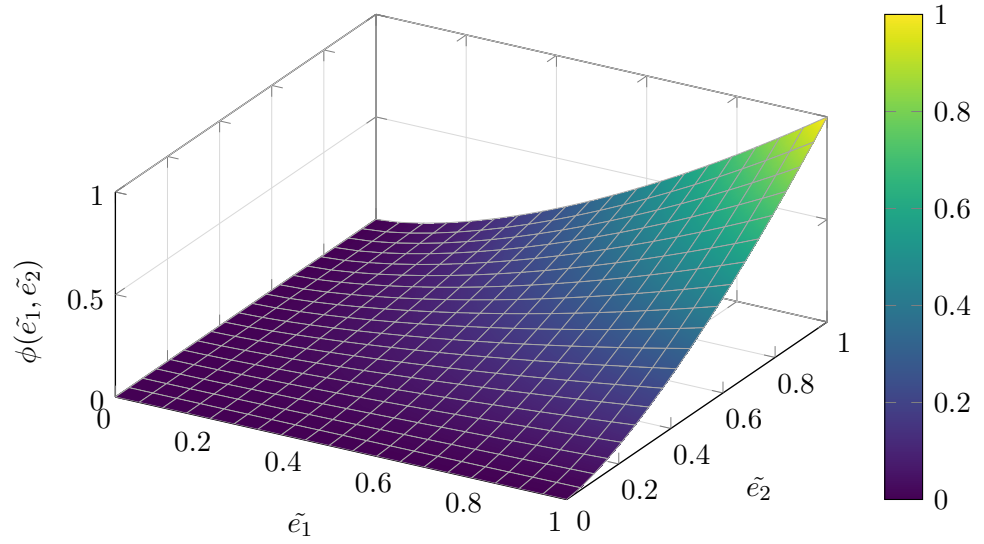
- *Projeção dos estoques para instantes futuros*. Navios já em atendimento afetam os níveis de estoque futuros.

Figura 11 – Gráfico da função $\phi(\tilde{e_1}, \tilde{e_2})$ conforme a Equação 4.2. a=2.



Fonte: Elaborada pelo autor.

Figura 12 – Gráfico da função $\phi(\tilde{e_1}, \tilde{e_2})$ conforme a Equação 4.3. a=1.5.



Fonte: Elaborada pelo autor.

- *Expectativa de disponibilidade de carga*. Cargas com baixo volume transportado têm menos margem para eventuais atrasos no atendimento.
- *Quantidade de cargas de outros navios*. O atendimento de navios com cargas muito grandes pode postergar excessivamente os atendimentos de navios com cargas críticas.
- *Ociosidade de berços*. A decisão de priorizar navios com cargas críticas não pode desconsiderar o tempo restante em que esses navios permanecerão disponíveis para atracação.

Isto torna a modelagem das métricas de estoque ainda mais desafiadora, tanto

para um único estoque quanto para vários estoques. Por isso, uma alternativa sugerida é recorrer a heurísticas para implementar ϕ. Essa abordagem tem vantagens fundamentais: fornece uma resposta rápida, permite incluir facilmente regras mais complexas e relaciona todas as variáveis à valoração empírica do grau de atenção necessária ao estoque.

4.3.3 Concatenação e Normalização

A observação o_t é finalizada por meio da concatenação e da normalização dos resultados das extrações de informações e métricas de estoque:

$$\mathbf{o}_t = \text{Norm}\left(\text{Concat}\left(\text{informações extraídas}, \text{métricas de estoque}\right)\right)$$

Concatenação (Concat) é o processo que une as diferentes variáveis, obtidas pela extração de informações e pelas métricas de estoque, fornecendo ao agente uma visão holística dos dados do sistema, de modo que ele capture correlações entre as variáveis.

Normalização (Norm) mapeia todas as variáveis para o intervalo $[0, 1] \in \mathbb{R}$. Isto previne a predominância de escala entre variáveis igualmente importantes e busca estabilidade numérica no processo de aprendizado.

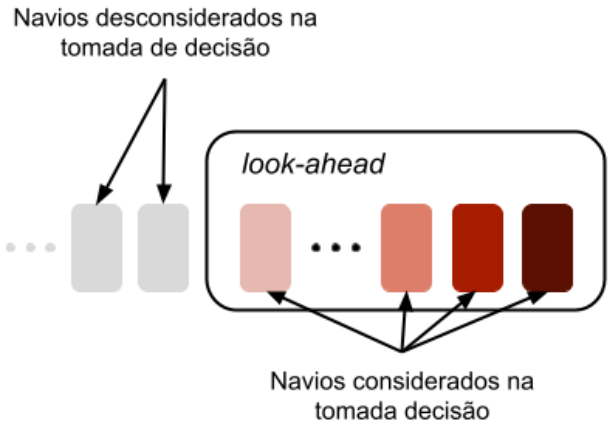
4.4 Ambiente e Agente

O ambiente é uma das duas entidades básicas do aprendizado por reforço. Ele representa o cenário, real ou simulado, no qual se pretende aprender a tomar boas decisões. Neste trabalho, o ambiente representa a dinâmica do processo de alocação de berços com restrições aos níveis de estoque, conforme descrito por meio do modelo matemático na Seção 2.1.

O BAP-RLIM é projetado para o caso *online* do problema de alocação de berços, em que não há acesso a todas as informações durante o processo de decisão. Comumente, o PAB, no caso *online*, apresenta informações parciais sobre os navios que demandam atendimento. Ao longo da evolução das decisões, as informações ora ausentes, ora imprecisas são definidas. Aqui, utiliza-se o mecanismo de lista de previsões, ou *look-ahead*, em que estão os ϵ navios observados com o tempo de chegada previsto mais próximo. Os demais navios programados para serem atendidos, mas que estão fora do *look-ahead*, são desconsiderados no processo de decisão no momento. Na Figura 13, ilustra-se o esquema de seleção de navios com *look-ahead*. Nota-se que o caso *offline*, em que todas as informações são conhecidas previamente, é facilmente adaptável.

Em cada tomada de decisão, um ou mais navios no *look-ahead* são selecionados para atracação e, automaticamente, retirados dele. A mesma quantidade de navios retirados é adicionada a partir dos navios anteriormente desconsiderados, tendo o menor

Figura 13 – Horizonte de planejamento rolante com *look-ahead*.



Fonte: Elaborada pelo autor.

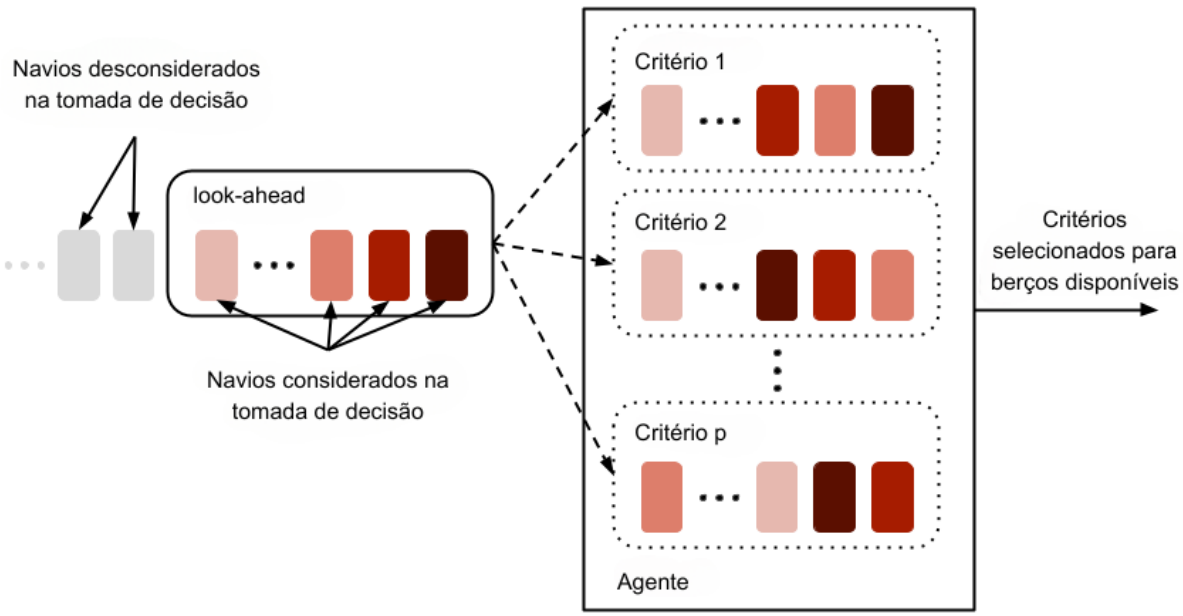
tempo de chegada esperado como critério. Ou seja, o horizonte de planejamento é rolante com o tamanho ϵ fixo.

O processo de alocação de um navio a um berço ocorre quando há disponibilidade de berço, ou seja, não há outro navio sendo atendido nele. Nesse contexto, cabe ao agente, em última instância, decidir qual navio deve ser atracado em qual berço, o que equivale a definir os valores de y_{ijl}, conforme a Seção 2.1.2.2. Porém, no BAP-RLIM, o navio não é selecionado diretamente. O agente seleciona um critério equivalente a uma regra de escolha do navio. A Figura 14 ilustra um esquema simplificado da seleção de critérios para os berços disponíveis no processo de tomada de decisão, com base nos navios no *look-ahead*.

Quando houver dois ou mais berços disponíveis simultaneamente, a decisão de atracação envolverá o mesmo número de critérios que o número de berços disponíveis. Essa abordagem visa atribuir ao agente a decisão de forma autônoma. Por outro lado, essa escolha leva ao surgimento de um conjunto de ações inválidas que exigem um tratamento cuidadoso, garantindo que a restrição de não sobreposição de navios seja cumprida, conforme descrito na Equação 2.4 do modelo matemático. Mais detalhes sobre o espaço de ações podem ser encontrados na Seção 5.1.2.

Para controlar o nível de estoque de diferentes tipos de carga, conforme a restrição do modelo matemático expressa pela Equação 2.5, é necessário calcular a quantidade de carga descarregada e o consumo do estoque ao longo do tempo. O cálculo da denominada taxa de descarregamento é fundamental para determinar o nível de estoque em cada unidade de tempo. Essa taxa é definida como o somatório da razão entre a quantidade de carga do navio e a vazão do berço que o atende ao longo de todos os navios que carregam o determinado tipo de carga e estão em atendimento, como mostrado pela Equação 4.4.

Figura 14 – Esquema simplificado de seleção de critérios com *look-ahead*.



Fonte: Elaborada pelo autor.

$$D_k = \sum_{i=1}^{N_k} \frac{q_{ik}}{v_{il}}, \tag{4.4}$$

Onde D_k é a taxa de descarregamento do tipo de carga k, N_k é o conjunto de navios em atendimento que carregam o tipo de carga k, q_{ik} é a quantidade de carga do navio i do tipo k, e v_{il} é a vazão do berço l que atende o navio i. A cada nova alteração dos navios em atendimento, D_k é recalculada.

Os níveis de estoque em cada unidade de tempo j (maré j), de cada tipo de carga k, são calculados somando o estoque na unidade de tempo anterior $e_k(j-1)$, à taxa de descarregamento do tipo de carga D_k e subtraindo o consumo do estoque c_k, como pode ser visto pela Equação 4.5.

$$e_k(j) = e_k(j-1) + D_k - c_k \tag{4.5}$$

A Equação 4.5 é necessária para controlar o estoque de diferentes tipos de cargas ao longo do tempo. Na proposta apresentada, sempre que a quantidade de uma determinada carga k torna-se igual a zero na unidade de tempo j, ou seja, $e_k(j) = 0$ (podendo ser outro limite de tolerância), é aplicada uma penalidade. Por fim, a restrição 2.3 não pode ser garantida, uma vez que o objetivo é aprender uma política de tomada de decisão em vez de definir um conjunto específico de atribuições de navios a berços.

Há algumas diferenças importantes entre o BAP-RLIM e o modelo matemático. O

BAP-RLIM pode ser aplicado em ambos os casos, *online* e *offline*, enquanto o modelo matemático é aplicável apenas no caso *offline*. A abordagem com *look-ahead* permite avaliar um número limitado, ϵ, de navios no caso *online*, quando $\epsilon < |N|$, ou o número total de navios, $\epsilon = |N|$, no caso *offline*.

Outra diferença importante entre as duas abordagens é a forma como a tomada de decisão considera a unidade de tempo. O modelo matemático define diretamente a maré de atracação para cada navio; em contrapartida, a abordagem de aprendizado por reforço determina apenas se um navio deve ou não ser atracado. No entanto, se um navio é atribuído a um berço antes de chegar ao porto, ou seja, com seu ETA maior que a maré atual ($a_i > TTW$ atual), o berço permanece bloqueado até o momento da chegada do navio ($a_i = TTW$ atual). Dessa forma, a restrição 2.2 do modelo matemático não é violada.

4.5 Política de decisão

Na formulação proposta, a política π atua como um seletor de critérios de seleção (ou priorização) de navios. O agente não seleciona diretamente o navio a ser atracado. Dado o estado observável $o_t = h(s_t)$, a política π escolhe um critério $p \in \mathcal{P}$ para $o \in \mathcal{O}$, onde $\mathcal{P}$ é o conjunto de critérios de seleção pré-definidos e $\mathcal{O}$ é o conjunto de estados observáveis.

$$\pi : \mathcal{O} \rightarrow \mathcal{P}$$

Na prática, cada critério p define uma fila de navios ordenados segundo sua regra subjacente, a partir dos navios no horizonte de planejamento definido pelo *look-ahead*. A Figura 14 ilustra este esquema simplificado, em que esses navios são virtualmente reordenados por critérios distintos, oferecendo ao tomador de decisão as respectivas filas. Na etapa seguinte do arcabouço, o critério escolhido é utilizado para selecionar o navio.

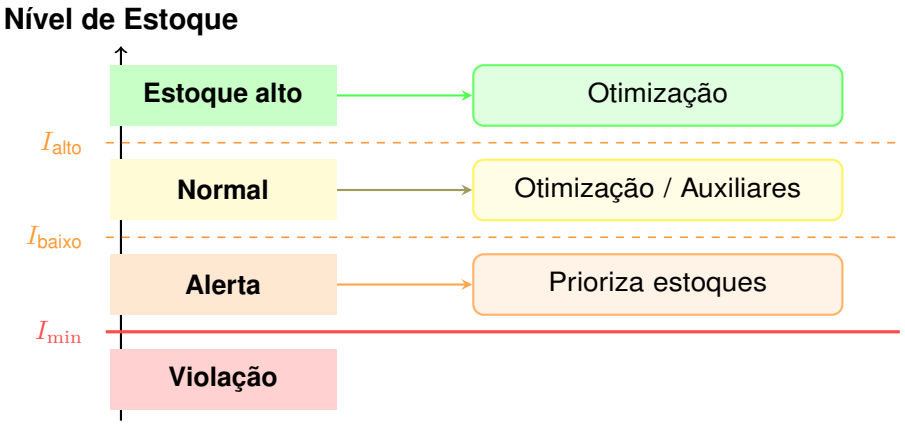
O conjunto de critérios $\mathcal{P}$ precisa estar coerente com os propósitos do BAP-RLIM. Então, é possível categorizar os critérios em *funcionais* e *não-funcionais*. Os critérios funcionais representam os objetivos centrais: controle de estoque e otimização de função objetivo. Os critérios não funcionais estão voltados a objetivos secundários ou auxiliares quando a decisão não prioriza o estoque e a otimização.

A lógica desejada para a política π busca garantir o controle de estoque pelo(s) critério(s) de estoque e condicionar o uso do(s) critério(s) de otimização a níveis de estoque seguros. Os critérios auxiliares podem atuar em conjunto com os critérios funcionais ou ser utilizados como *critério padrão* para zonas operacionais intermediárias.

Na Figura 15, é mostrado esse comportamento desejado, organizado em *zonas operacionais*, definidas por limites de estoque I, em um cenário de importação. O

limite I_{min} separa as zonas aceitáveis da zona de violação. Até um limite considerado baixo de estoque I_{baixo}, a lógica deve ser priorizar estoques críticos para sair da zona de alerta. Na zona entre os limites I_{alto} e I_{baixo}, o critério padrão (auxiliar) pode ser combinado com o critério de otimização. A lógica é buscar a otimização tomando cuidado com comportamentos que podem conduzir à zona de alerta. O estoque alto indica a possibilidade de maior esforço para otimizar a função objetivo.

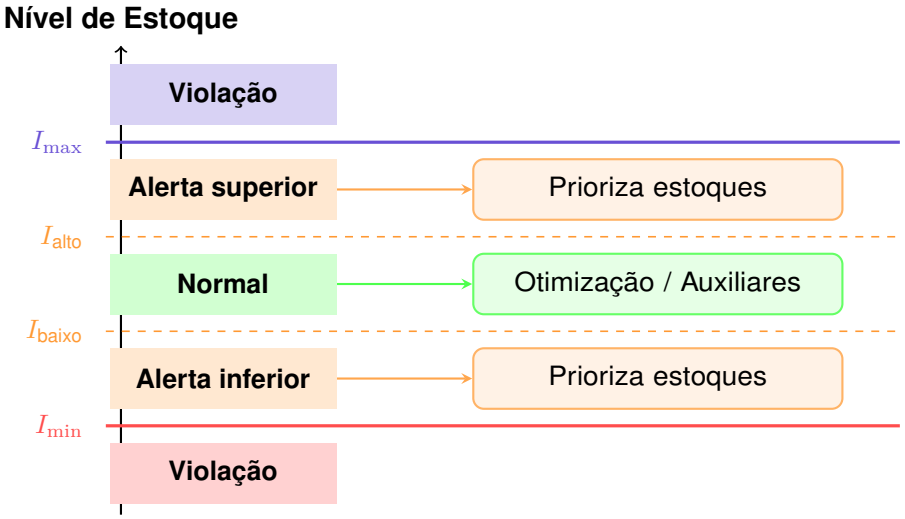
Figura 15 – Lógica desejada para a aplicação dos critérios em um cenário de importação.



Fonte: Elaborada pelo autor.

Na Figura 16 é mostrada a lógica de aplicação dos critérios desejada para o cenário de importação e exportação. Acima do limite I_{max} e abaixo do limite I_{min} há zonas de falha. A zona *normal* é a zona de segurança entre as zonas de *alerta superior* e *inferior*. Neste caso, os critérios de estoque atuam inversamente, dependendo de se os estoques estão em alerta superior ou inferior.

Figura 16 – Lógica da aplicação dos critérios desejada para um cenário de importação e exportação.



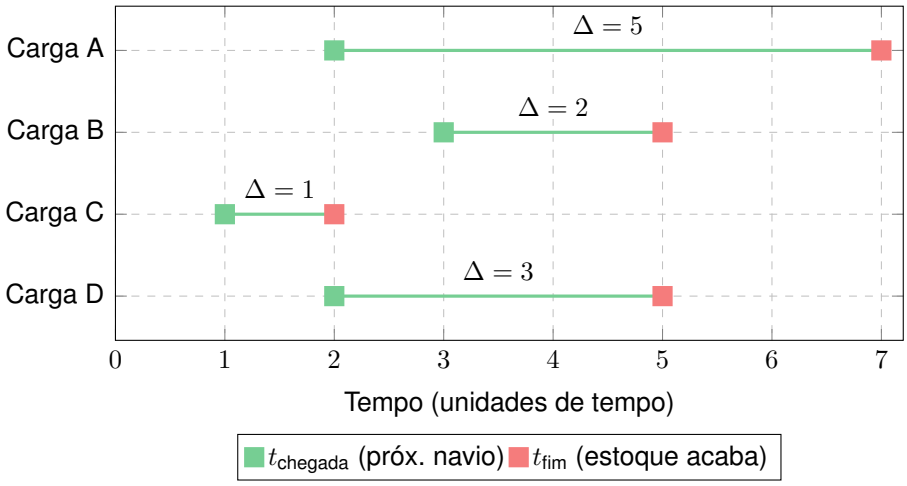
Fonte: Elaborada pelo autor.

4.5.1 Priorização de estoque

No BAP-RLIM, os critérios de estoque são fundamentais para garantir que os níveis de estoque sejam priorizados quando houver ameaça de violação. Alguns critérios levantados formam as seguintes filas:

- Uma fila de navios para cada tipo de carga transportada.
- Uma fila única contendo os navios associados à carga mais crítica no momento.
- Uma fila única contendo os navios associados à carga com maior risco futuro de colapso, isto é, com menor tempo de folga até o próximo atendimento necessário (considerando a chegada do próximo navio daquela carga). Na Figura 17 é exemplificado como a folga Δ é calculada. No exemplo, os navios com a carga do tipo C são priorizados.

Figura 17 – Chegada vs. esgotamento por carga (*Days of Supply* - DOS). Para cada carga, compara-se a chegada do próximo navio (ETA_i) com o instante projetado de esgotamento (DOS_k). A folga é $\Delta = \text{ETA}_i - \text{DOS}_k$.



Fonte: Elaborada pelo autor.

Os dois primeiros critérios atuam nos valores absolutos dos estoques para formar as filas. O primeiro critério oferece um poder de decisão maior ao agente. Em contrapartida, exige mais esforço no aprendizado. O último critério adiciona conhecimento empírico, conferindo inteligência maior ao agente.

4.5.2 Priorização de otimização

O problema de alocação de berços é, fundamentalmente, um problema de otimização. Desta forma, é necessário haver critérios na política π que conduzam às melhores soluções ou a soluções boas. A função objetivo do problema, ou as funções

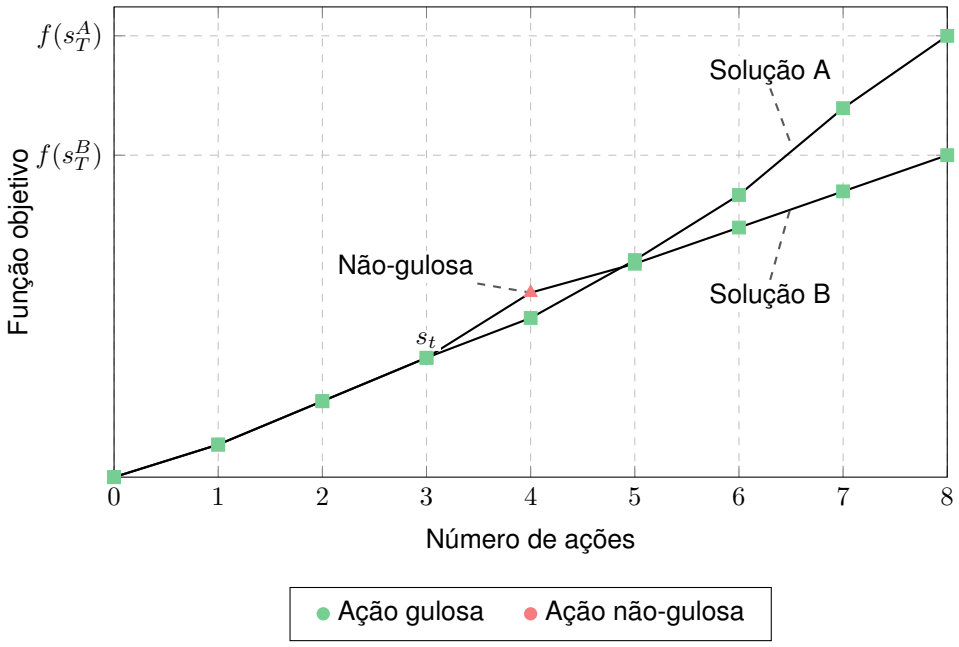
objetivo em um problema multiobjetivo, é determinante para a definição dos critérios. Lembrando que algumas das funções comuns na literatura são:

- Minimização do tempo de serviço (tempo de espera + tempo de atendimento do navio)
- Minimização do *Demurrage* (multa por atrasos no atendimento)
- Minimização do *Makespan* (período total de atendimento de todos os navios)

Os critérios empregam uma estratégia gulosa que levará à escolha do melhor navio no estado. Como a atração da solução para mínimos locais faz parte da natureza das estratégias gulosas, os critérios auxiliares podem desempenhar um papel adicional, juntamente com a política π, como mecanismo de exploração.

Na Figura 18, ilustra-se o comportamento desejado da política π, que aprende a usar um critério auxiliar que leva a um estado local pior, mas obtém uma solução final melhor do que a solução obtida com decisões estritamente gulosas. No estado s_t, a solução B seleciona uma ação não-gulosa que leva ao estado s^B_{t+1} com valor da função objetivo $f(s^B_{t+1})$ inferior a $f(s^A_{t+1})$, mas resulta em uma solução final $f(s^B_T) < f(s^A_T)$ (melhor).

Figura 18 – Duas soluções que evoluem majoritariamente por meio de ações gulosas. No estado s_t (ponto de bifurcação), a solução B executa uma única ação não-gulosa, pior naquele instante, mas conduz à solução final com menor função objetivo.



Fonte: Elaborada pelo autor.

Projetar critérios para fins de fuga de mínimos locais, todavia, é uma tarefa desafiadora. Um critério com uma lógica previsível favorece a generalização do

aprendizado. Técnicas para evitar mínimos locais, como perturbações estocásticas ou mesmo *desvios determinísticos*, introduzem complexidade no processo de aprendizado.

4.5.3 Critérios auxiliares

Como já foi dito, os critérios auxiliares podem assumir dois papéis básicos: (i) critério de apoio à otimização, já discutido na Seção 4.5.2, e (ii) critério padrão para situações de normalidade de estoque. O que se espera nessas situações é que o agente priorize métricas operacionais relacionadas à máxima vazão do atendimento nos berços, que não representam diretamente nem a otimização da função objetivo nem a garantia dos níveis de estoque. Alguns desses critérios padrão são:

- Menor tempo (instante) de conclusão do atendimento; e
- Menor tempo de chegada esperado.

Os critérios auxiliares tornam-se, portanto, imprescindíveis no processo de tomada de decisão, ao proporcionarem ao agente alternativas não gulosas e ao manterem os critérios funcionais mais simples e previsíveis.

4.5.4 Espaço de ações

O papel do agente, nesta proposta, é escolher um dos critérios implementados. Esse critério determina qual navio será alocado ao berço. Para cada berço, o número de critérios a serem implementados, juntamente com um critério adicional de 'não-operação' ou de 'espera', define o conjunto de ações possíveis.

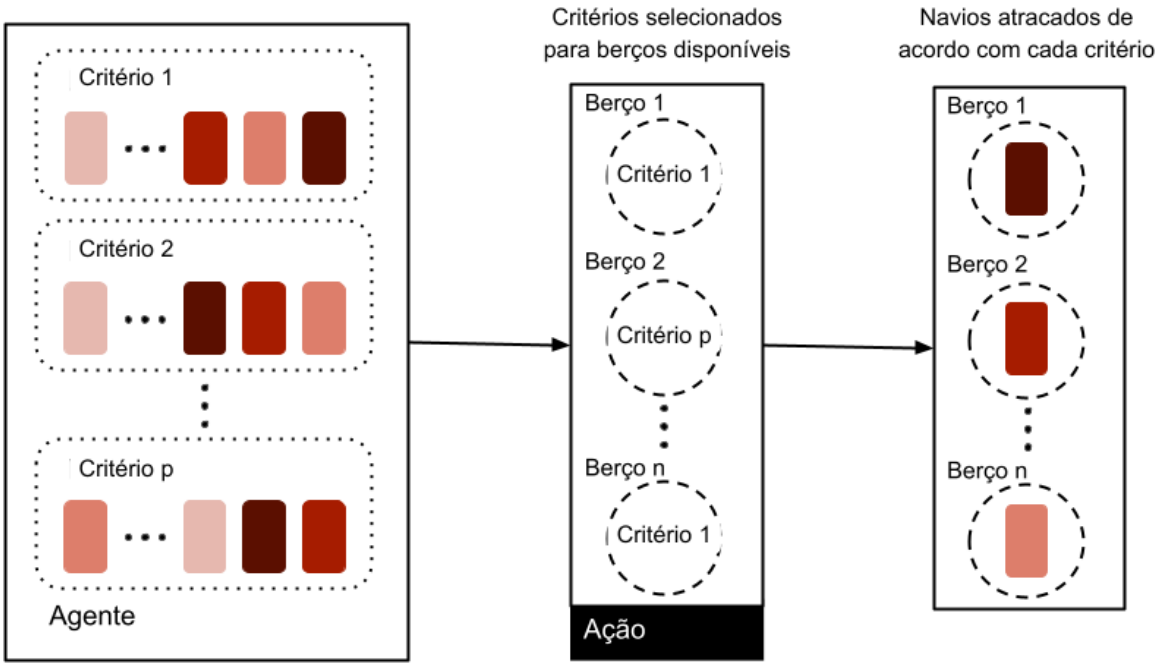
Uma ação tomada é representada na Figura 19. Um critério $p \in \mathcal{P}$ é adotado para cada berço $l \in L$. O mesmo critério pode ser aplicado a mais de um berço na mesma ação. Em caso de ocupação do berço, não é possível atracar nenhum navio; assim, surge a necessidade do critério de 'não-operação'. Dessa forma, o tamanho do espaço de ações pode ser obtido por $|\mathcal{P}|^{|L|}$.

4.5.4.1 Mapeamento de ações

Para simplificar o processo de representação do conjunto de ações, cada ação é mapeada para um número inteiro não negativo, conforme a função F_A, descrita em 4.6. $p_i \in \{0, 1, \ldots, |\mathcal{P}| - 1\}$ é o índice do critério e $\mathbf{p} = (p_1, p_2, \ldots, p_{|L|})$ é o vetor de critérios definidos para cada berço, ou, equivalentemente, uma ação.

$$F_A(\mathbf{p}) = \sum_{l=1}^{|L|} p_i |\mathcal{P}|^{|L|-l} \tag{4.6}$$

Figura 19 – Representação de uma ação tomada.



Fonte: Elaborada pelo autor.

Assim, as ações são codificadas (ou mapeadas) em uma sequência de inteiros no intervalo de 0 a $|\mathcal{P}|^{|L|} - 1$. Na Figura 20, ilustra-se o processo de mapeamento de cada ação como uma sequência de critérios adotados em cada berço, em um número inteiro não negativo, em um cenário com 3 berços e 3 critérios.

Neste esquema, o espaço de ações cresce rapidamente com o tamanho de $\mathcal{P}$ e, principalmente, de L. Como vantagem, mantém a formulação com apenas um agente; portanto, é mais simples. Uma alternativa seria empregar uma formulação multiagente.

4.5.4.2 Ações inválidas

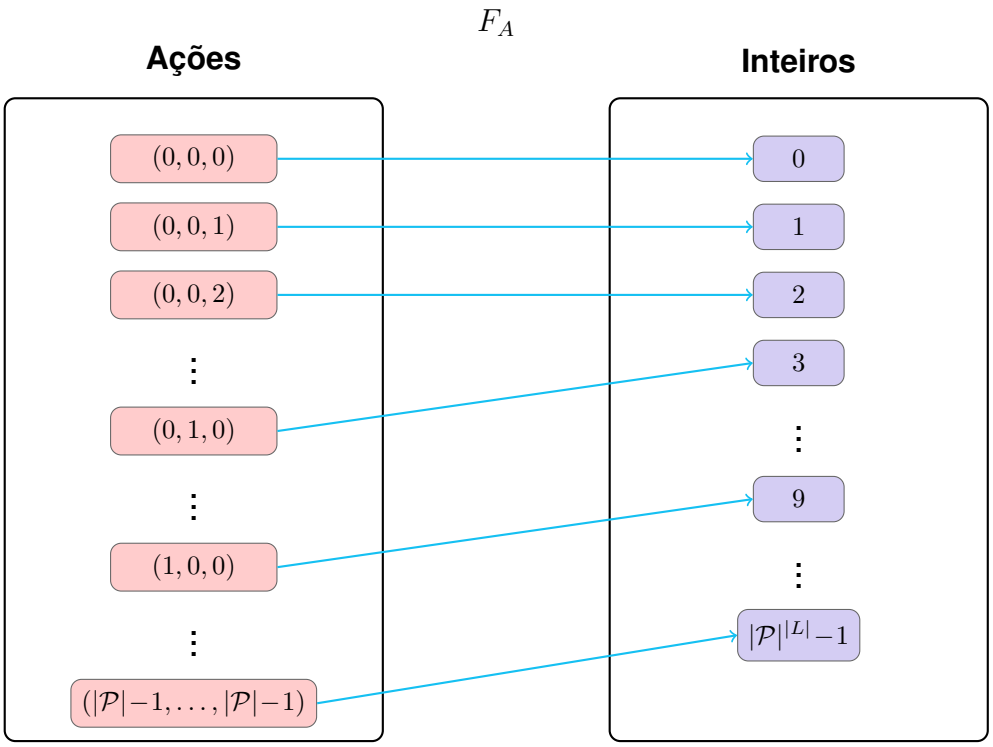
Essa abordagem, com uma ação podendo conter mais de uma atracação, proporciona maior facilidade no processo de aprendizagem; no entanto, é fácil observar que resulta em um subconjunto de ações inválidas. São identificados três tipos de ações inválidas:

Ação vazia: A ação 0, ou $(0, 0, \ldots, 0)$, que não gera nenhuma atracação;

Sobreposição: Ações que resultam na sobreposição de navios, ou seja, a atribuição de um navio a um berço que já está ocupado;

Duplicação: Ações que alocam o mesmo navio a mais de um berço.

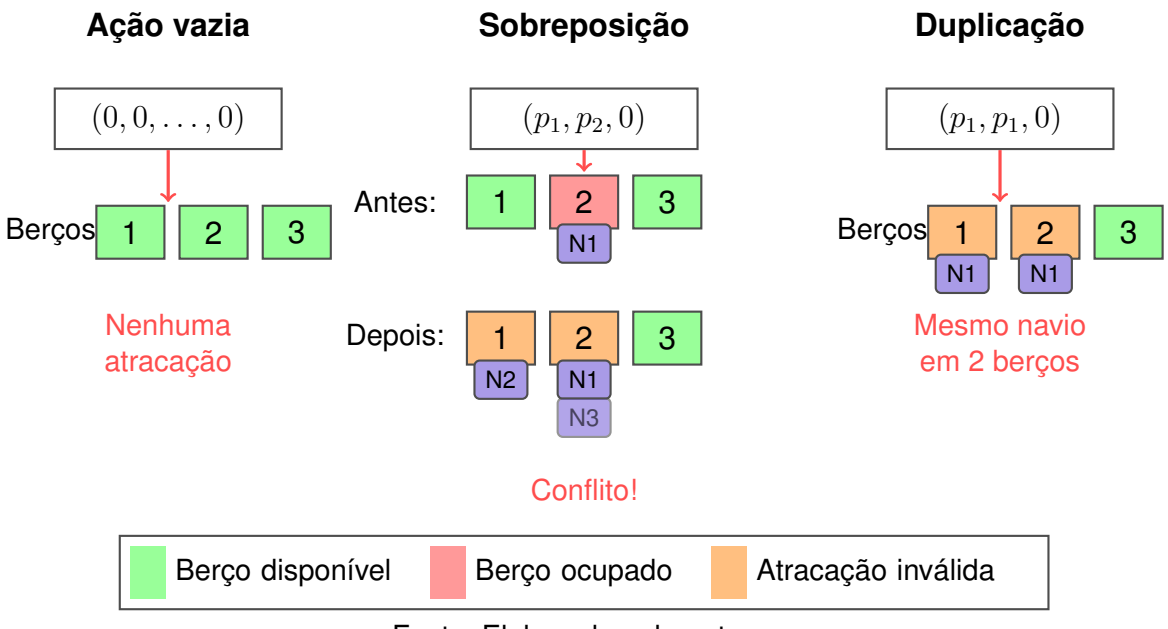
Figura 20 – Mapeamento de ações para inteiros. Cada ação, representada por um vetor de critérios $\mathbf{p} = (p_1, p_2, \ldots, p_{|L|})$, é mapeada para um único valor inteiro no intervalo $[0, |\mathcal{P}|^{|L|} - 1]$ por meio da função F_A (ver Equação 4.6).



Fonte: Elaborada pelo autor.

A duplicação pode ocorrer porque é possível que dois ou mais critérios distintos selecionem o mesmo navio para ser atracado em berços distintos. A Figura 21 apresenta os três tipos de ações inválidas em um cenário com três berços.

Figura 21 – Tipos de ações inválidas no processo de atracação.



Fonte: Elaborada pelo autor.

Para lidar com ações inválidas, é possível aplicar o mecanismo de atribuição de penalidades. Uma grande penalidade é aplicada à recompensa caso ocorra uma ação inválida. Uma forma simples de tratar a duplicação é controlar a seleção dos navios, eliminando das filas os navios já selecionados na mesma ação. Para tratar a ação vazia e a sobreposição, o mecanismo de mascaramento de ações é eficaz. Em cada estado, as ações consideradas inválidas são inviabilizadas, definindo o respectivo *q-value* como um valor muito ruim.

Projetar a função de recompensa é um dos maiores desafios do aprendizado por reforço; ao mesmo tempo, é uma tarefa crucial. O sucesso do aprendizado está condicionado a uma função de recompensa que se alinha à meta do aprendizado e é capaz de avaliar o progresso rumo a ela. Paralelamente a esses objetivos, a função de recompensa deve contornar problemas que comumente surgem, como o da *recompensa esparsa* e o de representar, em um sinal de recompensa, objetivos de alta complexidade (SUTTON; BARTO, 2018).

Neste arcabouço, a meta principal do aprendizado é garantir que os níveis de estoque não sejam ultrapassados. Em problemas de otimização, esse controle de estoque normalmente é tratado como uma restrição. No BAP-RLIM, essa restrição é relaxada, tornando-se um termo parcial ou até mesmo integral da função de recompensa.

As métricas de estoque, vistas na Seção 4.3.2, são candidatas a desempenhar um papel importante na função de recompensa ao fornecer informações sobre a segurança ou criticidade dos níveis de estoque no estado s_t. Além disso, conectam diretamente o estado s_t à função de recompensa, ou seja, à meta de aprendizado.

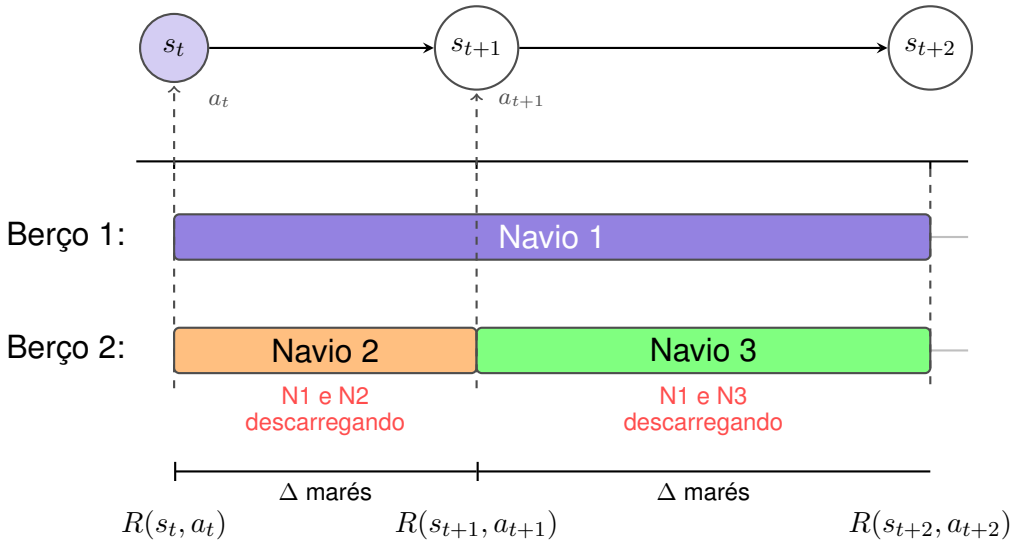
A otimização da função objetivo é a meta secundária no BAP-RLIM, porém fundamental no contexto do PAB. Em comparação com o controle de estoque, é algo desejável. Pelos mesmos motivos, os critérios de otimização são bons candidatos para integrar a função de recompensa.

Na Seção 4.5.4.2, discute-se como as ações inválidas são tratadas. Uma abordagem consiste em aplicar uma penalidade à função de recompensa. Neste caso, um termo específico para ações inválidas também é integrado à função de recompensa. A estrutura geral da recompensa é apresentada na Equação 4.7.

$$R(s_t, a) = w_1 \cdot R_{\text{estoque}}(s_t, a) + w_2 \cdot R_{\text{objetivo}}(s_t, a) + R_{\text{penalidade}}(s_t, a) \tag{4.7}$$

Onde w_1 e w_2 são pesos, R_{estoque} é o termo relacionado ao controle de estoque, R_{objetivo} é o termo baseado na função objetivo e $R_{\text{penalidade}}$ é o termo que impõe penalidade a ações inválidas.

Figura 22 – Problema de atribuição de crédito temporal. A ação a_t tem impactos observados ao longo de estados posteriores enquanto o navio 1 descarrega.



Fonte: Elaborada pelo autor.

A relação $w_1 > w_2$ é introduzida para refletir a prioridade entre a meta primária (controle de estoque) e a secundária (otimização do PAB). Os termos R_{estoque} e R_{objetivo} devem ser normalizados no intervalo $[0, 1]$, permitindo que os pesos w_1 e w_2 controlem efetivamente a importância relativa de cada componente. O termo $R_{\text{penalidade}}$ assume valores fortemente negativos, com o intuito de direcionar o aprendizado para evitar tais ações, ou zero, caso contrário.

Quando a abordagem de mascaramento de ações inválidas é utilizada, $R_{\text{penalidade}} = 0$ para todo par estado-ação. Considerando também que a otimização do PAB é uma meta secundária, a função de recompensa mínima é dada pela Equação 4.8.

$$R(s_t, a) = R_{\text{estoque}}(s_t, a) \tag{4.8}$$

Um grande desafio para o projeto da função de recompensa no BAP-RLIM é o tratamento do problema de *atribuição de crédito temporal* no controle de estoque. Toda a atracação de navios afeta os níveis de estoque; porém, os efeitos nos estoques podem ser observados ao longo de estados posteriores. Afinal, um navio pode ainda estar carregando/descarregando enquanto outros berços são liberados (estados posteriores); ou seja, o navio começa o atendimento no s_t e, nos estados $s_{t+1}, s_{t+2}, \ldots$ ainda está em atendimento. Então, creditar adequadamente a uma ação os efeitos que se estendem por múltiplos estados é uma tarefa que não é trivial. A Figura 22 ilustra este quadro.

4.5.5 Violação de estoque

No BAP-RLIM, o estoque violado é encarado como algo indesejável, porém, a ação que leva a essa situação é considerada válida. Portanto, as violações de estoque devem ser incorporadas à função de recompensa por meio do termo R_{estoque}, que desempenha esse papel, e não em $R_{\text{penalidade}}$.

5 Cenários do arcabouço

Neste capítulo, apresentam-se dois cenários do arcabouço BAP-RLIM, com o propósito de avaliar algumas de suas possibilidades. Os cenários distinguem-se entre si com base nas variações na formulação do arcabouço apresentadas no capítulo 4.

5.1 Cenário 1

Este primeiro cenário é voltado à experimentação do algoritmo DQN na modelagem do agente. O PAB considerado é um cenário de importação em que um navio transporta apenas um tipo de carga. Além disso, um conjunto de critérios de decisão é implementado.

5.1.1 Espaço de estados

Para o espaço de estados, adota-se uma codificação enxuta, alinhada à recompensa. Usam-se os seguintes atributos, organizados por categoria na Tabela 3.

Tabela 3 – Atributos do estado no cenário 1.

Atributo	**Descrição**
a_i	Tempo restante de chegada (ETA) do navio i
h_{il}	Tempo de atendimento do navio i no berço l (estimado)
nk_i	Tipo de carga do navio i
C_{ikl}	Contribuição do navio i ao estoque k se atendido no berço l (Eq. 5.3)
r_l	Tempo de atendimento restante no berço l
bk_l	Tipo de carga atualmente atendida no berço l
c_k	Taxa de consumo/produção do estoque k (unid./tempo)
ξ_k	Medida de *estoque seguro* do estoque k (def. na Eq. 5.2)
t_k	Tempo estimado até colapso do estoque k (se aplicável)

Fonte: Elaborada pelo autor.

Assim, o número de atributos de um estado s_t depende das cardinalidades dos conjuntos L, K e N, sendo determinado pela Equação 5.1, em que B é o conjunto de berços, K é o conjunto de tipos de carga e ϵ é o tamanho do horizonte de planejamento, ou seja, o número de navios presentes no *look-ahead*.

$$\text{Número de atributos} = |B| \cdot 2 + |K| \cdot 3 + \epsilon \cdot 6 \tag{5.1}$$

5.1.1.1 Medida de estoque seguro

Como métrica de estoque, define-se a medida de *estoque seguro*, que visa fornecer uma informação sobre a segurança de cada estoque projetada sob os navios no *look-ahead*. Este indicador baseia-se na quantidade de unidades de tempo (marés) que restam até o estoque colapsar quando o navio mais próximo com o tipo de carga k chegar. A Equação 5.2 mostra como o *estoque seguro* é calculado.

$$\xi_k = \begin{cases} \frac{e_k}{c_k} - \eta_{\text{max}}, & \text{se } \forall \text{ navio } i \text{ no } \textit{look-ahead}, \text{ navio } i \text{ não transporta carga } k \\ \frac{e_k}{c_k} - \eta_k, & \text{caso contrário} \end{cases} \tag{5.2}$$

onde:

- e_k é o nível atual do estoque do tipo k.
- c_k é a taxa de consumo do estoque do tipo k.
- η_k é o tempo de chegada restante do próximo navio que transporta o tipo de carga k.
- η_{max} é o maior tempo de chegada restante entre todos os navios na lista de navios próximos (*look-ahead*).

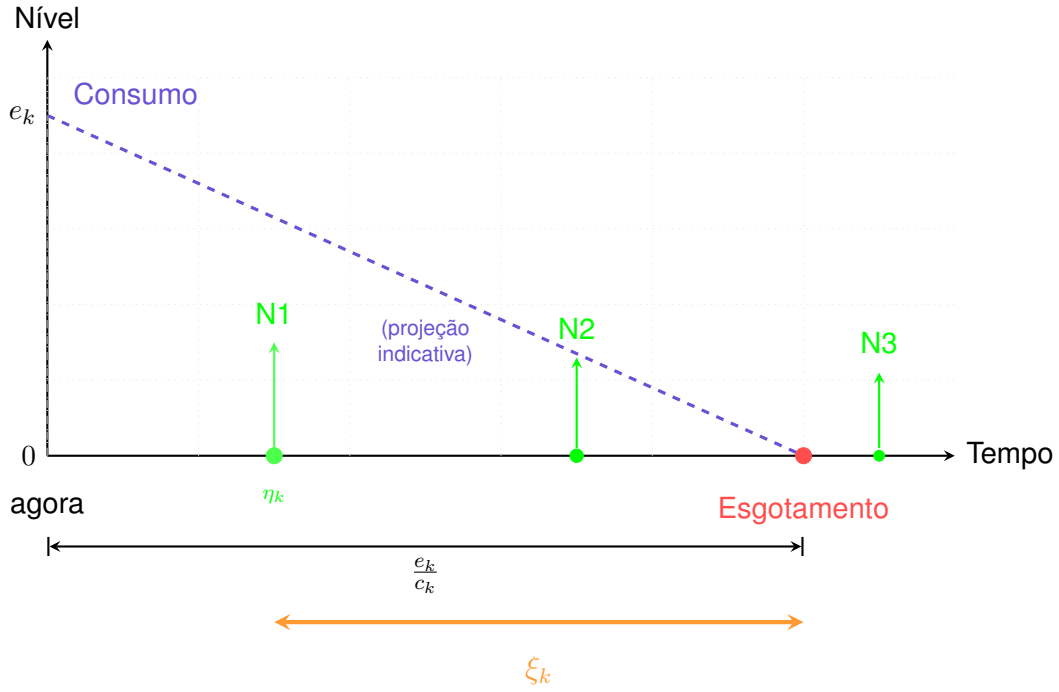
O termo $\frac{e_k}{c_k} - \eta_{\text{max}}$ equivale a considerar o tempo de chegada do último no *look-ahead* como uma projeção para o cálculo do estoque seguro k, tendo em vista que eventualmente pode ocorrer de nenhum navio no *look-ahead* transportar o tipo de carga k. A Figura 23 demonstra esta métrica.

Valores de estoque seguro ξ_k distantes de zero indicam que os estoques estão em uma margem segura, portanto, são desejáveis. Os valores próximos a zero indicam situação de risco alto de colapso e o próximo navio com o tipo de carga k deve ser priorizado. É importante observar que este indicador desconsidera os navios em atendimento, fornecendo, portanto, um indicador aproximado de margem temporal.

5.1.1.2 Medida de contribuição do navio

A *contribuição do navio* é outra medida projetada para fornecer mais informações ao agente e tem por objetivo mensurar a "importância" do navio em relação ao tipo de estoque em cada unidade de tempo. A contribuição C_{ik} é dada pela diferença entre a quantidade de carga descarregada pelo navio i e a quantidade de carga do tipo k consumida pelo porto durante o atendimento em cada um dos berços l e pode ser dada

Figura 23 – Medida de estoque seguro ξ_k. O indicador mede a margem temporal entre o esgotamento projetado e a chegada do próximo navio, fornecendo uma métrica aproximada para orientar decisões conscientes do estoque.



Fonte: Elaborada pelo autor.

pela Equação 5.3. É importante lembrar que cada navio transporta exatamente um tipo de carga.

$$C_{ikl} = q_{ik} - c_k \cdot h_{il} \tag{5.3}$$

5.1.2 Espaço de ações

5.1.2.1 Critérios de seleção de navios

Os critérios de seleção de navios foram escolhidos empiricamente e representam possibilidades de escolha diferentes que devem ser aprendidas, com o objetivo de serem tomadas em situações distintas. Neste cenário, foram utilizados um critério de estoque e um critério auxiliar. Além de um critério adicional de não-operação (ver Seção 4.5), que indica que não haverá atribuição de um navio ao berço. Nenhum critério de otimização foi aplicado.

- Estoque seguro (critério de estoque)
- Tempo de conclusão (critério auxiliar)
- Não-operação.

5.1.2.1.1 Critério de estoque seguro

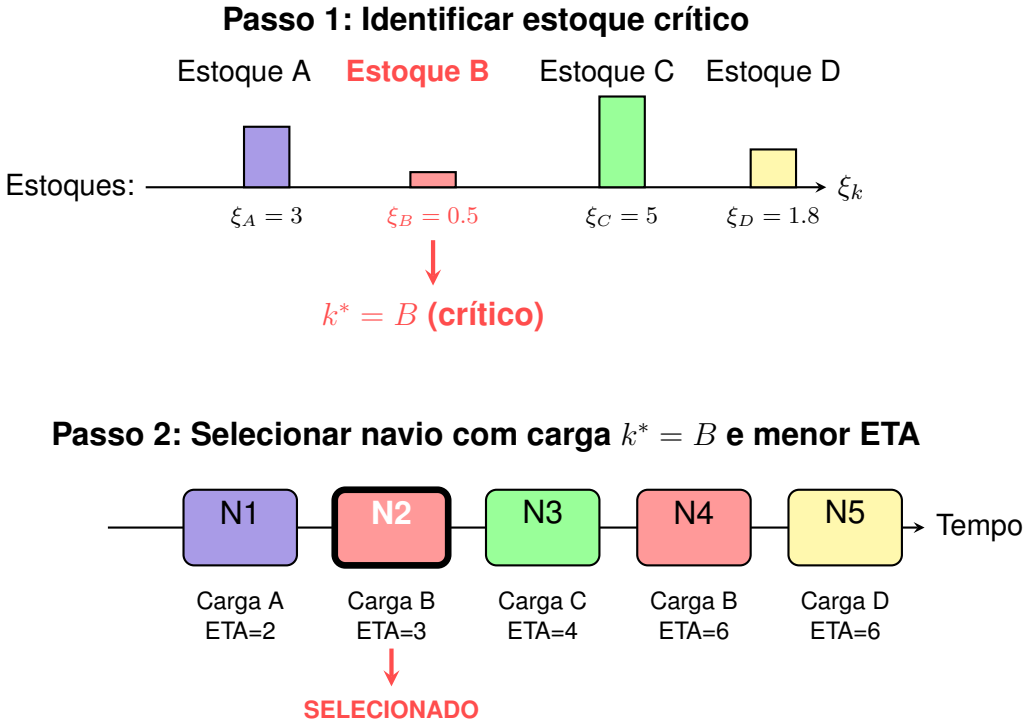
O critério de estoque seguro $p_{\text{estoque}}(C_t)$ (Equação 5.5) baseia-se na medida de estoque seguro apresentada na Seção 5.1.1.1. Esse critério identifica o estoque com menor ξ_k e seleciona o primeiro navio (menor ETA) com a carga k. C_t é o conjunto de navios candidatos, presentes no *look-ahead*. Se não há navios com a carga k em C_t, então o navio com menor ETA é selecionado.

$$k^* = \arg\min_{k \in K} \xi_k \tag{5.4}$$

$$p_{\text{estoque}}(C_t) = \arg \min_{i \in C_t : k_i = k^*} \text{ETA}_i \tag{5.5}$$

Na Figura 24 é ilustrado o mecanismo realizado em dois passos no critério de estoque seguro.

Figura 24 – Critério de seleção por estoque seguro. O processo identifica o estoque mais crítico (menor ξ_k) e seleciona o navio com menor ETA que transporta aquele tipo de carga.



Fonte: Elaborada pelo autor.

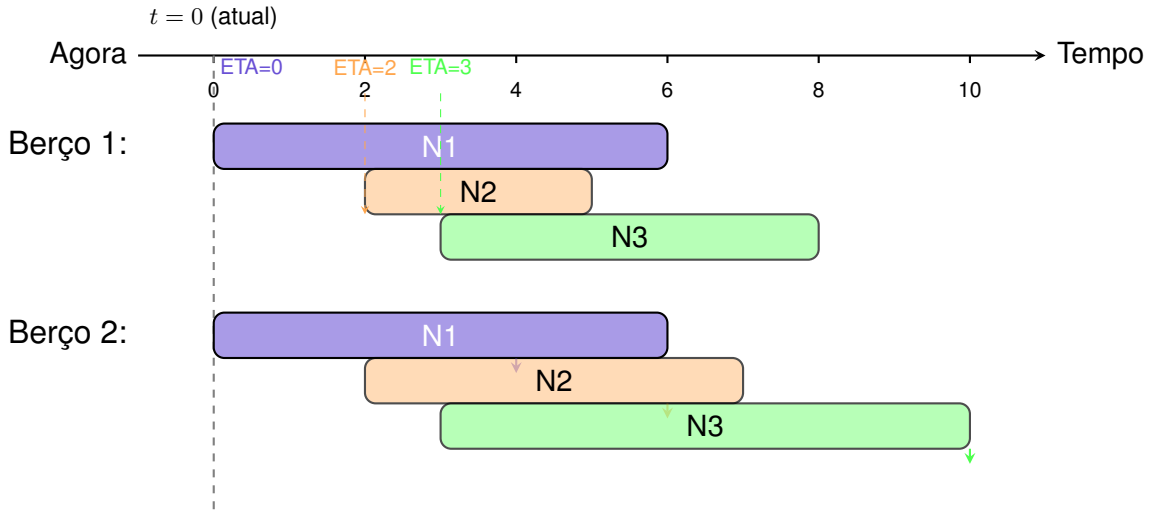
5.1.2.2 Critério de tempo de conclusão

O critério de tempo de conclusão consiste em uma estratégia simples: o navio que, caso seja atendido, desatracará primeiro será o escolhido. É importante notar que, para

cada berço, esse tempo pode variar de um mesmo navio para outro. Portanto, na prática, o número de filas de tempo de conclusão é igual ao número de berços.

A Figura 25 ilustra duas situações distintas em dois berços em que esse critério é aplicado. No berço 1, o navio 1 (N1) tem o tempo de atendimento maior e o tempo de conclusão menor do que o navio 2 (N2). Portanto, o navio 2 deve ser selecionado para o berço 1. No berço 2, o mesmo navio 1 também tem tempo de atendimento maior, porém com tempo de conclusão inferior ao do navio 2. Desta forma, no berço 2, o navio 1 deve ser selecionado.

Figura 25 – Critério de tempo de conclusão. Para cada par (navio, berço), calcula-se quando o navio desatracaria. Seleciona-se o par com menor tempo de conclusão.



Fonte: Elaborada pelo autor.

O atraso do navio 1, visualizado no berço 1, pode, intuitivamente, levar a um prejuízo no tempo de serviço total. Todavia, possibilita que o navio 3, que eventualmente pode ser importante para o estoque, seja atendido antes. Além disso, em uma situação de ETAs iguais, o navio com atendimento mais rápido é escolhido, o que é, indiretamente, uma boa estratégia gulosa para minimizar o tempo de serviço total ou o *makespan*. A Equação 5.6 descreve o critério de tempo de conclusão.

$$p_{\text{tempo}}(C_t, l) = \arg\min_{i \in C_t} (\text{ETA}_i + h_{il}) \tag{5.6}$$

Onde C_t é o conjunto de navios candidatos no estado s_t, l é o berço e $\text{ETA}_i + h_{il}$ é o tempo de conclusão do navio i.

5.1.3 Recompensa

Neste cenário, avaliam-se duas funções de recompensa. A primeira recompensa pode ser dividida em 3 partes: (i) a primeira parte é uma função não linear cujo objetivo é

a manutenção de estoques "saudáveis". Para isso, ela penaliza fortemente ações que levam a estados com níveis de estoque críticos e pioras nos estoques críticos. O critério de *estoque seguro* é a base para esse cálculo e é expresso por meio da Equação 5.10.

Para construir esta primeira parte da função de recompensa, os seguintes passos são adotados. Considere $\xi = [\xi_1, \xi_2, \ldots, \xi_{|K|}]$, onde $|K|$ é o número de estoques, as medidas obtidas pelo critério de estoque seguro. O erro diferencial para a medida de estoque seguro é dado por $\Delta\xi_k = \{\xi'_1 - \xi_1, \xi'_2 - \xi_2, \ldots, \xi'_{|K|} - \xi_{|K|}\}$.

O estoque seguro xi_k é potencializado comparativamente aos outros estoques, dando maior força a níveis mais seguros. O estoque seguro potencializado, $\hat{\xi}_k$, é calculado como:

$$\hat{\xi}_k = \begin{cases} \delta_1 \cdot \xi'_i, & \text{se } \xi_i < 1 \\ \delta_2 \cdot \xi'_i, & \text{se } 1 \leq \xi_i < 3 \\ \delta_3 \cdot \xi'_i, & \text{se } 3 \leq \xi_i < 5 \\ \xi'_i, & \text{caso contrário} \end{cases} \tag{5.7}$$

onde $0 < \delta_1 < \delta_2 < \delta_3 < 1$.

Dois termos são calculados para compor o valor final desta primeira parte:

$$E_{\text{total}} = \sum_{k=1}^{|K|} \frac{1}{\max(\hat{\xi}_k, 0) + 1} \tag{5.8}$$

$$E_{\Delta} = \sum_{k=1}^{|K|} \frac{1}{\max(\hat{\xi}_k, 0) + 1} \quad \text{se } \Delta\xi_k < 0 \tag{5.9}$$

O termo final E é dado por:

$$E = 1 - \min(\alpha \cdot E_{\text{total}} + \beta \cdot E_{\Delta}, E_{\text{max}}) \tag{5.10}$$

onde α, β e E_{max} devem ser ajustados com $\alpha < \beta$ e $0 < E_{\text{max}} < 1$ é utilizado apenas para evitar saturação de E.

A segunda parte da função de recompensa penaliza a escolha de navios que ainda não chegaram. Este tipo de escolha pode não ser muito intuitiva, mas, para garantir os níveis de estoque, às vezes pode ser necessário que nenhum navio entre e, consequentemente, evitar a postergação do atendimento a navios prioritários. O termo de penalidade, D, por tempo ocioso, Δt_{ocioso}, é dado pela Equação 5.11. A constante ϕ deve ser ajustada a um valor próximo de zero e $\phi < 0$, tendo em vista que este sub-objetivo

não pode sobrepor o objetivo principal da função de recompensa:

$$D = \phi \cdot \Delta t_{\text{ocioso}} \tag{5.11}$$

Finalmente, é aplicada uma penalidade caso um dos níveis de estoque colapse. Este valor deve ser ajustado, mas, neste cenário, foi arbitrado em -1. Desta forma, a função de recompensa R é dada pela Equação 5.12:

$$R = \begin{cases} -1 & \text{, caso haja falha de estoque} \\ \max(E + D, 0) & \text{, caso contrário} \end{cases} \tag{5.12}$$

5.2 Cenário 2

Este cenário também se aplica a importação. Diferentemente do primeiro cenário, os navios podem carregar todos os tipos de carga de uma só vez. Um caso mais geral em portos graneleiros. Este cenário é modelado para avaliar a aplicação da rede neural recorrente LSTM. Novos critérios de decisão e de recompensa são aplicados, e os resultados são comparados aos obtidos com um *solver* comercial.

5.2.1 Espaço de estados

O espaço de estados desse segundo cenário é similar ao do primeiro, variando apenas alguns atributos utilizados, conforme a Tabela 4. Não é utilizada nenhuma métrica específica de estoque. As informações mais diretas sobre o esgotamento do estoque estão presentes nos atributos e_k e t_k.

Tabela 4 – Configuração do espaço de estados no cenário 2.

Atributo	**Descrição**
a_i	Tempo restante até a chegada (ETA) do navio i
h_{il}	Tempo de atendimento do navio i no berço l
q_{ik}	Quantidade do tipo de carga k no navio i
r_l	Tempo de atendimento restante no berço l
w_l	Tempo restante que o berço l deve aguardar até o navio selecionado atracar
μ_l	Vazão (taxa de atendimento) do berço l
c_k	Taxa de consumo do tipo de carga k
e_k	Nível atual do estoque do tipo de carga k
t_k	Tempo estimado até o esgotamento do estoque k

Fonte: Elaborada pelo autor.

5.2.2 Espaço de ações

5.2.2.1 Critérios de seleção de navios

Neste cenário, foram utilizados um critério de estoque, um critério de otimização e um critério auxiliar. Além de um critério adicional de não-operação (ver Seção 4.5).

- Urgência de estoque (critério de estoque)
- Tempo de serviço (critério de otimização)
- Tempo de conclusão (critério auxiliar)
- Não-operação.

5.2.2.2 Critério de urgência de estoque

O critério de urgência de estoque é concebido para cenários em que navios transportam mais de uma carga. Um navio que carrega quantidades maiores de carga com estoque em uma *região crítica* (ou próxima a ela) tem prioridade em relação aos demais navios. Esse critério calcula um escore para cada navio i com base na urgência projetada dos estoques que ele transporta. O estoque projetado $e_k^{\text{proj}(i)}$ e a *urgência* $w_k(i)$, que atribui um peso a cada carga k ao navio i, são obtidos pelas equações 5.13 e 5.14.

$$e_k^{\text{proj}}(i) = e_k - c_k \cdot a_i \tag{5.13}$$

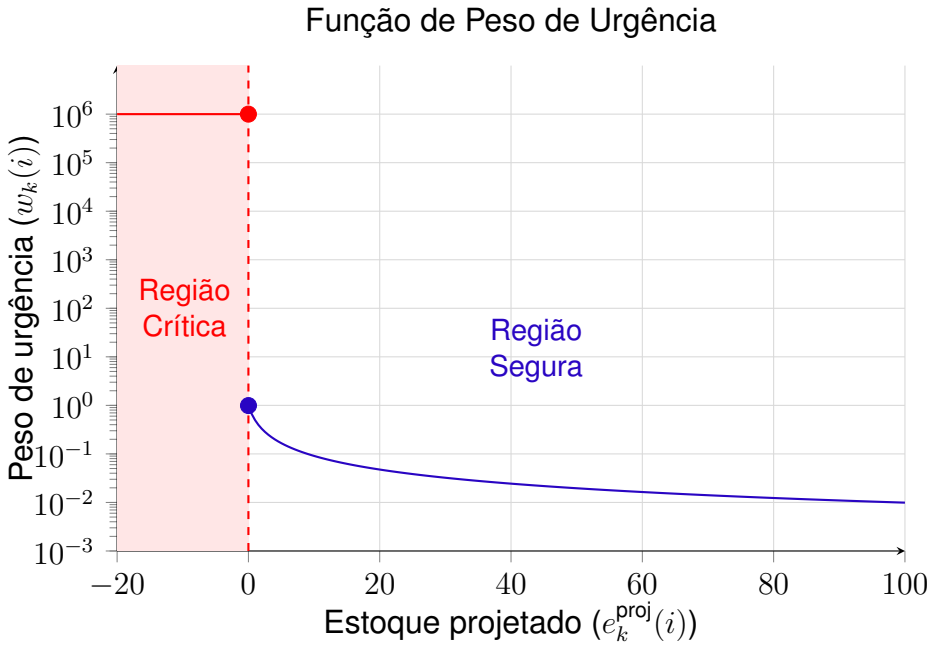
$$w_k(i) = \begin{cases} 10^6, & \text{se } e_k^{\text{proj}}(i) \leq 0 \\ \frac{1}{e_k^{\text{proj}}(i)+1}, & \text{caso contrário} \end{cases} \tag{5.14}$$

$$\tag{5.15}$$

A urgência é inversamente proporcional ao nível de estoque projetado: estoques críticos (próximos de zero ou abaixo de zero) recebem pesos muito altos, enquanto estoques seguros recebem pesos muito baixos. Na Figura 26, ilustra-se o comportamento de $w_k(i)$. A constante 10^6 é arbitrária e deve ser definida como um número muito grande.

Os navios são então ordenados decrescentemente por escore, priorizando aqueles que transportam cargas mais urgentes. O escore S_i é a soma das cargas q_{ik} ponderada pela respectiva urgência w_k. O navio com maior S_i é, portanto, selecionado, conforme as equações 5.16 e 5.17.

Figura 26 – Função de peso de urgência $w_k(i)$ em função do estoque projetado $e_k^{\text{proj}}(i)$.



Fonte: Elaborada pelo autor.

$$S_i = \sum_{k \in K} q_{ik} \cdot w_k(i) \tag{5.16}$$

$$p_{\text{urgência}}(C_t) = \arg\max_{i \in C_t} S_i \tag{5.17}$$

5.2.2.3 Critério de tempo de serviço

Como critério de otimização, utiliza-se, neste cenário, o tempo de serviço. Este objetivo é selecionado porque *o tempo de serviço total* (soma dos tempos de serviço de todos os navios) é um dos objetivos mais comuns na literatura sobre o problema de alocação de berços e porque é o mesmo utilizado no modelo matemático apresentado na Seção 2.1.2. Para a obtenção do tempo de serviço do navio, somam-se o tempo de espera e o tempo de atendimento no berço atracado e o navio com menor tempo de serviço é selecionado, conforme as equações 5.18 e 5.19.

$$t_{\text{serv}}(i, l) = \max\{0, t - \text{ETA}_i\} + h_{il} - 1 \tag{5.18}$$

$$p_{\text{serviço}}(C_t, l) = \arg\min_{i \in C_t} t_{\text{serv}}(i, l) \tag{5.19}$$

onde:

- ETA_i é o tempo de chegada do navio i;

- t é o instante atual de decisão;
- $t_{\text{serv}}(i, l)$ é o tempo de serviço total do navio i no berço l;
- h_{il} é o tempo de atendimento do navio i no berço l;
- C_t é o conjunto de navios candidatos no instante t;
- $p_{\text{serviço}}(C_t, l)$ retorna o navio selecionado para o berço l.

5.2.2.4 Critério de tempo de conclusão

Esse critério é uma estratégia em que o agente escolhe o navio que, se atendido, desatracará primeiro. Este mesmo critério já foi utilizado no cenário 1 e mais detalhes são discutidos na Seção 5.1.2.2.

5.2.3 Recompensa

Uma função de recompensa foi desenvolvida para avaliar o BAP-RLIM com a estrutura DQN+LSTM. Essa função de recompensa visa premiar ações com menor tempo de serviço médio e penalizar ações com níveis de estoque críticos. É possível,portanto, dividir essa recompensa em dois componentes: (i) o tempo de serviço médio dos navios selecionados na ação e (ii) uma função não linear que mede a criticidade dos níveis de estoque.

O primeiro componente é normalizado por uma função sigmoide, conforme definido na Equação 5.20, em que $\bar{T}$ é o tempo médio de serviço dos navios alocados na ação corrente, calculado como $\bar{T} = \frac{1}{|A_t|} \sum_{i \in A_t} t_{\text{serv}}(i)$, onde A_t é o conjunto de navios selecionados, r é o ponto de inflexão e η é a taxa de crescimento.

$$\mathcal{T} = \frac{1}{1 + e^{-\eta(\bar{T} - r)}} \tag{5.20}$$

O segundo componente, a *criticidade do estoque*, $\mathcal{C}_{\text{max}}$, depende dos níveis atuais de estoque, $e_k(t)$, e da variação negativa em relação à ação anterior, $\Delta_k(t) = e_k(t) - e_k(t-1)$. Para cada tipo de carga k, calculam-se duas componentes de criticidade: a criticidade de nível, baseada no estoque atual (Equação 5.21), e a criticidade de variação, quando há queda no estoque (Equação 5.22).

$$\mathcal{C}_{\text{nível}} = \frac{1}{1 + e_k(t)} \tag{5.21}$$

$$\mathcal{C}_{\text{variação}} = \frac{1}{1 + e^{R \cdot \left(\frac{\Delta_k(t)}{-\Delta_k(t) + \kappa} + \xi \right)}} \cdot \mathcal{C}_{\text{nível}} \tag{5.22}$$

em que R, κ e ξ são parâmetros de $\mathcal{C}_{\text{variação}}$ que ajustam, respectivamente, a inclinação, a suavidade e o deslocamento da curva.

A criticidade combinada para cada item k é:

$$\mathcal{C}_k = \alpha_{\text{nível}} \cdot \mathcal{C}_{\text{level},k} + \beta_{\text{variação}} \cdot \mathcal{C}_{\text{variação},k}$$

A máxima criticidade entre os tipos de carga é então dada por:

$$\mathcal{C}_{\text{max}} = \max_{k \in K} \mathcal{C}_k$$

Finalmente, a função de recompensa global R é definida na Equação 5.23:

$$R = 1 - \alpha \cdot \mathcal{T} - \beta \cdot \mathcal{C}_{\text{max}} \tag{5.23}$$

onde $\alpha_{\text{nível}} + \beta_{\text{variação}} = 1$ e $\alpha + \beta = 1$.

6 Metodologia

6.1 Ambiente Virtual

Foi desenvolvido um ambiente virtual com o simulador, conforme descrito no Capítulo 4, em Python, utilizando a biblioteca *TF-Agents* de aprendizado por reforço, baseada em *TensorFlow* (GUADARRAMA et al., 2018). Foi criado um ambiente customizado, implementado a partir da interface *PyEnvironment*, adequado ao treinamento de agentes com algoritmos de aprendizado por reforço disponíveis na biblioteca *TF-Agents*. Uma versão do ambiente está disponível em (BARROS, 2026a).

6.1.1 Simulador

O simulador desempenha um papel crucial na dinâmica do sistema. Ele realiza um conjunto de tarefas até chegar à condição de solicitar uma nova decisão de atracação. Nesse processo, um conjunto de variáveis do ambiente é atualizado e retornado ao ambiente. O ambiente, então, extrai as informações correspondentes ao novo estado atual e calcula a recompensa antes de enviá-las ao agente, aguardando a próxima ação. A Tabela 5 relaciona as principais tarefas realizadas pelo simulador do BAP-RLIM.

6.1.1.1 Configurações

O simulador permite a configuração de parâmetros, de modo a tornar possível a realização da simulação para qualquer instância, com base no problema apresentado. Entre os principais parâmetros estão: navios, berços e estoques, além de parâmetros gerais. Os navios, o *look-ahead*, os berços e os estoques são os principais componentes do simulador, pois determinam as dimensões de cada instância do problema. Na Tabela 6, são listados os parâmetros de configuração do simulador.

6.1.1.1.1 Navios

Cada navio contém dados intrínsecos, como ETA, carga e tempo de atendimento em cada berço. A lista de navios presentes na simulação é predefinida e organizada em navios atracados e não atracados. A vazão é o atributo a ser configurado para cada berço, todavia, este valor é transformado em tempos de atendimento para cada navio, calculado previamente conforme a Equação 6.1.

Tabela 5 – Principais tarefas realizadas pelo Simulador do BAP-RLIM

Tarefa	**Descrição**
Inicialização e Controle de Episódios	
Reiniciar episódio	Reseta o simulador para um novo episódio com nova instância do problema (navios, berços, estoques) e zera contadores de desempenho
Atualização de Estado	
Processar atracações	Valida lista de atracações solicitadas (disponibilidade de berços, duplicação de navios), executa atracações válidas e calcula recompensa total do *step*
Avançar tempo	Avança o tempo discreto (TTW) até que pelo menos um berço fique disponível, atualizando todas as variáveis do sistema
Atualizar níveis de estoque	Atualiza níveis dos estoques considerando taxas de consumo do estoque e descarregamento simultâneo de navios atracados
Geração de Observações	
Gerar observação do estado	Constrói vetor de observação
Atualizar janela *look-ahead*	Mantém lista ordenada dos próximos ϵ navios disponíveis para alocação
Validação e Verificação	
Verificar colapso de estoque	Identifica violações de restrições quando nível de estoque atinge zero, sinalizando término prematuro do episódio
Cálculo de Desempenho	
Calcular tempo de serviço	Computa tempo total de serviço de cada navio

Fonte: Elaborada pelo autor.

6.1.1.1.2 Berços

Os berços contêm as variáveis necessárias para gerenciar a atracação dos navios, como o tempo restante até que o berço seja liberado.

6.1.1.1.3 Estoques

Os estoques, além dos dados de consumo de carga, contêm informações sobre o nível de estoque: o nível atual e o nível mínimo tolerável.

$$h_{il} = \left\lceil \frac{q_{ik}}{v_l} \right\rceil \tag{6.1}$$

onde:

- h_{il} é o tempo de atendimento do navio i no berço l.
- q_{ik} é a quantidade de carga do navio i do tipo k.
- v_l é a vazão do berço l.

Tabela 6 – Parâmetros de configuração do simulador: navios e estoques

Parâmetro	**Descrição**
`id`	Identificador do navio
`arrival_time`	Tempo de chegada esperado (ETA) do navio
`handling_times`	Lista de tempos de atendimento do navio em cada berço
`cargo_type`	Tipo de carga que o navio está transportando
`cargo_quantity`	Quantidade de carga no navio
`level`	Nível atual de estoque
`consumption_rate`	Taxa de consumo do estoque
`minimum_level`	Nível mínimo permitido de estoque

Fonte: Elaborada pelo autor.

6.1.1.1.4 Parâmetros gerais

Os parâmetros gerais são várias configurações essenciais que definem o funcionamento básico do simulador, entre elas: penalidades por ações inválidas, o número máximo de navios e o tamanho do *look-ahead*.

6.1.1.2 Atualizações

O funcionamento do simulador pode ser dividido em duas etapas principais: inicialização e atualização. A inicialização consiste na instanciação do simulador e na configuração dos parâmetros descritos na Seção 6.1.1.1.

A etapa de atualização constitui o núcleo do simulador, responsável por processar as ações do agente e determinar o próximo estado do sistema. Esta etapa recebe como entrada uma ação do agente, representada por uma lista de atracações (pares navio-berço), e executa o seguinte fluxo de operações, ilustrado na Figura 27:

1. **Recepção da ação**: O simulador recebe a decisão de atracação proposta pelo agente, ou seja, uma lista de pares navio-berço;
2. **Validação da ação**: Verifica-se se a ação proposta é válida em relação às restrições operacionais. Caso haja restrições violadas, aplicam-se as penalidades e as variáveis de ambiente são retornadas;
3. **Atualização dos navios**: Os navios selecionados para atracação são associados aos berços designados. Cada navio tem seu estado atualizado para refletir o início do atendimento. O tempo total que cada navio permanecerá no sistema é registrado, considerando tanto o tempo de manuseio da carga quanto eventuais tempos de espera;
4. **Atualização dos berços**: Registram-se os navios nos respectivos berços e atualizam-se os tempos restantes de atendimento em cada berço;

5. **Cálculo de recompensas**: Calcula-se a métrica de desempenho principal com base nos tempos de serviço dos navios atracados;

6. **Avanço temporal**: O tempo do simulador avança incrementalmente até que pelo menos um berço fique disponível para novas atracações. Durante este avanço, a cada unidade de tempo, três atualizações são realizadas: os tempos restantes para a chegada dos navios são decrementados; os tempos restantes de operação nos berços ocupados são reduzidos; e os níveis de estoque são recalculados, considerando as taxas de descarregamento dos navios em operação e as taxas de consumo de cada estoque;

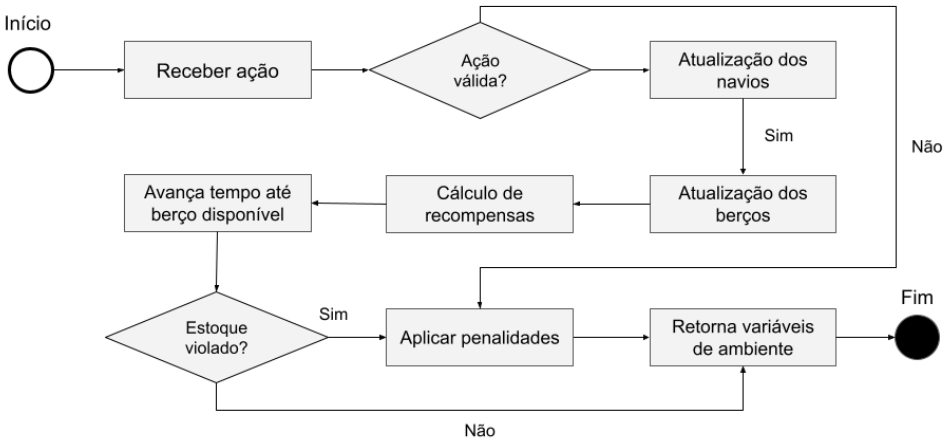
7. **Verificação de estoque**: Identificam-se violações nos níveis de estoque que possam ter ocorrido. Se houver violações, penalidades são aplicadas e as variáveis de ambiente são retornadas;

8. **Aplicação de penalidades**: Caso haja violações, as penalidades correspondentes são aplicadas;

9. **Retorno ao ambiente**: O novo estado e as métricas são retornados, encerrando o ciclo.

Figura 27 – Fluxograma do processo de atualização no simulador.

Fonte: Elaborada pelo autor.



Fonte: Elaborada pelo autor.

Este ciclo se repete a cada decisão do agente, permitindo a simulação contínua das operações portuárias ao longo do horizonte de planejamento.

6.1.1.3 Violações de estoque

Um dos processos mais importantes dentro do simulador é a verificação de violações nos níveis de estoque. O objetivo é fazer a verificação descrita por meio da Equação 2.5, ou seja, não permitir que os níveis de estoque das cargas armazenadas no

porto decaiam para abaixo do mínimo tolerável. Esta verificação é feita passo a passo ao longo do avanço do tempo do simulador. A atualização dos níveis de estoque é feita a cada unidade de tempo, conforme a Equação 6.2.

$$e_k(t+1) = \max(e_k(t) + D_k(t) - c_k, 0) \tag{6.2}$$

onde:

- $e_k(t)$ é o nível de estoque atual k no tempo t;
- $D_k(t)$ é a taxa de descarregamento de carga k entre todos os navios em atendimento no tempo t;
- c_k é a taxa de consumo do estoque k.

Caso $e_k(t) \leq e_{\text{min},k}$ para qualquer t, em que $e_{\text{min},k}$ é o nível mínimo de estoque permitido para a carga k, configura-se uma violação de estoque.

6.2 Instâncias

Os dados utilizados nos experimentos consistem em um conjunto de instâncias empregadas no treinamento do agente e nos testes de desempenho da formulação proposta. Essas instâncias são baseadas no formato do modelo matemático apresentado na Seção 2.1.2 e proposto por Silva (2021).

Neste trabalho, foram usadas apenas instâncias cujas soluções viáveis são conhecidas, obtidas por *solver* de programação linear, o que permite uma melhor avaliação da qualidade das soluções obtidas pela abordagem proposta neste trabalho. Um exemplo dessas instâncias é apresentado na Tabela 7, extraída de Silva (2021).

No treinamento, uma instância é selecionada como instância de referência, a partir da qual serão geradas dinamicamente instâncias para o processo de treinamento. Outras instâncias são usadas para comparar os resultados obtidos com as políticas aplicadas.

6.3 DQN

Para a derivação de soluções a partir da formulação proposta, foi utilizado o algoritmo DQN (Deep Q-Network) como estratégia de aprendizado por reforço profundo. Uma introdução ao DQN pode ser vista na Seção 2.2.2. Este algoritmo foi escolhido por ter apresentado resultados promissores recentemente em problemas distintos de tomada de decisão sequencial. No Capítulo 3, uma amostra de trabalhos de otimização na área de operações portuárias é apresentada. A DQN foi utilizada por meio da biblioteca TF-Agents do software Tensorflow (ABADI et al., 2016).

Tabela 7 – Exemplo de instância a partir dos conjuntos N, M, K, L que definem o tamanho do problema

set $N :=$ 1 2 3 4 5 6 7 8 9 10;
set $M :=$ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 $\cdots$ 35;
set $K :=$ rawMatter1 rawMatter2 rawMatter3 rawMatter4;
set $L :=$ 1 2 3;
param $v :=$
1 5
2 4
3 2;
param $a :=$
1 1
2 2
$\vdots$
10 11;
param $e :=$
rawMatter1 49
$\vdots$
rawMatter4 57;
param $ck :=$
rawMatter1 3
$\vdots$
rawMatter4 2;
param q : rawMatter1 rawMatter2 rawMatter3 rawMatter4 :=
1 6 0 0 0
$\vdots$
10 10 0 0 0;

Fonte: Adaptado de (SILVA, 2021).

6.3.1 Rede Neural

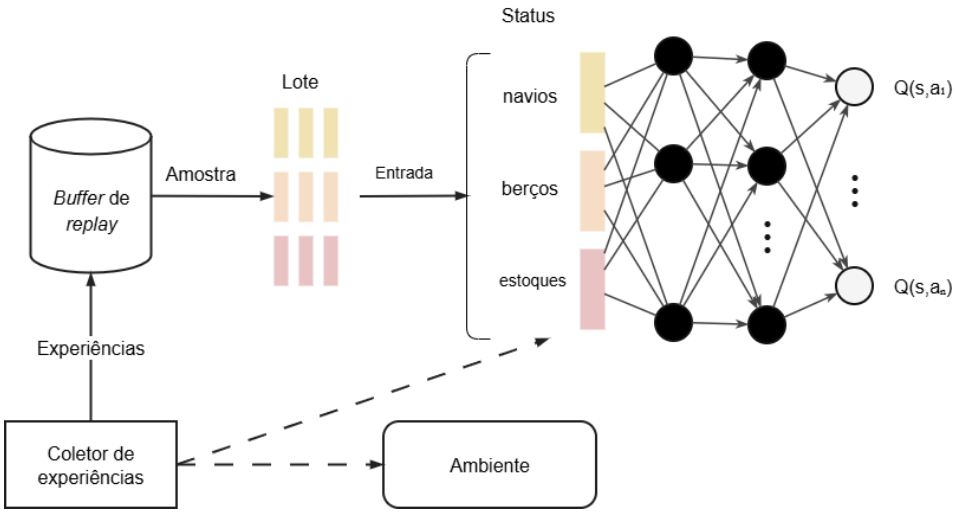
O modelo de rede neural utilizado possui uma arquitetura sequencial. A camada de entrada contém o número de atributos de um estado e pode ser calculado pela Equação 5.1. São três camadas ocultas e densas, com 64 nós em cada, e função de ativação ReLU. Por fim, tem-se uma camada de saída densa com o número de unidades equivalente ao número de ações possíveis na formulação do BAP-RLIM.

A Figura 28 ilustra o esquema básico do DQN utilizado no BAP-RLIM. O *buffer* de *replay* contém as experiências coletadas ao longo do treinamento e dele são amostradas as entradas, em lotes de tamanho 64. Esses lotes são submetidos à rede neural em cada passo (*step*). Cada entrada é composta pelas atributos de um estado, ou seja, pelas informações de status dos navios, dos berços e dos estoques, conforme descrito na Seção 5.1.1.

6.3.2 LSTM

A arquitetura de rede neural Long Short-Term Memory (LSTM) é um tipo de rede neural recorrente (RNN) projetada para modelar dependências de longo prazo em

Figura 28 – Esquema básico da DQN integrada ao BAP-RLIM.



Fonte: Elaborada pelo autor.

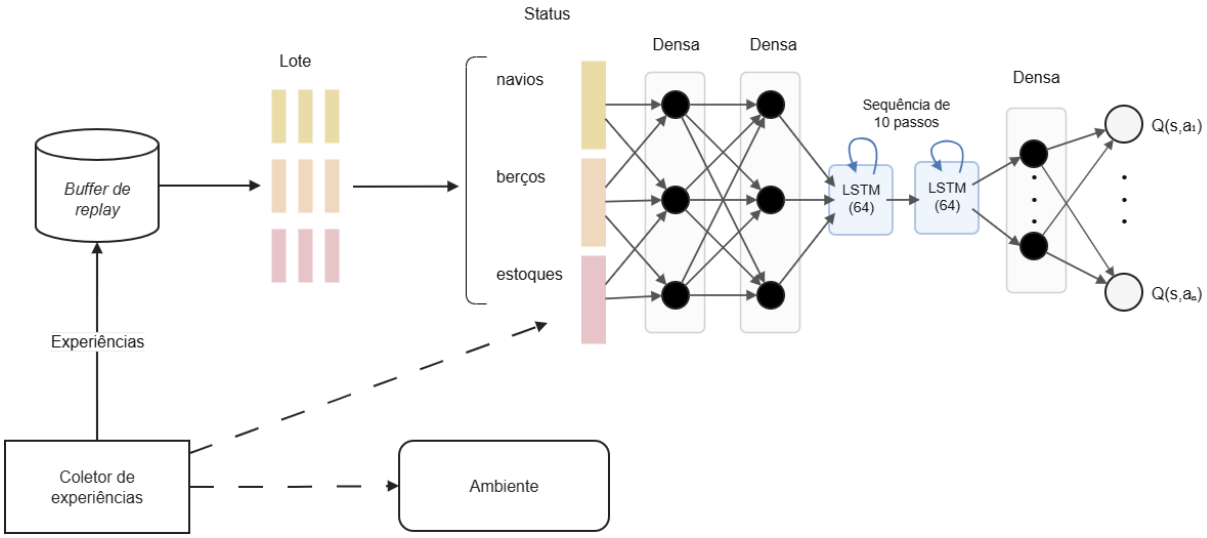
dados sequenciais. Ela utiliza portas de entrada, de esquecimento e de saída que regulam o fluxo de informação ao longo do tempo. Esse mecanismo pode preservar ou esquecer, de forma seletiva, informações por longos intervalos, diferentemente das redes recorrentes tradicionais, e, assim, evita o problema do gradiente que se desvanece em tarefas sequenciais (HOCHREITER; SCHMIDHUBER, 1997) (GERS; SCHMIDHUBER; CUMMINS, 2000).

6.3.3 Arquitetura DQN + LSTM

Para aproveitar a capacidade da LSTM de processar sequências, emprega-se, no cenário 2, um esquema que a integra ao framework DQN, baseado na abordagem proposta por (HAUSKNECHT; STONE, 2015) e implementado por meio da *LSTMEncodingNetwork* da biblioteca TF-Agents (GUADARRAMA et al., 2018). A Figura 29 ilustra o esquema básico do DQN com LSTM utilizado no BAP-RLIM. O *replay buffer* armazena as experiências coletadas durante o treinamento como fragmentos de episódios de 10 passos. A partir desse buffer, são selecionados lotes (*batches*) de 32 amostras como entradas. Esses lotes são processados pela LSTM a cada passo. Cada entrada é composta pelas informações de estado dos navios, berços e estoques, conforme descrito na Seção 5.2.1.

Em cada passo de tempo, o vetor de observação é processado por duas camadas totalmente conectadas com 75 e 40 unidades, respectivamente, ambas utilizando a função de ativação tangente hiperbólica. A saída é então passada por uma LSTM empilhada com duas camadas de 64 unidades cada, seguida de uma camada totalmente conectada com ativação tangente hiperbólica. Uma camada totalmente conectada final produz os valores Q, de dimensionalidade igual ao número de ações disponíveis.

Figura 29 – Esquema básico do DQN+LSTM integrado ao BAP-RLIM.



Fonte: Elaborada pelo autor.

7 Resultados

7.1 *Softwares* e *Hardware*

Os experimentos foram realizados utilizando as bibliotecas *Tensorflow* (versão 2.15.0) e *TF-Agents* (versão 0.18.0) em ambiente *Python* (versão 3.15.11), executados em um computador com sistema operacional *Microsoft Windows 10*. O hardware utilizado foi um processador *Intel Core i7* de 4 núcleos com 8 GB de memória principal.

7.2 Cenário 1

7.2.1 Instâncias de treinamento

As instâncias de treinamento utilizadas nos experimentos computacionais do *cenário 1* são geradas dinamicamente com base em um cenário específico do problema de alocação de berços descrito na Seção 2.1.2, com estrutura exemplificada na Seção 6.2. Este cenário corresponde a uma *instância de referência* com 30 navios e 2 berços, resumida na Tabela 8. No entanto, os atributos que definem a estrutura do porto considerado são o número de berços, o número de tipos de carga e a vazão dos berços. Cada instância gerada contém 90 navios.

Tabela 8 – Resumo da instância de referência

Descrição	**Valores**
Número de navios	30
Número de berços	2
Número de tipos de carga	4
Vazão dos berços	{1: 5, 2: 7}
Níveis de estoque iniciais	{1: 30, 2: 39, 3: 61, 4: 21}

Fonte: Elaborada pelo autor.

Para gerar os atributos tempos de chegada, níveis de estoque iniciais e quantidades de carga transportada pelos navios, foram utilizadas distribuições discretas uniformes, com os intervalos relacionados na Tabela 16, nos quais são definidas faixas de valores a mais e a menos a partir do valor da instância de referência. O objetivo do uso dessas distribuições é adicionar ruído aos valores originais.

Para cada navio da instância de referência, é gerado um navio com atributos cujos valores são regidos pelas regras acima. Como a instância de referência contém apenas 30 navios, o processo é repetido a cada 30 navios gerados, a partir do primeiro, com o

Tabela 9 – Intervalos para tempos de chegada, níveis de estoque iniciais e quantidades de carga

Atributo	**Intervalo**
Tempos de chegada	(-10, +10)
Níveis de estoque iniciais	(-20, +20)
Quantidades de carga transportada pelos navios	(-10, +10)

Fonte: Elaborada pelo autor.

tempo de chegada e a quantidade de carga ajustados. A Equação 7.1 mostra como o cálculo do tempo de chegada ajustado é feito para cada navio i gerado e a Equação 7.2 mostra como o cálculo da quantidade de carga ajustada é feito.

$$a_i^{(\varsigma)} = \max\left(a_i + (a_{\text{max}} + \left\lfloor \frac{a_{\text{max}}}{|N|} \right\rfloor) \cdot \varsigma + D(a^-, a^+), 0\right) \tag{7.1}$$

onde:

- a_i é o tempo de chegada do navio i na instância de referência.
- a_{max} é o maior tempo de chegada gerado.
- $|N|$ é o número de navios gerados.
- ς é o número de repetições dos 30 navios.
- $D(a^-, a^+)$ é o valor gerado pela distribuição uniforme discreta entre os limites a^- e a^+, sendo $a^- = -10$ e $a^+ = 10$, conforme Tabela 16.

$$q_{ik}^{(\varsigma)} = \max\left(\min\left(q_{ik} + D(q^-, q^+), q_{\text{max}}\right), q_{\text{min}}\right) \tag{7.2}$$

onde:

- q_{ik} é a quantidade de carga do tipo k do navio i na instância de referência.
- ς é o número de repetições dos 30 navios.
- $D(q^-, q^+)$ é o valor gerado pela distribuição uniforme discreta entre os limites q^- e q^+, sendo $q^- = -10$ e $q^+ = 10$, conforme Tabela 16.
- q_{min} e q_{max} são os limites máximo e mínimo permitidos para os valores de quantidade de carga.

Os níveis de estoque iniciais são mais simples de gerar, tendo em vista que são únicos para as instâncias; portanto, não têm relação com a quantidade de navios. A Equação 7.3 mostra como estes valores são gerados.

$$e_k^{(0)} = \max\left(\min\left(e_k + D(e^-, e^+), e_{\text{max}}\right), e_{\text{min}}\right) \tag{7.3}$$

onde:

- e_k é o nível de estoque inicial para o tipo de carga k na instância de referência.
- $D(\text{ruído}^-, \text{ruído}^+)$ é o valor gerado pela distribuição uniforme discreta entre os limites e^- e e^+, sendo $e^- = -10$ e $e^+ = 10$, conforme Tabela 16.
- e_{min} e e_{max} são os limites mínimo e máximo permitidos para os valores de nível de estoque.

Por fim, os valores de taxa de consumo dos estoques $c_k^{(0)}$ são gerados de forma independente da instância de referência, respeitando apenas limites mínimo e máximo definidos. Assim, os níveis de consumo gerados são dados por $c_k^{(0)} = D(c_{\text{min}}, c_{\text{max}})$, onde $c_{\text{min}} = 2$ e $c_{\text{max}} = 4$, conforme Tabela 10.

Na Tabela 10, estão relacionados os parâmetros mais gerais, que delimitam os dados das instâncias de treinamento: q_{min} e q_{max}, e_{min} e e_{max}, e c_{min} e c_{max}. Na mesma tabela, tem-se as informações sobre o tamanho do *look-ahead*, ou seja, o número de navios considerados em cada decisão, e o número de navios atracados necessários para o término de cada episódio.

Tabela 10 – Parâmetros gerais

Parâmetro	**Descrição**	**Valor**
ϵ	Quantidade de navios no *look-ahead*	10
N_{max}	Número de navios atracados necessários por episódio	60
q_{min}	Quantidade mínima de carga transportada por navio	5
q_{max}	Quantidade máxima de carga transportada por navio	50
e_{min}	Nível mínimo permitido de estoque gerado	38
e_{max}	Nível máximo permitido de estoque gerado	75
c_{min}	Taxa de consumo mínima	2
c_{max}	Taxa de consumo máxima	4

Fonte: Elaborada pelo autor.

O Algoritmo 2 descreve, em alto nível, a sequência lógica de etapas apresentadas nesta seção para a geração de instâncias.

Algorithm 2 Geração de Instâncias de Treinamento

1: **Início**
2: **Enquanto** houver navios a gerar **faça**
3: Selecionar navio de referência correspondente
4: Ajustar tempo de chegada com base nas repetições e adicionar ruído
5: Ajustar quantidades de carga de cada tipo e adicionar ruído
6: **Fim**
7: Gerar níveis de estoque iniciais adicionando ruído aos valores de referência
8: Gerar taxas de consumo aleatoriamente
9: **return** Instância gerada

Os parâmetros selecionados com 2 berços, 4 tipos de carga e *look-ahead* de 10 navios, conforme a Equação 5.1, resultam em um espaço de estados com 76 variáveis observadas. Assim como esses parâmetros definem o espaço de estados, também determinam o número de ações possíveis. Com base na Seção 5.1.2, observa-se que o número de ações varia conforme o número de berços e de critérios utilizados. Nestes experimentos, portanto, conforme visto na Seção 5.1.2, o número de ações é dado por $3^2 = 9$.

A escolha adequada dos parâmetros para a criação de instâncias é fundamental para garantir a representatividade do modelo em relação aos problemas reais. Entretanto, é importante ressaltar que essa escolha também pode impactar diretamente a dimensão do espaço de estados observado pelo modelo. Neste contexto, é preciso encontrar um equilíbrio entre a representatividade do modelo e sua capacidade de processamento.

7.2.2 Hiperparâmetros da DQN

Diversos testes foram realizados para escolher a configuração dos parâmetros e hiperparâmetros para o uso do método DQN, que gerasse os melhores resultados. Este procedimento foi feito por meio de *tentativa e erro* e os valores selecionados estão resumidos na Tabela 11.

7.2.3 Experimentos

Os experimentos foram conduzidos com o objetivo de verificar a capacidade da proposta BAP-RLIM de tomar decisões de atracação que evitem falhas de estoque no PAB descrito na Seção 2.1.2. Para isso, um conjunto de instâncias foi criado exclusivamente para os testes, e os resultados obtidos foram comparados com três critérios de decisão apresentados na Seção 5.1.

Tabela 11 – Principais hiperparâmetros da DQN

Hiperparâmetro	**Valor**
Número total de passos de treinamento	50000
Número de passos de coleta inicial antes do treinamento começar	10000
Passos de coleta por iteração durante o treinamento	4
Capacidade do *buffer* de repetição (*replay buffer*)	10000
Quantidade de neurônios das camadas ocultas	(64, 64, 64)
Tamanho do lote de experiências para treinamento	64
Fator de desconto (γ)	0,99
Probabilidade (ϵ_{greedy}) de escolher uma ação aleatória de exploração (*exploration*)	0,15
Número de passos para atualizar a rede-alvo	10000
Taxa de aprendizado (α) inicial	0,001
Número de passos para decaimento	10000
Taxa de decaimento	0,95
Otimizador	Adam

Fonte: Elaborada pelo autor.

7.2.3.1 Instâncias de testes

Foram geradas 70 instâncias, cada uma com 50 navios, para os experimentos. Todas essas instâncias contêm uma solução viável; ou seja, em cada uma delas há pelo menos uma sequência de atracações de todos os 50 navios, sem que ocorra falha de estoque. Elas foram avaliadas por meio de uma combinação de decisões de um agente previamente treinado e de decisões aleatórias. As decisões aleatórias ocorreram com uma probabilidade de 20%. A quantidade de 70 instâncias foi arbitrada como limite devido à dificuldade de obter tantas com esse nível de aleatoriedade, sem falhas de estoque. Esse conjunto de instâncias está disponível em <https://doi.org/10.5281/zenodo.19411605> e descrito em (BARROS, 2026b).

Foi introduzido um nível de aleatoriedade maior do que o das instâncias de treinamento, com o objetivo de avaliar o poder de generalização do agente treinado. Da instância de referência, apenas o número de berços, o número de tipos de carga e as vazões foram usados e podem ser verificados na Tabela 8. Estes parâmetros são importantes, pois descrevem uma estrutura de porto básica, com baixa expectativa de mudanças, na qual vários cenários, em relação aos demais parâmetros, podem ser derivados. Os demais parâmetros são gerados por meio de distribuições uniformes discretas que seguem os limites mínimo e máximo na Tabela 12.

Tabela 12 – Limites mínimo e máximo das instâncias de testes

Parâmetro	**Descrição**	**Valor**
a_{min}	Tempo de chegada mínimo	0
a_{max}	Tempo de chegada máximo	50
q_{min}	Quantidade mínima de carga transportada por navio	10
q_{max}	Quantidade máxima de carga transportada por navio	50
e_{min}	Nível mínimo de estoque	20
e_{max}	Nível máximo de estoque	70
c_{min}	Taxa de consumo mínima	2
c_{max}	Taxa de consumo máxima	4

Fonte: Elaborada pelo autor.

7.2.4 Análise dos resultados

Os experimentos computacionais foram realizados com o objetivo de avaliar a eficácia da formulação do PAB proposta como um problema de aprendizado por reforço para controlar os níveis de estoque no porto por meio das decisões de atracação. Os experimentos buscam comparar o BAP-RLIM às demais abordagens para: (i) verificar a capacidade de tomar decisões que não incorram em falhas de estoque, (ii) verificar a capacidade do BAP-RLIM de manter os níveis de estoque em níveis mais seguros e (iii) verificar o desempenho do BAP-RLIM em cenário de incertezas em relação ao tempo de chegada esperado.

7.2.4.1 Experimento 1

No primeiro experimento, verifica-se o percentual de simulações sem falhas de estoque organizados em quatro subconjuntos de navios para quatro abordagens diferentes, sendo elas, além do BAP-RLIM, os três tipos de critérios de decisão descritos no Capítulo 6: *First-In First-Served* (FCFS) (navio com menor ETA é chamado primeiro), menor tempo de conclusão (navio com menor tempo de saída do berço é chamado primeiro), e estoque seguro (navio mais breve que atende o tipo de carga com maior risco de entrar em colapso é chamado primeiro). As 70 instâncias foram utilizadas neste experimento. Os subconjuntos com 20, 30, 40 e 50 navios são extraídos delas para o experimento.

Na Tabela 13, estão resumidos os dados obtidos pelas simulações. Na Figura 30, apresenta-se o resultado das simulações, ilustrado por meio de um gráfico de barras. Como se pode ver, os critérios FCFS e *tempo de conclusão* apresentam baixa capacidade de garantir que os níveis de estoque não sejam violados, ficando abaixo de 36% nas simulações sem falhas de estoque com 20 navios e decaem até ficar abaixo de 15% à medida que o número de navios chega a 50. Desta forma, fica claro que os dois

critérios não devem ser usados isoladamente para o objetivo de controlar os níveis de estoque. Mais do que isso, torna-se evidente que dois critérios de usos tão comuns em estratégias operacionais reais e em cenários simulados para fins acadêmicos do problema de alocação de berços são insuficientes para o controle de estoque porque não consideram informações relacionadas aos níveis de estoque.

Tabela 13 – Porcentagem de simulações sem falhas de estoque

Nº de navios	Critério	Sem falhas de estoque (%)
20	FCFS	34,2
20	Tempo de conclusão	34,2
20	Estoque seguro	85,7
20	BAP-RLIM	98,5
30	FCFS	25,7
30	Tempo de conclusão	21,4
30	Estoque seguro	81,4
30	BAP-RLIM	95,7
40	FCFS	18,5
40	Tempo de conclusão	15,7
40	Estoque seguro	78,5
40	BAP-RLIM	92,8
50	FCFS	12,8
50	Tempo de conclusão	14,2
50	Estoque seguro	72,8
50	BAP-RLIM	88,5

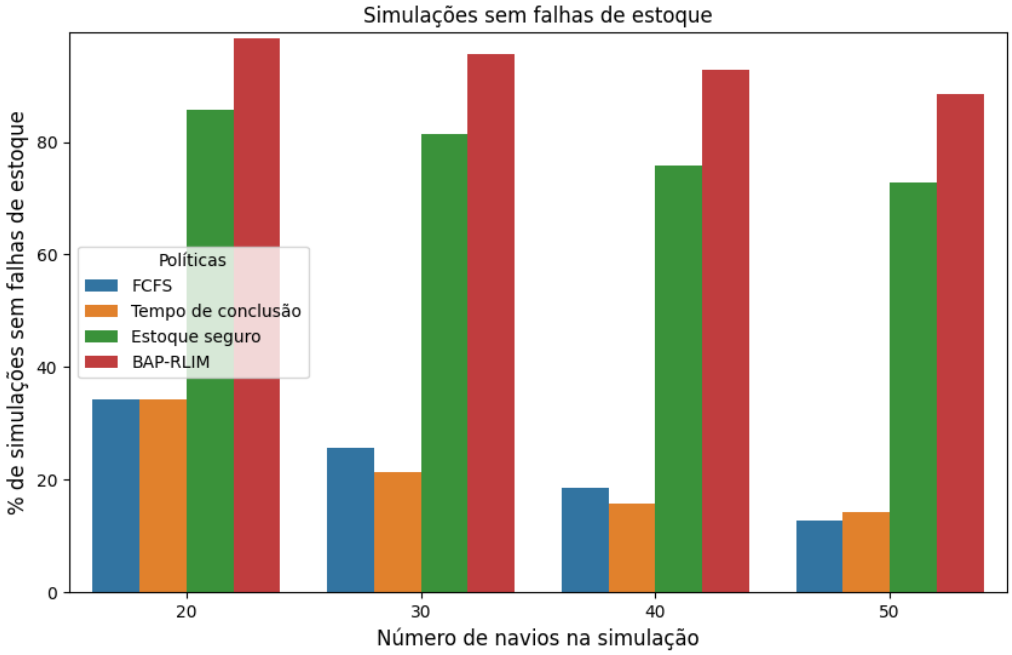
Fonte: Elaborada pelo autor.

O critério *estoque seguro*, por sua vez, utiliza informações relacionadas aos níveis de estoque para decidir o próximo navio a ser chamado. Esta abordagem apresenta resultados bem melhores, variando de 85,7% no subconjunto com 20 navios até 72,8% no subconjunto com 50 navios. Todavia, a garantia dos níveis de estoque é uma restrição rígida, como mencionado na Seção 2.1.2 e, por isso, deve ser assegurada ao máximo. Assim, o critério *estoque seguro* isoladamente pode ser considerado insuficiente para este objetivo. Por outro lado, fica evidente que a abordagem BAP-RLIM proposta obteve resultado consideravelmente superior na tarefa de garantir soluções sem falhas de estoque, conforme pode ser verificado por meio da Tabela 13, o que comprova ser uma proposta promissora para o objetivo apresentado.

Uma sequência de gráficos de linhas é apresentada, por meio das figuras 31, 32 e 33, para análise das evoluções dos níveis de estoque ao longo dos processos de decisão em simulações com 30 navios.

Na Figura 31, uma instância é solucionada por meio das 4 abordagens sem

Figura 30 – Simulações sem falhas de estoque



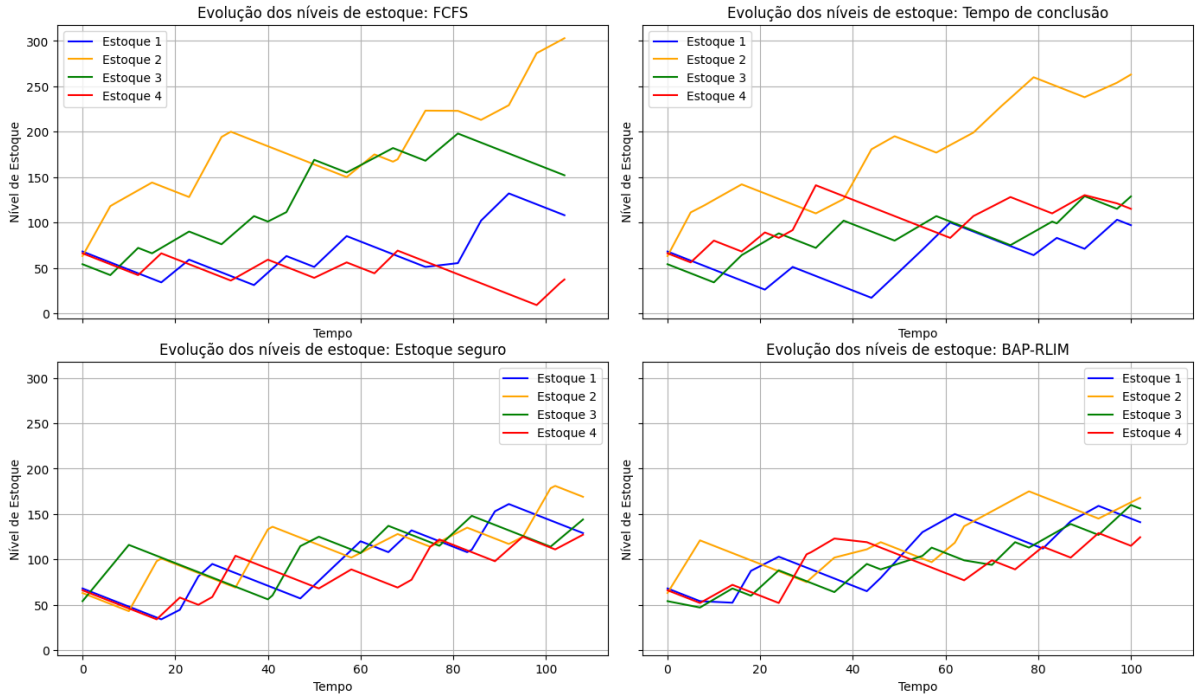
Fonte: Elaborada pelo autor.

apresentar falhas de estoque. É possível observar, porém, que os critérios FCFS e *Tempo de conclusão*, apesar de terem concluído a simulação sem falhas, permitiram que os níveis de estoque 4 e 1, respectivamente, caíssem para níveis perigosos, próximos a 0, em determinados momentos. Por outro lado, o critério *Estoque seguro* e o BAP-RLIM não apenas evitaram falhas, mas também mantiveram os estoques em níveis seguros, concentrados em patamares homogêneos.

A Figura 32 é o resultado do mesmo experimento em outra instância. Neste caso, os critérios FCFS e *Tempo de conclusão* não foram capazes de evitar o colapso de todos os estoques. Enquanto o critério *Estoque seguro* e o BAP-RLIM mantiveram os estoques novamente em níveis concentrados.

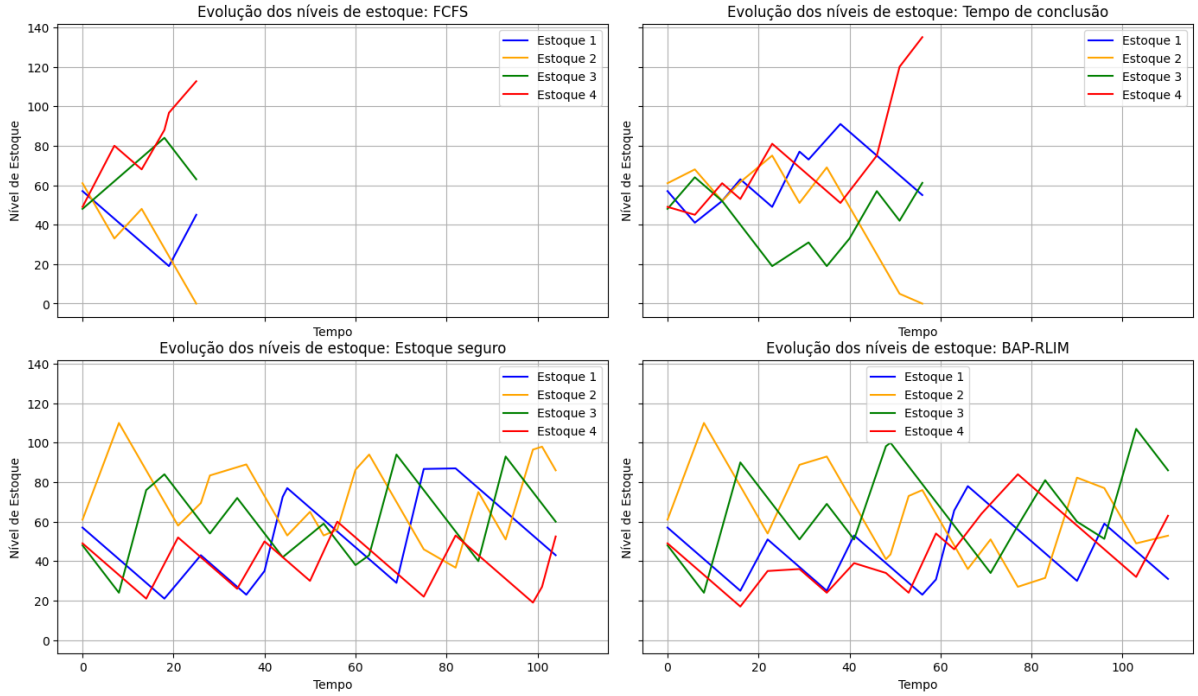
Finalizando a sequência de gráficos de linhas, a Figura 33 demonstra um caso em que apenas o BAP-RLIM evitou falhas no estoque ao longo da simulação. Esse resultado corrobora a Tabela 13, ao evidenciar que apenas um critério ou uma heurística pode ser difícil de identificar regras mais complexas para garantir decisões adequadas no controle de estoque. Mostra também que uma combinação desses critérios simples, por sua vez, pode ser promissora quando um agente inteligente aprende a usá-los. Mais importante ainda, os critérios adotam a abordagem gulosa, ou seja, consideram apenas os dados atuais na tomada de decisão, ignorando consequências futuras. Isso se torna especialmente grave, tendo em vista que o PAB com controle de estoque possui a característica conhecida como dependência de longo prazo, ou seja, as decisões

Figura 31 – Evolução dos níveis de estoque sem falhas.



Fonte: Elaborada pelo autor.

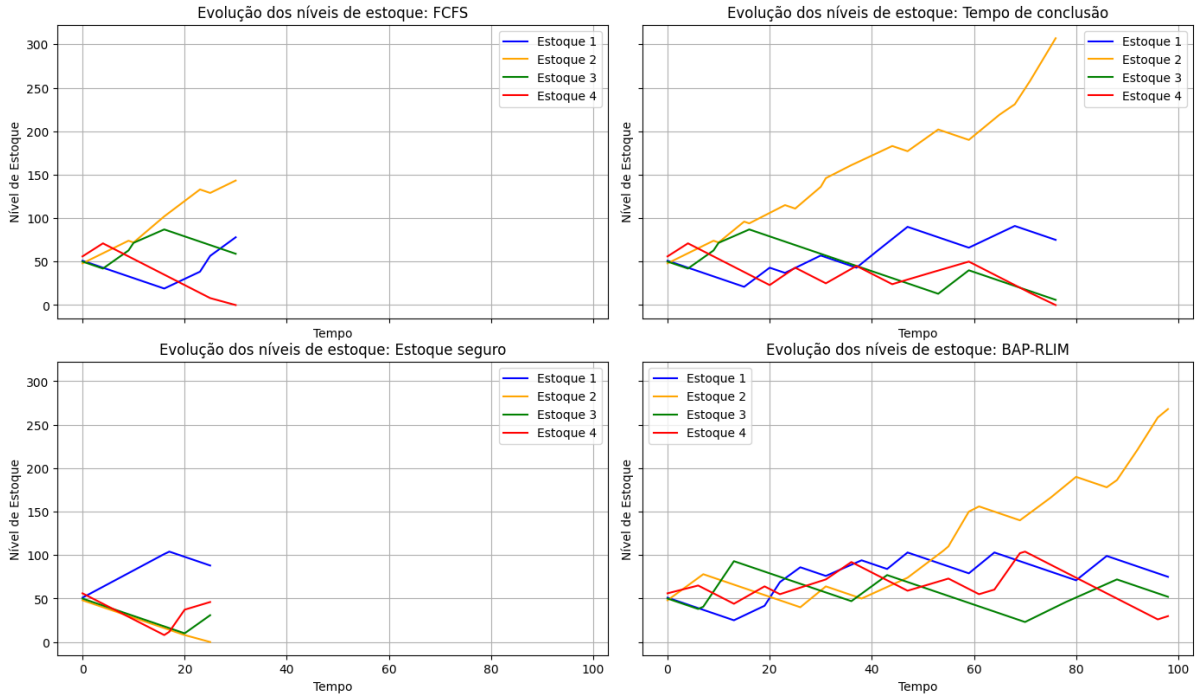
Figura 32 – Evolução dos níveis de estoque sem falhas nos critérios de *estoque seguro* e BAP-RLIM.



Fonte: Elaborada pelo autor.

atuais de atracação persistem e afetam fortemente decisões e estados futuros. Esta característica é mais uma evidência de que a escolha de solução de aprendizado por reforço é uma boa alternativa.

Figura 33 – Evolução dos níveis de estoque sem falhas com BAP-RLIM



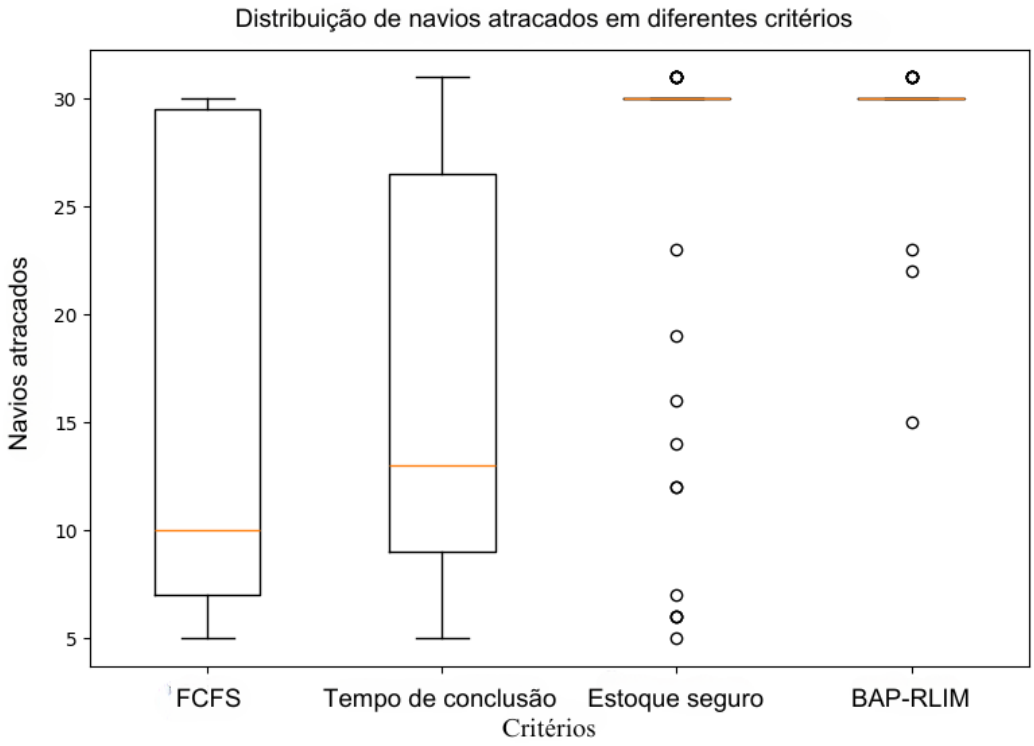
Fonte: Elaborada pelo autor.

Uma outra maneira de avaliar a qualidade das decisões tomadas pelo BAP-RLIM em comparação às outras abordagens é a quantidade de atracações que foram realizadas antes da finalização da simulação. A simulação encerra-se quando o número de navios definido, no caso, 30, for atingido ou quando houver colapso em pelo menos um dos níveis de estoque. As figuras 34 e 35 apresentam duas perspectivas distintas sobre a mesma avaliação.

A Figura 34 demonstra que, para este conjunto de instâncias, os critérios FCFS e *Tempo de conclusão* não encaminham um padrão de soluções claro, além de apresentarem medianas muito distantes da referência de 30 navios, com 10 e 13 navios atracados, respectivamente. O critério *Estoque seguro* e o BAP-RLIM, por outro lado, tiveram exatamente 30 navios como mediana, ficando concentrados nesse valor, como pode ser visto por meio da Figura 35, onde outros resultados surgem como *outliers*.

Este experimento contribui para demonstrar a efetividade da abordagem baseada em aprendizado por reforço com emprego de critérios de seleção de navios como possíveis ações de decisão de atracação. A hipótese verificada é que um agente inteligente é capaz de aprender a usar um conjunto de critérios gulosos simples, que isoladamente não são tão efetivos, para o controle de níveis de estoque em um problema de alocação de berços. Os resultados para as instâncias geradas indicam que a proposta, BAP-RLIM, foi efetiva para esta a avaliação.

Figura 34 – Distribuição de navios atracados sob diferentes critérios.



Fonte: Elaborada pelo autor.

7.2.4.2 Experimento 2

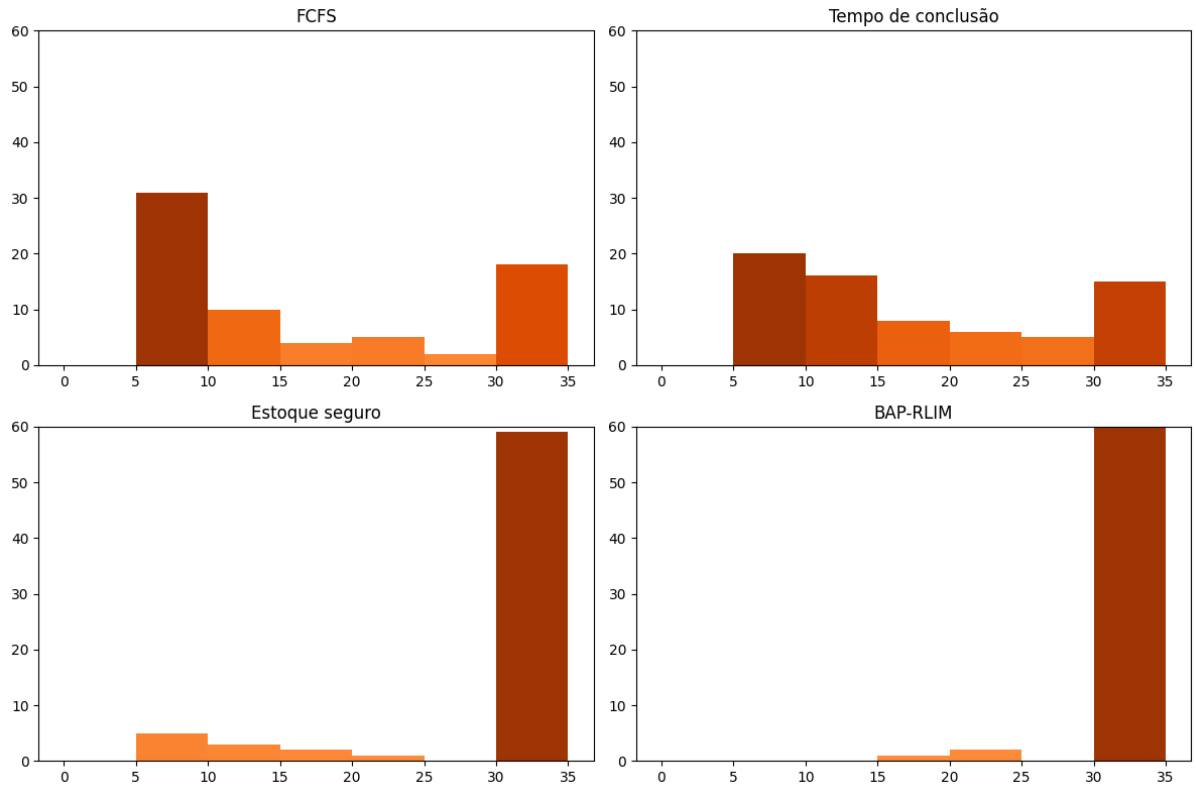
Para avaliar o comportamento do BAP-RLIM diante de incertezas, comparado aos demais critérios no cenário 1, realizou-se o seguinte experimento: ao conjunto de 70 instâncias criadas aplicaram-se atrasos ruídos ao tempo de chegada esperado dos navios não atracados ao longo do processo de decisão. Estes ruídos equivalem a eventuais atrasos dos navios e quanto mais tempo falta para a chegada do navio, maior é o atraso esperado (imprecisão).

Para isto, utilizou-se a distribuição exponencial, com a média baseada no tempo de chegada restante do navio i, ajustado por um fator f, para obter o novo tempo de chegada, $a_i^{\text{ruído}}$, conforme equações 7.4-7.6. Assim, a cada novo estado, ou seja, toda vez que houver um berço disponível, os navios não atracados possivelmente recebem um novo valor de tempo de chegada.

$$\text{atraso}_i \sim e^{\left(\frac{1}{\mu_i}\right)} \tag{7.4}$$

$$\mu_i = f \cdot a_i^{\text{restante}} \tag{7.5}$$

Figura 35 – Distribuição de frequências de navios atracados por critérios.



Fonte: Elaborada pelo autor.

$$a_i^{\text{ruído}} = a_i + \text{atraso}_i \tag{7.6}$$

A Tabela 14 contém um resumo de um conjunto de execuções com ruídos. Cada uma das 70 instâncias foi executada 20 vezes para um total de 40 navios com o critério *estoque seguro* e com o BAP-RLIM. O fator de ruído foi definido como $f = 0,1$. O percentual de simulações sem falhas (%A) e a média (Méd), o desvio padrão (DP), o mínimo (Min) e o máximo (Max) de navios atracados são relacionados.

É fácil notar que o desempenho do BAP-RLIM foi muito superior ao desempenho do critério de estoque seguro. Enquanto o BAP-RLIM obteve 100% de execuções sem falhas de estoque em 45 instâncias (64,2%), o critério de estoque seguro obteve esse resultado em apenas 34 instâncias (48,5%). Por meio da Tabela 13, observando que os resultados são de uma única execução e sem ruídos, o BAP-RLIM diminuiu as execuções sem falhas de estoque em 28,6 pontos percentuais e o critério de estoque diminuiu em 30,0 pontos percentuais. Além disso, é possível notar um comportamento mais homogêneo do BAP-RLIM pelos valores de mínimo, máximo e desvio padrão obtidos.

Na Tabela 15, tem-se a comparação dos resultados da execução do BAP-RLIM para f=0,3 e f=0,5 também. Como pode ser facilmente verificado, há uma perda de desempenho à medida que o fator f aumenta.

Tabela 14 – Desempenho de BAP-RLIM e Estoque Seguro para instâncias com ruído

Inst.	Estoque Seguro					BAP-RLIM				
	%A	Méd	DP	Min	Max	%A	Méd	DP	Min	Max
1	100.0	40.3	0.47	40	41	100.0	40.0	0.0	40	40
2	100.0	40.0	0.0	40	40	100.0	40.2	0.41	40	41
3	90.0	38.25	5.79	19	41	100.0	40.05	0.22	40	41
4	85.0	35.3	12.26	1	40	100.0	40.1	0.31	40	41
5	95.0	38.3	7.84	5	41	95.0	39.6	1.79	32	40
6	40.0	20.9	19.6	1	40	100.0	40.05	0.22	40	41
7	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40
8	55.0	28.65	15.49	1	40	90.0	39.8	1.2	35	41
9	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
10	0.0	3.45	4.7	1	20	90.0	38.05	6.48	14	41
11	100.0	40.0	0.0	40	40	95.0	38.5	6.95	9	41
12	0.0	1.5	1.36	1	7	50.0	33.25	9.94	13	41
13	100.0	40.0	0.0	40	40	100.0	40.15	0.37	40	41
14	90.0	37.55	9.02	1	41	100.0	40.05	0.22	40	41
15	100.0	40.2	0.41	40	41	100.0	40.0	0.0	40	40
16	0.0	1.7	1.66	1	7	100.0	40.1	0.31	40	41
17	95.0	38.75	6.31	12	41	95.0	39.2	4.3	21	41
18	100.0	40.05	0.22	40	41	40.0	26.05	11.99	10	40
19	100.0	40.0	0.0	40	40	80.0	37.25	6.0	22	41
20	40.0	22.45	19.81	1	41	100.0	40.2	0.41	40	41
21	0.0	5.05	6.77	1	23	75.0	35.85	7.71	19	41
22	0.0	6.65	6.43	1	21	75.0	36.7	6.71	19	41
23	0.0	2.15	4.45	1	21	45.0	26.65	12.95	10	41
24	100.0	40.0	0.0	40	40	95.0	38.45	7.17	8	41
25	100.0	40.15	0.37	40	41	100.0	40.2	0.41	40	41
26	100.0	40.0	0.0	40	40	100.0	40.15	0.37	40	41
27	100.0	40.0	0.0	40	40	100.0	40.1	0.31	40	41
28	100.0	40.15	0.37	40	41	100.0	40.05	0.22	40	41
29	100.0	40.0	0.0	40	40	100.0	40.1	0.31	40	41
30	0.0	3.25	5.01	1	19	100.0	40.15	0.37	40	41
31	5.0	19.9	17.35	1	40	100.0	40.05	0.22	40	41
32	100.0	40.1	0.31	40	41	90.0	38.2	5.72	21	41
33	100.0	40.0	0.0	40	40	100.0	40.1	0.31	40	41
34	90.0	36.7	10.48	1	41	95.0	39.95	0.22	39	40
35	0.0	1.2	0.7	1	4	25.0	22.65	12.18	8	41
36	20.0	20.1	17.81	1	40	55.0	33.25	12.0	10	40
37	100.0	40.2	0.41	40	41	90.0	39.45	2.7	28	41
38	0.0	1.8	2.35	1	10	25.0	13.75	15.55	5	40
39	85.0	35.7	12.04	1	41	100.0	40.1	0.31	40	41
40	0.0	1.15	0.67	1	4	75.0	37.85	5.38	22	41
41	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
42	85.0	35.3	12.26	1	40	100.0	40.2	0.41	40	41
43	100.0	40.1	0.31	40	41	100.0	40.05	0.22	40	41
44	100.0	40.0	0.0	40	40	100.0	40.05	0.22	40	41
45	100.0	40.0	0.0	40	40	90.0	37.45	8.03	13	41
46	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
47	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
48	90.0	37.9	8.61	2	41	100.0	40.0	0.0	40	40
49	95.0	38.35	7.38	7	40	100.0	40.0	0.0	40	40
50	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40
51	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40
52	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
53	100.0	40.0	0.0	40	40	100.0	40.1	0.31	40	41
54	20.0	16.8	17.68	1	41	100.0	40.1	0.31	40	41
55	100.0	40.1	0.31	40	41	100.0	40.15	0.37	40	41
56	100.0	40.1	0.31	40	41	100.0	40.15	0.37	40	41
57	0.0	1.45	1.39	1	7	100.0	40.05	0.22	40	41
58	5.0	12.75	14.97	1	40	60.0	37.75	3.45	31	41
59	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40
60	0.0	1.3	1.13	1	6	100.0	40.0	0.0	40	40
61	0.0	5.95	8.55	1	31	100.0	40.05	0.22	40	41
62	0.0	1.4	1.19	1	6	100.0	40.15	0.37	40	41
63	0.0	4.6	3.9	1	14	90.0	37.35	8.34	12	41
64	75.0	32.05	15.94	1	40	95.0	40.1	0.31	40	41
65	5.0	27.25	11.79	1	40	100.0	40.1	0.31	40	41
66	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
67	100.0	40.05	0.22	40	41	100.0	40.0	0.0	40	40
68	100.0	40.0	0.0	40	40	90.0	39.55	1.57	35	41
69	80.0	33.2	14.52	1	41	85.0	38.9	2.94	31	41
70	95.0	38.75	5.59	15	40	100.0	40.05	0.22	40	41

Fonte: Elaborada pelo autor.

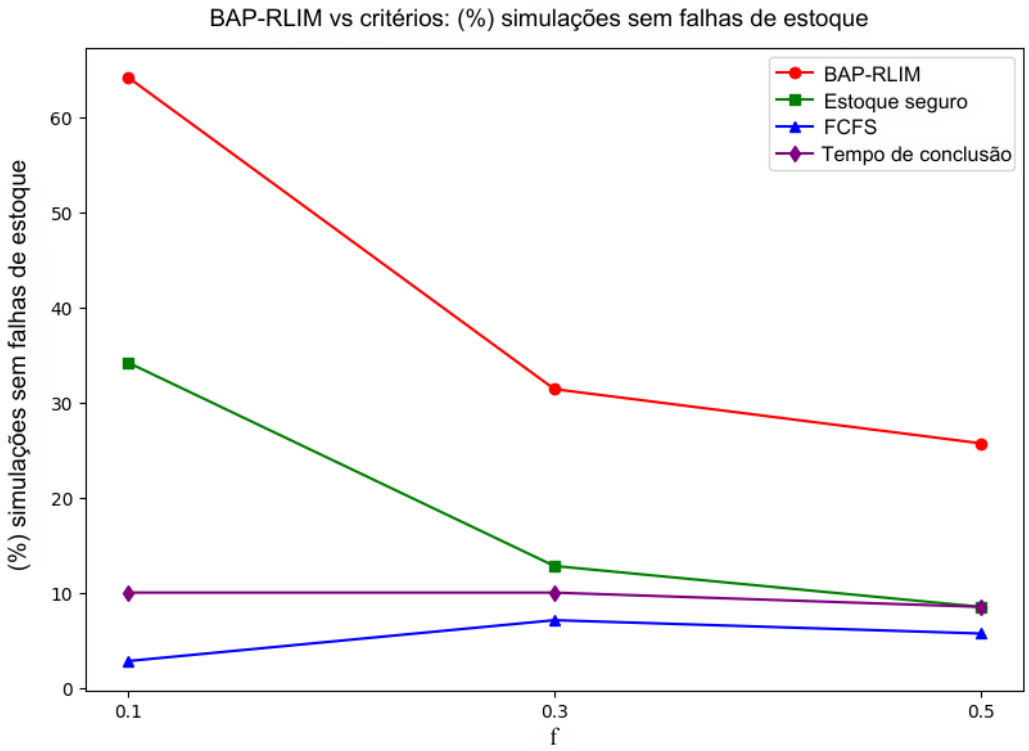
Tabela 15 – Comparação para diferentes variações do BAP-RLIM (f=0.1, f=0.3 e f=0.5).

Inst.	f=0,1					f=0,3					f=0,5				
	%A	Méd	DP	Min	Max	%A	Méd	DP	Min	Max	%A	Méd	DP	Min	Max
1	100.0	40.0	0.0	40	40	100.0	40.1	0.31	40	41	100.0	40.05	0.22	40	41
2	100.0	40.2	0.41	40	41	100.0	40.05	0.22	40	41	100.0	40.15	0.37	40	41
3	100.0	40.05	0.22	40	41	90.0	39.9	1.02	36	41	80.0	38.7	3.57	25	41
4	100.0	40.1	0.31	40	41	100.0	40.1	0.31	40	41	90.0	36.75	10.35	6	41
5	95.0	39.6	1.79	32	40	30.0	29.2	10.07	9	40	30.0	30.95	8.22	16	40
6	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41	90.0	40.1	0.31	40	41
7	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40	100.0	40.1	0.31	40	41
8	90.0	39.8	1.2	35	41	75.0	39.4	2.01	31	40	85.0	39.4	1.64	35	41
9	100.0	40.05	0.22	40	41	100.0	40.1	0.31	40	41	100.0	40.1	0.31	40	41
10	90.0	38.05	6.48	14	41	55.0	32.1	12.79	6	40	65.0	34.6	11.24	6	41
11	95.0	38.5	6.95	9	41	60.0	28.8	14.54	8	40	35.0	22.7	13.77	8	41
12	50.0	33.25	9.94	13	41	0.0	17.9	4.9	12	31	5.0	17.25	7.06	6	40
13	100.0	40.15	0.37	40	41	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
14	100.0	40.05	0.22	40	41	85.0	36.65	8.33	17	41	85.0	36.65	8.33	17	41
15	100.0	40.0	0.0	40	40	35.0	35.55	4.17	31	41	55.0	34.6	9.52	9	40
16	100.0	40.1	0.31	40	41	85.0	37.8	7.92	5	40	75.0	38.8	2.57	32	40
17	95.0	39.2	4.3	21	41	55.0	26.7	16.56	6	41	50.0	23.0	17.44	6	40
18	40.0	26.05	11.99	10	40	10.0	15.7	8.88	10	40	0.0	13.85	3.75	9	21
19	80.0	37.25	6.0	22	41	30.0	26.75	11.03	4	41	10.0	21.85	8.18	5	40
20	100.0	40.2	0.41	40	41	100.0	40.05	0.22	40	41	100.0	40.0	0.0	40	40
21	75.0	35.85	7.71	19	41	45.0	28.3	11.72	12	41	55.0	30.05	11.76	9	41
22	75.0	36.7	6.71	19	41	80.0	37.25	6.25	19	41	50.0	32.8	9.06	18	40
23	45.0	26.65	12.95	10	41	0.0	7.4	1.79	7	15	0.0	8.85	4.06	7	22
24	95.0	38.45	7.17	8	41	65.0	36.1	6.19	20	41	25.0	27.65	9.76	6	40
25	100.0	40.2	0.41	40	41	5.0	19.75	6.68	10	40	0.0	21.75	4.89	14	31
26	100.0	40.15	0.37	40	41	85.0	36.65	8.37	12	40	80.0	35.65	9.3	12	41
27	100.0	40.1	0.31	40	41	30.0	23.55	11.99	7	40	30.0	22.65	12.76	7	40
28	100.0	40.05	0.22	40	41	95.0	39.2	4.54	20	41	95.0	38.7	6.53	11	41
29	100.0	40.1	0.31	40	41	5.0	31.25	5.06	24	40	0.0	29.7	3.26	25	35
30	100.0	40.15	0.37	40	41	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40
31	100.0	40.05	0.22	40	41	15.0	34.8	5.63	21	40	25.0	33.25	7.16	20	41
32	90.0	38.2	5.72	21	41	95.0	39.65	1.57	33	40	100.0	40.05	0.22	40	41
33	100.0	40.1	0.31	40	41	100.0	40.15	0.37	40	41	100.0	40.1	0.31	40	41
34	95.0	39.95	0.22	39	40	30.0	30.05	13.95	7	41	55.0	31.4	14.47	7	40
35	25.0	22.65	12.18	8	41	0.0	17.55	8.97	4	30	0.0	14.85	8.42	6	32
36	55.0	33.25	12.0	10	40	15.0	18.35	12.77	7	40	0.0	19.05	9.65	7	39
37	90.0	39.45	2.7	28	41	65.0	38.55	3.75	27	41	60.0	38.05	4.41	27	41
38	25.0	13.75	15.55	5	40	60.0	28.65	16.36	4	40	40.0	23.6	18.13	4	41
39	100.0	40.1	0.31	40	41	100.0	40.0	0.0	40	40	100.0	40.05	0.22	40	41
40	75.0	37.85	5.38	22	41	65.0	32.75	11.39	4	40	45.0	32.85	7.83	21	40
41	100.0	40.05	0.22	40	41	75.0	32.3	13.8	9	41	80.0	34.0	12.57	9	41
42	100.0	40.2	0.41	40	41	95.0	38.65	6.75	10	41	85.0	36.8	8.02	15	41
43	100.0	40.05	0.22	40	41	100.0	40.0	0.0	40	40	100.0	40.05	0.22	40	41
44	100.0	40.05	0.22	40	41	100.0	40.1	0.31	40	41	100.0	40.05	0.22	40	41
45	90.0	37.45	8.03	13	41	0.0	14.35	7.13	10	40	5.0	13.3	6.62	10	40
46	100.0	40.05	0.22	40	41	70.0	39.3	1.89	34	41	55.0	38.4	6.48	11	41
47	100.0	40.05	0.22	40	41	95.0	39.75	1.12	35	40	95.0	39.95	0.22	39	40
48	100.0	40.0	0.0	40	40	75.0	37.25	4.99	28	40	65.0	37.2	4.48	24	40
49	100.0	40.0	0.0	40	40	100.0	40.0	0.0	40	40	90.0	37.6	7.94	7	40
50	100.0	40.0	0.0	40	40	100.0	40.15	0.37	40	41	95.0	38.45	7.17	8	41
51	100.0	40.0	0.0	40	40	100.0	40.15	0.37	40	41	100.0	40.05	0.22	40	41
52	100.0	40.05	0.22	40	41	100.0	40.1	0.31	40	41	100.0	40.1	0.31	40	41
53	100.0	40.1	0.31	40	41	100.0	40.1	0.31	40	41	100.0	40.0	0.0	40	40
54	100.0	40.1	0.31	40	41	95.0	39.95	0.51	38	41	95.0	40.0	0.32	39	41
55	100.0	40.15	0.37	40	41	25.0	21.5	13.57	5	40	10.0	13.2	10.77	4	41
56	100.0	40.15	0.37	40	41	40.0	36.05	3.99	29	40	25.0	32.6	10.68	4	41
57	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41	100.0	40.05	0.22	40	41
58	60.0	37.75	3.45	31	41	30.0	35.3	4.29	30	41	55.0	37.25	3.71	31	40
59	100.0	40.0	0.0	40	40	90.0	37.85	6.62	18	40	90.0	38.45	5.42	18	41
60	100.0	40.0	0.0	40	40	95.0	38.55	6.96	9	41	95.0	38.3	7.84	5	41
61	100.0	40.05	0.22	40	41	90.0	38.45	4.95	24	41	85.0	38.05	6.31	16	41
62	100.0	40.15	0.37	40	41	100.0	40.0	0.0	40	40	100.0	40.05	0.22	40	41
63	90.0	37.35	8.34	12	41	75.0	34.1	11.06	12	41	60.0	31.25	11.35	12	41
64	95.0	40.1	0.31	40	41	45.0	28.15	14.5	5	40	35.0	28.15	14.67	5	41
65	100.0	40.1	0.31	40	41	70.0	37.05	8.91	11	40	70.0	40.05	0.39	39	41
66	100.0	40.05	0.22	40	41	100.0	40.2	0.41	40	41	100.0	40.15	0.37	40	41
67	100.0	40.0	0.0	40	40	90.0	37.4	8.11	10	40	75.0	33.25	12.31	9	41
68	90.0	39.55	1.57	35	41	100.0	40.1	0.31	40	41	90.0	39.3	2.83	28	41
69	85.0	38.9	2.94	31	41	0.0	19.05	7.51	9	38	0.0	18.15	6.55	5	34
70	100.0	40.05	0.22	40	41	20.0	17.35	12.73	7	40	30.0	17.6	15.11	7	40

Fonte: Elaborada pelo autor.

A Figura 36 faz um comparativo entre o BAP-RLIM e os demais critérios na perspectiva do percentual de instâncias em que não houve falhas de estoque em nenhuma simulação. Fica evidente a tendência de queda à medida que f aumenta, mas demonstra também como a proposta BAP-RLIM indica ser promissora para este cenário de incertezas de tempo de chegada em comparação às demais critérios.

Figura 36 – (%) simulações sem falhas de estoques



Fonte: Elaborada pelo autor.

7.3 Cenário 2

7.3.1 Instâncias e Treinamento

As instâncias baseiam-se no modelo matemático apresentado na Seção 2.1.2, organizadas em 35, 40, 45 e 50 navios; 4 e 5 berços; e 4, 5 e 6 tipos de carga, resultando em um total de 20 instâncias. Todas as instâncias utilizadas neste cenário estão disponíveis em (SILVA; OLIVEIRA, 2022).

Para o treinamento, seleciona-se uma instância de referência para cada combinação de número de berços e de tipos de carga, totalizando 6 processos de treinamento. Todas as instâncias de referência possuem 30 navios. Utilizam-se outras instâncias com 35, 40, 45 e 50 navios para avaliar os agentes treinados. Seu desempenho é comparado ao de soluções ótimas obtidas pelo modelo de programação linear inteira.

Durante o treinamento, cada instância de referência é repetida, após o último navio, duas vezes para formar um episódio contendo 90 navios, com os tempos de chegada dos navios adicionais ajustados adequadamente. Contudo, os atributos dos navios listados na Tabela 16 estão sujeitos a ruído uniformemente distribuído dentro dos intervalos especificados. A Tabela 17 resume os parâmetros gerais empregados no treinamento.

Tabela 16 – Intervalos para os tempos de chegada, os níveis iniciais de inventário e as quantidades de carga.

Atributo	**Intervalo**
Tempos de chegada	(-10, +10)
Níveis iniciais de inventário	(-20, +20)
Quantidades de carga transportadas pelos navios	(-10, +10)

Fonte: Elaborada pelo autor.

Tabela 17 – Parâmetros gerais de treinamento e do ambiente portuário.

Parâmetro	**Valor**
Treinamento	
Número de iterações	10,000
Passos iniciais de coleta	2,000
Tamanho do lote	32
Comprimento da sequência (LSTM)	10
Capacidade do *replay buffer*	2,000
Intervalo de atualização da *target network*	2,000 passos
Hiperparâmetros	
Taxa de aprendizado	0.04 a 0.026
Fator de desconto (γ)	0.99
ϵ-greedy	1.0 a 0.05
Arquitetura da Rede	
Camadas LSTM	(64, 64)
Tamanho da camada totalmente conectada final	número de ações
Ambiente Portuário	
Tamanho do *look-ahead*	10

Fonte: Elaborada pelo autor.

7.3.2 Experimentos

Ao longo do treinamento, as 20 instâncias apresentadas na Tabela 18 foram resolvidas pelos agentes correspondentes para avaliar sua capacidade de obter soluções viáveis que respeitem as restrições de inventário, ao mesmo tempo em que apresentam bom desempenho no tempo total de serviço.

A Tabela 18 apresenta os resultados computacionais, em que os melhores tempos totais de serviço obtidos durante o treinamento são comparados aos produzidos pelo *solver* Gurobi, versão 9.5.1. A tabela também informa os tempos de execução do solver,

limitados a 4.800 segundos, e destaca a melhor solução encontrada para cada instância em ambas as abordagens.

Tabela 18 – Resultados computacionais

	BAP-RLIM	Gurobi v 9.5.1		Best
Instance	Solution	Solution	Time (s)	Solution
35N.4B.5P	1.098,0	**1.086,0	1.273,3	1.086,0
35N.4B.6P	1.239,0	**1.234,0	2.114,7	1.234,0
35N.5B.4P	1.224,0	**1.210,0	508,5	1.210,0
35N.5B.6P	1.606,0	**1.597,0	1.555,2	1.597,0
40N.4B.4P	1.805,0	**1.664,0	4.800,5	1.664,0
40N.4B.5P	2.223,0	**2.181,0	4.800,5	2.181,0
40N.4B.6P	2.183,0	**2.151,0	4.804,6	2.114,0
40N.5B.4P	2.035,0	**2.015,0	4.800,4	2.015,0
40N.5B.5P	2.357,0	**2.272,0	4.804,6	2.272,0
40N.5B.6P	**2.603,0	2.661,0	4.808,7	2.603,0
45N.4B.4P	1.748,0	**1.699,0	1.493,5	1.699,0
45N.4B.5P	**1.588,0	1.590,0	4.800,8	1.588,0
45N.4B.6P	**2.199,0	2.246,0	4.800,9	2.199,0
45N.5B.5P	**2.136,0	2.101,0	4.800,6	2.136,0
45N.5B.6P	**2.545,0	2.661,0	4.800,6	2.545,0
50N.4B.4P	2.013,0	**1.987,0	3.024,0	1.987,0
50N.4B.5P	**2.194,0	2.413,0	4.891,5	2.194,0
50N.4B.6P	**2.726,0	2.755,0	4.812,6	2.726,0
50N.5B.5P	2.703,0	**2.685,0	4.800,9	2.685,0
50N.5B.6P	**3.349,0	3.526,0	4.800,9	3.349,0

Fonte: Elaborada pelo autor.

Todas as instâncias foram solucionadas pela abordagem BAP-RLIM, sem violações de inventário, o que cumpriu, portanto, o objetivo de viabilidade. Observa-se que os resultados obtidos pelo BAP-RLIM superaram os obtidos pelo Gurobi em diversas instâncias. A abordagem de melhor desempenho para cada instância está marcada com (**). Vale ressaltar que, nos seis casos em que o solver finalizou antes do limite de tempo, o Gurobi apresentou soluções superiores, conforme o esperado. Em contraste, das 14 instâncias restantes, oito foram melhor resolvidas pelo BAP-RLIM.

8 Conclusão

Neste trabalho, propôs-se o tratamento do problema de alocação de berços (PAB) em portos graneleiros, sob restrições de maré e de níveis de estoque, com base no complexo portuário de São Luís. Utilizaram-se técnicas de aprendizado por reforço, visando garantir o controle dos níveis de estoque diante de incertezas nos tempos de chegada.

Inicialmente, o PAB foi descrito e formulado como um problema de aprendizado por reforço, mapeando todos os elementos do modelo matemático de programação linear inteira que formaliza o problema. No entanto, foram necessários alguns ajustes para adequar o problema às técnicas de aprendizado por reforço.

O *software* simulador foi implementado para refletir os estados a partir do recebimento das atracações e, por conseguinte, a evolução do cenário até o próximo estado, quando um ou mais berços voltam a estar disponíveis. O simulador foi projetado com um parâmetro para aplicação de ruídos nos tempos de chegada.

Buscou-se um método capaz de fornecer modelos que aprendem a não violar as restrições impostas pelo cenário portuário. Além disso, o processo de decisão é orientado por um conjunto de critérios de atendimento distintos, em que a decisão consiste em determinar qual critério utilizar em cada estado do ambiente.

Inicialmente, foi utilizada a técnica *Deep Q-Network*, que tem sido aplicada com sucesso em problemas de otimização combinatória. Posteriormente, foi proposta uma abordagem que combina DQN com a arquitetura LSTM.

Os resultados obtidos demonstram que o BAP-RLIM apresenta desempenho superior em relação às demais abordagens no controle dos níveis de estoque. No primeiro experimento do Cenário 1, o método atingiu taxas de simulações sem falhas entre 88,5% e 98,5%, enquanto os critérios FCFS e tempo de conclusão permaneceram abaixo de 35% e decresceram para cerca de 13% a 15% com o aumento do número de navios. O critério de estoque seguro apresentou desempenho intermediário, variando entre 72,8% e 85,7%, ainda inferior ao BAP-RLIM. Em instâncias com 30 navios, o BAP-RLIM e o critério de estoque seguro apresentaram uma mediana de 30 navios atracados, ao contrário das outras abordagens, que apresentaram maior variabilidade e desempenho inferiores. No experimento com incerteza, o BAP-RLIM também se destacou em robustez, alcançando 100% de execuções sem falhas em 64,2% das instâncias, contra 48,5% do critério de estoque seguro. Apesar da redução do desempenho decorrente da introdução de ruídos, o método manteve maior estabilidade e consistência nos resultados.

No Cenário 2, todos os casos foram resolvidos pelo BAP-RLIM sem violações de estoque, garantindo viabilidade em 100% das instâncias. Em termos de desempenho, a abordagem superou o *solver* Gurobi em 8 das 14 instâncias em que o tempo limite foi atingido. Esses resultados indicam que o BAP-RLIM é competitivo em relação a métodos exatos, especialmente em instâncias mais complexas, mantendo soluções de boa qualidade com menor dependência de tempo computacional elevado.

De forma geral, os resultados quantitativos indicam que o BAP-RLIM supera consistentemente as demais abordagens, apresentando maior taxa de sucesso, melhor qualidade das soluções e maior robustez diante de incertezas. Todavia, a estrutura BAP-RLIM apresenta algumas limitações ou dificuldades a serem enfrentadas. O número de ações do BAP-RLIM é $|P|^{|L|}$, onde $|P|$ é o número de critérios utilizados e $|L|$ é o número de berços. Em instâncias com um grande número de berços, esse crescimento exponencial pode dificultar o treinamento e comprometer a escalabilidade do método. Além disso, ao contrário de outras abordagens na literatura, a decisão de atracar um navio não exige que este já esteja presente no porto (área de fundeio), embora a atracação, de fato, ocorra apenas após a chegada. Essa flexibilidade pode ser necessária em situações críticas de estoque, mas tende a gerar atrasos subsequentes e deve ser utilizada com cautela.

Como trabalhos futuros, pretende-se aprofundar a investigação do uso de outras arquiteturas de redes neurais, como *Transformers*, visando lidar de forma mais eficaz com o problema de atribuição de crédito temporal presente no PAB descrito neste trabalho. Em relação ao tratamento das incertezas, pretende-se investigar a presença de atrasos nos tempos de atendimento, bem como considerar variações nas taxas de consumo dos estoques. Além disso, planeja-se evoluir o método proposto para uma abordagem multiagente, com o objetivo de melhorar a escalabilidade do modelo em instâncias com maior número de berços, bem como mitigar as dificuldades associadas ao crescimento do espaço de ações decorrente do número de critérios de decisão.

Referências

ABADI, M.; BARHAM, P.; CHEN, J.; CHEN, Z.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; IRVING, G.; ISARD, M. et al. Tensorflow: A system for large-scale machine learning. **12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)**, 2016. Citado na página 74.

AGHALARI, A.; NUR, F.; MARUFUZZAMAN, M. A bender's based nested decomposition algorithm to solve a stochastic inland waterway port management problem considering perishable product. **International Journal of Production Economics**, Elsevier, v. 229, p. 107863, 2020. Citado na página 18.

AI, T.; HUANG, L.; SONG, R.; HUANG, H.; JIAO, F.; MA, W. An improved deep reinforcement learning approach: A case study for optimisation of berth and yard scheduling for bulk cargo terminal. **Advances in Production Engineering & Management**, v. 18, n. 3, 2023. Citado 2 vezes nas páginas 32 e 35.

Alcoa Brasil. **Fact Sheet: Alumar**. 2023. Disponível em: <https://www.alcoa.com/brasil/pt/pdf/brasil-alumar-fact-sheet.pdf>. Citado na página 17.

ANTAQ. **Agência Nacional de Transportes Aquaviários**. Agência Nacional de Transportes Aquaviários, 2022. Disponível em: <http://anuario.antaq.gov.br/ea/index.html>. Citado na página 16.

ANTAQ. **Estatístico Aquaviário**. 2023. Disponível em: <https://web3.antaq.gov.br/ea/sense/movport.html#>. Citado na página 17.

ANTAQ. **Agência Nacional de Transportes Aquaviários**. Agência Nacional de Transportes Aquaviários, 2024. Disponível em: <https://web3.antaq.gov.br/ea/sense/index.html#pt>. Citado na página 17.

BARROS, V. H. **BAP-RLIM Simulation Environment**. 2026. Disponível em: <https://github.com/victorhugobs/bap-rlim/>. Citado na página 70.

BARROS, V. H. **Dataset for the Berth Allocation Problem with Inventory Control**. Zenodo, 2026. Disponível em: <https://doi.org/10.5281/zenodo.19411605>. Citado na página 82.

BARROS, V. H.; COSTA, T. S.; OLIVEIRA, A. C.; LORENA, L. A. Model and heuristic for berth allocation in tidal bulk ports with stock level constraints. **Computers & Industrial Engineering**, Elsevier, v. 60, n. 4, p. 606–613, 2011. Citado 2 vezes nas páginas 17 e 24.

BELOV, G.; BOLAND, N. L.; SAVELSBERGH, M. W.; STUCKEY, P. J. Logistics optimization for a coal supply chain. **Journal of Heuristics**, Springer, v. 26, n. 2, p. 269–300, 2020. Citado 2 vezes nas páginas 18 e 19.

BIERWIRTH, C.; MEISEL, F. A survey of berth allocation and quay crane scheduling problems in container terminals. **European Journal of Operational Research**, Elsevier, v. 202, n. 3, p. 615–627, 2010. Citado na página 17.

BIERWIRTH, C.; MEISEL, F. A follow-up survey of berth allocation and quay crane scheduling problems in container terminals. **European Journal of Operational Research**, Elsevier, v. 244, n. 3, p. 675–689, 2015. Citado na página 17.

BINGHAM, C.; MIKKELSEN, I. T. **Understanding Maritime Decarbonization's Impacts on Trade Costs to Unlock a Just Transition**. 2023. Artigo nº 107. Disponível em: <https://unctad.org/news/transport-newsletter-article-no-107-understanding-maritime-decarbonization>. Citado na página 16.

CERVELLERA, C. et al. Policy optimization for berth allocation problems. In: IEEE. **2021 International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2021. p. 1–6. Citado 3 vezes nas páginas 32, 35 e 36.

CHANG, S.-C.; LIN, M.-H.; TSAI, J.-F. An optimization approach to berth allocation problems. **Mathematics**, MDPI, v. 12, n. 5, p. 753, 2024. Citado na página 17.

CORDEAU, J.-F.; LAPORTE, G.; LEGATO, P.; MOCCIA, L. Models and tabu search heuristics for the berth-allocation problem. **Transportation science**, INFORMS, v. 39, n. 4, p. 526–538, 2005. Citado 2 vezes nas páginas 22 e 23.

DAI, Y.; LI, Z.; WANG, B. Optimizing berth allocation in maritime transportation with quay crane setup times using reinforcement learning. **Journal of Marine Science and Engineering**, MDPI, v. 11, n. 5, p. 1025, 2023. Citado 2 vezes nas páginas 33 e 35.

EMAP. **Porto do Itaqui: Movimentação de Cargas**. 2023. Disponível em: <https://www.portodoitaqui.com/porto-do-itaqui/operacoes-portuarias/movimentacao-de-carga>. Citado na página 17.

FILOM, S.; AMIRI, A. M.; RAZAVI, S. Applications of machine learning methods in port operations–a systematic literature review. **Transportation Research Part E: Logistics and Transportation Review**, Elsevier, v. 161, p. 102722, 2022. Citado 2 vezes nas páginas 19 e 31.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with lstm. **Neural computation**, MIT press, v. 12, n. 10, p. 2451–2471, 2000. Citado na página 76.

GUADARRAMA, S.; KORATTIKARA, A.; RAMIREZ, O.; CASTRO, P.; HOLLY, E.; FISHMAN, S.; WANG, K.; GONINA, E.; WU, N.; KOKIOPOULOU, E.; SBAIZ, L.; SMITH, J.; BARTóK, G.; BERENT, J.; HARRIS, C.; VANHOUCKE, V.; BREVDO, E. **TF-Agents: A library for Reinforcement Learning in TensorFlow**. 2018. <https://github.com/tensorflow/agents>. [Online; accessado 22-Agosto-2023]. Disponível em: <https://github.com/tensorflow/agents>. Citado 2 vezes nas páginas 70 e 76.

HAUSKNECHT, M. J.; STONE, P. Deep recurrent q-learning for partially observable mdps. In: **AAAI fall symposia**. [S.l.: s.n.], 2015. v. 45, p. 141. Citado na página 76.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 76.

IMAI, A.; NAGAIWA, K.; TAT, C. W. Efficient planning of berth allocation for container terminals in asia. **Journal of Advanced transportation**, Wiley Online Library, v. 31, n. 1, p. 75–94, 1997. Citado na página 23.

IMAI, A.; NISHIMURA, E.; PAPADIMITRIOU, S. The dynamic berth allocation problem for a container port. **Transportation Research Part B: Methodological**, Elsevier, v. 35, n. 4, p. 401–417, 2001. Citado na página 22.

JIN, X.; DUAN, Z.; SONG, W.; LI, Q. Container stacking optimization based on deep reinforcement learning. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 123, p. 106508, 2023. Citado na página 18.

KOLLEY, L.; RÜCKERT, N.; KASTNER, M.; JAHN, C.; FISCHER, K. Robust berth scheduling using machine learning for vessel arrival time prediction. **Flexible Services and Manufacturing Journal**, Springer, p. 1–41, 2022. Citado 3 vezes nas páginas 19, 31 e 35.

LEÓN, A. D. de; LALLA-RUIZ, E.; MELIÁN-BATISTA, B.; MORENO-VEGA, J. M. A machine learning-based system for berth scheduling at bulk terminals. **Expert Systems with Applications**, Elsevier, v. 87, p. 170–182, 2017. Citado 2 vezes nas páginas 31 e 35.

LI, B.; YANG, C.; YANG, Z. Multiple container terminal berth allocation and joint operation based on dueling double deep q-network. **Journal of Marine Science and Engineering**, MDPI, v. 11, n. 12, p. 2240, 2023. Citado 3 vezes nas páginas 19, 34 e 35.

LI, C.; WU, S.; LI, Z.; ZHANG, Y.; ZHANG, L.; GOMES, L. Intelligent scheduling method for bulk cargo terminal loading process based on deep reinforcement learning. **Electronics**, MDPI, v. 11, n. 9, p. 1390, 2022. Citado 2 vezes nas páginas 32 e 35.

LIM, A. The berth planning problem. **Operations research letters**, Elsevier, v. 22, n. 2-3, p. 105–110, 1998. Citado na página 22.

LIU, C.; XIANG, X.; ZHANG, C.; ZHENG, L. A decision model for berth allocation under uncertainty considering service level using an adaptive differential evolution algorithm. **Asia-Pacific Journal of Operational Research**, World Scientific, v. 33, n. 06, p. 1650049, 2016. Citado na página 18.

LV, Y.; ZOU, M.; LI, J.; LIU, J. Dynamic berth allocation under uncertainties based on deep reinforcement learning towards resilient ports. **Ocean & Coastal Management**, Elsevier, v. 252, p. 107113, 2024. Citado 4 vezes nas páginas 18, 19, 33 e 35.

MEHDI, E. R. E.; ILYAS, H.; FRANÇOIS, S. et al. Incremental lns framework for integrated production, inventory, and vessel scheduling: Application to a global supply chain. **Omega**, Elsevier, v. 116, p. 102821, 2023. Citado na página 18.

MNIH, V.; KAVUKCUOGLU, K.; SILVER, D.; GRAVES, A.; ANTONOGLOU, I.; WIERSTRA, D.; RIEDMILLER, M. Playing atari with deep reinforcement learning. **arXiv preprint arXiv:1312.5602**, 2013. Citado na página 29.

MNIH, V.; KAVUKCUOGLU, K.; SILVER, D.; RUSU, A. A.; VENESS, J.; BELLEMARE, M. G.; GRAVES, A.; RIEDMILLER, M.; FIDJELAND, A. K.; OSTROVSKI, G. et al. Human-level control through deep reinforcement learning. **nature**, Nature Publishing Group, v. 518, n. 7540, p. 529–533, 2015. Citado na página 29.

RODRIGUES, F.; AGRA, A. Berth allocation and quay crane assignment/scheduling problem under uncertainty: A survey. **European Journal of Operational Research**, Elsevier, v. 303, n. 2, p. 501–524, 2022. Citado 2 vezes nas páginas 22 e 23.

SCHEPLER, X.; ABSI, N.; FEILLET, D.; SANLAVILLE, E. The stochastic discrete berth allocation problem. **EURO Journal on Transportation and Logistics**, Elsevier, v. 8, n. 4, p. 363–396, 2019. Citado na página 19.

SILVA, J. R. S. e; OLIVEIRA, A. C. M. de. **Instances for the Berth Allocation Problem in Tidal Bulk Ports with Inventory Control (BAPTBI)**. 2022. *Mendeley Data*, version 3. Disponível em: <https://data.mendeley.com/datasets/58ph43s6h4/3>. Citado na página 92.

SILVA, J. R. Silva e. **Dynamic berth allocation problem for tidal bulk ports with inventory level constraints**. Tese (dissertation) — Universidade Federal do Maranhão, Ago 2021. Disponível em: <https://tedebc.ufma.br/jspui/handle/tede/3995>. Citado 3 vezes nas páginas 24, 74 e 75.

STEENKEN, D.; VOSS, S.; STAHLBOCK, R. Container terminal operation and operations research-a classification and literature review. **OR spectrum**, Springer, v. 26, p. 3–49, 2004. Citado na página 17.

SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: An introduction**. [S.l.]: MIT press, 2018. Citado 3 vezes nas páginas 27, 28 e 56.

TWILLER, J. van; SIVERTSEN, A.; PACINO, D.; JENSEN, R. M. Literature survey on the container stowage planning problem. **European Journal of Operational Research**, Elsevier, 2023. Citado na página 17.

United Nations Conference on Trade and Development. **Review of Maritime Transport 2021**. [S.l.]: United Nations, 2021. (Review of maritime transport). ISBN 978-92-1-113026-3. Citado na página 16.

United Nations Conference on Trade and Development (UNCTAD). **Review of Maritime Transport 2023: Towards a Green and Just Transition**. Geneva: United Nations, 2023. 157 p. (Review of Maritime Transport). ISBN 978-92-1-002886-8. Disponível em: <https://unctad.org/publication/review-maritime-transport-2023>. Citado na página 16.

WTO Secretariat. **World Trade Report 2021**. Genève, Switzerland: World Trade Organization, 2021. Citado na página 16.

WTO Secretariat. **World Trade Report 2023**. Genève, Switzerland: World Trade Organization, 2023. Citado na página 18.

XIANG, X.; LIU, C. An expanded robust optimisation approach for the berth allocation problem considering uncertain operation time. **Omega**, Elsevier, v. 103, p. 102444, 2021. Citado na página 19.

ZHANG, Y.; BAI, R.; QU, R.; TU, C.; JIN, J. A deep reinforcement learning based hyper-heuristic for combinatorial optimisation with uncertainties. **European Journal of Operational Research**, Elsevier, v. 300, n. 2, p. 418–427, 2022. Citado 2 vezes nas páginas 31 e 35.